e-ISSN: 2148-7456 | a peer-reviewed online journal | hosted by DergiPark

# International Journal of Assessment Tools in Education

Volume: 12 Issue: 2 June 2025

https://dergipark.org.tr/en/pub/ijate
https://ijate.net/index.php/ijate

e-ISSN: 2148-7456

Volume 12 Issue 2 June 2025

**Editor** Dr. Safiye BILICAN DEMIR
**Address** Department of Educational Sciences, Kırıkkale University, Ankara yolu Km 7. Yahşihan 71450, Kirikkale, Türkiye

**E-mail** ijate.editor@gmail.com
safiyebilican@gmail.com

**Publisher** Dr. Izzet KARA
**Address** Pamukkale University, Education Faculty, Kinikli Campus, 20070, Denizli, Türkiye

**Phone** +90 258 296 1036
**Fax** +90 258 296 1200
**E-mail** ikara@pau.edu.tr
ijate.info@gmail.com

**Journal Contact** Dr. Eren Can AYBEK
**Address** Department of Educational Sciences, Pamukkale University, Faculty of Education, Kinikli Yerleskesi, Denizli, 20070, Türkiye
**Phone** +90 258 296 31050
**Fax** +90 258 296 1200
**E-mail** erencanaybek@gmail.com

**Address** Dr. Anil KANDEMIR
Department of Educational Sciences, Agri Ibrahim Cecen University, Faculty of Education, Agri, Türkiye
akandemir@agri.edu.tr

**Frequency** 4 issues per year (March, June, September, December)
**Online ISSN** 2148-7456
**Website** https://dergipark.org.tr/en/pub/ijate
https://ijate.net/index.php/ijate

**Cover Design** IJATE

International Journal of Assessment Tools in Education (IJATE) is a peer-reviewed and academic online journal. The scientific and legal responsibility for manuscripts published in our journal belongs to the authors(s).

# International Journal of Assessment Tools in Education

***International Journal of Assessment Tools in Education*** (IJATE) accepts original research on the design, analysis and use of evaluation along with assessment to enhance comprehension of the performance and quality of stakeholders in educational settings. IJATE is pleased to receive discriminating theoretical and empirical manuscripts (quantitative or qualitative) which could direct significant national and international argumentations in educational policy and practice.

IJATE, as an online journal, is hosted by DergiPark [TUBITAK-ULAKBIM (The Scientific and Technological Research Council of Türkiye)].

In IJATE, there are no charges under any procedure for submitting or publishing an article.

## Indexes and Platforms:

- **Emerging Sources Citation Index (ESCI)**
- **TR Index (ULAKBIM),**
- **Education Resources Information Center (ERIC),**
- **EBSCOhost,**
- **SOBIAD Citation Index,**
- **MIAR (Information Matrix for Analysis of the Journals),**
- **idealonline,**

**Editor**

Dr. Safiye BILICAN DEMIR, *Kırıkkale University*, Türkiye

**Section Editors**

Dr. Ebru BALTA, *Ağrı İbrahim Çeçen Üniversitesi*, Türkiye
Dr. Nisu BORAN, *Pamukkale University*, Türkiye
Dr. Sumeyra SOYSAL, *Necmettin Erbakan University*, Türkiye
Dr. Ömer Faruk ŞEN, *Kırıkkale University*, Türkiye
Dr. Selma SENEL, *Balikesir University*, Türkiye
Dr. Esin YILMAZ KOGAR, *Nigde Omer Halisdemir University*, Türkiye

**Editorial Board**

Dr. Beyza AKSU DUNYA, *Bartın University*, Türkiye
Dr. Stanislav AVSEC, *University of Ljubljana*, Slovenia
Dr. Utkun AYDIN, *University of Glasgow*, United Kingdom
Dr. Duru BAYRAM, *Eindhoven University of Technology,* The Netherlands
Dr. Kelly D. BRADLEY, *University of Kentucky*, United States
Dr. Okan BULUT, *University of Alberta*, Canada
Dr. William W. COBERN, *Western Michigan University*, United States
Dr. Nuri DOGAN, *Hacettepe University*, Türkiye
Dr. Selahattin GELBAL, *Hacettepe University*, Türkiye
Dr. Qiwei HE, *Georgetown University*, United States
Dr. Lim HOOI LIAN, *School of Educational Studies University Sains*, Malaysia
Dr. Tugba KARADAVUT, *Izmir Democracy University*, Türkiye
Dr. Orhan KARAMUSTAFAOGLU, *Amasya University*, Türkiye
Dr. Hulya KELECIOGLU, *Hacettepe University*, Türkiye
Dr. Hakan KOGAR, *Akdeniz University,* Türkiye
Dr. Seongyong LEE, *Hannam University*, South Korea
Dr. Sunbok LEE, *University of Houston*, United States
Dr. Hamzeh MORADI, *Sun Yat-sen Universit*, China
Dr. Nesrin OZTURK, *Izmir Democracy University*, Türkiye
Dr. Turan PAKER, *Pamukkale University*, Türkiye
Dr. Hossein SALARIAN, *University of Tehran*, Iran
Dr. Ragip TERZI, *Harran University*, Türkiye
Dr. Turgut TURKDOGAN, *Pamukkale University*, Türkiye
Dr. Hongwei YANG, *University of West Florida*, United States
Dr. Ozen YILDIRIM, *Pamukkale University*, Türkiye

**English Language Editors**

Dr. Hatice ALTUN, *Pamukkale University*, Türkiye
Ahmet KUTUK, *Akdeniz University*, Türkiye

**Editorial Assistant**

Dr. Asiye BAHTIYAR, *Pamukkale University,* Türkiye

Dr. Neslihan Tugce OZYETER, *Kocaeli University*, Türkiye

PhDc. Ibrahim Hakki TEZCI, *Akdeniz University*, Türkiye

**Technical Assistant**

Dr. Eren Can AYBEK, *Pamukkale University*, Türkiye

Dr. Anil KANDEMIR, *Agri Ibrahim Cecen University*, Türkiye

## CONTENTS

***Research Articles***

Using self-organizing map and rating scale model in examining internal structure of scales

**Page: 180-197** **PDF**

Sinan Bekmezci, Nuri Doğan

The relationship between teachers' organizational commitment and career anxiety

**Page: 198-221** **PDF**

Tuğba Alagöz, Yusuf İnandı, Yüksel Gündüz

A systematic review on the effect of lesson study on entrepreneurial competencies of foreign language teachers

**Page: 222-239** **PDF**

Nalan Özbey, Duygu İleritürk

A systematic review of the recommendations of theses on fine arts high school music departments in Türkiye

**Page: 240-276** **PDF**

Haşim Vapur, Ülkü Sevim Şen

Teacher perceived emotional job demands scale: Validity, reliability and invariance study

**Page: 277-292** **PDF**

Alper Uslukaya, Yılmaz Arslan, Füsun Yükrük, Oktay Ağ, Mehmet Akif Evcimik

Preschool teachers' awareness and views of children's self-regulation scales: A professional development program's effect

**Page: 293-316** **PDF**

Asiye İvrendi

A review of automatic item generation techniques leveraging large language models

**Page: 317-340** **PDF**

Bin Tan, Nour Armoush , Elisabetta Mazzullo, Okan Bulut, Mark Gierl

Exploring number of response categories in factor analysis: Implications for sample size

**Page: 341-352** **PDF**

Fatih Orçan

Project guidance self-efficacy scale: Validity and reliability study

**Page: 353-366** **PDF**

Perihan Güneş, Esra Çakırlar Altuntaş, Miraç Yılmaz

The perception scale for curriculum leadership: Validity and reliability study

**Page: 367-386** **PDF**

Cemile Keski, İlkay Doğan Taş

Examining the level of culturally responsive school culture

**Page: 387-413** **PDF**

Gürkan Sarıdaş, Funda Nayır

The effect of polytomous item ratio on ability estimation in multistage tests

**Page: 414-429** **PDF**

Hasibe Yahsi Sari, Hülya Kelecioğlu

Investigating the relationship among gifted students' spatial reasoning skills, their conceptual understanding of astronomy and academic achievement

**Page: 430-456** **PDF**

Pelin Ertekin, Mustafa Serdar Köksal

Validity and reliability study of the school technology leadership scale according to teachers' perceptions

**Page: 457-475** **PDF**

Tugba Ögücü, Aytac Gogus

Assessing the reliability of the Beck Depression Inventory adapted for Turkish culture: A meta-analysis

**Page: 476-488** **PDF**

Yasemin Duygu Esen, Filiz İsnaç , Hamide Deniz Gülleroğlu

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 180–197
https://doi.org/10.21449/ijate.1388256
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# Using self-organizing map and rating scale model in examining internal structure of scales

**Sinan M. Bekmezci**[1*], **Nuri Doğan**[2]

[1]Manisa Celal Bayar University, Faculty of Education, Mathematics and Science Education, Manisa, Türkiye
[2]Hacettepe University, Faculty of Education, Department of Educational Sciences, Ankara, Türkiye

**ARTICLE HISTORY**
*Received: Nov. 9, 2023*
*Accepted: Nov. 29, 2024*

**Keywords:**
Scale development,
Internal structure,
Self-organizing map,
Rating scale modal,
Exploratory factor analysis.

**Abstract:** This study compares the psychometric properties of scales developed using Exploratory Factor Analysis (EFA), Self-Organizing Map (SOM), and Andrich's Rating Scale Model (RSM). Data for the research were collected by administering the "Statistical Attitude Scale" trial form, previously used in a separate study, to 808 individuals. First, EFA, SOM and RSM were applied to decide the number of dimensions of the scale, and to select items. Subsequently, Confirmatory Factor Analysis (CFA) was used to the forms obtained from different methods and their CFA fit indices were compared. The analysis revealed variations in the number of dimensions and item distribution across different methods. Results indicated that the form generated using SOM exhibited the highest fit indices. Furthermore, the CFA fit indices of the form created with RSM were found to be satisfactory, offering detailed insights into both items and individuals.

## 1. INTRODUCTION

Acquiring and measuring knowledge, skills, and affective qualities are extremely important in education. Measurement procedures related to the affective qualities expected to be possessed by the individual find a vast place in academic studies and international exams. For this reason, obtaining statistically significant results from the data collection tools used in the measurement and evaluation processes is important to determine the level of these affective qualities. In this context, the psychometric properties of the scores obtained from the data collection tool, such as validity and reliability, should be aligned with the standards.

There are different definitions and classifications for validity in literature. In one approach, validity is defined as the degree to which a test can measure the desired feature without confusing it with other features (Ebel & Frisbie, 1991). Validity, as defined by the American Psychological Association, refers to the extent to which evidence and theory support the interpretations of test scores for their intended uses (AERA, 2014, p.11). According to AERA (2014), there are four validity evidences: test content, response processes, internal structure, relations to other variables. The internal structure includes the number of dimensions of the test, whether the structure is the same among different groups, which requires statistical calculations in its presentation. Depending on the type of test, the internal structure may be influenced by

*CONTACT: Sinan M. BEKMEZCİ ✉ sinan.bekmezci@cbu.edu.tr Manisa Celal Bayar University, Faculty of Education, Mathematics and Science Education, Manisa, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

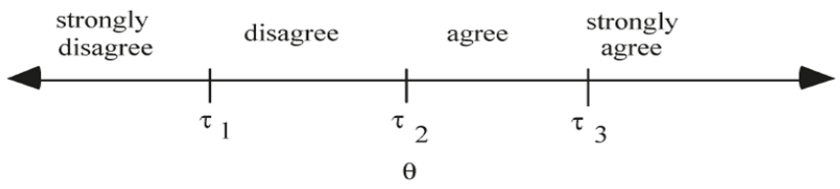
**e-ISSN: 2148-7456**

differences in item types within the test, type of item response category, unexpected nonlinear relationships between items, non-normal distribution of data, unexpected relationships between factors, and size of the sample. For this reason, when developing a measurement tool, it is thought to be critical to choose a statistical technique appropriate to the nature of the measurement tool to reveal evidence about the internal structure in a meaningful way. Fava and Velicer (1992) and Wood *et al*. (1996) have suggested that failure to accurately determine the number of dimensions may lead to the merging of dimensions with each other and item loading onto incorrect factors. The current study focuses on the internal construct validity evidence of Likert-type scales.

Scales can be employed to measure individuals' abstract features, called 'structure', that cannot be directly observed (Erkuş, 2012; Nunally & Bernstein, 1994). Likert-type scales typically consist of items with ordinal and multi-category responses. To statistically prove the validity of the scores obtained from the Likert-type scales, it is important to determine the number of dimensions of the scales and select the most qualified items. During the development process, various algorithms or models rooted in different theories may be applied for item selection and dimensionality determination. Factor analysis is commonly used in dimension determination and item selection (Anastasi, 1976; Crocker & Algina, 2008). The literature includes various comparative studies on techniques for determining the number of dimensions in Likert-type scales and selecting items. Examples of such studies include Şimşek's (2006) comparison of EFA with cluster analysis and multi-dimensional scaling, Doğan and Başokçu's (2010) comparison of factor analysis with stepwise cluster analysis, and Kozan's (2016) study comparing factor analysis, cluster analysis, and multi-dimensional scaling. In addition to classical methods, Item Response Theory (IRT) models also provide extensive opportunities for examining the psychometric properties of the Likert-type scale.

In this context, since the Likert-type scale typically consist of polytomous item, it would be reasonable to use polytomous IRT models to estimate non-linear relationships between the respondent's propensity level and the likelihood of responding in a specific category (Embretson & Reise, 2000). In IRT, model selection is made considering the dimensionality of the measurement tool (single or multiple), the categorization of item responses (dichotomous or polytomous), and the parameters to be estimated. Graded response model, partial credit model, rating scale model, and generalized partial credit model are among the models that can be used in the analysis of Likert-type ordinal polytomous measurement instruments (De Ayala, 2009; Embretson & Reise, 2000; Ostini & Nering, 2006). Polytomous IRT models reflect two types of conditional probabilities: (a) the probability of responding in a specific category and (b) the probability of responding positively instead of negatively at a particular boundary between two categories (Ostini & Nering, 2006). Current study, RSM (Andrich, 1978a; Andrich, 1978b) was employed. The reason for using RSM in this study is that the category threshold values calculated in the RSM process remain the same for all items in the scale (Bond & Fox, 2015). In addition to the fact that the calculated boundary values in RSM remain the same for all items, a boundary value equal to one less than the number of categories is calculated for all items in the scale. Figure 1 illustrates the representation of an item with four response categories and three thresholds.

**Figure 1.** *Category thresholds.*




According to Figure 1, an individual can respond positively to the first and second thresholds (leading to Category 3), respond negatively to the first and second thresholds (leading to

Category 1), or respond positively to the first threshold and negatively to the second threshold (leading to Category 2) (De Ayala, 2009). In RSM, each item has its own estimated item difficulty, and all items share a common threshold structure (Bond & Fox, 2015), the fact that all items share a common threshold structure distinguishes RSM from other polytomous IRT models.

In recent years, with the rapid advancement of information technologies, there has been a significant increase in the amount of data that can be collected. Among these large datasets, previously unknown relationships may exist, requiring more advanced analysis to find these relationships. Additionally, sometimes when the assumptions of classical methods cannot be met, the analysis results may not be sufficiently dependable. Considering such reasons, it has been thought that artificial neural network (ANN) algorithms, which are part of machine learning, can also be beneficial. ANN has proven successful in various fields, including object recognition, natural language processing, numerical prediction, and classification (Bramer, 2020), for purposes such as prediction, clustering, and classification (Eğrioğlu *et al*., 2019).

The ability of artificial neural networks to create complex and nonlinear models is also considered an advantage (Çevik & Tabaru-Örnek, 2020). Murtagh and Hernandez-Pajares (1995) call SOM a well-known neural network that is closely related to cluster analysis. Kohonen (1982), who discovered SOM, said that SOM is a network where neurons in a two-dimensional grid are individually adjusted for each input signal, particularly in large datasets and for different input patterns. Kohonen (2014) stated that while producing low-dimensional projections from high-dimensional data, SOM preserves similarity relationships. SOM is a nonlinear statistical technique and can be used to analyze multivariate data for exploration and clustering purposes, working with fewer traditional statistical principles, and requiring less in-depth knowledge than in statistical and multivariate analyses (Ferles *et al*. 2021; Galvan, *et al*., 2020; Galvan *et al*., 2020; Galvan *et al*., 2021; Haykin, 2009). The unsupervised learning model of SOM enables clustering in situations where groups are unknown. Competitive learning rule is used to perform unsupervised learning (Haykin, 2009). SOM selects a subset of "models," each consisting of a data vector and its best-matching model, called the "winner," along with spatial neighbors, and adjusts them for better matching (Kohonen, 2014). Distance measures are used to find the similarity between the input vector and neurons (Miljković, 2017). These distance measures include Euclidean, correlation, cosine similarity, and Manhattan distance (Grajciarova *et al*., 2012).

When reviewing the literature, it is evident that while some studies exclusively employed IRT models, others compared classical models with IRT models. For instance, Krishnan and Idris (2018), Cantó Cerdán *et al*. (2021), İlhan and Güler (2018), Peixoto *et al*. (2021), Takasaki *et al*. (2021) and Karlin and Karlin (2018) utilized various Rasch models in scale development studies. Additionally, Karlin and Karlin (2018) contended that besides assessing test validity, the Rasch model also offers an opportunity to evaluate students. In their studies, Uysal (2015) and İlhan and Güler (2017) compared classical test theory and IRT, with İlhan and Güler (2017) reported a strong correlation and relative agreement between ability measures obtained from Classical Test Theory and Rasch analysis. In a few rare studies, it is seen that SOM is also used. For instance, Kiang, and Kumar (2001) found that SOM outperformed pre-rotation factor analysis, providing more accurate insights into the latent structure, particularly in skewed data. Francis (2001) suggested the use of ANN models for dimension reduction in data with nonlinear relationships, comparing them to factor analysis in their study. Sadesky (2007) aimed to determine test structure using SOM and highlighted its ability to accurately preserve contiguity and proximity relations, suggesting further research applications in educational assessment. Tezbaşaran (2016) and Eriş Hasırcı (2019) employed the SOM method and reported its utility in item selection to determine the dimension-item structure of tests measuring affective features with polytomous items. In this context, it can be observed that IRT models may be effective in

revealing the internal structure of Likert-type scales. Additionally, it appears that SOM can be useful in determining the number of dimensions and selecting items for the Likert-type scales.

Despite the advantages identified in the studies, it is thought that there is a significant gap in research on determining the psychometric properties of IRT models and artificial neural networks, especially in the context of Likert-type scales. In addition, it is thought that comparative studies evaluating the effectiveness of these methods and clarifying their advantages and disadvantages are lacking. For this purpose, the current study was designed to compare different techniques that can be used in the test development process. In this context, it is thought that this study can help researchers who want to develop a Likert-type scale in the process of dimension determination and item selection.

For this purpose, the study aims to answer the following question: "What are the confirmatory factor analysis CFA fit indices for test forms developed from items selected based on the results of EFA, SOM and RSM?"

## 2. METHOD

In the initial step of the study, the pilot version of the Statistical Attitude Scale was administered. In the second step, variations of the Statistical Attitude Scale were created using EFA (1 dimension - 32 item), SOM (1 dimension - 16 item, 2 dimension - 26 item), and RSM (1 dimension - 25 item, 1 dimension - 32 item). The fit indices obtained by separately applying CFA to these 5 models were compared, and the form with the best-fit values was revealed. In this context, the research has descriptive features in terms of revealing scale structures to be obtained from different techniques (Grove *et al*., 2012) and relational features in terms of comparing CFA fit indexes of scale structures obtained from different techniques (Büyüköztürk *et al*., 2017).

### 2.1. Study Group

The Attitude Scale Towards Statistics trial form used in the study was administered to 808 university students and graduates. While 26% (*N*: 211) of the group is male, 74% (*N*: 597) were female. While 32% (*N*: 257) of the participants stated that they had not taken a statistics course before, 68% (*N*: 551) stated that they had taken a statistics course. "Which of the following best describes your academic field?" It is seen that 36% (*N*: 289) quantitative, 43% (*N*: 347) equal weight (quantitative and qualitative), 18% (*N*: 149) qualitative, and 3% (*N*: 23) foreign language answers were given to the question.

### 2.2. Data Collection

Data were collected in the fall semester of 2020 after obtaining the approval from the Hacettepe University Ethics Committee. During the research process, the scale was planned to be administered face-to-face. However, when the implementation process was started, the scale trial form was delivered to the participants online due to the national and international COVID-19 pandemic.

### 2.3. Instruments

There are 40 items in this scale, which were prepared to determine the level of attitude towards statistics. Twenty of the items are positive, and 20 of them are negative statements. Negatively worded statements were reverse scored. The scale trial form and the descriptive statistics of the items in the scale are in Appendix A. Item pairs m9-m40 and m1-m31 in the scale were used as control items. Before the analyses were performed, the polychoric correlation value between m1-m31 (.69) and m9-m40 (.77) item pairs was significant at .01 significance level; it was determined that there is a relationship above the medium level. Therefore, control items (m31 and m40) with weaker psychometric properties were excluded from the scale in the subsequent analyses.

### 2.4. Data Analysis

Analyzes were performed with MATLAB, Winsteps, JASP, Factor, and R software. To provide cross-validation before the analyses were performed, the data set was randomly divided into three groups of 33%, and EFA was performed with the JASP program over three groups. After these analyses, it was found that there was no significant difference between the factor loadings of the same items in different forms. Afterward, the polychoric correlation matrix was used in the EFA performed with the entire data set. In addition, estimations were made using the unweighted least squares method since the item responses were in the ordering scale and the responses were not normally distributed. Parallel Analysis was also used to decide the number of dimensions. The scale's psychometric properties were determined in line with the variance rates explained because of the analysis, item factor load, and dimensionality convergence index.

SOM analyses were performed in the MATLAB software. SOM works by creating maps consisting of "nxn" neurons, where "n" is the number of neurons, each neuron represents a dimension, and which items are in which dimensions, and this distribution can change at different "n" values. In the study, before the final evaluation according to SOM, the item and size distributions in the maps consisting of 4x4, 3x3, and 2x2 neurons were examined, and it was determined that the model consisting of 3x3 neurons was ideal and the items to be selected for the final test were determined through this model.

RSM analysis was performed using Winsteps software. In the RSM analysis, the items selected for the final form were chosen according to their infit and outfit values. Accordingly, a form was obtained by removing the items with at least one of the infit or outfit values above "1.4" (Bond & Fox, 2015). Then, the items with at least one of the infit or outfit values above "1" were removed, and another form was created.

Since each of the techniques discussed in the research performs its own calculations, instead of directly comparing the results, CFA was applied via the JASP software to the different models established, and each model's fit and error indices were compared. In CFA, diagonally weighted least squares estimation (DWLS), which is the most reliable parameter estimation method for ordinal data and generally when the variables are not normally distributed, was used (Brown, 2006; Flora & Curran, 2004; Li, 2016).

## 3. FINDINGS

### 3.1. Findings Related to EFA

The results of KMO and Bartlett Sphericity Test obtained from EFA are given in Table 1. A KMO value above .90 indicates that the sample size is appropriate for factoring, and that the Bartlett test is statistically significant, showing that the inter-item correlation matrix is suitable for factor extraction (Şencan, 2005). As shown in Table 1, data were suitable for EFA.

**Table 1.** *KMO-Bartlett's sphericity test results.*

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin Measure of Sampling Adequacy | | .978 |
| Bartlett's Test of Sphericity | Approx. Chi-Square | 27891.0 |
| | *df* | 703 |
| | Sig | .000 |

Table 2 shows the eigenvalues obtained using the unweighted least squares technique and the explained variance percentages. Çokluk *et al.* (2012) stated that one of the points to be taken into consideration when deciding on the number of factors is the percentage of variance explained and that the number of factors can be determined as "1" if the percentage of variance explained by the eigenvalues after the percentage of variance explained by the first eigenvalue is significantly reduced. As shown in Table 2, the first eigenvalue has a 59% contribution to the

explained variance. In comparison, the second eigenvalue has a relatively low contribution of 9%. According to Table 2, the scale is one-dimensional.

**Table 2.** *Explained variance based on eigenvalues.*

| Factor | Eigenvalue | Proportion of Variance | Cumulative Proportion of Variance |
|---|---|---|---|
| 1 | 22.26430 | .58590 | .58590 |
| 2 | 3.56160 | .09373 | .67963 |
| 3 | 1.62018 | .04264 | .72227 |
| 4 | 0.91090 | .02397 | |

Table 3 shows the matrix of unrotated factor loadings. The reason for using unrotated factor loadings in the analysis process is that there is no similar process in the analyses of SOM and RSM. Since the last item in Table 3 is 39, there seem to be 39 items in the analysis, but there are 38 items in the analysis. This is because the item numbers were kept the same to avoid confusion when the analysis was performed after item 31 and item 40 were subtracted and compared with the items selected by other methods. The same is true for Table 4. Table 3 shows that item 3, item 15, and item 16 communality values are less than 0.50. And item 34 and item 38 give the highest loading values to the third factor. In addition, the PA technique with one-dimensional limitation was applied to support the one-dimensionality assessment based on Table 2 and Table 3.

**Table 3.** *Unrotated loading matrix.*

| Item | F1 | F2 | F3 | Communality | Item | F1 | F2 | F3 | Communality |
|---|---|---|---|---|---|---|---|---|---|
| 1 | .839 | -.227 | -.084 | .763 | 20 | .840 | .232 | -.106 | .771 |
| 2 | .819 | -.238 | -.087 | .736 | 21 | .843 | .140 | -.098 | .740 |
| 3 | .525 | .402 | .110 | **.449** | 22 | .766 | -.360 | -.094 | .726 |
| 4 | .757 | .229 | -.010 | .626 | 23 | .789 | -.388 | -.083 | .780 |
| 5 | .862 | -.221 | -.020 | .792 | 24 | .831 | -.190 | .035 | .727 |
| 6 | .774 | -.344 | -.088 | .726 | 25 | .841 | -.332 | -.053 | .820 |
| 7 | .715 | .375 | .071 | .656 | 26 | .776 | -.426 | -.060 | .786 |
| 8 | .768 | .240 | -.201 | .688 | 27 | .775 | .292 | .052 | .689 |
| 9 | .739 | .276 | -.218 | .669 | 28 | .728 | .305 | -.149 | .646 |
| 10 | .760 | .260 | -.247 | .706 | 29 | .831 | .204 | -.040 | .735 |
| 11 | .770 | -.417 | -.032 | .767 | 30 | .750 | -.269 | .008 | .635 |
| 12 | .837 | -.133 | .064 | .723 | 32 | .862 | .287 | -.012 | .826 |
| 13 | .843 | .300 | -.016 | .801 | 33 | .762 | .339 | .139 | .714 |
| 14 | .812 | .304 | .052 | .755 | 34 | .514 | .018 | **.621** | .650 |
| 15 | .538 | .351 | .036 | **.414** | 35 | .783 | .377 | -.013 | .755 |
| 16 | .302 | .326 | .083 | **.204** | 36 | .630 | -.170 | .444 | .623 |
| 17 | .769 | -.317 | .005 | .691 | 37 | .803 | -.389 | .005 | .796 |
| 18 | .868 | -.186 | .008 | .788 | 38 | .514 | -.042 | **.629** | .662 |
| 19 | .863 | .210 | -.011 | .789 | 39 | .741 | -.372 | -.026 | .688 |

Table 4 shows the PA factor loadings matrix with a one-dimensional boundary. Table 4 has shown that the communality values of item 3, item 15, item 16, item 34, item 36 and item 38 were less than 0.5. In addition, when the one-dimensional convergence indices obtained as a result of PA are examined, the unidimensionality fit index (UniCo), which is desired to be more than 0.95, was found to be 0.971, the explained common variance index (ECV), which is desired to be more than 0.85, was found to be 0.866, the mean index of the residual absolute loadings

of the item, which was required to be less than 0.30, was found to be 0.284. These results also support the interpretation that the scale is one-dimensional.

**Table 4.** *Unrotated loading matrix according to single factor limitation.*

| Item | F1 | Communality | Item | F1 | Communality |
|---|---|---|---|---|---|
| 1 | .868 | .754 | 20 | .843 | .710 |
| 2 | .846 | .715 | 21 | .841 | .707 |
| 3 | .538 | **.289** | 22 | .798 | .638 |
| 4 | .749 | .561 | 23 | .822 | .675 |
| 5 | .882 | .778 | 24 | .830 | .690 |
| 6 | .804 | .647 | 25 | .865 | .749 |
| 7 | .728 | .530 | 26 | .821 | .673 |
| 8 | .779 | .607 | 27 | .789 | .623 |
| 9 | .759 | .576 | 28 | .738 | .544 |
| 10 | .780 | .609 | 29 | .838 | .702 |
| 11 | .808 | .654 | 30 | .760 | .578 |
| 12 | .846 | .716 | 32 | .885 | .783 |
| 13 | .854 | .730 | 33 | .780 | .608 |
| 14 | .830 | .689 | **34** | **.514** | **.264** |
| 15 | .544 | **.295** | 35 | .806 | .650 |
| **16** | .303 | **.092** | 36 | .630 | **.397** |
| 17 | .787 | .620 | 37 | .824 | .679 |
| 18 | .872 | .760 | **38** | **.518** | **.268** |
| 19 | .869 | .754 | 39 | .770 | .593 |

McDonald's Omega reliability coefficient of the form was calculated as .97, and Cronbach's Standardized Alpha reliability coefficient was calculated as .97. These values provide evidence that the results obtained from the measurement tool are reliable.

As a result, it is likely appropriate to exclude item 16, item 34, and item 38 from the test, whose factor loadings in Table 3 are not high enough and/or give high loads to more than one dimension. In addition to these items, item 3, item 15, and item 36, whose common variance values were less than .5 according to the PA results made with one-dimensionality limitation, were not included in the final form. Thus, a one-dimensional and 32-item scale was obtained.

### 3.2. Findings Related to SOM

Figure 2 shows a SOM graph consisting of 3x3 neurons, n=3. It says which dimension and how many items are in the hexagons. The neuron and hence the size numbering is done in a specific order, with the lower left neuron being the first neuron and the upper right neuron being the ninth neuron. According to Figure 2, there are 16 items in neuron number eight, 10 items in neuron number three, 3 items in neuron number one, 2 items each in neurons 6 and 7, and 1 item in neurons 2, 4, and 5. The fact that the inside of a neuron is fully colored indicates that the neuron is saturated, without the color being important (different programs may have different color representations). We can consider the saturation of the neuron as the network receives as much material as a neuron can receive at a given number of pieces of training. Accordingly, the eighth neuron was fully saturated, the third neuron was nearing saturation, and the remaining neurons were not saturated According to Figure 2, items are collected in 2 main dimensions.

**Figure 2.** *Graph of hits (item neuron distribution) for SOM analysis.*

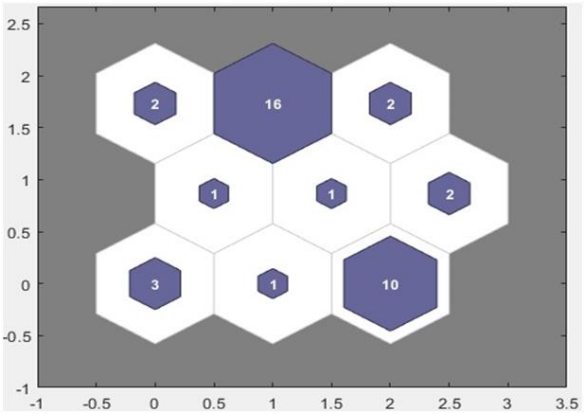



Figure 3 gives the color representation of the distance between neurons. By examining the "Neighboring Weight Distances" between neurons in Figure 3 and the items in the scale, items in close dimensions (neurons) can be considered together. The fact that the colors go from light to dark shows that the distance between the dimensions increases. In this case, the dimensions 3-6 and 5-9 are close to each other and the items in these dimensions can be combined. The researcher can decide by considering the items' distances and the theoretical structure.

**Figure 3.** *Graph of hits (item neuron distribution) for SOM analysis.*

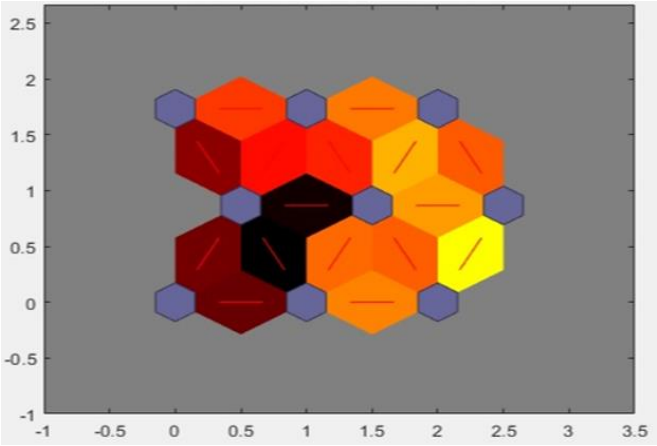


Table 5 shows which item is included in which dimension. When Figure 2 and Figure 3 are evaluated together, a 1- or 2- dimensional structure can be created by considering the theoretical background of the scale. According to Table 5, items 34, 36, and 38 belong to the first dimension; item 24 to the second dimension; item 15 to the fourth dimension; item 5 to the fifth dimension; items 1 and 2 to the sixth dimension; items 3 and 16 to the seventh dimension; and items 12 and 18 to the ninth dimension.

**Table 5.** *SOM substances neuron distribution.*

| Dimensions | Item |
|---|---|
| 1 | 34, 36, 38 |
| 2 | 24 |
| 3 | 6, 11, 17, 22, 23, 25, 26, 30, 37, 39 |
| 4 | 15 |
| 5 | 5 |
| 6 | 1, 2 |
| 7 | 3, 16 |
| 8 | 4, 7, 8, 9, 10, 13, 14, 19, 20, 21, 27, 28, 29, 32, 33, 35 |
| 9 | 12, 18 |

According to Figure 1, a one-dimensional form comprising 16 items was created with items 4, 7, 8, 9, 10, 13, 14, 19, 20, 21, 27, 28, 29, 32, 33, and 35 located in the eighth dimension, which is saturated. Another two-dimensional form with 26 items was created by adding items 6, 11, 17, 22, 23, 25, 26, 30, 37 and 39 in the third dimension to the items in the eighth dimension.

### 3.3. Findings Related to RSM

Rasch analysis examines dimensionality through primary components analysis of Rasch measurement residuals. Since EFA was performed during the research process and RSM dimensionality results were also highly like these results, the dimensionality outputs of RSM were not included in the findings.

Table 6 shows descriptive statistics, and reliability values of 804 people who were determined as non-extreme (outlier) individuals. As shown in Table 6, fit statistics of the person' mean are very close to 1, and the data on the individuals provide the model-data fit. In addition, The Cronbach Alpha reliability coefficient is relatively high as .97. The fact that the separation index of individuals is more than 2 (Linacre, 2012) indicates that students with different attitudes toward statistics are effectively differentiated.

**Table 6.** *Summary of 804 measured (non-extreme) person.*

| | Total Score | Measure | Model S. E. | Infit Mnsq | Infit Zstd | Outfit Mnsq | Outfit Zstd |
|---|---|---|---|---|---|---|---|
| Mean | 116.6 | 0.0674 | 0.2039 | 1.05 | -0.3 | 1.09 | -0.2 |
| P. Sd. | 33.3 | 1.2829 | 0.0762 | 0.63 | 2.4 | 0.79 | 2.3 |
| S. Sd. | 33.4 | 1.2837 | 0.0762 | 0.63 | 2.4 | 0.79 | 2.3 |
| Max. | 189.0 | 5.2893 | 1.0037 | 4.08 | 8.4 | 9.90 | 9.0 |
| Min. | 39.0 | -5.0817 | 0.1736 | 0.19 | -5.9 | 0.20 | -5.7 |
| Real RMSE .2502 | | | Separation 5.03 | | | Person Reliability .96 | |
| True SD 1.2583 | | | | | | | |
| Model RMSE .2177 | | | Separation 5.81 | | | Person Reliability .97 | |
| True SD 1.2643 | | | | | | | |
| S.E. of Person Mean = 0.0453 | | | | | | | |
| Person Raw Score-to-Measure Correlation = .96 | | | | | | | |
| Cronbach Alfa (KR-20) Person Raw Score “Test” Reliability = 0.97 SEM = 5.41 | | | | | | | |

**Table 7.** *Item measurement reports.*

| | Total Score | Measure | Model S. E. | Infit Mnsq | Infit Zstd | Outfit Mnsq | Outfit Zstd |
|---|---|---|---|---|---|---|---|
| Mean | 2478.0 | 0.0000 | 0.0417 | 1.01 | -0.9 | 1.09 | -0.1 |
| P. Sd. | 477.4 | 0.7979 | 0.0018 | 0.35 | 5.2 | 0.50 | 4.7 |
| S. Sd. | 483.9 | 0.8086 | 0.0018 | 0.36 | 5.3 | 0.51 | 4.8 |
| Max. | 3362.0 | 1.2794 | 0.0486 | 2.28 | 9.9 | 3.13 | 9.9 |
| Min. | 1715.0 | -1.6056 | 0.0398 | 0.62 | -9.1 | 0.60 | -8.4 |
| Real RMSE 0.0445 | | | Separation 17.92 | | | Item Reliability 1.00 | |
| True SD 0.7966 | | | | | | | |
| Model RMSE 0.0417 | | | Separation 19.11 | | | Item Reliability 1.00 | |
| True SD 0.7968 | | | | | | | |
| S.E. of Item Mean = 0.1312 | | | | | | | |
| Item Raw Score-to-Measure Correlation =-1.00 | | | | | | | |

Table 7 shows descriptive statistics, and reliability values of the items. The items' mean infit and outfit values in Table 7 are very close to 1. Item reliability value is 1. The item reliability index indicates the consistency of item placements when the same items are given to another

sample with similar behavior (Bond & Fox, 2015). In addition, since the item reliability level is relatively high, there is no need to look at the item separation index.

Table 8 gives category threshold values, category infit and outfit values, and descriptive statistics of categories. The threshold values in Table 8 and the infit and outfit values of these values show us that the item categorizations work properly. As shown the fit statistics in Table 8 which between .5 and 1.5, indicating that the categories work properly (Linacre, 2014). Therefore, the answer categories and the five-point Likert rating work properly.

**Table 8.** *Category threshold values.*

| Category | % | Observed Average | Sample Expect | Infit | Outfit | Andrich Threshold | Category Measure | Estimate Disc. |
|---|---|---|---|---|---|---|---|---|
| 1 | 18 | -1.742 | -1.71 | 1.01 | 1.24 | NONE | -2.56 | |
| 2 | 17 | -0.7128 | -0.746 | 1.01 | 1.07 | -1.14 | -1.10 | 1.10 |
| 3 | 24 | 0.0168 | 0.0149 | 0.84 | 0.81 | -0.69 | -0.05 | 1.00 |
| 4 | 22 | 0.8005 | 0.7728 | 0.86 | 0.92 | 0.48 | 1.08 | 1.00 |
| 5 | 19 | 1.7060 | 1.738 | 1.21 | 1.38 | 1.35 | 2.69 | 0.91 |

Table 9 gives the item measure values, which can be interpreted as RSM item difficulty and infit-outfit. The measurement value gives information about the item's difficulty level (the level of having the measured feature) (Linacre, 2012). Small measure values in Likert-type scale items indicate that the state of expressing strongly disagree and disagree is high, and the state of agreeing and strongly agree is low. According to Table 9, item 3 has the lowest difficulty, while item 23 has the highest. Table 9 also shows that items 3, 15, 16, 34, 38, and 39 have infit or outfit values of 1.4 or higher, indicating they are not compatible with the model. Consequently, these items were removed, resulting in a one-dimensional form with 32 items.

**Table 9.** *Rating scale model results.*

| Item | Measure | Infit | Outfit | Item | Measure | Infit | Outfit |
|---|---|---|---|---|---|---|---|
| 1 | 0.7568 | 0.6355 | 0.6802 | 20 | -0.2739 | 0.7668 | 0.7792 |
| 2 | 0.9528 | 0.7613 | 0.928 | 21 | -0.0963 | 0.8723 | 0.829 |
| 3 | -1.6056 | 1.5467 | 1.6677 | 22 | 0.9562 | 0.9055 | 0.8517 |
| 4 | -0.7262 | 1.0874 | 1.056 | 23 | 1.2794 | 0.8178 | 0.8317 |
| 5 | 0.2884 | 0.6798 | 0.6612 | 24 | 0.4397 | 0.6994 | 0.7171 |
| 6 | 1.1631 | 0.8623 | 1.0584 | 25 | 0.9339 | 0.7293 | 0.8302 |
| 7 | -1.1194 | 1.2334 | 1.1815 | 26 | 1.2663 | 0.8653 | 0.9299 |
| 8 | -0.13 | 0.9958 | 0.9749 | 27 | -.4788 | 1.0177 | 1.1069 |
| 9 | -0.3889 | 1.1492 | 1.3611 | 28 | -.3608 | 1.102 | 1.1208 |
| 10 | -0.0547 | 0.8923 | 0.94 | 29 | -0.2398 | 0.8915 | 0.8378 |
| 11 | 1.1233 | 0.8498 | 0.8813 | 30 | 0.6808 | 0.9141 | 0.9163 |
| 12 | -0.0324 | 0.8049 | 0.8832 | 32 | -0.5719 | 0.8127 | 0.7784 |
| 13 | -0.846 | 0.9183 | 0.8535 | 33 | -0.9942 | 0.9958 | 0.9914 |
| 14 | -0.5277 | 0.9456 | 0.8961 | 34 | -1.061 | 1.7331 | 1.8537 |
| 15 | -0.4822 | 1.8003 | 2.5511 | 35 | -0.7474 | 0.9677 | 0.9364 |
| 16 | -0.6113 | 2.2792 | 3.131 | 36 | -.2903 | 1.3273 | 1.3758 |
| 17 | 0.7319 | 0.7922 | 0.8461 | 37 | 1.0965 | 0.7163 | 0.7583 |
| 18 | 0.1916 | 0.6234 | 0.6017 | 38 | -1.169 | 1.5985 | 1.5629 |
| 19 | -0.2414 | 0.8702 | 0.8517 | 39 | 1.1886 | 0.9938 | 1.48 |

In addition, infit and outfit statistics are desired to be 1.00 because it shows perfect match of between data and modal (Güler *et al*., 2018). For this reason, an attempt was made to ensure perfect data-modal compatibility by considering that items with at least one of the infit and outfit values greater than '1' could be deleted. Accordingly, another one-dimensional 25-item form was created by removing items numbered 3, 4, 6, 7, 9, 15, 16, 27, 28, 34, 36, 38, 39, of which at least one of the infit, and outfit values are greater than 1.

### 3.4. CFA Findings of Different Models

Below is a table of critical values for CFA indexes and tables of CFA indexes of different forms. Table 10 presents the critical values for CFA as suggested by Schermelleh-Engel *et al*. (2003).

**Table 10.** *Critical values for CFA indices.*

| Indices | Good Fit | Acceptable Fit |
|---|---|---|
| $\chi^2/df$ | $0 \leq \chi^2/df \leq 2$ | $2 < \chi^2/df \leq 3$ |
| RMSEA | $.00 \leq$ RMSEA $\leq .05$ | $.05 <$RMSEA $\leq .08$ |
| NFI | $.95 \leq$ NFI $\leq 1.00$ | $.90 \leq$ NFI $< .95$ |
| NNFI | $.97 \leq$ NNFI $\leq 1.00$ | $.95 \leq$ NNFI $< .97$ |
| CFI | $.97 \leq$ CFI $\leq 1.00$ | $.95 \leq$ CFI $< .97$ |
| GFI | $.95 \leq$ GFI $\leq 1.00$ | $.90 \leq$ GFI $< .95$ |
| SRMR | $.00 \leq$ SRMR $\leq .05$ | $.05 <$SRMR $\leq .10$ |

Table 11 shows the fit indices of different models. Considering the fit indices of the one-dimensional and 32-item model obtained from EFA in Table 11, $\chi^2/df$, and RMSEA values are not within acceptable limits. SRMR value shows acceptable fit. CFI, NFI, NNFI, and GFI values show a good fit. The fit indices of the one-dimensional and 16-item SOM model show good fit. In addition, the fit indices of the two-dimensional and 26-item SOM model also show good fit. As shown in Table 11, there was no difference of more than .01 between the fit indices of the two models generated through SOM. It can be stated that there is no statistically significant difference between the models. This comment was made because, according to Cheung and Rensvold (2002), a difference of less than .01 between the indices indicates measurement invariance. When the items in the two-dimensional model created by SOM were examined, it was seen that there were items consisting of positive expressions in one dimension and items consisting of negative expressions in the other dimension.

**Table 11.** *Fit indices of different models.*

| | EFA | SOM | | RSM | |
|---|---|---|---|---|---|
| Indices | 1 D. 32 item | 1 D. 16 item | 2 D. 26 item | 1 D. 25 item | 1 D. 32 item |
| $\chi^2/df$ | 7.133 | 1.493 | 1.493 | 6.668 | 6.627 |
| CFI | .977 | .999 | .999 | .981 | .978 |
| NNFI | .975 | .998 | .999 | .979 | .977 |
| NFI | .973 | .996 | .995 | .978 | .975 |
| RMSEA | .087 | .025 | .019 | .084 | .084 |
| GFI | .977 | .997 | .996 | .982 | .979 |
| SRMR | .093 | .040 | .039 | .088 | .090 |

The one-dimensional 25-item model obtained with the RSM shows that the $\chi^2/df$ value is not within acceptable limits, although the RMSEA and SRMR values are within the limit. Additionally, the CFI, NFI, NNFI, and GFI values indicate a good fit. The one-dimensional 32-item model obtained with the RSM shows that the $\chi^2/df$ value is not within acceptable limits, while the RMSEA and SRMR values are within the limit. Additionally, the other indices

indicate a good fit. When comparing the fit indices of the two models obtained by RSM, it is observed that there is no difference greater than 0.01 in all values except for $\chi^2/df$. However, both models' $\chi^2/df$ values are outside the acceptable range.

According to Table 11, there is a difference of more than 0.01 between the fit indices of both the one-dimensional and two-dimensional models obtained through SOM and the fit indices of the one-dimensional model obtained through EFA. SOM achieved higher fit values than EFA with the additional 16 items it removed, in addition to the 6 items eliminated according to EFA, while maintaining the same number of dimensions. This may indicate that SOM is more sensitive to item selection than EFA. In cases where application time is important, SOM may be more useful to measure with the same sensitivity with a shorter form. In addition, it is seen that some fit indices of the one-dimensional model obtained from SOM are more than .01 different from some of the fit indices of one-dimensional models obtained by RSM. In this context, the model obtained from SOM indicates a significantly better fit than the models obtained with RSM. When comparing the fit indices of the one-dimensional 32-item models obtained from RSM and EFA, we can see that all the fit statistics are almost the same except for $\chi^2/df$. Out of the 6 items removed according to RSM and EFA, only one of them shows a difference. The RSM shows good fit values with one dimension of 25 items.

## 4. DISCUSSION and CONCLUSION

The fit indices of the CFA applied to the different forms were examined. According to the results, $\chi^2/df$ in the fit indices of the CFA applied to the form obtained by EFA, as well as indices other than SRMR, indicate a good fit. Additionally, all the fit indices obtained from CFA applied to two different forms obtained by SOM indicate good fit. Similarly, the fit indices obtained from the CFA applied to two forms obtained with RSM also indicate a good fit.

The high fit indices of the scale construct obtained from SOM indicate that SOM effectively distributes the dimensional items during test development. This finding is consistent with the results of Tezbaşaran (2016) and Eriş Hasırcı (2019), suggesting that SOM can serve as an alternative to EFA. In his study aiming to determine the structure of multiple-choice tests, Sadesky (2007) also stated that SOM could be effectively used to represent test data. Additionally, Stambuk *et al*. (2007), Astel *et al*. (2007), Chattopadhyay *et al*. (2011), and Das *et al*. (2016) made similar comments, highlighting SOM's suitability for large dataset analysis and result visualization compared to cluster analysis and principal component analysis. Sorrosal Forradellasa *et al*. (2012) used SOM to identify the most relevant items in the data and concluded that the methodology they employed in the study was appropriate.

One of the remarkable results here is that although the data did not strictly meet the assumption of normal distribution in the analysis process, the fit index values of the models belonging to SOM are significantly higher than the fit index values of the models obtained from RSM and EFA when estimations in CFA are made using maximum likelihood instead of diagonally weighted least squares. In this context, it can be interpreted that SOM is more effective in explaining skewed data and creating a better model from these data. This aligns with Francis's (2001) findings suggesting that ANN models are suitable for dimension reduction in datasets with nonlinear relationships, and with the observations of Kiang and Kumar (2001), indicating that SOM effectively detects the structure of non-normally distributed data. In addition, when the final scales were examined on an item basis, it was seen that SOM created models with higher CFA fit indices with the items it deleted in addition to the items that RSM and EFA deleted from the test. Accordingly, with this result, it can be said that SOM performs a more sensitive item selection process.

When examining the CFA fit indices, it becomes evident that RSM, and consequently, IRT models, can be effectively employed in the process of test development. Since the infit and outfit values of the items obtained from the RSM exceed 1.4, it is observed that only one item differs when comparing the items removed from the test and those derived from the EFA results.

Çelen (2008) further suggests that IRT can be viewed as an alternative approach to attaining similar outcomes in the test development process. Among the techniques utilized in this study, RSM offers more detailed statistics on both item and individual levels compared to EFA and SOM and model-data inconsistencies also offer detailed insights into relevant items and individuals. Similar to the findings of Karlin and Karlin (2018), the Rasch model not only validates the test but also enhances student evaluation, providing more accurate results than raw scores. Krishnan and Idris (2018) also observed that the Rasch model effectively determines psychometric properties by identifying and correcting items that compromise validity, rather than re-evaluating all test items. This approach, as demonstrated by Cantó-Cerdán *et al*. (2021), Sözer and Kahraman (2021), Peixoto *et al*. (2021), and Takasaki *et al*. (2021), proves beneficial in scale development and adaptation.

When the findings obtained throughout the study are generally evaluated, it can be stated that SOM and RSM are quite useful techniques in revealing the internal structure of the test. Considering the items of the scale used in the research, it has been observed that items with positive and negative expressions tend to go in two separate dimensions. However, when the theoretical foundations of this structure are considered and items are limited to a single dimension, it has been observed that the model belonging to SOM gives higher fit indices than the one-dimensional model obtained from EFA. In this case, studies can be conducted to examine the effect of deleting more items according to SOM on the test content of the scale. Additionally, studies can be conducted to determine whether the tendency of items with positive and negative expressions to go in two separate dimensions is due to the measurement of different sub-factors of the items or is related to the grammatical structure used in expressing the items. A real data set was used in this study. A simulation study can be conducted to test the performance of SOM in case of different levels of deviation from normality in the used data.

## Acknowledgments

This research is part of thesis “Examination of the Construct Validity of Likert Type Scales with Data Mining, Artificial Neural Networks, Factor Analysis and Rasch Model”.

## Limitations

This scale was applied online due to the COVID-19 Pandemic. For this reason, the sample's level of representation of university students and graduates may be weak.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Hacettepe University Ethics Committee, 35853172-300.

## Contribution of Authors

**Sinan M. Bekmezci**: Research Design, Methodology, Data Collection, Data Analysis, and Writing. **Nuri Doğan**: Research Design, Methodology, Data Collection, Supervision, Critical Review.

## Orcid

Sinan M. Bekmezci https://orcid.org/0000-0001-5190-1894
Nuri Doğan https://orcid.org/0000-0001-6274-2016

## REFERENCES

AERA. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.

Anastasi, A. (1976). *Psychological testing*. The Macmillan Company, Collier Macmillan Limited.

Andrich, D. (1978a). Application of a psychometric model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2*(4), 581-594.

Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Astel, A., Tsakovski, S., Barbieri, P., & Simeonov, V. (2007). Comparison of self-organizing maps classification approach with cluster and principal components analysis for large environmental data sets. *Water Research, 41*, 4566-4578. https://doi:10.1016/j.watres.2007.06.030

Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). Routledge.

Bramer, M. (2020). *Principles of data mining*. Springer.

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. The Guildford Press.

Büyüköztürk, Ş., Çakmak, E.K., Akgün, Ö.E., Karadeniz, Ş., & Demirel, F. (2017). *Bilimsel araştırma yöntemleri [Scientific research methods].* Pegem Akademi.

Cantó-Cerdán, M., Cacho-Martínez, P., Lara-Lacárcel, F., & García-Muñoz, A. (2021). Rasch analysis for development and reduction of symptom questionnaire for visual dysfunctions (SQVD). *Scientific Reports, 11*(14855), 1-10. https://doi.org/10.1038/s41598-021-94166-9

Chattopadhyay, M.J., Dan, P.K., & Majumdar, S. (2011). Principal Component Analysis and Self Organizing Map for Visual Clustering of Machine-Part Cell Formation in Cellular Manufacturing System. https://arxiv.org/pdf/1201.5524.pdf

Crocker, L.M., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.

Çelen, Ü. (2008). Klasik test kuramı ve madde tepki kuramı yöntemleriyle geliştirilen iki testin geçerlilik ve güvenilirliğinin karşılaştırılması. [Comparison of validity and reliability of two tests developed by classical test theory and item response theory]. *Elementary Education Online, 7*(3), 758-768.

Çevik, M., & Tabaru-Örnek, G. (2020). Comparison of MATLAB and SPSS software in the prediction of academic achievement with artificial neural networks: Modeling for elementary school students. *International Online Journal of Education and Teaching, 7*(4), 1689-1707.

Çokluk, Ö., Şekercioğlu G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları. [SPSS and LISREL applications of multivariate statistics for social sciences].* Pegem Akademi.

Das, G., Chattopadhyay, M., & Gupta, S. (2016). A comparison of self-organizing maps and principal components analysis. *International Journal of Market Research, 58*(6). 1-20. https://doi.org/10.2501/IJMR-2016-039

De Ayala, R.J. (2009). *The theory and practice of item response theory*. Guilford Press.

Doğan, N., & Başokçu, O. (2010). İstatistik Tutum Ölçeği İçin Uygulanan Faktör Analizi ve Aşamalı Kümeleme Analizi Sonuçlarının Karşılaştırılması. [Comparison of factor analysis and progressive cluster analysis results applied to the statistics attitude scale]. *Journal of Measurement and Evaluation in Education and Psychology, 1*(2), 65-71.

Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of educational measurement*. Englewood Cliffs, NJ: Prentice- Hall International.

Eğrioğlu, E., Yolcu, U., & Baş, E. (2019). *Yapay sinir ağları öngörü ve tahmin uygulamaları* [*Artificial neural networks prediction and prediction applications*]. Nobel yayınları.

Embretson, S.E., & Reise, S. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.

Eriş Hasırcı, H.M. (2019). *Öz düzenlemeli haritalar yöntemi ile elde edilen yapı geçerliği kanıtlarının faktör analizi ve kümeleme analizi ile karşılaştırılması. [Comparison of*

*construct validity proofs obtained from self-organizing maps method with the factor analysis and clustering analysis]* [Unpublished doctoral dissertation]. Ankara University.

Erkuş, A. (2012). *Psikolojide ölçme ve ölçek geliştirme-1: Temel kavramlar ve işlemler. (Measurement and Scale Development in Psychology- Part 1: Basic Concepts and Processes)*. Pegem Akademi.

Fava, J.L., & Velicer, W.F. (1992). The effects of over extraction on factor and component analysis. *Multivariate Behavioral Research, 27*, 387-415.

Ferles, C., Papanikolaou, Y., Savaidis, S.P., & Mitilineos, S.A. (2021). Deep learning self-organizing map of convolutional layers. *Computer Science and Information Technology (CS & IT) Computer Science Conference*, 25-32. https://doi.org/10.5121/csit.2021.110303

Flora, D.B., & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Francis, L. (2001). Neural Network Demystified. *2001 Winter Forum Ratemaking Discussion Papers and Data Management/Data Quality/Data Technology Call Papers*. www.cacact.org./pubs/forum/01wforum/01wf253.pdf.

Galvan, D., Cremasco, H., Gomes Mantovani, A.C., Bona, E., Killner, M., & Borsato, D. (2020). Kinetic study of the transesterification reaction by artificial neural networks and parametric particle swarm optimization. *Fuel 267*, *117221*. https://doi.org/10.1016/j.fuel.2020.117221

Galvan, D., Effting, L., Cremasco, H., & Conte-Junior, C.A. (2020). Can socioeconomic, health, and safety data explain the spread of COVID-19 outbreak on Brazilian Federative Units? *International Journal of Environmental Research and Public Health*, *17*(8921), 2-16. https://doi.org/10.3390/ijerph17238921

Galvan, D., Effting, L., Cremasco H., & Conte-Junior, C.A. (2021). The spread of the Covid-19 outbreak in Brazil: An overview by Kohonen self-organizing map networks. *Medicina*, *57*(*235)*, 2-19. https://doi.org/10.3390/medicina57030235

Güler, N., İlhan, M., & Taşdelen Teker, G. (2018). Comparing the scale values obtained from pairwise comparison scaling method and Rasch analysis. *Inonu University Journal of the Faculty of Education*, *19*(1), 31-48. https://doi.org/10.17679/inuefd.400386

Grajciarova, L., Mares, J., Dvorak P., & Prochazka, A. (2012). Biomedical image analysis using self-organizing maps. *Technicka*, *5*, 166.

Grove, S.K., Burns, N., & Gray, J.R. (2012). *The practice of nursing research: Appraisal, synthesis, and generation of evidence* (7th Ed.). Elsevier.

Haykin, S. (2009). *Neural networks and learning machines* (3rd Edition). Pearson Education.

İlhan, M., & Güler, N. (2017). Likert tipi ölçeklerde klasik test kuramı ile Rasch analizinden elde edilen yetenek kestirimleri arasındaki uyumun test edilmesi. [Testing the agreement between ability estimations made through classical test theory and the ones made through Rasch analysis in Likert type scales]. *Ege Journal of Education*, *18*(1). 244-265.

İlhan, M., & Güler, N. (2018). Likert tipi ölçeklerde Rasch modelinin kullanımı: olumsuz değerlendirilme korkusu ölçeği-öğrenci formu (ODKÖ-ÖF) üzerinde bir uygulama. [The use of Rasch model in Likert type scales: An application on the fear of negative evaluation scale-student form (FNE-SF)]. *Trakya Journal of Education*, *8*(4), 756-775.

Karlin, O., & Karlin S. (2018). Making better tests with the Rasch measurement model. *InSight: A Journal of Scholarly Teaching*, *13*, 76-100.

Kiang, M.Y., & Kumar, A. (2001). An evaluation of self-organizing map networks as a robust alternative to factor analysis in data mining applications. *Information Systems Research*, *12*(2), 177-194.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological*

*Cybernetics*, *43*, 59–69.
Kohonen, T. (2014). *MATLAB Implementations and Applications of the Self-Organizing Map*. Unigrafia Oy.
Kozan, S. (2016). *Madde sınıflamada faktör analizi, çok boyutlu ölçekleme ve kümeleme analizlerinin karşılaştırılması [Comparison of factor analysis, multidimensional scaling and cluster analysis in item classification]* [Unpublished master's thesis]. Mersin University.
Krishnan, S., & Idris, N. (2018). Using partial credit model to improve the quality of an instrument. *International Journal of Evaluation and Research in Education*, *7*(4), 313-316.
Li, C-H. (2016). Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*, 936–949. https://doi.org/10.3758/s13428-015-0619-7
Linacre, J.M. (2012). *Many-facet Rasch measurement: Facets tutorial*. http://www.winsteps.com/a/ftutorial2.pdf
Linacre, J.M. (2014). *A user's guide to FACETS Rasch-model computer programs*. http://www.winsteps.com/a/facets-manual.pdf
Miljković, D. (2017). Brief review of self-organizing maps. *MIPRO/CTS*, 1252-1257
Murtagh, F., & Hernandez-Pajares, M. (1995). The Kohonen self-organizing map method: An assessment. *Journal of Classification*, *12*(2), 165-190
Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd. ed.). McGraw-Hill.
Ostini, R., & Nering, M.L. (2006). *Polytomous item response theory models* (3rd. Ed). Sage Publications.
Peixoto, E.M., Zanini, D.S., & Andrade J.M. (2021). Cross-cultural adaptation and psychometric properties of the Kessler distress scale (K10): an application of the rating scale model. *Psicologia: Reflexão e Crítica*, *34*(21), 2-10. https://doi.org/10.1186/s41155-021-00186-9
Sadesky, G.S. (2007). *Determining the structure in test performance* [Unpublished doctoral dissertation]. Alberta University.
Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, *8*(2), 23-74.
Sorrosal Forradellasa, M.T., Barberà-Mariné, M.G., Fernández Bariviera, A., & Garbajosa-Cabello, M.J. (2012). Advantages of using self-organizing maps to analyse student evaluations of teaching. *Fuzzy Economic Review*. https://doi.org/10.25102/fer.2012.01.03
Sözer, E., & Kahraman, N. (2021). Investigation of psychometric properties of Likert items with the same response categories using polytomous item response theory models. *Journal of Measurement and Evaluation in Education and Psychology*, *12*(2), 129-146. https://doi.org/10.21031/epod.819927
Stambuk, A., Stambuk, N., & Konjevoda P. (2007). Application of Kohonen self- organizing maps (SOM) based clustering for the assessment of religious motivation. 29th International Conference on Information Technology Interfaces, Croatia.
Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlik [Reliability and validity in social and behavioral measurements] (1st ed.)*. Seçkin Yayınları.
Şimşek, D. (2006). *Kümeleme analizi, çok boyutlu ölçekleme, doğrulayıcı ve açıklayıcı faktör analizi ile elde edilen yapı geçerliği kanıtlarının karşılaştırılması. [A comprasion of construct validity evidences obtained through cluster analysis, multidimensional scaling, confirmatory, and exploratory factor analysis]* [Unpublished master's thesis]. Hacettepe University.
Takasaki, H., Kawazoe, S., Miki, T., Chiba H., & Godfrey, E. (2021). Development and validity assessment of a Japanese version of the exercise adherence rating scale in participants with

musculoskeletal disorders. *Health and Quality Life Outcomes*, *19(169)*, 2-8. https://doi.org/10.1186/s12955-021-01804-x.

Tezbaşaran, E. (2016*). Temel bileşenler analizi ve yapay sinir ağı modellerinin ölçek geliştirme sürecinde kullanılabilirliğinin incelenmesi. [An investigation on usability of principal component analysis and artificial neural network models in the process of scale development]* [Unpublished doctoral dissertation]. Mersin University.

Uysal, M. (2015). *Araştırma özyeterlik ölçeğinin psikometrik özelliklerinin klasik test kuramı ve madde tepki kuramına göre incelenmesi. [An investigation of psychometric properties of research self-efficacy scale according to classical test theory and item response theory]* [Unpublished master's thesis] Gazi University.

Wood, J.M., Tataryn, D.J., & Gorsuch, R.L. (1996). Effects of under- and over extraction on principal axis factor analysis with varimax rotation. *Psyhological Methods, 1*(4), 354-365.

## APPENDIX

**Appendix A.** Descriptive Statistics for the Statistical Attitude Scale.

| Item | Mean | Std. Deviation | Skewness | S.E. of Skewness | Kurtosis | S.E. of Kurtosis | Item | Mean | Std. Deviation | Skewness | S.E. of Skewness | Kurtosis | S.E. of Kurtosis |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2.49 | 1.18 | .30 | .086 | -.81 | .172 | 21 | 3.15 | 1.38 | -.18 | .086 | -1.18 | .172 |
| 2 | 2.35 | 1.20 | .50 | .086 | -.75 | .172 | 22 | 2.34 | 1.17 | .46 | .086 | -.76 | .172 |
| 3 | 4.16 | 1.03 | -1.13 | .086 | .73 | .172 | 23 | 2.12 | 1.11 | .70 | .086 | -.35 | .172 |
| 4 | 3.62 | 1.28 | -.62 | .086 | -.67 | .172 | 24 | 2.73 | 1.22 | .13 | .086 | -.96 | .172 |
| 5 | 2.85 | 1.28 | .04 | .086 | -1.07 | .172 | 25 | 2.36 | 1.22 | .51 | .086 | -.75 | .172 |
| 6 | 2.20 | 1.15 | .58 | .086 | -.64 | .172 | 26 | 2.13 | 1.14 | .73 | .086 | -.36 | .172 |
| 7 | 3.88 | 1.25 | -.94 | .086 | -.16 | .172 | 27 | 3.44 | 1.32 | -.46 | .086 | -.88 | .172 |
| 8 | 3.17 | 1.33 | -.22 | .086 | -1.08 | .172 | 28 | 3.35 | 1.28 | -.37 | .086 | -.91 | .172 |
| 9 | 3.37 | 1.33 | -.38 | .086 | -1.00 | .172 | 29 | 3.26 | 1.37 | -.29 | .086 | -1.11 | .172 |
| 10 | 3.12 | 1.24 | -.17 | .086 | -.94 | .172 | 30 | 2.55 | 1.18 | .027 | .086 | -.86 | .172 |
| 11 | 2.23 | 1.13 | .57 | .086 | -.58 | .172 | 31 | 3.59 | 1.37 | -.60 | .086 | -.90 | .172 |
| 12 | 3.10 | 1.32 | -.15 | .086 | -1.06 | .172 | 32 | 3.51 | 1.35 | -.54 | .086 | -.87 | .172 |
| 13 | 3.70 | 1.33 | -.74 | .086 | -.64 | .172 | 33 | 3.80 | 1.21 | -.80 | .086 | -.21 | .172 |
| 14 | 3.47 | 1.33 | -.53 | .086 | -.85 | .172 | 34 | 3.84 | 1.16 | -.89 | .086 | .09 | .172 |
| 15 | 3.44 | 1.34 | -.43 | .086 | -.97 | .172 | 35 | 3.63 | 1.27 | -.65 | .086 | -.61 | .172 |
| 16 | 3.53 | 1.22 | -.40 | .086 | -.80 | .172 | 36 | 3.30 | 1.21 | -.29 | .086 | -.77 | .172 |
| 17 | 2.51 | 1.12 | .32 | .086 | -.72 | .172 | 37 | 2.25 | 1.12 | .54 | .086 | -.60 | .172 |
| 18 | 2.92 | 1.24 | .00 | .086 | -1.00 | .172 | 38 | 3.91 | 1.07 | -.94 | .086 | .39 | .172 |
| 19 | 3.26 | 1.42 | -.33 | .086 | -1.16 | .172 | 39 | 2.18 | 1.16 | .64 | .086 | -.56 | .172 |
| 20 | 3.28 | 1.30 | -.39 | .086 | -.91 | .172 | 40 | 3.50 | 1.28 | -.52 | .086 | -.76 | .172 |

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 198–221
https://doi.org/10.21449/ijate.1472350
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# The relationship between teachers' organizational commitment and career anxiety

**Tuğba Alagöz**[1*], **Yusuf İnandı**[2], **Yüksel Gündüz**[3]

[1]Ministry of National Education, Niğde, Türkiye
[2]Mersin University, Faculty of Education, Department of Educational Sciences, Mersin, Türkiye
[3]Ondokuz Mayıs University, Faculty of Education, Department of Educational Sciences, Samsun, Türkiye

**ARTICLE HISTORY**
*Received: Apr. 23, 2024*
*Accepted: Dec. 24, 2024*

**Keywords:**
Career anxiety,
Teachers,
Organizational commitment.

**Abstract:** In this study, it was aimed to examine the relationship between teachers' organizational commitment and career anxiety. A cross-sectional correlational design belonging to the relational survey model was used. A simple random sampling method was utilized to determine the sample. The sample consisted of 384 teachers. Personal Information Form, Teachers' Career Anxiety Scale, and Organizational Commitment Scale were applied to collect the data. In data analysis, descriptive statistics, Pearson correlation and regression analysis were used. In the study, teachers' perceptions of career anxiety were found to be high, and their perceptions of organizational commitment were found to be moderate. A negative and lower significant relationship was found between teachers' career anxiety and organizational commitment. It was determined that teachers' career anxiety scores could be significantly predicted by their organizational commitment responses. It was concluded that in order to increase the organizational commitment of teachers, it is necessary to realize their career anxiety reasons and levels. Suggestions were made to prevent the factors that hinder organizational commitment and to decrease the career anxiety of teachers.

## 1. INTRODUCTION

Throughout history, as humanity has taken significant steps towards modernization, societal structures have been influenced by the developments and changes brought about by this process of modernization, giving rise to the need for the reconstruction of society. The foundation of this reconstruction lies in education. To achieve the desired goals in education, effective communication among all stakeholders is crucial. One of the most important stakeholders in this process is teachers. The motivation and well-being of teachers can significantly impact the effectiveness of the education system. However, teachers face various challenges, particularly concerning their career paths and organizational commitment. These challenges include uncertainty in professional advancement, job security anxieties, and a perceived lack of support. These factors can increase teachers' career anxiety and decrease their commitment to their institutions, leading to reduced job satisfaction and productivity. Understanding the complex relationship between teachers' career anxieties and organizational commitment is crucial for

*CONTACT: Tuğba ALAGÖZ ✉ tugbaalagoz87@gmail.com Ministry of National Education, Niğde, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

educators, administrators, and policymakers. By addressing the challenges faced by teachers, educational institutions can foster a supportive environment that promotes strong organizational commitment among teachers, ultimately enhancing the overall quality of education.

Teachers, who directly impact every aspect of education, play an active role in the reconstruction of society (Yaman *et al*., 2022). Teachers, who hold a crucial role, see the students they educate as a reflection of themselves. This situation underscores the need to enhance the quality of the teaching profession for the qualitative development of the education system in the reconstruction of society. Teachers, entrusted with the critical mission of shaping the future through education (Jacobson-Lundeberg, 2016), need to be open to development, skillful, and capable of easily adapting to innovations in order to successfully fulfill their missions. The acquisition of the knowledge and skills that teachers need can only be achieved through their professional development. The professional development of teachers, specialization in their fields, or gaining different perspectives is possible through transitioning between various career stages. It is expected that the perceptions, ideas, concerns, anxieties, and expectations of teachers will undergo changes at each different career stage. Therefore, it is useful to explain how the concept of a career is defined and what it means for teachers.

In the 1970s, the concept of career, which gained popularity and began to be studied, is expressed as 'Carrierre' in French and 'Career' in English. According to Klatt, Murdick, and Schuster (1985, p.380), career implies progress and is defined as upward mobility in the organizational or professional hierarchy. Similarly, Bayraktaroğlu (2006, p.137) describes a career as an individual's progression through skill and experience acquisition over the course of their professional life.

In today's context, the concept of a career, defined as a way for individuals to achieve self-actualization, the highest level in Maslow's hierarchy of needs (Çelenk, 2022), can be interpreted as encompassing aspects such as "from the early stages of their career and learning journey, individuals initially aim to learn, but over time, they also aim to acquire a profession, and this goal shapes their education and time" (Sarsıkoğlu, 2019). Therefore, the career cycle of teachers is seen as a process that begins with choosing the teaching profession. A review of the literature reveals that the majority of studies on career anxiety are conducted with university students and adults at the beginning of their career development (Hardin *et al*., 2006; Pisarik *et al*., 2017; Şahin *et al*., 2019). However, considering the ever-changing employment policies, sociological, and economic conditions, it is likely that individuals who continue their career journey also experience career anxiety. Particularly in educational organizations, teachers with diverse needs and characteristics experience different anxieties at each stage of their careers (Aydın, 2018).

The teaching profession has always stood out for its profound impact on the lives of students. Consequently, teachers actively monitor all developments in the changing world to provide the highest benefit to their students. However, in this process of monitoring, teachers often face numerous challenges and pressures. Scovel (1978) stated that anxiety can be defined in two different ways: "debilitating anxiety," which hinders the learning process and impedes individual success, and "facilitating anxiety," which leads individuals to achieve success beyond their natural capabilities. While facilitating anxiety motivates individuals to learn and encourages active participation in the learning process, debilitating anxiety causes excessive distress and self-doubt (Scovel, 1978). Indeed, it is possible to define anxiety as a driving force that motivates individuals to take action. Although anxiety generally has a negative connotation, as long as it remains at a healthy level, it can, contrary to popular belief, act as a driving force that leads individuals towards success. Fuller (1969) evaluates this situation from a developmental perspective and argues that occupational anxiety should be experienced and resolved in order to become a better teacher. In this context, facilitating anxiety can be considered an element that increases teacher motivation.

Nevertheless, among these challenges, career anxiety emerges as a critical factor among teachers, becoming a genuine and widespread sources of concern in today's educational environment. Career anxiety manifests itself among teachers due to various factors, including education reforms, standardized tests, classroom management, the challenges of meeting constantly evolving curricula, and addressing different student needs (Engin & İpek, 2020; İnandı & Gılıç, 2020; Karatepe *et al*., 2021). Collectively, these factors create a storm of stress and anxiety. This anxiety significantly impacts the well-being of teachers, job satisfaction, and their overall commitment to educational institutions. Indeed, when teachers' commitment to the school weakens, it negatively affects the degree to which the school achieves its goals and becomes an obstacle to its success (Celep, 2000).

In recent years, a rich body of literature has established the role of teachers in improving student outcomes (Akkurt & Karabağ-Köse, 2019; Lipesa, 2018; McLean & Connor, 2015; Sammons, Toth & Sylva, 2017; Wenner & Campbell, 2017). There is also ample evidence in the literature that school administrators' impact on students is mediated through teachers (Çağlar & Çınar, 2021; Franklin, 2016; Yakışıklı *et al*., 2023). In this context, it can be argued that one of the ways school principals influence teachers is through their organizational commitment. Therefore, it is observed that highly motivated teachers also demonstrate high organizational commitment (Ertürk & Aydın, 2016; Memişoğlu & Kaya, 2017).

In the field of education, organizational commitment is a critical component of a teacher's overall professional experience. Allen and Meyer (1996) define organizational commitment as the degree to which employees adopt the goals and norms of an organization, expressed as the emotional attachment felt by employees towards the organization and their desire to continue existing in the organization. Dağlı *et al*. (2018) view organizational commitment as a factor that increases employees' participation in the organization, commitment to organizational values and goals, and organizational effectiveness. Thus, for teachers, organizational commitment signifies the loyalty, dedication, and devotion they show towards their schools.

This research comprehensively examines the organizational commitment of teachers, considering three dimensions: emotional, normative, and continuance commitment (Meyer *et al*., 1991). Each dimension contributes to the general connection a teacher has with the school. For instance, emotional commitment reflects a teacher's emotional commitment to the organization, which can be equated to an emotional commitment to the school. A teacher with strong emotional commitment will continue to take on tasks within or outside their job description as long as their values align with those of the organization (Demirel, 2009). Factors such as job satisfaction, organizational justice, and organizational citizenship have positive effects on emotional commitment, while stress-related factors like anxiety and burnout have negative effects on the emotional connection teachers feel towards their schools.

On the other hand, normative commitment is defined as the ethical or moral obligation felt by an employee to stay in the organization (Meyer & Allen, 1991). As education professionals, teachers are expected to feel a high degree of normative commitment, viewing staying in school and fulfilling their duties as a necessity. However, if teachers perceive their dedication as unrewarded, their normative commitment may decrease.

The third commitment dimension explored in this study is continuance commitment, which sees staying in the organization as a conscious awareness related to costs (Örücü & Kışlalıoğlu, 2014). In other words, continuance commitment prevents the loss of investments, such as time and effort, made by an employee in the organization if they were to leave. Additionally, the incentives and rewards provided by the organization can enhance an employee's commitment to the organization. However, the employee would need to sacrifice all these rewards if they were to leave (Jaros *et al*., 1993). For teachers, this situation may lead to fluctuating continuance commitment as they weigh the personal and professional costs of working in an environment that could cause them concern against the benefits of alternative job opportunities elsewhere.

All three dimensions come together to shape a teacher's commitment to the institution. Insufficient organizational commitment of a teacher may not only affect their perceptions of the teaching profession and the communication structure within the school (İnandı & Gılıç, 2020; Meriç *et al*., 2019) but also lead to a decrease in teacher effectiveness, subsequently impacting the productivity of the school. The productivity of a school, in turn, is demonstrated through the societal contribution of the students it educates.

Given that teachers play a crucial role in shaping the future of society, the recognition of teaching as a special profession is not coincidental. However, teachers cannot always demonstrate high motivation while practicing their profession. In the modern educational environment, teachers frequently encounter complexities and challenges that can affect their well-being and career trajectories. In other words, career issues have become a widespread problem among teachers, manifesting in various ways and significantly impacting their commitment to the educational institutions they serve. This research aims to examine the complex relationship between teachers' career anxiety and organizational commitment. It is considered important to shed light on the factors that influence teachers' career anxiety and organizational commitment, both for teachers and their educational institutions, and to highlight the potential consequences. Understanding the interaction between career anxiety and organizational commitment is crucial for education leaders, administrators, and policymakers. This understanding can facilitate the development of strategies to reduce career anxiety and improve teacher support systems. Ultimately, it can create an environment where the likelihood of educators' success is higher, leading to an increase in the quality of education provided to students.

Research Questions:

1. To what extent do teachers perceive their career anxiety?
2. Is there a significant difference in teachers' opinions on career anxiety concerning gender, marital status, age, and educational status?
3. To what extent do teachers perceive their organizational commitment?
4. Is there a significant difference in teachers' opinions on organizational commitment concerning gender, marital status, age, and educational status?
5. Is there a significant relationship between teachers' opinions on career anxiety and organizational commitment?
6. Do teachers' opinions on career anxiety significantly predict their organizational commitment?

## 2. METHOD

### 2.1. Research Model

This study employs a cross-sectional correlational survey model as it aims to reveal the relationship between the career anxiety and organizational commitment levels of teachers working in public schools in Niğde province. Cross-sectional survey studies, with large-scale and representative sampling, allow for taking snapshots of different samples, i.e., simultaneous analysis, enable comparison of different groups, and are both retrospective and prospective (Cohen *et al*., 2007). In this study, the data collection tools used addressed the career anxieties and organizational commitment experienced by teachers in terms of different variables and relationships based on data collected only once and made inferences about the future based on their current situation. For these reasons, a cross-sectional correlational survey research design was preferred in this study. Cross-sectional surveys are preferred in educational research due to their quick implementation, ability to show an integrated pattern, and high participation rate (Rose & Sullivan, 1993, p.184-188).

## 2.2. Population and Sample

The study's population consists of 4,663 teachers serving in public schools in Niğde province during the 2023-2024 academic year. The sample of the study comprises 384 teachers selected through simple random sampling. Simple random sampling ensures that each unit has an equal chance of being included in the sample, and the selection of one does not affect the chances of others being chosen (Koçak & Arun, 2006), making it a preferred method for sample selection. The calculation of the sample size was based on the random sampling size table recommended by Cohen *et al*. (2021), determining that a sample size of 357 units is sufficient for an universe size of 5000, with a 95% confidence level and a 5% margin of error in homogeneous conditions. Thus, it can be stated that the sample size collected for the research is adequate. Demographic characteristics of the research group are presented in Table 1.

**Table 1.** *Demographic characteristics of the sample.*

| Variable | Group | *n* | % |
|---|---|---|---|
| Gender | Female | 272 | 70.8 |
| | Male | 112 | 29.2 |
| Marital Status | Married | 303 | 78.9 |
| | Single | 81 | 21.1 |
| Age | 20-29 | 30 | 7.8 |
| | 30-39 | 168 | 43.8 |
| | 40-49 | 144 | 37.5 |
| | 50 and above | 42 | 10.9 |
| Education Level | Bachelor's Degree | 257 | 66.9 |
| | Master's Degree | 115 | 29.9 |
| | Doctorate | 12 | 3.1 |

*Note. n* =384

As presented in Table 1, 272 (70.8%) of the participants are female and 112 (29.2%) are male. In terms of marital status, 303 (78.9%) participants were married and 81 (21.1%) were single. In terms of age variable, 30 (7.8%) of the participants are between the ages of 20-29, 168 (43.8%) are between the ages of 30-39, 144 (37.5%) are between the ages of 40-49 and 42 (10.9%) are 50 years and over. In terms of education level, there are 257 (66.9%) participants with bachelor's degrees, 115 (29.9%) participants with master's degrees and 12 (3.1%) participants with doctorate degrees.

## 2.3. Data Collection Tools

Within the scope of the research, Personal Information Form, Teachers' Career Anxiety Scale and Organizational Commitment Scale were used as data collection tools.

### 2.3.1. *Personal information form*

The researchers created a personal information form to determine the participants' demographic characteristics, including gender, education level, marital status, and age.

### 2.3.2. *Teacher's Career Anxiety Scale*

To determine teachers' career anxiety, the Teachers' Career Anxiety Scale developed by Alagöz *et al*. (2024) was used. It is a 5-point Likert-type scale, consisting of 18 items categorized into four sub-dimensions: organizational-based, colleague-based, individual-based, and work-life balance. The lowest possible score from the scale is 18, and the highest is 90. Higher scores indicate higher career anxiety. The four sub-dimensions together explain 58.6% of the total variance. The reliability of the scale, measured by Cronbach's Alpha coefficient, is 0.89 for the

entire Teachers' Career Anxiety Scale, and for the sub-dimensions, it is 0.81 for organizational-based, 0.87 for colleague-based, 0.78 for individual-based, and 0.66 for work-life balance. For this study, reliability analysis yielded Cronbach's alpha values of 0.87 for the entire scale, 0.82 for the first dimension, 0.86 for the second dimension, 0.76 for the third dimension, and 0.61 for the fourth dimension. As it is not stated in the original form of scale, Table 2 provides the necessary information for evaluating the total scores.

**Table 2.** *Interpretation of scores for the teachers' career anxiety scale.*

| Dimension | Number of Items | Very Low | Low | Moderate | High | Very High |
|---|---|---|---|---|---|---|
| Organizational-based | 7 | 7-12.5 | 12.6-18.1 | 18.2-23.7 | 23.8-29.3 | 29.4-35 |
| Colleague-based | 4 | 4-7.1 | 7.2-10.3 | 10.4-13.5 | 13.6-16.7 | 16.8-20 |
| Individual-based | 4 | 4-7.1 | 7.2-10.3 | 10.4-13.5 | 13.6-16.7 | 16.8-20 |
| Work-Life Balance-based | 3 | 3-5.3 | 5.4-7.7 | 7.8-10.1 | 10.2-12.5 | 12.6-15 |
| Teachers' Career Anxiety Scale | 18 | 18-32.3 | 32.4-46.7 | 46.8-61.1 | 61.2-75.5 | 75.6-90 |

*Note.* The Teachers' Career Anxiety Scale had four sub-scales (Organizational-based with 7 items, Colleague-based with 4 items, Individual-based with 4 items, Work-Life Balance-based with 3 items).

These score ranges help interpret the level of career anxiety for each dimension and the overall scale. Participants can be categorized into different levels based on their total scores and scores on each sub-dimension of the Teachers' Career Anxiety Scale.

### 2.3.3. *Organizational Commitment Scale*

In this research, the Organizational Commitment Scale, developed by Meyer *et al*. (1993), and adapted into Turkish by Dağlı *et al*. (2017), was employed. The scale encompasses three dimensions: Emotional Commitment, Normative Commitment, and Continuance Commitment, comprising a total of 18 items. It explains approximately 53% of the total variance.

Upon analyzing the reliability of the scale through Cronbach Alpha values, the following results were obtained: .80 for the first dimension, .73 for the second dimension, .80 for the third dimension, and .88 for the overall scale. For this study, the calculated Cronbach Alpha coefficient was .89 for the entire scale, .82 for the first dimension, .72 for the second dimension, and .81 for the third dimension. As it is not stated in the original form of scale, the necessary information for the evaluation of total scores is detailed in Table 3. The values provided represent the ranges for the evaluation of total scores in each dimension of the Organizational Commitment Scale.

**Table 3.** *Interpretation of scores for the organizational commitment scale.*

| Dimension | Number of Items | Very Low | Low | Moderate | High | Very High |
|---|---|---|---|---|---|---|
| Emotional Commitment | 6 | 6-10.7 | 10.8-15.5 | 15.6-20.3 | 20.4-25.1 | 25.2-30 |
| Normative Commitment | 6 | 6-10.7 | 10.8-15.5 | 15.6-20.3 | 20.4-25.1 | 25.2-30 |
| Continuance Commitment | 6 | 6-10.7 | 10.8-15.5 | 15.6-20.3 | 20.4-25.1 | 25.2-30 |
| Organizational Commitment Scale | 18 | 18-32.3 | 32.4-46.7 | 46.8-61.1 | 61.2-75.5 | 75.6-90 |

*Note.* The Organizational Commitment Scale had three sub-scales (Emotional Commitment with 6 items, Normative Commitment with 6 items, Continuance Commitment with 6 items).

## 2.4. Data Collection

Permissions were obtained from the Mersin University Ethics Committee and the Niğde Provincial Directorate of National Education to conduct the research. In the data collection

process, a personal information form and scale items were transferred to an online form to reach teachers working in districts and villages. The online form was configured to be answered only once, with no unanswered items. The link to the online form was distributed to 4663 teachers in 384 schools in Niğde through school principals. Of these teachers, 384 filled out the online form. This process took approximately 15 days during the spring semester of the 2023-2024 academic year.

### 2.5. Data Analysis

SPSS 24 software was used for the analysis of research data. Initially, analyses were conducted to detect missing data and outliers. While no missing data was found in the data to be analyzed, Mahalanobis outlier distances were examined for outlier detection, and it was observed that the data set did not contain outliers. Skewness-Kurtosis values were examined to assess the normal distribution of the 384-unit data set. The results of normality distribution tests are presented in Table 4 below.

**Table 4.** *The skewness and kurtosis values of the variables.*

| Scales | Skewness | Kurtosis |
|---|---|---|
| Organizational-based | -1.039 | 1.374 |
| Colleague-based | 0.317 | -0.930 |
| Individual-based | 0.403 | -0.382 |
| Work-Life Balance-based | -0.137 | -0.194 |
| Career Anxiety Scale | -0.100 | 0.375 |
| Emotional Commitment | -0.629 | 0.545 |
| Continuance Commitment | -0.171 | -0.071 |
| Normative Commitment | 0.141 | -0.137 |
| Organizational Commitment | -0.260 | 0.025 |

According to Table 4, the Skewness-Kurtosis values of the variables fall within the range of +1.5 to -1.5 (Tabachnick & Fidell, 2013), assuming normal distribution of the data, the analysis continued. Therefore, t-tests were conducted for gender and marital status variables, one-way analysis of variance (ANOVA) for education level, age, and years of service at the school, Pearson correlation analysis for determining the relationship between teachers' career anxiety and organizational commitment, and multiple linear regression analysis to determine if teachers' career anxiety predicts organizational commitment significantly. Cohen's *d* is a measure of effect size that quantifies the difference between two means in terms of standard deviations. It provides an indication of how large or small the difference between groups is (Cohen, 1988): $0.2 \leq |d| < 0.5$: Small effect; $0.5 \leq |d| < 0.8$: Medium effect; $|d| \geq 0.8$: Large effect. When calculating the effect size for correlation statistics, Kotrlik and Williams (2003) state that the correlation coefficient itself is an effect size value. In the interpretation of effect sizes for correlation coefficients (*r*), Hinkle, *et al*. (1979) emphasize that the relationship between 0-.30 is very low, between .30-.50 is low, between .50-.70 is medium, between .70-.90 is high and between 0.90-1.00 is very high. For effect size calculations in regression analyses, the $R^2$ value, which is the multiple regression coefficient, can be examined (Özsoy & Özsoy, 2013). According to Cohen (1988), these effect size results ($R^2$): 0.0196 can be interpreted as small; 0.1300 as medium; 0.2600 as large effect size. In this study, in tables that represent the analysis the largest effect size has been interpreted.

## 3. RESULTS

The findings obtained in the study are presented under this heading. The results of the descriptive statistics, independent samples of t-test, one-way of ANOVA, Pearson correlation

and multiple linear regression analysis were presented in order. For all analyses the results related to teachers' career anxiety were given first and the results related to teachers' organizational commitment were given later.

First of all, descriptive statistical analysis was conducted to determine teachers' perceptions of career anxiety and organizational commitment. In descriptive statistics, minimum score, maximum score, total score, standard deviation and level information are calculated. The Findings of the descriptive statistical analysis are presented in Table 5.

**Table 5.** *Findings of descriptive statistical analysis.*

| Scales | *n* | *Min.* | *Max.* | *M* | *SD* | Level |
|---|---|---|---|---|---|---|
| Organizational-based | 384 | 7 | 35 | 28.31 | 4.94 | Moderate |
| Colleague-based | 384 | 4 | 20 | 11.97 | 4.23 | Moderate |
| Individual-based | 384 | 4 | 20 | 11.77 | 3.52 | Moderate |
| Work-Life Balance-based | 384 | 3 | 15 | 9.91 | 2.55 | Moderate |
| Career Anxiety Scale | 384 | 24 | 90 | 61.97 | 11.47 | High |
| Emotional Commitment | 384 | 6 | 30 | 20.73 | 5.10 | High |
| Normative Commitment | 384 | 6 | 30 | 16.45 | 5.03 | Moderate |
| Continuance Commitment | 384 | 6 | 30 | 18.96 | 4.76 | Moderate |
| Organizational Commitment | 384 | 19 | 83 | 52.46 | 11.93 | Moderate |

*Note*. Interpretation of scores stated as levels in Table 5 were presented in Table 2 for the teachers' career anxiety scale and in Table 3 for the organizational commitment scale.

The findings of the descriptive statistical analysis for the variables are presented in Table 5. The data set consists of 384 units, and the minimum (Min.) and maximum (Max.) values, mean (Mean), standard deviation (*SD*), and assessment level (Level) are reported for each variable. The assessment level indicates the categorization based on the mean score, representing the level of each variable (e.g., High, Moderate).

According to Table 5, the lowest score obtained for the organizational-based dimension is 7, and the highest score is 35. Teachers' average score for the organizational-based dimension is $M = 28.31$, with a standard deviation of 4.94. Therefore, it is determined that teachers strongly agree with the statements in the organizational-based dimension. For the colleague-based dimension, the lowest score is 4, and the highest score is 20. Teachers' average score for the dimension is $M = 11.97$, with a standard deviation of 4.23. Thus, teachers moderately agree with the statements in the colleague-based dimension. The individual-based dimension has a minimum score of 4 and a maximum score of 20. The average score for the dimension is $M = 11.77$, with a standard deviation of 3.52. Accordingly, teachers moderately agree with the statements in the individual-based dimension. The total score for the Teacher Career Anxiety Scale ranges from 24 to 90. The average score for the entire scale is $M = 61.97$, with a standard deviation of 11.47. Hence, teachers exhibit a high level of agreement with statements related to career anxiety. In light of these findings, it can be said that teachers experience career anxiety, and organizational career anxiety is relatively more prominent than other types of anxiety. When examining teachers' views on organizational commitment in Table 5, the emotional commitment dimension has a minimum score of 6 and a maximum score of 30. The average score for the dimension is $M = 20.73$, with a standard deviation of 5.10. Thus, teachers strongly agree with the statements in the emotional commitment dimension. For the normative commitment dimension, the lowest score is 6, and the highest score is 30. The average score for the dimension is $M = 16.45$, with a standard deviation of 5.03. Therefore, teachers' participation in statements in the normative commitment dimension is at a moderate level. In the continuance commitment dimension, the lowest score is 6, and the highest score is 30. The average score for the dimension is $M = 18.96$, with a standard deviation of 4.76. Hence,

teachers' level of participation in statements in the continuance commitment dimension is at a moderate level. for the Organizational Commitment Scale total, the lowest score is 19, and the highest score is 83. The average score for the entire scale is $M = 52.46$, with a standard deviation of 11.93. Accordingly, teachers exhibit a moderate level of agreement with statements related to organizational commitment.

**Table 6.** *t-test results for gender variable.*

| Scales | Gender | *n* | *M* | *SD* | *df* | *t* | *p* | *Cohen's d* |
|---|---|---|---|---|---|---|---|---|
| Organizational-based | Female | 272 | 28.81 | 4.92 | 382 | 3.144 | .002* | 0.37 |
| | Male | 112 | 27.08 | 4.08 | | | | |
| Colleague-based | Female | 272 | 11.96 | 4.25 | 382 | -0.077 | .938 | |
| | Male | 112 | 12.00 | 4.21 | | | | |
| Individual-based | Female | 272 | 11.62 | 3.48 | 382 | -1.320 | .188 | |
| | Male | 112 | 12.14 | 3.60 | | | | |
| Work-Life Balance-based | Female | 272 | 10.12 | 2.45 | 382 | 2.465 | .014* | 0.28 |
| | Male | 112 | 9.41 | 2.72 | | | | |
| Career Anxiety Scale | Female | 272 | 62.5 | 11.43 | 382 | 1.454 | .147 | |
| | Male | 112 | 60.65 | 11.51 | | | | |
| Emotional Commitment | Female | 272 | 20.54 | 5.11 | 382 | -1.139 | .255 | |
| | Male | 112 | 21.19 | 5.05 | | | | |
| Normative Commitment | Female | 272 | 16.21 | 4.97 | 382 | -1.474 | .141 | |
| | Male | 112 | 17.04 | 5.14 | | | | |
| Continuance Commitment | Female | 272 | 18.81 | 4.76 | 382 | -0.928 | .354 | |
| | Male | 112 | 19.31 | 4.76 | | | | |
| Organizational Commitment Scale | Female | 272 | 51.86 | 11.99 | 382 | -1.520 | .129 | |
| | Male | 112 | 53.90 | 11.71 | | | | |

*$p < .05$

To determine whether teachers' career anxiety and organizational commitment perceptions meaningfully differ according to gender and marital status variables, we conducted an independent samples *t*-test. Table 6 presents the t-test results for gender variable. As presented in Table 6, as a result of the independent sample t-test on whether teachers' career anxiety views differ according to gender; it is seen that there is a significant difference in favor of female teachers for the sub-dimensions of the scale in terms of organization ($t$ (382)=- 3.144, $p$<.05) and work-life balance ($t$ (256)= 2.465, $p$<.05). The largest effect sizes are observed for organizational-based ($d$ = 0.37) and work-life balance-based ($d$ = 0.28), both of which are small but meaningful. These suggest that females show slightly higher levels in these areas compared to males. There is no significant difference between the whole and other sub-dimensions of the Teachers' Career Anxiety Scale. According to the results in Table 6, teachers' views on organizational commitment do not differ according to gender variable in terms of both the whole scale and sub-dimensions.

As presented in Table 7, as a result of the independent sample t-test on whether teachers' career anxiety views differ according to marital status; the sub-dimensions of the scale are organizational-based ($t$(382)= 2.244, $p$<.05), colleague ($t$(382)=2.890, $p$<.05), work-life balance ($t$(382)=3.802, $p$<.05) and the whole scale ($t$(382)=2.907, $p$<.05) significantly differed in favor of married teachers.

The largest effect sizes are observed for work-life balance-based ($d$ =0.48), colleague-based ($d$=0.36), career anxiety scale ($d$=36) and organizational-based ($d$ = 0.28), all of which are small but meaningful. These suggest that married participants show slightly higher career anxiety levels in these areas compared to single participants. Variables such as emotional commitment,

normative commitment, continuance commitment, and the overall organizational commitment scale exhibit negligible differences between married and single participants. In other words, the analysis revealed no statistically significant differences in the individual-based sub-dimension of the scale according to marital status.

**Table 7.** *t-test results for marital status variable.*

| Scales | Marital Status | *n* | *M* | *SD* | *df* | *t* | *p* | *Cohen's d* |
|---|---|---|---|---|---|---|---|---|
| Organizational-based | Married | 303 | 28.60 | 4.61 | 382 | 2.244 | .025* | 0.28 |
| | Single | 81 | 27.22 | 5.93 | | | | |
| Colleague-based | Married | 303 | 12.29 | 4.24 | 382 | 2.890 | .004* | 0.36 |
| | Single | 81 | 10.77 | 4.00 | | | | |
| Individual-based | Married | 303 | 11.78 | 3.55 | 382 | 0.094 | .925 | |
| | Single | 81 | 11.74 | 3.43 | | | | |
| Work-Life Balance-based | Married | 303 | 10.16 | 2.49 | 382 | 3.802 | .000* | 0.48 |
| | Single | 81 | 8.97 | 2.54 | | | | |
| Career Anxiety Scale | Married | 303 | 62.84 | 11.33 | 382 | 2.907 | .004* | 0.36 |
| | Single | 81 | 58.71 | 11.47 | | | | |
| Emotional Commitment | Married | 303 | 20.70 | 5.08 | 382 | -0.208 | .835 | |
| | Single | 81 | 20.83 | 5.19 | | | | |
| Normative Commitment | Married | 303 | 16.27 | 4.96 | 382 | -1.396 | .164 | |
| | Single | 81 | 17.14 | 5.24 | | | | |
| Continuance Commitment | Married | 303 | 18.94 | 4.86 | 382 | -0.109 | .913 | |
| | Single | 81 | 19.01 | 4.40 | | | | |
| Organizational Commitment Scale | Married | 303 | 52.26 | 11.98 | 382 | -0.625 | .533 | |
| | Single | 81 | 53.19 | 11.82 | | | | |

*$p < .05$

The results presented in Table 7 indicate that the marital status variable does not exert a notable influence on teachers' perceptions of organizational commitment, as observed across the entire scale and its constituent sub-dimensions.

To determine whether teachers' career anxiety and organizational commitment perceptions meaningfully differ according to age and educational variables, we conducted one-way ANOVA. The ANOVA results for age variable is presented in Table 8. As depicted in Table 8, as a result of a one-way analysis of variance on whether teachers' views on career anxiety differ significantly according to age variable; no significant difference was found for the whole scale, organization-based, colleagues-based and work-life balance-based sub-dimensions, while a significant difference was found for individual-based ($F_{(3;\ 383)}$= 4.682, $p$<.05). It is seen that this difference is in favor of teachers between the ages of 30-39 (B), 40-49 (C) and 50 and over (D) for teachers between the ages of 30-39 (B).

Again according to Table 8, when the teachers' views on organizational commitment are analyzed in terms of age variable, it is seen that in the emotional commitment sub-dimension ($F_{(3;\ 383)}$= 4.125, $p$<.05) there is a significant difference between teachers aged 30-39 (B) and teachers aged 50 and above (D) in favor of teachers aged 50 and above (D). While there is a significant difference for normative commitment sub-dimension ($F_{(3;\ 383)}$= 3.070, $p$<.05) according to the age variable, there is a significant difference for continuance commitment ($F_{(3;\ 383)}$= 5.302, $p$<.05) and the whole Organizational Commitment Scale ($F_{(3;\ 383)}$= 5.236, $p$<.05) between teachers aged 30-39 (B) and those aged 40-49 (C) in favor of those aged 40-49 (C) and between teachers aged 30-39 (B) and those aged 50 and over (D) in favor of those aged 50 and over (D).

**Table 8.** *ANOVA results for the age variable.*

| Scales | Age | *n* | *M* | *SD* | *df* | *F* | *p* | *Cohen's d* | Significant difference |
|---|---|---|---|---|---|---|---|---|---|
| Organizational-based | 20-29 | 30 | 27.80 | 5.99 | 3<br>3803<br>83 | 2.359 | .071 | | |
| | 30-39 | 168 | 29.05 | 4.49 | | | | | |
| | 40-49 | 144 | 27.80 | 5.21 | | | | | |
| | 50 and above | 42 | 27.42 | 4.70 | | | | | |
| Colleague-based | 20-29 | 30 | 11.96 | 4.18 | 3<br>380<br>383 | .748 | .524 | | |
| | 30-39 | 168 | 12.11 | 4.32 | | | | | |
| | 40-49 | 144 | 11.61 | 4.25 | | | | | |
| | 50 and above | 42 | 12.64 | 3.83 | | | | | |
| Individual-based | 20-29 | 30 | 12.36 | 3.38 | 3<br>380<br>383 | 4.682 | .003* | 0.50 | B>C<br>B>D |
| | 30-39 | 168 | 12.40 | 3.73 | | | | | |
| | 40-49 | 144 | 11.22 | 3.18 | | | | | |
| | 50 and above | 42 | 10.69 | 3.36 | | | | | |
| Work-Life Balance-based | 20-29 | 30 | 9.06 | 2.63 | 3<br>380<br>383 | 2.208 | .087 | | |
| | 30-39 | 168 | 10.17 | 2.51 | | | | | |
| | 40-49 | 144 | 9.93 | 2.59 | | | | | |
| | 50 and above | 42 | 9.42 | 2.33 | | | | | |
| Career Anxiety Scale | 20-29 | 30 | 61.20 | 12.39 | 3<br>380<br>383 | 2.451 | .063 | | |
| | 30-39 | 168 | 63.75 | 11.08 | | | | | |
| | 40-49 | 144 | 60.59 | 11.73 | | | | | |
| | 50 and above | 42 | 60.19 | 10.81 | | | | | |
| Emotional Commitment | 20-29 | 30 | 21.56 | 4.24 | 3<br>380<br>383 | 4.125 | .007* | -0.38 | D>B |
| | 30-39 | 168 | 19.71 | 5.66 | | | | | |
| | 40-49 | 144 | 21.44 | 4.46 | | | | | |
| | 50 and above | 42 | 21.78 | 4.76 | | | | | |
| Normative Commitment | 20-29 | 30 | 18.03 | 5.64 | 3<br>380<br>383 | 3.070 | .028* | 0.44 | A>D |
| | 30-39 | 168 | 15.76 | 5.01 | | | | | |
| | 40-49 | 144 | 16.54 | 4.63 | | | | | |
| | 50 and above | 42 | 17.78 | 5.55 | | | | | |
| Continuance Commitment | 20-29 | 30 | 20,16 | 4.29 | 3<br>380<br>383 | 5.302 | .001* | 0.47 | A>B<br>D>B<br>C>B |
| | 30-39 | 168 | 17.91 | 4.82 | | | | | |
| | 40-49 | 144 | 19.54 | 4.52 | | | | | |
| | 50 and above | 42 | 20.28 | 4.92 | | | | | |
| Organizational Commitment Scale | 20-29 | 30 | 55.80 | 10.83 | 3<br>380<br>383 | 5.236 | .001* | 0.48 | A>B<br>D>B<br>C>B |
| | 30-39 | 168 | 49.90 | 12.55 | | | | | |
| | 40-49 | 144 | 53.74 | 10.83 | | | | | |
| | 50 and above | 42 | 55.90 | 12.02 | | | | | |

*$p < .05$. The groups of the participants according to age variable were abbreviated from A to D to represent the significant difference among groups. A= 20-29 age; B= 30-39 age; C= 40-49 age; D= more than 50.

Also, Cohen's *d* effect size calculations for the scales and age groups reveal several interesting trends. The individual-based scale showed more variation, with small-to-moderate differences, the largest being between the 20-29 and 50+ groups ($d$ = 0.50). The work-life balance-based scale revealed negative effects, with the largest difference ($d$ = -0.44) between the 20-29 and 30-39 groups, suggesting younger individuals perceive worse balance. For the career anxiety scale, small positive effects were noted, the most significant being between the 30-39 and 50+ groups ($d$=0.32). In emotional commitment, moderate effects appeared, such as a negative effect ($d$ =-0.38) between the 30-39 and 50+ groups. The normative commitment scale showed the largest effect ($d$=0.44) between the 20-29 and 30-39 groups, while for continuance commitment, the most substantial difference ($d$ = 0.47) was also between these groups. Lastly, the organizational commitment scale had a small-to-moderate effect ($d$ = 0.48) between the 20-29 and 30-39 groups. Overall, age-related differences were more pronounced in scales like individual-based and normative commitment, highlighting nuanced shifts in perceptions across age groups.

According to Table 9, when teachers' views on career anxiety are examined in terms of educational level, there is a significant difference between bachelor's (A) and master's (B) graduates in favor of master's graduates (B); between bachelor's (A) and doctorate (C) graduates in favor of doctorate graduates (C); between master's (B) and doctorate (C) graduates in favor of doctorate graduates in the individual-based dimension of the scale ($F_{(2;\ 383)}$= 12.126, $p$<.05). In the whole Career Anxiety Scale of Teachers ($F_{(2;\ 383)}$= 4.148, $p$<.05), there is a significant difference between bachelor (A) and doctorate (C) graduates in favor of doctorate graduates (C).

As depicted in Table 9 in terms of teachers' organizational commitment, for the emotional commitment sub-dimension ($F_{(2;\ 383)}$= 10.769, $p$<.05), there is a significant difference between bachelor's (A) and doctorate (C) graduates in favor of bachelor's (A) graduates, and between master's (B) and doctorate (C) graduates in favor of master's (B) graduates. In the normative commitment dimension of organizational commitment ($F_{(2;\ 383)}$= 8.669, $p$<.05), there is a significant difference between bachelor's (A) and doctorate (C) graduates in favor of bachelor's (A) graduates.

For the third sub-dimension of the organizational commitment scale, continuance commitment ($F_{(2;\ 383)}$= 4.723, $p$<.05), there is a significant difference between bachelor's (A) and doctorate (C) graduates in favor of bachelor's (A) graduates. In the whole Organizational Commitment Scale ($F_{(2;\ 383)}$= 10.228, $p$<.05), it is concluded that there is a significant difference between bachelor's (A) and doctorate (C) graduates in favor of bachelor's (A) graduates.

Moreover, Cohen's *d* calculations reveal significant differences in several scales between education levels (BA, MA, and PhD). PhD holders scored substantially higher in individual-based ($d$ = -1.31) and work-life balance ($d$ = -0.68) scales, indicating stronger individual-based perceptions and better work-life balance compared to BA holders. However, BA holders showed much greater emotional ($d$ = 1.36), normative ($d$ = 0.99), continuance ($d$ = 0.78), and organizational commitment ($d$ = 1.21), reflecting their stronger connections and obligations to their organizations. Career anxiety was also notably higher for PhD holders compared to BA holders ($d$ = -0.85). Comparisons between MA and PhD holders generally showed moderate differences, while BA vs. PhD comparisons exhibited large effect sizes, highlighting significant contrasts in perceptions and commitments across education levels. These findings underscore the nuanced relationship between education level and various organizational and personal scales.

**Table 9.** *ANOVA results for the variable of education level.*

| Scales | Education level | *n* | *M* | *SD* | *df* | *F* | *p* | *Cohen's d* | Significant difference |
|---|---|---|---|---|---|---|---|---|---|
| Organizational-based | BA (A) | 257 | 28.10 | 4.86 | 2<br>381<br>383 | 2.318 | .100 | | |
| | MA (B) | 115 | 28.48 | 5.22 | | | | | |
| | PhD (C) | 12 | 31.16 | 2.91 | | | | | |
| Colleague-based | BA (A) | 257 | 12.13 | 4.09 | 2<br>381<br>383 | 0.758 | .469 | | |
| | MA (B) | 115 | 11.57 | 4.51 | | | | | |
| | PhD (C) | 12 | 12.41 | 4.46 | | | | | |
| Individual - based | BA (A) | 257 | 11.29 | 3.28 | 2<br>381<br>383 | 12.126 | .000* | -1.31<br>-0.85 | C>A<br>C>B |
| | MA (B) | 115 | 12.44 | 3.75 | | | | | |
| | PhD (C) | 12 | 15.58 | 3.08 | | | | | |
| Work-Life Balance-based | BA (A) | 257 | 9.735 | 2.49 | 2<br>381<br>383 | 3.305 | .038* | -0.68 | C>A |
| | MA (B) | 115 | 10.16 | 2.66 | | | | | |
| | PhD ( C) | 12 | 11.41 | 2.108 | | | | | |
| Career Anxiety Scale | BA (A) | 257 | 61.26 | 11.03 | 2<br>381<br>383 | 4.148 | .017* | -0.85 | C>A |
| | MA (B) | 115 | 62.66 | 12.19 | | | | | |
| | PhD (C) | 12 | 70.58 | 10.66 | | | | | |
| Emotional Commitment | BA (A) | 257 | 21.25 | 4.59 | 2<br>381<br>383 | 10.769 | .000* | 1.36<br>0.96 | A>C<br>B>C |
| | MA (B) | 115 | 20.19 | 5.38 | | | | | |
| | PhD (C) | 12 | 14.75 | 8.19 | | | | | |
| Normative Commitment | BA (A) | 257 | 17.08 | 4.82 | 2<br>381<br>383 | 8.669 | .000* | 0.99 | A>C |
| | MA (B) | 115 | 15.48 | 5.05 | | | | | |
| | PhD (C) | 12 | 12.25 | 6.09 | | | | | |
| Continuance Commitment | BA (A) | 257 | 19.38 | 4.54 | 2<br>381<br>383 | 4.723 | .009* | 0.78 | A>C |
| | MA (B) | 115 | 18.33 | 5.07 | | | | | |
| | PhD (C) | 12 | 15.83 | 4.72 | | | | | |
| Organizational Commitment | BA (A) | 257 | 53.93 | 10.94 | 2<br>381<br>383 | 10.228 | .000* | 1.21 | A>C |
| | MA (B) | 115 | 50.41 | 12.61 | | | | | |
| | PhD (C) | 12 | 40.41 | 16.42 | | | | | |

*Note.* $^{*}p < 0.05$. The groups of the participants according to education level variable were abbreviated as from A to C to represent the significant difference among groups. A= BA refers to Bachelor of Arts; B= MA refers to Master of Arts; C= PhD refers ro Doctor of Philosophy degrees.

Correlation analysis was performed to determine whether there was a significant relationship between teachers' perceptions of career anxiety and organizational commitment. The results of the correlation analysis are presented in Table 10. As presented in Table 10, there is a negative and low level significant relationship between teachers' views on career anxiety and affective commitment ($r$=-.204, $p$<.01); normative commitment ($r$=-.161, $p$<.01); continuance commitment ($r$=-.131, $p$<.01) and organizational commitment ($r$=-.197, $p$<0.01). There is a negative and low level significant relationship between the organizational-based dimension of the Career Anxiety Scale and affective commitment ($r$=-.171, $p$<.01), normative commitment ($r$=-.185, $p$<.01) and organizational commitment ($r$=-.181, $p$<.01). There is a medium level negative relationship between the individual-based sub-dimension of the Career Anxiety Scale and affective commitment ($r$=-.309, $p$<.01), a low level negative relationship between normative commitment ($r$=-.243, $p$<.01) and continuance commitment ($r$=-.225, $p$<.01). It is concluded that there is a significant negative relationship between the individual-based sub-dimension and the whole Organizational Commitment Scale ($r$=-.301, $p$<.01). Furthermore, in the light of the explanations made in the data analysis section, the correlation coefficient values were interpreted with reference to Hinkle *et al*. (1979) to represent the effect size for correlation statistics.

**Table 10.** *Correlation analysis on the relationship between career anxiety and organizational commitment of teachers.*

| | OB | CB | IB | WLBB | CAS | Emotional | Normative | Continuance | OCS | *M* | *SD* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| OB | 1 | | | | | | | | | 28.31 | 4.94 |
| CB | .451** | 1 | | | | | | | | 11.97 | 4.23 |
| IB | .358** | .309** | 1 | | | | | | | 11.77 | 3.52 |
| WLBB | .495** | .479** | .384** | 1 | | | | | | 9.91 | 2.55 |
| **CAS** | .818** | .765** | .661** | .731** | 1 | | | | | 61.97 | 11.47 |
| Emotional | -.171** | -.080 | -.308** | -.025 | -.204** | 1 | | | | 20.73 | 5.10 |
| Normative | -.185** | .009 | -.243** | -.046 | -.161** | .600** | 1 | | | 16.45 | 5.03 |
| Continuance | -.097 | .005 | -.225** | -.100 | -.131* | .496** | .629** | 1 | | 18.96 | 4.76 |
| **OCS** | -.181** | -.027 | -.301** | -.073 | -.197** | .801** | .887** | .845** | 1 | 52.46 | 11.93 |

*Note.* *$p$<.05, **$p$<.01. OB=Organizational-based; CB=Colleague-based; IB=Individual-based; WLBB=Work-life Balance-based; CAS=Career Anxiety Scale; OCS=Organizational Commitment Scale

In order to determine whether teachers' opinions on career anxiety significantly predict their organizational commitment levels, multiple linear regression was utilized. Also, for the effect size calculations in regression analyses, the $R^2$ value is examined. Table 11 shows the results regarding the prediction of organizational commitment by teachers' career anxiety. According to Table 11, teachers' views on career anxiety significantly predicted their organizational commitment ($R$=.335, $R^2$=.112, $p$<.05). Teachers' views on career anxiety explained approximately 11% of the variance of their views on organizational commitment. The order of importance of the predictor variables is individual ($\beta$=-.306), organization ($\beta$=-.153), colleagues ($\beta$=.102), work-life balance ($\beta$=.071). However, only the organizational ($t$=-2.600, $p$<.05) and individual ($t$=-5.670, $p$<.05) dimensions of teachers' career anxiety are significant predictors of organizational commitment.

**Table 11.** *Multiple linear regression analysis results regarding the prediction of organizational commitment by teachers' career anxiety.*

| Scale | Variable | *B* | *SE* | *β* | *t* | *p* |
|---|---|---|---|---|---|---|
| Organizational Commitment | Constant | 68.361 | 3.491 | | 19.580 | .000 |
| | Organizational-based | -0.370 | 0.142 | -.153 | -2.600 | .010* |
| | Colleague-based | 0.288 | 0.163 | .102 | 1.771 | .077 |
| | Individual-based | -1.035 | 0.183 | -.306 | -5.670 | .000* |
| | Work-life balance-based | 0.334 | 0.282 | .071 | 1.181 | .238 |
| *R*= 335 | $R^2$= .112 | *F*(4,379)= 11.940 | *p*=.000 | | | |

*$p<0.05$

According to Table 12, teachers' views on career anxiety significantly predicted their views on the affective commitment dimension of organizational commitment (*R*=.345, $R^2$ =.119, *p*<.05). Teachers' views on career anxiety explained approximately 12% of the variance related to affective commitment. The order of importance of the predictor variables is individual (*β*=-.306), work-life balance (*β*=.167), organization (*β*= -.139) and colleagues (*β*=.002). In addition, among the dimensions of teachers' career anxiety, the organizational -based dimension (*t*=-2.365, *p*<.05), individual-based dimension (*t*=-6.029, *p*<.05) and work-life balance-based dimension (*t*=-2.776, *p*<.05) are significant predictors of emotional commitment.

**Table 12.** *Multiple linear regression analysis results regarding the prediction of emotional commitment dimension of organizational commitment by teachers' career anxiety.*

| Scale | Variable | *B* | *SE* | *β* | *t* | *p* |
|---|---|---|---|---|---|---|
| Emotional commitment | Constant | 26.966 | 1.487 | | 18.140 | .000 |
| | Organizational-based | -0.143 | 0.061 | -.139 | -2.365 | .019* |
| | Colleague-based | 0.003 | 0.069 | .002 | 0.038 | .970 |
| | Individual-based | -0.469 | 0.078 | -.324 | -6.029 | .000* |
| | Work-life balance-based | 0.334 | 0.120 | .167 | 2.776 | .006* |
| *R*= 345 | $R^2$= .119 | *F*(4,379)= 12.820 | *p*=.000 | | | |

*$p<0.05$

According to Table 13, teachers' views on career anxiety are a significant predictor of their views on the normative commitment dimension of organizational commitment (*R*=.306, $R^2$ =.093, *p*<.05). Teachers' views on career anxiety explain approximately 10% of the variance of normative commitment. The order of importance of the predictor variables is individual (*β*= -.244), organization (*β*= -.198), colleagues (*β*=.135) and work-life balance (*β*=.081). In addition, among the dimensions of teachers' career anxiety, organizational -based dimension (*t*=-3.328, *p*<.05), colleague-based dimension (*t*=2.310, *p*<.05) and individual-based dimension (*t*=-4.491, *p*<.05) are significant predictors of normative commitment.

**Table 13.** *Multiple linear regression analysis results regarding the prediction of normative commitment dimension of organizational commitment by teachers' career anxiety.*

| Scale | Variable | *B* | *SE* | *β* | *t* | *p* |
|---|---|---|---|---|---|---|
| Normative commitment | Constant | 22.768 | 1.487 | | 15,311 | .000 |
| | Organizational-based | -0.202 | 0.061 | -.198 | -3.328 | .001* |
| | Colleague-based | 0.160 | 0.069 | .135 | 2.310 | .021* |
| | Individual-based | -0.349 | 0.078 | -.244 | -4.491 | .000* |
| | Work-life balance-based | 0.160 | 0.120 | .081 | 1.333 | .183 |
| *R*= 306 | $R^2$= .093 | *F*(4,379)= 9.758 | *p*=.000 | | | |

*$p<.05$

According to Table 14, teachers' views on career anxiety are a significant predictor of their views on the continuance commitment dimension of organizational commitment ($R$=.246, $R^2$ =.061, $p$<.05). Teachers' views on career anxiety explain approximately 6% of the variance related to continuance commitment. The order of importance of the predictor variables is individual-based ($\beta$= -.227), colleague-based ($\beta$=.118), work-life balance-based ($\beta$=-.047) and organizational -based ($\beta$= -.046). In addition, among the dimensions of teachers' career anxiety, colleague-based dimension ($t$=1.986, $p$<.05) and individual-based dimension ($t$=-4.098, $p$<.05) are significant predictors of continuance commitment.

**Table 14.** *Multiple linear regression analysis results regarding the prediction of continuance commitment dimension of organizational commitment by teachers' career anxiety.*

| Scale | Variable | *B* | *SE* | *β* | *t* | *p* |
|---|---|---|---|---|---|---|
| Continuance commitment | Constant | 23.092 | 1.432 | | 16.120 | .000 |
| | Organizational-based | -0.044 | 0.058 | -.046 | -0.751 | .453 |
| | Colleague-based | 0.133 | 0.067 | .118 | 1.986 | .048* |
| | Individual-based | -0.307 | 0.075 | -0227 | -4.098 | .000* |
| | Work-life balance-based | -0.087 | 0.116 | -.047 | -0.751 | .453 |
| $R$= 246 $R^2$= .061 $F$(4,379)= 6.127 $p$=.000 | | | | | | |

*$p$<.05

## 4. DISCUSSION and CONCLUSION

The findings of this study indicate that teachers experience career anxiety, particularly as a result of organizational factors such as limited opportunities for professional development, obstacles to academic engagement, and uncertainty created by administrators (Amil, 2015; İnandı, 2009; Köstek, 2007; Swanson & Tokar, 1991). This is consistent with existing research on career barriers, although Gündüz (2010) posited that these anxieties are relatively uncommon. Female teachers reported higher levels of anxiety related to organizational issues and work-life balance than males, which may be attributed to societal expectations and the fulfilment of multiple roles (İnandı *et al*., 2009; King *et al*., 2013; Yang *et al*., 2009;). Furthermore, married teachers, particularly women, are more likely to experience career anxiety due to the limitations imposed by family responsibilities (Halaçlı, 2018; İnandı & Tunç, 2012; Sükun, 2022).

The age of the teacher also has an impact on levels of anxiety, with those aged between 30 and 39 experiencing the highest levels of anxiety, potentially due to their career stage and the uncertainties that they face about the future. This is at odds with the findings of some literature, which suggests that younger individuals prioritize career development (Marinakou & Giousmpasoglou, 2019), while other studies have found that younger teachers experience less stress (Akyüz Ünlü, 2014; Yang *et al*., 2009). A correlation has been identified between higher education levels and increased anxiety. Doctoral graduates are found to experience greater anxiety than those with a bachelor's degree. This is thought to be due to heightened competition and limited opportunities in academia (Aktan, 2020; Alabaş *et al*., 2012; Başaran, 2020; Köse & İnce, 2022). It is of particular importance to note that this study has identified a negative correlation between career anxiety and organizational commitment. When teachers feel that they are not being adequately supported in pursuing their career aspirations, their commitment to the organization in question is likely to diminish. This is influenced by a number of factors at both the individual and organizational levels (Coyle-Shapiro & Morrow, 2006; Weng *et al*., 2010).

Job stress is another critical factor that interacts with organizational commitment. Studies have shown a negative relationship between organizational commitment and job stress, indicating that higher stress levels can reduce a teacher's commitment to their organization (Taran & Fathi, 2017). This relationship is further complicated by job satisfaction, which can mediate the effects

of job stress on organizational commitment. Teachers who experience job satisfaction despite stress may maintain higher levels of commitment (Wang et.al, 2020). Additionally, factors such as inadequate professional development, low job satisfaction and excessive workload have been identified as contributors to this anxiety and reduced commitment (Güzeller & Çelikel, 2020).

This highlights the necessity for educational institutions to establish supportive environments that recognize the value of teachers, facilitate professional growth, and encourage a healthy work-life balance. In summary, teachers' career anxieties are complex and shaped by a range of factors, including organizational dynamics, gender, marital status, age and educational attainment. The development of comprehensive strategies to cultivate a supportive and fulfilling environment is essential to address these anxieties, with the ultimate objective of benefiting both teachers and the quality of education.

The findings of this study indicate that teachers generally exhibit high levels of affective commitment to their organizations, but only moderate levels of normative and continuance commitment. This indicates that although teachers exhibit a strong sense of belonging and identification with their schools, they may lack awareness of the potential consequences of leaving their position and a corresponding sense of obligation to remain in their role. This finding is consistent with the results of other studies in the field (Aksanaklu & İnandı, 2018; Canpolat, 2011; Demir & İnandı, 2022; Doğan & Çelik, 2019; Hamidi, 2023; Özyavuz, 2021; Sezgin, 2010).

Furthermore, the study revealed that there were no significant differences in organizational commitment based on gender or marital status. However, age was identified as a significant factor, with teachers in the 30-39 age group exhibiting lower levels of commitment than their older counterparts. This may be attributed to the fact that older teachers exhibit a stronger sense of loyalty to their institution, potentially due to their longer tenure, greater integration with the organizational culture, and increased likelihood of occupying leadership roles. Additionally, they may be more focused on retirement planning and subsequent career stages, which may make them less likely to leave their current position (Cengiz, 2001; Çakır, 2001; Memişoğlu & Kaya, 2017; Önder, 2023).

It is noteworthy that teachers with a Bachelor's degree exhibited higher levels of organizational commitment than those with a postgraduate degree. This is consistent with previous research indicating a negative correlation between educational level and organizational commitment (Çakır, 2001; Dağlı & Gençdal, 2018; Gürbüz, 2012; Steers, 1977). This may be attributed to the fact that teachers with higher levels of education tend to have more ambitious career aspirations and expectations that may not be fulfilled by their current organization, which in turn makes them more receptive to alternative job opportunities.

The study revealed a statistically significant negative correlation between teachers' career anxiety and organizational commitment. Individual and organizational career anxieties were identified as robust predictors of this commitment. This finding is consistent with the results of previous studies by Weng *et al*. (2010) and Coyle-Shapiro and Morrow (2006), which emphasize the influence of individual and organizational factors on career development and organizational commitment. In other words, career success, both subjective and objective, is closely linked to organizational commitment. Teachers who perceive themselves as successful in their careers tend to exhibit higher levels of commitment. Factors such as career commitment, professional commitment, and perceived organizational support significantly contribute to teachers' subjective career success, which in turn can enhance their organizational commitment (Ingarianti *et al*., 2022). This suggests that fostering a supportive environment that acknowledges and rewards career success can mitigate career anxiety and bolster commitment. On the other hand, in the event that teachers' expectations from their organization are not met, they may experience career anxiety and a subsequent reduction in their organizational commitment. In other words, teachers' career anxieties have the potential to negatively impact

their organizational commitment, which may in turn give rise to a range of adverse outcomes. Teachers experiencing elevated levels of career anxiety may demonstrate diminished job satisfaction, motivation, fulfilment, and performance, which are all beneficial outcomes of organizational commitment (Soysal & Selvi, 2023). This can result in a reduction in interest and enthusiasm for the teaching profession, which may ultimately lead to a decline in productivity. Furthermore, career anxiety can increase the likelihood of teachers engaging in negative behaviors, including seeking alternative employment, leaving their current roles and absenteeism (Akgül, 2020). When teachers experience feelings of uncertainty, insecurity, or lack of support in relation to their career progression, their commitment to their current roles may diminish, prompting them to consider alternative opportunities. It is therefore incumbent upon educational institutions to prioritize the reduction of teachers' career anxieties and the enhancement of their organizational commitment. It is incumbent upon school administrators to play a pivotal role in mitigating career anxiety and fostering organizational commitment. This may be achieved by providing professional development opportunities, offering guidance on career paths and creating a supportive work environment. This will consequently result in enhanced job satisfaction, motivation, and performance among teachers, which will in turn have a beneficial impact on the quality of education provided to students. Conversely, career anxiety resulting from factors such as inadequate job descriptions, excessive workload, lack of support, or professional uncertainties can lead to a reduction in organizational commitment and an increased willingness to leave the organization if alternative opportunities arise. Such anxieties may impede teachers' ability to integrate and feel a sense of belonging within the organization (Güzeller & Çelikel, 2020; Perveen *et al*., 2023). Furthermore, uncertainty regarding career progression, job descriptions, and evolving educational policies can also contribute to anxiety and erode teachers' commitment. Teachers who feel their anxieties are not adequately addressed or who lack opportunities for professional development may question their loyalty to the organization (Güzeller & Çelikel, 2020). This emphasizes the significance of fostering a supportive school environment that offers transparent career pathways, professional growth opportunities, and a sense of worth for teachers. The fulfilment of these requirements has the potential to alleviate career-related anxiety and cultivate a stronger sense of organizational commitment, which in turn can yield benefits for both teachers and the educational system.

### 4.1. Limitations of the Study

This study is not without limitations, but it offers valuable insights that should inform future research. One limitation of the present study is that a cross-sectional correlational model was employed to examine the relationship between teachers' organizational commitment and career anxiety. With the attributions of different variables advanced, statistical analysis can be conducted. Another limitation of this study is its limited geographical scope, as it was conducted in only one province. By expanding the study to include teachers from various cities, the generalizability of the findings can be ensured.

### 4.2. Implications for Future Research

In order to address the concerns of teachers with regard to their career development and to foster a sense of belonging within their schools, a number of recommendations can be implemented. It is recommended that schools implement structured professional development programmes that enhance teachers' skills and knowledge, thereby facilitating their career advancement. Moreover, it is of the utmost importance to provide guidance and support regarding career paths and opportunities, potentially through tailored career counselling or mentorship programmes. It is of the utmost importance to establish transparent communication channels through which teachers can freely express their concerns and suggestions. It is incumbent upon school administrators to maintain open and regular communication with teachers, providing clear and transparent information regarding career advancement policies, performance evaluations, and decision-making processes.

It is equally important to create a supportive environment, which can be achieved through the establishment of peer support networks, mentoring programmes and the fostering of a positive organizational climate in which teachers feel valued and appreciated. Such measures can be further reinforced through the implementation of policies that support a healthy work-life balance, recognize teacher achievements, and promote a positive school environment. It is also imperative to address the underlying causes of teachers' career-related anxieties, such as the avoidance of overload and the setting of realistic goals. It would be prudent for educational institutions to be mindful of the distinctive challenges faced by female teachers and to consider implementing measures such as flexible work arrangements or childcare provisions. The enhancement of teachers' organizational commitment can be facilitated by their involvement in decision-making processes and the provision of a voice in the shaping of the school's future.

Further research is required to gain a deeper understanding of the factors influencing teachers' career anxieties and organizational commitment. It is further recommended that interventions, such as professional development programmes or support groups, be developed with the specific aim of alleviating career anxieties and boosting organizational commitment. Finally, the implementation of policies that address these anxieties, including flexible work arrangements and childcare provisions, can contribute to the creation of a more supportive and fulfilling work environment for teachers. By taking these steps, it is possible to create a more positive and productive environment for teachers, which will lead to increased job satisfaction, motivation, and ultimately, an improved quality of education for students.

## Acknowledgments

The study has not been funded by any organization or project. Moreover, the study has not been previously presented at a conference or a scholarly meeting.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Mersin University Social Sciences Ethics Committee, 05.03.2024-60.

## Contribution of Authors

**Tuğba Alagöz**: Literature review, Resources, Methodology, Data collection, Data analysis and Writing-original draft. **Yusuf İnandı**: Supervision, Literature review, Writing-review & editing. **Yüksel Gündüz**: Supervision, Writing-review & editing and Validation.

## Orcid

Tuğba Alagöz https://orcid.org/0000-0003-4300-2076
Yusuf İnandı https://orcid.org/0000-0003-2760-0957
Yüksel Gündüz https://orcid.org/0000-0002-4710-8444

## REFERENCES

Akgül, R. (2020). *Association teacher's perceptual mobbing level with organizational socialization and organizational commitment and organizational loneliness levels* [Master thesis, Fırat University]. Council of Higher Educaion National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=4J_FzTwlrMCH4qBROpXPH_uTGiA8mEG6lrYu3PSXaTMqwDy5CGpRufYGvgWFEt51

Akkurt, Z., & Karabağ Köse, E. (2019). Examination of student achievement in terms of school, teacher, and family-related variables. *PAU Journal of Education, 47,* 1-16. https://doi.org/10.9779/pauefd.451853

Aksanaklu, P., & İnandı, Y. (2018). The investigation of the relationship between organizational commitment and burnout levels of school administrators. *Mersin University*

*Journal of the Faculty of Education*, *14*(13), 937-955. https://doi.org/10.17860/mersinefd.469769

Aktan, O. (2020). Evaluation of postgraduate education in terms of teachers' career development. *Higher Education and Science Journal*, *10*(3), 596-607. https://doi.org/10.5961/jhes.2020.419

Akyüz Ünlü, B. (2014). *Kadınların kariyer engelleri ile kariyer tatmini ve yaşam doyumu ilişkisi [The relationship between career obstacles, career satisfaction, and life satisfaction of women]* [Master thesis, Mehmet Akif Ersoy University]. Council of Higher Education National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=gyLHMouPes-CvnhRcjQsKR_8__-81SZ1iuZldFHVri4i0qBsgWkSCfq4m8dC_jFd

Alabaş, R., Kamer, T., & Polat, Ü. (2012). Postgraduate education in teachers' career development: Preference reasons and problems encountered in the process. *E-International Journal of Educational Research, 3*(4), 89-107. https://www.e-ijer.com/tr/pub/issue/8020/105368

Alagöz, T., Fakiroğlu, M., & İnandı, Y. Development of teachers' career anxiety scale. *International Journal of Active Learning*, *8*(1), 21-41. https://doi.org/10.48067/ijal.1405582

Allen, N.J., & Meyer, J.P. (1996). Affective, continuance, and normative commitment to the organization: An examination of construct validity. *Journal of Vocational Behavior, 49*, 252-276. https:// doi.org/10.1006/jvbe.1996.0043

Amil, O. (2015). Examination of the relationship between employees' perceived career obstacles and perceived performance in terms of various variables. *OPUS-Turkey Journal of Social Policy and Labor Studies, 5*(9), 7-26. https://dergipark.org.tr/en/pub/opus/issue/22698/242307

Baştürk, R. (2007) Investigation of test anxiety levels of pre-service teachers taking civil servant selection examination. *Fırat University Journal of Social Science*, *17*(2), 163-176.

Bayraktaroğlu, S. (2006). *İnsan kaynakları yönetimi [Human resource management]*. Gazi Kitabevi.

Bozoğlu, G. (2011). *In the view of organizational culture and organizational commitment a comparative research at educational sector in Yalova* [Master thesis, Yalova University]. Council of Higher Education National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=EEdeQgIdFRxX5NbvVau-Ai_MWReXYGPUm6JrjSxgiYXl5Ha4eIBoj9btafTlRh2P

Canpolat, C. (2011). *Relationships between teacher motivation and organizational commitment through the implementation of teacher career steps* [Master thesis, Fırat University]. Council of Higher Education National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=zD1B0cW7zVr3VcnZjitVXplEjr3WvlrgepGYE1ijE9-B53Jr7ga0_tdHTviTEDin

Celep, C. (2000). Teachers organizational commitment in educational organizations. *National FORUM of Teacher Education Journal*, *10*(3), 3-24. https://eric.ed.gov/?id=ED452179

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Erlbaum.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge Falmer.

Cohen, L., Manion, L., & Morrison, K. (2021). *Research methods in education* (E. Dinç and K. Kıroğlu, Trans. Ed.). Pegem Academy.

Coyle-Shapiro, J., & Morrow, P.C. (2006). Organizational and client commitment among contracted employees. *Journal of Vocational Behavior*, *68*, 416–431. https://eprints.lse.ac.uk/825/1/JVB_2006.pdf

Çağlar, Ç., & Çınar, H. (2021). The effect of school principals 'communication competencies on teachers' organizational commitment. *E-International Journal of Educational Research, 12*(1), 157-170. https://doi.org/10.19160/ijer.831956

Çakır, Ö. (2001). *The concept of job involvement and influencing factors* [Unpublished master thesis]. Dokuz Eylül University.

Çelenk, O. (2022). New career models. *Adnan Menderes University Journal of Social Sciences Institute, 9*(1), 118-127. https://doi.org/10.30803/adusobed.1096400

Dağlı, A., & Gençdal, G. (2018). Examination of teachers' perceptions of organizational commitment according to some variables. *Electronic Journal of Educational Sciences, 7*(14), 164-175. https://dergipark.org.tr/tr/download/article-file/585859

Dağlı, A., Elçiçek, Z., & Han, B. (2018). Adaptation of the organizational commitment scale to Turkish: Validity and reliability study. *Electronic Journal of Social Sciences, 17*(68), 1765-1777. https://doi.org/10.17755/esosder.445932

Demir, N., & İnandı, Y. (2022). Investigating the relationship between school administrators' organizational justice and teachers' organizational commitment and trust. *Pamukkale University Journal of Education,* 57, 1-29. https://doi.org/10.9779.pauefd.1093082

Demirel, Y. (2009). A conceptual approach to the relationship between organizational commitment and counterproductive behaviors for productivity. *Istanbul Commerce University Journal of Social Sciences, 8*(15), 115-132. https://ticaret.edu.tr/uploads/kutuphane/dergi/s15/115-132.pdf

Doğan, H., & Çelik, K. (2019). The relationship between the power usage styles of school administrators and teachers' organizational commitments and job performances. *Education and Science, 44*(198), 37-55. https://doi.org/10.15390/EB.2019.7943

Engin, A., & İpek, M. (2020). Determination of teachers' stress sources, coping methods, and levels of school-family-life conflict. *Aydın Health Journal, 6*(2), 183-218. https://dergipark.org.tr/en/pub/asder/issue/56058/770280

Franklin, K. (2016). *Examining the ımpact of leadership styles on the motivation of U.S. teachers* (Thesis No: 3028) [Doctoral dissertation, Walden University]. https://scholarworks.waldenu.edu/dissertations/3028

Fuller, F.F. (1969) Concerns of teachers: A developmental conceptualization. *American Educational Research Journal, 6*(2), 207-226.

Gündüz, Y. (2010). Öğretmen algılarına göre kadın öğretmenlerin kariyer engellerinin incelenmesi [Examination of career obstacles of female teachers according to teacher perceptions]. *Abant İzzet Baysal University Journal, 10*(1), 133-148. https://dergipark.org.tr/tr/pub/aibuefd/issue/1498/18131

Gürbüz, R. (2012). *The relationship of perceived organizational support and organizational trust with organizational commitment* [Master thesis, Atatürk University]. Council of Higher Education National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=RYan9_S-Z7Eir3xdWGXBiK7x8FkUtvJL2iqCvzHM8RZLobZCe05V8VLUiwoZGv15

Güzeller, C.O., & Çeliker, N. (2020). Examining the relationship between organizational commitment and turnover intention via a meta-analysis. *International Journal of Culture, Tourism and Hospitality Research, 14*(1), 102-120. https://doi.org/10.1108/IJCTHR-05-2019-0094

Hamidi, R., Barari, R., Sahebdel, F., & Bayat, F. (2024). Evaluating the model of causal relations between organizational identity and organizational commitment in hospital nursing staff through the mediation of organizational culture. *Employee Responsibilities and Rights Journal, 36*(1), 79-100.

Halaçlı, B. (2018). *Career development of female staff: A comparative study in the education sector: Case of Denizli* [Master thesis, Pamukkale University]. Council of Higher Education National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=hcgrYffRbz0Z44UJEuLtwR_W-VIdHvZMVZmIDlp1CtETrz2xHzCREs3KkJc3F4_w

Hardin, E.E., Varghese, F.P., Tran, U.V., & Carlson, A.Z. (2006). Anxiety and career exploration: Gender differences in the role of self-construal. *Journal of Vocational Behavior, 69*(2), 346-358.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1979). *Applied statistics for the behavioral sciences*. Rand McNally College Publishing.

Ingarianti, T., Suhariadi, F., Fajrianthi, F., & Kristiana, I. (2022). The Effect of Antecedents of Teachers' Subjective Career Success. *International Journal of Environmental Research and Public Health*, 19. https://doi.org/10.3390/ijerph191711121

İnandı, Y. (2009). The barriers to career advancement of female teachers in Turkey and their levels of burnout. *Social Behavior and Personality, 37*(8), 1143-1152. https://doi.org/10.2224/sbp.2009.37.8.1143

İnandi, Y., Özkan, S., Peker, S., & Atik, Ü. (2009). Career development obstacles of female teachers. *Mersin University Journal of Faculty of Education, 5*(1). https://dergipark.org.tr/tr/download/article-file/160943

İnandı, Y., & Tunç, B. (2012). The relationship between career obstacles of female teachers and their job satisfaction. *Journal of Educational Sciences Research, 2*(2), 202-222. https://arastirmax.com/tr/system/files/438/13_11.pdf

İnandı, Y., Evşen Gün, M., & Gılıç, F. (2017). The study of relationship between women teachers' career barriers and organizational silence: Viewpoint of women and men teachers. *European Journal of Contemporary Education, 6*(3), 542-556. https://doi.org/10.13187/ejced.2017.3.542

İnandı, Y., Tunç, B., & Kılavuz, T. (2018). Women's Career Barriers and Professional Alienation: The Teachers' Case in Turkey. *International Online Journal of Educational Sciences, 10*(3), 133-146. https://doi.org/10.15345/iojes.2018.03.009

İnandı, Y., & Gılıç, F. (2020a). Developing teachers' career barriers scale. *Mersin University Journal of the Faculty of Education, 16* (1), 15-31. https://doi.org/10.17860/mersinefd.669825

İnandı, Y., & Gılıç, F. (2020b). The relationship between school administrators' leadership styles and organisational cynicism from teachers' perspectives. *South African Journal of Education, 41*(1),1-12. https://doi.org/10.15700/saje.v41n1a1825

Jacobson-Lundeberg, V. (2016). Pedagogical implementation of 21st century skills. *Educational Leadership and Administration: Teaching and Program Development, 27*, 82-100. https://eric.ed.gov/?id=EJ1094407

Jaros, S.J., Jermier, J.M., Koehler, J.W., & Sincich, T. (1993). Effects of continuance, affective, and moral commitment on the withdrawal process: An evaluation of eight structural equation models. *The Academy of Management Journal, 36*(5), 951-995. https://doi.org/10.2307/256642

Karatepe, R., İnandı, Y., & Akar Karatepe, D. (2021). Academicians' views on career barriers and academic alienation. *Education Quarterly Reviews, 4*(2), 152-165. https://doi.org/10.31014/aior.1993.04.02.207

Kaya, O. (2007). *Organizational commitment: An application in the general directorate of security* [Master thesis, Gazi University]. Council of Higher Education National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=ePX_SaJ0b35Gq45swKG3lJPD6g2WeJXTDnNL_X2OoF81I4RwMCDF1Q8uhgdwv4Ij

King, D., Rucker, K.J., Duncan, D.W., & Boland, K.J. (2013). *A comparison of job stressors experienced by male and female beginning agriculture teachers.* University of Georgia. General Publications. https://hdl.handle.net/10724/35958

Klatt, L., Murdick, R., & Schuster, F. (1985). *Human resource management*. A Belland Howell Company.

Kotrlik, J.W., & Williams, H.A. (2003). The incorporation of effect size in information technology, learning, and performance research. *Information Technology, Learning, and Performance Journal*, *21*(1), 1-7.

Köse, A., & İnce, S. (2022). The career barriers faced by female administrators and teachers. *Inonu University Journal of the Faculty of Education, 23*(3), 1717-1755. https://doi.org/10.17679/inuefd.1141438

Köstek, H. (2007). *Career obstacles of female teachers according to the opinions of administrators and teachers working in formal primary schools* [Master thesis, Yıldız Technical University]. Council of Higher Education National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=wBmNpkQC9Nhi90NLW7E7-VksVHocJ4k-3F4C51MCrULzg9t3YabGhcP34fYHEBfM

Lipesa, H.A. (2018). The role of teacher leadership in student achievement in mathematics. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3276372

Marinakou, E., & Giousmpasoglou, C. (2019). Talent management and retention strategies in luxury hotels: evidence from four countries. *International Journal of Contemporary Hospitality Management, 31*(10), 3855-3878. https://doi.org/10.1108/IJCHM-10-2018-0862

McLean, L., & Connor, M.C. (2015). Depressive symptoms in third-grade teachers: relations to classroom quality and student achievement. *Child Development, 86*(3), 945-954. https://doi.org/10.1111/cdev.12344

Meriç, E., Çiftçi, D.Ö., & Yurtal, F. (2019). A study on the relationship between perceived organizational support and work engagement. *Kastamonu Journal of Education, 27*(1), 65-74. https://doi.org/10.24106/kefdergi.2296

Meyer, J.P., & Allen, N.J. (1984). Testing the "side-bet theory" of organizational commitment: Some methodological considerations. *Journal of Applied Psychology, 69*(3), 372. https://doi.org/10.1037/0021-9010.69.3.372

Meyer, J.P., & Allen, N.J. (1991). A three-component conceptualization of organizational commitment. *Human Resource Management Review, 1*(1), 61-89. https://doi.org/10.1016/1053-4822(91)90011-Z

Meyer, J.P., Allen, N.J., & Smith, C.A. (1993) Commitment to organizations and occupations: extension and test of a three-component conceptualization. *Journal of Applied Psychology, 78*, 538-551. https://doi.org/10.1037/0021-9010.78.4.538

Örücü, E., & Kışlalıoğlu, S. (2014). A field study on organizational commitment. *International Journal of Management Economics and Business, 10*(22), 45-67. https://dergipark.org.tr/en/download/article-file/1145885

Özsoy, S., & Özsoy, G. (2013). Effect Size Reporting in Educational Research. *Ilkogretim Online, 12*(2), 334-346. https://dergipark.org.tr/en/pub/ilkonline/issue/8585/106644

Özyavuz, V.M. (2021). *Examination of the relationship between teachers' emotional intelligence levels and organizational commitments* [Master thesis, Istanbul Sabahattin Zaim University]. Council of Higher Education National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=tqUiYt63sTQLTpozMJ92QuSylaXyvbRK4t2X0H305rs-ucZWPLKzId9OQOuzwiQn

Perveen, S., Riaz, A., & Bano, S. (2023). Organizational commitment, psychological capital, and job burnout among teachers: Closer focus on job sector. *Qlantic Journal of Social Sciences*. https://doi.org/10.55737/qjss.570993992

Pisarik, C.T., Rowell, P.C., & Thompson, L.K. (2017). A phenomenological study of career anxiety among college students. *The Career Development Quarterly, 65*(4), 339-352.

Rose, D., & Sullivan, O. (1993) *Introducing data analysis for social scientists*. Open University Press.

Sammons, P., Toth, K., & Sylva, K. (2017). The drivers of academic success for 'bright' but disadvantaged students: A longitudinal study of AS and A-level outcomes in England. *Studies in Educational Evaluation, 57*, 31-41. https://doi.org/10.1016/j.stueduc.2017.10.004

Sarsıkoğlu, F.A. (2019). *The factors explaning career adaptability of university students: A suggested model* [Doctoral dissertation, Gazi University]. Council of Higher Educaion National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=npGs9H39x7G6401x51yqpJpXmAZ1VRdVHYMHE3QfeimluF0dJqiFhR1_QkE6j6zB

Sezgin, F. (2010). School culture as a predictor of teachers' organizational commitment. *Education and Science, 35*(156), 142-159.

Soysal, H., & Selvi, Ö. (2023). An analysis on the relation between the organizational justice perception and organizational commitment levels of the teachers. *Gümüşhane University E-Journal of Communication Faculty, 11*(1), 803-825.

Sükun, T.N. (2022). *The relationship between female teachers' career barriers and levels of resilience* [Master thesis, Mersin University]. Council of Higher Educaion National Thesis Center. https://tez.yok.gov.tr/UlusalTezMerkezi/TezGoster?key=CG8WvdvvxJP04Unr7Yecf-antJYCDQWEulD5-erXhPAPGElc_kC3nnlIEqcUIE5F

Swanson, J.L., & Tokar, D.M. (1991). College students' perceptions of barriers to career development. *Journal of Vocational Behavior, 38*(1), 92-106. http://doi.org/10.1016/0001-8791(91)90020-M

Şahin, S., Özaydın, S., & Siyez, D.M. (2019). Adaptation of the adult career concerns ınventory-short form into Turkish: A validity and reliability study. *Journal of Career Psyhological Councile, 2*(2), 166-193. https://dergipark.org.tr/tr/pub/kpdd/issue/51379/623567

Tabachnick, B.G., & Fidell, L.S. (2015). *Using multivariate statistics*. Pearson.

Taran, T., & Fathi, S. (2017). The relationship between organizational commitment and job stress and job satisfaction of teachers employed in Dehdasht department of education, Iran. *Journal of Interdisciplinary Research, 7*(1), 44-49. http://www.magnanimitas.cz/ADALTA/070101/papers/A_007.pdf

Wang, P., Chu, P., Wang, J., Pan, R., Sun, Y., Yan, M., Jiao, L., Zhan, X., & Zhang, D. (2020). Association between job stress and organizational commitment in three types of Chinese university teachers: Mediating effects of job burnout and job satisfaction. *Frontiers in Psychology, 11, 1-12*. https://doi.org/10.3389/fpsyg.2020.576768

Weng, Q., McElroy, J.C., Morrow, P.C., & Liu, R. (2010). The relationship between career growth and organizational commitment. *Journal of Vocational Behavior, 77*(3), 391-400. https://psycnet.apa.org/doi/10.1016/j.jvb.2010.05.003

Wenner, J.A., & Campbell, T. (2017). The theoretical and empirical basis of teacher leadership a review of the literature. *Review of Educational Research, 87*(1). https://doi.org/10.3102/0034654316653478

Yang, X., Ge, C., Hu, B., Chi, T., & Wang, L. (2009). Relationship between quality of life and occupational stress among teachers. *Public Health, 123*, 750-755. http://doi.org/10.1016/j.puhe.2009.09.018

Yakışıklı, A., Yavuz, S., Asmaz, C.B., & Karagedik, H. (2023). Okul yöneticilerinin liderlik stillerinin öğretmen motivasyonuna etkisi [The impact of school administrators' leadership styles on teacher motivation]. *Ulusal Eğitim Dergisi [Journal of National Education], 3*(5), 864–878. https://www.uleder.com/index.php/uleder/article/view/262

Yaman, S., Incebacak Bal, B., & Sarışan Tungaç, A. (2022). Stakeholder views on determining teacher qualifications. *Journal of Ahmet Keleşoğlu Education Faculty, 4*(2), 376-397. https://doi.org/10.15700/saje.v39n1a1670

Yiğit, Ö., Ayaz, M.A., Yiğit, A., & Güngör, E. (2022). Examination of the relationship between teachers' union satisfaction and organizational loyalty. *Cihanşümul Academy Journal of Social Sciences, 3*(4), 20-29.

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 222–239
https://doi.org/10.21449/ijate.1553522
journal homepage: https://dergipark.org.tr/en/pub/ijate *Research Article*

# A systematic review on the effect of lesson study on entrepreneurial competencies of foreign language teachers

**Nalan Özbey**[1*], **Duygu İleritürk**[2]

[1]Kafkas University, Faculty of Education, Department of Educational Sciences, Kars, Türkiye
[2]Ataturk University, Faculty of Education, Department of Educational Sciences, Erzurum, Türkiye

**ARTICLE HISTORY**
*Received: Sep. 20, 2024*
*Accepted: Dec. 30, 2024*

**Keywords:**
English language teaching,
Entrepreneurship,
Entrepreneurship competence,
Lesson study,
Systematic review.

**Abstract:** The objective of this study is to analyze literature research on the lesson study model of English teachers and their entrepreneurship competencies. So, systematic review approach was used in the study. The systematic research was conducted in Web of Science and Scopus databases, covering studies between 2019 and 2024. The systematic review procedure of the study is guided by the PRISMA flow diagram. The analysis examined data from 17 papers that satisfied the specified criteria for inclusion in this study. Within the parameters of the initial research inquiry, we identified patterns in study of lesson study and competences related to entrepreneurship. It can be asserted that lesson study, which converts teachers' practical knowledge into professional knowledge during their professional service processes, enhances their professional skills, and empowers them to teach in alignment with future demands, can also contribute to the long-term development of entrepreneurial competencies in teachers' professional lives, as its cyclical process evolves into professional learning.

## 1. INTRODUCTION

In order to effectively adapt to the changes of the 21st century, implement necessary adjustments, and achieve success, it is imperative to possess specific abilities. The objective of the education system to cultivate persons with integrated competences implies that education is the most efficient method of obtaining these competencies and converting them into patterns of behavior. Given this rationale, it is crucial for each nation to evaluate its educational institutions in accordance with progress and transformation. Each country has developed national qualification frameworks with distinct structures and contents, which will greatly enhance the development and operation of their education systems.

Core competencies, the overarching and shared objectives of all educational programmes within the European Union, pertain to the fundamental abilities that every individual will develop during their formal education. The European Commission introduced core competences in EU countries, which encompass the fundamental knowledge, skills, and attitudes necessary to prepare students for personal, professional, and socioeconomic

*CONTACT: Duygu İLERİTÜRK ✉ duygu.birbiri@atauni.edu.tr Ataturk University, Faculty of Education, Department of Educational Sciences, Erzurum, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

development, with the explicit objective of fostering social engagement and contribution. This framework utilizes "qualification type identifiers" to specify the qualification types that are included. Key competences are a critical component that qualification type identifiers should include. Entrepreneurship competence is identified as one of the foremost competencies (Ministry of National Education [MEB], 2018).

The cultivation of entrepreneurial skills has a significant impact on the economic strength of nations and is crucial for their integration into the global power structure. In our nation, it is imperative to nurture persons who possess the ability to convert obstacles into opportunities, promote unrestricted innovation, and harness their entrepreneurial potential. Hence, nations have endeavored to include entrepreneurship education into their educational curricula, ranging from primary education to higher education, with the aim of fostering an entrepreneurial mindset and, consequently, cultivating persons with entrepreneurial traits (Çiçek & Durna, 2012). The curriculum of our country includes entrepreneurship competencies, even if it is taught as a foreign language.

The Directorate General for European Union and Culture's commission report highlights the necessity of improving foreign language instruction to fully realize the cultural, social, and economic potential of the European Union (Aldemir & Er, 2012). The Turkey 2023 education vision has established specific goals for foreign language education, including offering online and in-person training to foreign language teachers in line with lifelong learning. Additionally, it aims to ensure that these teachers are native English speakers and implement foreign language education policies and teaching methods that promote qualified foreign language education. The establishment of a national foreign language education council aims to define certain criteria, classroom methodologies, and teacher proficiencies (MEB, 2019). The Ministry of National Education has undertaken research on optimizing long-term planning and increasing efficiency in professional development and foreign language instruction.

Banjica (2020) asserts that contemporary teaching necessitates teachers to update their professional credentials. The primary professional responsibility of teachers is to facilitate the acquisition of new habits and behaviors among learners. During the instructional process, every teacher aspires to exhibit competent and effective teaching. Essential components of teacher competences include their knowledge, abilities, and the desired outcomes. What the teacher knows is indicative of his expertise and pedagogical practices. The actions and behaviors of a teacher demonstrate their abilities, attitudes, and values. The objectives that the teacher aims to accomplish are indicative of their motivation and enthusiasm in teaching (Banjica, 2020). Teachers necessitate assistance from professional development models throughout their tenure to sustain their continuous learning and augment their 21st century competencies. Several professional development approaches, including the lesson study model, include comparable stages and are designed to facilitate lifelong education for teachers. In contrast to these paradigms, lesson study emphasizes the ongoing and adaptable growth of teachers' professional skills.

Lesson study is a cooperative professional development method where educators gather data to create a lesson plan, deliver a lesson, watch students, and restructure the lesson. Teachers are interested in implementing more efficient learning strategies that might enable students to get better and more meaningful learning results (Stepanek *et al.*, 2007). Lesson study is a form of classroom research that enables educators to examine the process of teaching and learning within the setting of a functioning classroom. It is the systematic procedure by which teachers document their work, including the design of the lesson, their explanation and experience of the lesson's functioning, and the knowledge and skills they acquire and impart during the lesson (Akiba & Wilkinson, 2016; Dudley, 2015; Lewis & Lee, 2017; Ono & Ferreira, 2010).

The notion of lesson study is commonly understood as the distributed responsibility, cooperative effort in the preparation of a lesson plan, implementation, observation of lessons,

monitoring and assessment of teaching, reflection on practice, and subsequent planning (İleritürk, 2021). Given these factors, lesson study is an autonomous and participatory activity. All participants collaborate and provide an equitable contribution to one another in all decision-making processes (Dudley, 2015; Lewis & Lee, 2017; Stepanek *et al*., 2007). Educators consider the examination of students' intended learning and cognitive processes to be a crucial and stimulating aspect of class preparation. The lesson study groups explore the process by which students acquire knowledge and construct their own learning experiences. Furthermore, this review serves to reveal the areas of interest among pupils. Through observation during lesson study, teachers enhance their own comprehension and acquire insight into how to establish connections between various subjects and expand upon them. Nevertheless, teachers dedicate time to scrutinize and contemplate the curriculum and numerous instructional resources. Simultaneously, teachers convert their expertise in the subject into authentic and significant experiences for students (Stepanek *et al*., 2007).

Classroom-based professional development utilizes lesson study, which offers a diverse array of characteristics. The key aspects among these qualities are the allocation of time for engagement with pertinent teachers, provision of administrative support, and development of an action plan. The lesson study is structured into a four-stage cycle. The first stage of the cycle involves creating the program and establishing its objectives. In this phase, educators contemplate the long-term objectives for students' acquisition of knowledge and growth, engage in the creation of programs and standards, and ascertain subjects that will captivate kids.

The planning stage is designated as the second phase. The planning phase entails the meticulous selection or reorganisation of the research route. An effective plan must incorporate the subsequent elements:

- Overarching objectives
- Anticipated students’ feedbacks
- Data acquisition strategy
- One conceptual framework of the learning cycle
- Rationale for the employed methodology

The third phase of the lesson study is commonly referred to as the research course. A single individual among the group of teachers conducts the research lesson, while the remaining members of the group observe and gather data. The concluding phase of the cycle of lesson study is reflection. The observers convene a formal instructional session during the phase of reflection. During this meeting, observers exchange the knowledge acquired from the class. The collected data is used to enhance student learning, subject material, course and unit design, and improve learning and teaching practices. During the reflection phase, students document all components of the lesson study cycle and generate fresh inquiries for future learning (Lewis & Hurd, 2011; Lewis & Lee, 2017).

Upon reviewing the literature, it was seen that the capabilities of educators in professional development and the success and skills of students improved after completing the lesson study process (İleritürk, 2021; Kanellopoulou & Darra, 2019; Uştuk & Çomoğlu, 2019; Willems & Bossche, 2019). The contributions of lesson study include fostering collaboration, promoting empathy among students and teachers, enhancing the creative and critical thinking abilities of teachers and students, enabling teachers to develop innovative and researcher identities, and fostering courage in both students and teachers (Akalın & Kılıç, 2022). It may be argued that the obtained competences include entrepreneurial skills. Lesson Study entails collaborative efforts among educators to design, execute, observe, and assess a lesson. Collaborative work fosters the enhancement of entrepreneurial competencies, including innovation, problem-solving, risk-taking, and leadership. In this process, educators collaboratively explore how to successfully communicate entrepreneurship concepts. Entrepreneurship necessitates practical abilities over theoretical understanding. Lesson study facilitates the observation of instructional

practices in an authentic classroom setting, enabling the direct evaluation and enhancement of teaching approaches that foster entrepreneurial abilities through feedback. The lesson study cycle entails utilizing each observation and assessment to improve the following portion of the lesson. Ongoing practice and the cultivation of feedback are essential elements of entrepreneurial education. This parallelism illustrates that lesson study serves as an exemplary example for cultivating entrepreneurial skills. Therefore, the researchers think that lesson study is a more suitable Professional development model for entrepreunership skills because of its nature being collaborative, cyclical and practice-based approach.

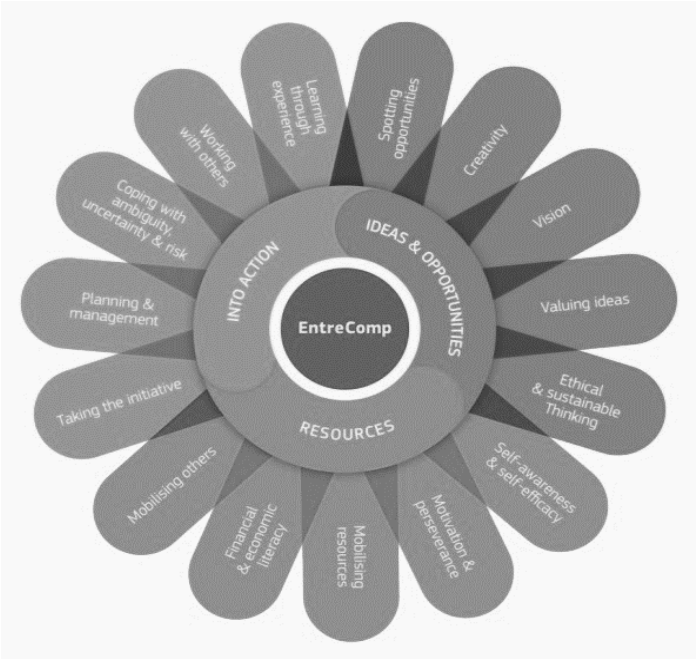
An entrepreneur is an individual who establishes a new business despite risk and uncertainty with the aim of maximising profit and growth by recognising significant opportunities and assembling the required resources to capitalise on them (Swanson, 2017; Zimmerer & Scarborough, 2008). An entrepreneur is an individual who possesses the ingenuity to combine resources in a unique manner and is prepared to assume risk and/or uncertainty associated with the undertaken action (Hisrich *et al.*, 2012). Entrepreneurship may be succinctly defined as the utilisation of available resources to actively seek for and exploit possibilities without any restrictions (Brooks, 2009; Swanson, 2017). A well acknowledged definition of entrepreneurship is the act of identifying and capitalizing on opportunities (Verma & Chaurasia, 2020). An analysis of entrepreneurial opportunity sources refers to the systematic identification, assessment, and exploitation of opportunities, together with the individuals who identify, assess, and profit from them (Shane & Venkataraman, 2000). Entrepreneurship can be defined as the systematic identification of resources, generation of ideas, and implementation of creative solutions, thereby converting opportunities into distinctive and valuable commodities (Özbey, 2022).

Advocates of the pedagogy approach contend that all instructional disciplines should include entrepreneurship as a pedagogical approach, prioritizing individual growth rather than entrepreneurship as a career, rather than teaching it as an independent topic (Moberg, 2014). The English language curriculum incorporates entrepreneurship within its curriculum, rather than teaching it as an independent module.

*Entrepreneurship encompasses the capacity to transform concepts into tangible outcomes by means of ingenuity, originality, and willingness to take risks, together with the competencies to strategize and oversee initiatives. Entrepreneurial competence refers to the capacity to effectively seize chances and ideas and convert them into valuable outcomes for others. Entrepreneurial competency requires awareness of the various contexts and possibilities for transforming ideas into action in personal, social, and professional spheres, as well as comprehension of the processes by which these chances emerge. Individuals must possess knowledge and comprehension of the methodologies required to strategize and oversee projects that encompass both systematic procedures and limited resources. Proficient communication and negotiation skills, together with the capacity to handle uncertainty, ambiguity, and risk, are crucial for making well-informed decisions. Entrepreneurial mentality is defined by qualities such as initiative, agency, invention, future-oriented perspective, bravery, and persistence in attaining objectives. It encompasses the aspiration to inspire others throughout the process, prioritizing ideas, empathy, concern for individuals and the environment, and assuming accountability through ethical methods (Turkish Qualification Framework [TYÇ], 2015).*

Figure 1 illustrates 15 competences that reflect the essential knowledge, abilities, and attitudes related to entrepreneurship. As stated by the European Commission in 2016, the competencies include identifying opportunities, creativity, vision, valuing ideas, ethical and sustainable thinking, self-awareness and self-efficacy, motivation and perseverance, mobilizing resources, financial and economic literacy, mobilizing others, taking initiative, planning and managing, skepticism and risk, teamwork, and learning through experience.

**Figure 1.** *EntreComp Cycle: 3 competenciencies field and 15 competences (McCallum et al., 2018).*




The assessment of these dimensions indicates a robust association between entrepreneurial aspects and the professional competences aimed at in the lesson study. The identification of opportunities, which is one of the entrepreneurship competencies and constitutes the starting point of the process, is defined as "identifying and capturing opportunities to create value by exploring the social, cultural, and economic image and identifying needs and challenges" (European Commission, 2016). Essentially, the process begins with the identification and definition of opportunities, followed by their utilization. Creativity is another competency in entrepreneurship. Creativity is the ability to generate and create valuable ideas by combining and matching different information, perspectives, and know-how (Zhang & Zhang, 2018). It begins with recognizing a problem or foreseeing an opportunity. Understanding the situation and considering the issues then leads to the design of new connections and the revelation of possible new component combinations (Akanbi & Aun, 2015). Entrepreneurial behavior is defined as creative thinking (invention) (Verma & Chaurasia, 2020).

Ames and Runco (2005) discovered a connection between diverse thinkers and successful entrepreneurs and identified evidence showing that successful entrepreneurs tend to generate more ideas and can produce valuable, useful products and services thanks to creativity. Teamwork can be defined as working with others through cooperation and communication to achieve a common goal. Members must understand their role in the team, be able to work on tasks independently of others, be flexible, and strive for the same goal in order to work effectively as a team. According to Vygotsky (1978), learning is a social process. According to Warhuus (2016), learners not only learn based on their developmental stages but also from the context they are in; they learn in groups and collectively create ideas. Another competency associated with lesson study is the ability to learn from experience. It involves using any initiative as a learning opportunity to create value, learning with others, including peers and mentors, and reflecting on success and failure (McCallum *et al.*, 2018). One of the entrepreneurial competencies is self-awareness and self-efficacy. According to Rizal (2014), self-knowledge is self-understanding, self-introspection, and self-discovery. According to Günhan and Başer (2007), self-efficacy is the individual's self-confidence. The European Commission's entrepreneurship qualification framework says that self-sufficiency and self-awareness mean "thinking about needs and wants in the short, medium, and long term, figuring out and judging strengths and weaknesses of oneself and others, and believing in the power to change the course of events despite uncertainty, disruptions, and temporary failures"

(EntreComp, 2016). Skepticism and risk-taking are characteristics of entrepreneurial people. Individuals who are entrepreneurs take more risks than those who are not entrepreneurs. People with a high tendency to take risks are more willing to make decisions in uncertain situations (Raab *et al.*, 2005). Smejkal and Raise (2006) define risk as "probability, the possibility of loss or failure, the variability of possible outcomes or their success, the difference between expected and actual outcomes, the probability of an outcome different from the expected, and the situation in which the quantitative magnitude of the event is subject to a certain probability distribution." Risk can be defined as "the danger of negative change in the target, the possibility of loss or profit" (Kozubikava *et al.*, 2017). Another entrepreneurial competency is mobilizing resources, among other things. Mobilizing resources is defined as “getting and managing the material, non-material, and digital resources needed to turn ideas into action," while mobilizing others is defined as “getting the support needed to achieve valuable results” (European Commission, 2016). Perseverance is a very pertinent entrepreneurial skill that might be applied to the lesson study. In the context of adversity, perseverance can be defined as the capacity of an individual to endure with effort (Hodge *et al.*, 2018). The entrepreneurial competency of valuing ideas involves the process of assessing the social, cultural, and economic worth of an idea, acknowledging its potential for value creation, and identifying suitable strategies to maximize its potential (McCallum *et al.*, 2018).

Given the significance of fostering entrepreneurial skills through lesson study, it is widely accepted that the systematic review approach should be used to provide the results of existing research in the literature. However, although there are many studies on the professional development of foreign language teachers (Coşkun, 2017; Goh & Fang, 2017; Gözel, 2017; Heppt *et al.*, 2022; Karataş *et al.*, 2017; Liu, 2021; Saykal & Uluçınar-Sağır, 2022; Xiao *et al.*, 2020; Zhiyong *et al.*, 2020), there are very few studies on the relationship between entrepreneurship and lesson study in the 21st century competency fields (Capella-Paris *et al.*, 2023; Rahmatullah *et al.*, 2022; Widiastuti *et al.*, 2023). The objective of this study is to analyze literature research on the lesson study model of English teachers and their entrepreneurship competencies. In line with this purpose, answers to the following questions were sought.

- What are the tendencies about the effect of foreign language teachers' lesson study models on their entrepreneurial skills?
- What are the results about advanteges and disadvantges of effect of lesson study on entrepreneurship competences?
- What are the results about foreign language teachers’ lesson study practices and entrepreneurship competencies?

## 2. METHOD

This study employed a systematic review methodology to assess the current literature. A systematic review entails a methodical and thorough approach to identifying, selecting, and analyzing studies to address specific research inquiries or resolve certain issues (Karçam, 2013). A systematic review, in contrast to a conventional literature review, adheres to a stringent process to guarantee transparency and reduce prejudice. The review method encompasses well delineated inclusion and exclusion criteria, quality evaluation, and synthesis of results.

### 2.1. Search Strategy

The systematic review was performed utilizing two primary databases: Web of Science and Scopus. These databases were chosen for their extensive coverage, superior quality standards, and interdisciplinary characteristics, rendering them suitable for systematic reviews (Mulrow, 1994). The literature review was confined to papers published from 2019 to 2024 to emphasize current research, specifically investigating the effects of phenomena such as pandemics and digital transitions. Search terms; “lesson study”, “lesson study and English”, “lesson study and

foreign language teaching", "lesson study and English teachers", "entrepreneurship competencies", "entrepreneurship competencies and English", "entrepreneurship competencies and foreign language teaching", "Entrepreneurship competencies and English teachers", "lesson study and entrepreneurship", "lesson study and entrepreneurship competencies" were selected.

### 2.2. Study Selection

The review incorporated studies according to the specified inclusion criteria:

- Released from 2019 until 2024.
- Concentrate on lesson study, entrepreneurial competencies, or both, within the realm of foreign language instruction.
- Research assessing the efficacy of lesson study in improving entrepreneurship competencies among foreign language educators.
- Articles subjected to peer review and published in English.

Studies were excluded if they:

- Failed to cover the fundamental principles of lesson study or entrepreneurial competencies.
- We are not concentrating on foreign language teaching.
- Do not constitute original research (e.g., review articles, opinion pieces).
- Were not subjected to peer review.

The criteria were established to guarantee the inclusion of only the most pertinent, high-caliber papers in the review, hence enhancing the validity and application of the results. The systematic review procedure of the study is guided by the PRISMA flow diagram shown in Figure 2.

### 2.3. Data Extraction and Coding

Two independent researchers conducted data extraction by methodically coding the experiments according to preset criteria. The categories encompassed the year of publication, study methods, and educational setting (sample type and educational background). Intercoder reliability was evaluated by the Cohen's kappa coefficient, resulting in a value of 0.85, signifying a substantial level of agreement among the coders. All differences were addressed through dialogue.

### 2.4. Risk of Bias Assessment

A risk of bias assessment was performed on each included study to ensure methodological rigor, utilizing the Cochrane Risk of Bias Tool (Higgins *et al*., 2011). This instrument evaluates possible biases across the specified domains:

- Selection bias arises from inadequate random sequence generation and insufficient allocation concealment.
- Performance bias refers to the lack of blinding of participants and personnel in a study.
- Detection bias refers to the influence of outcome assessors' knowledge on the assessment of results, which can be mitigated through blinding.
- Attrition bias refers to the systematic differences in outcomes between participants who complete a study and those who drop out, leading to incomplete outcome data.
- Reporting bias refers to the selective reporting of outcomes in research studies.

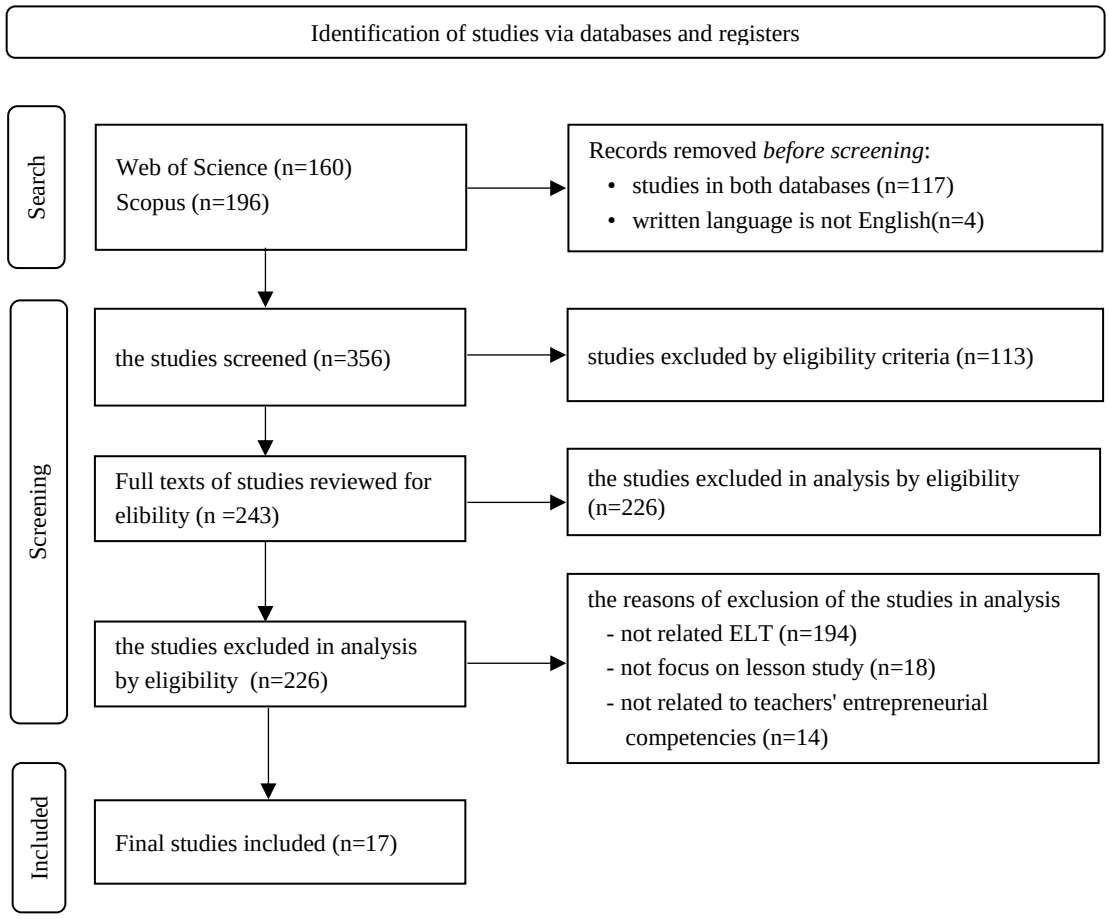
We assessed each study for the existence of these biases, classifying the risk of bias as low, high, or uncertain for each category. We subsequently assessed the overall risk of bias by aggregating these individual evaluations. We classified research exhibiting evident methodological deficiencies or numerous high-risk domains as high risk, whereas those conforming to rigorous procedures across all domains were deemed low risk. In the synthesis of results, we evaluated the overall quality of the evidence. This evaluation took into account the risk of bias as well as the precision, consistency, and directness of the study results. The synthesis assigned diminished importance to research exhibiting a significant risk of bias or

inconsistent results. Following the risk of bias evaluation, the researchers developed a summary table that delineates the methodological quality of each study incorporated in the review. Consequently, it is thought to enhance transparency and facilitate comprehension of the evidence's credibility.

### 2.5. Data Analysis

Content analysis was used to synthesize and analyze the data from the selected studies. The data were classified based on the impact of lesson study on improving the entrepreneurial competencies of foreign language educators. The analysis utilized the PRISMA flow diagram shown in Figure 2, offering a clear visual representation of the study selection process.

**Figure 2.** *Data analysis process according to PRISMA.*




## 3. RESULTS

Please present the findings/results in this section. This section should give significant results obtained from the study clearly and concisely. Please present the findings/results in this section. This section should give significant results obtained from the study clearly and concisely.

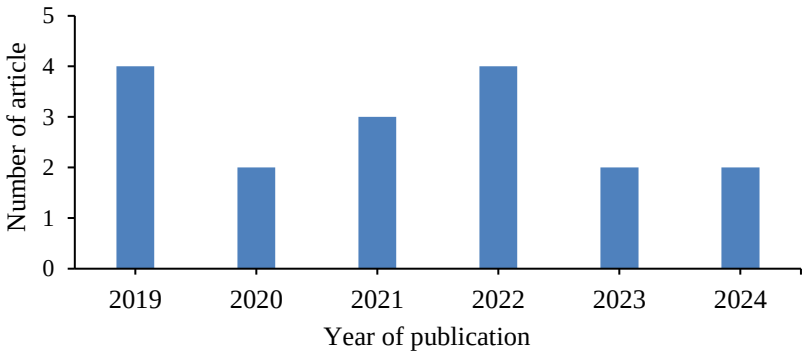
### 3.1. The Results About the Tendencies of Lesson Study and Entrepreneurship Competences

The research inquiry was about "What are the tendencies about the effect of foreign language teachers' lesson study models on their entrepreneurial skills?". The categories created in this context yielded the following findings:

#### 3.1.1. *Distribution of published years*

The study examined the articles used in data analysis based on their publication year. Figure 4 displays the findings derived from this approach. Figure 3 displays the findings of the published articles for the years 2019 (n = 4), 2020 (n = 2), 2021 (n = 3), 2022 (n = 4), 2023 (n = 2), and 2024 (n = 2).

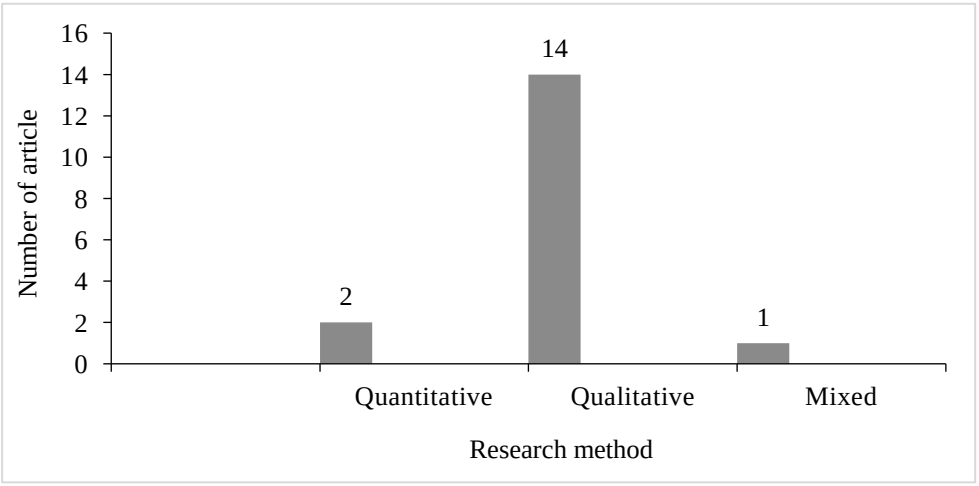
**Figure 3.** *Distribution of published years of the research.*




### 3.1.2. *Methods used in research*

The study examined the articles included in the data analysis in accordance with the methods employed. Figure 4 illustrates the results obtained in this direction. Figure 4 reveals that the articles predominantly prefer the qualitative method (n = 14). Two studies (n = 2) used the quantitative method, while only one study (n = 1) used the mixed method.

**Figure 4.** *The methods used in research.*




### 3.1.3. *Sample groups*

Analysis of the examined studies reveals that English teachers make up the majority of the sample groups. The sample groups of included studies was incorporate with other sample groups. Therefore, they were shown in Figure 5.

**Figure 5.** *Sample groups of research.*

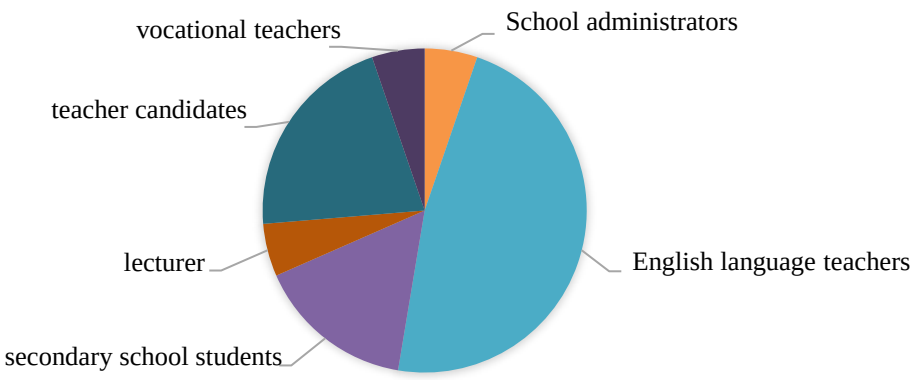
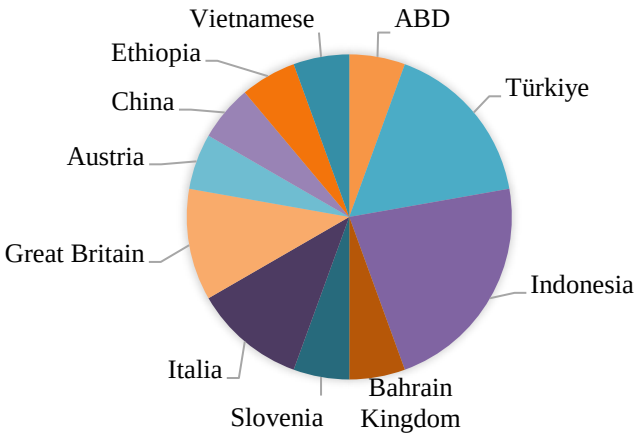


Figure 5 shows the sample groups used in the research. Other sample groups used in the research are "secondary school students, teacher candidates, school administrators, lecturers, and vocational teachers," respectively.

### 3.1.4. *Countries where the research took place*

When the countries where the studies were conducted in line with the examined studies were analyzed, it was found that the majority of these studies were conducted in "Indonesia, the Great Britain, Turkey, and Italy." In addition to this, countries such as the Kingdom of Bahrain, Slovenia, USA, China, Ethiopia, and Vietnamese also host studies. Figure 6 illustrates the distribution of countries where research has taken place.

**Figure 6.** *The countries where the research are held.*



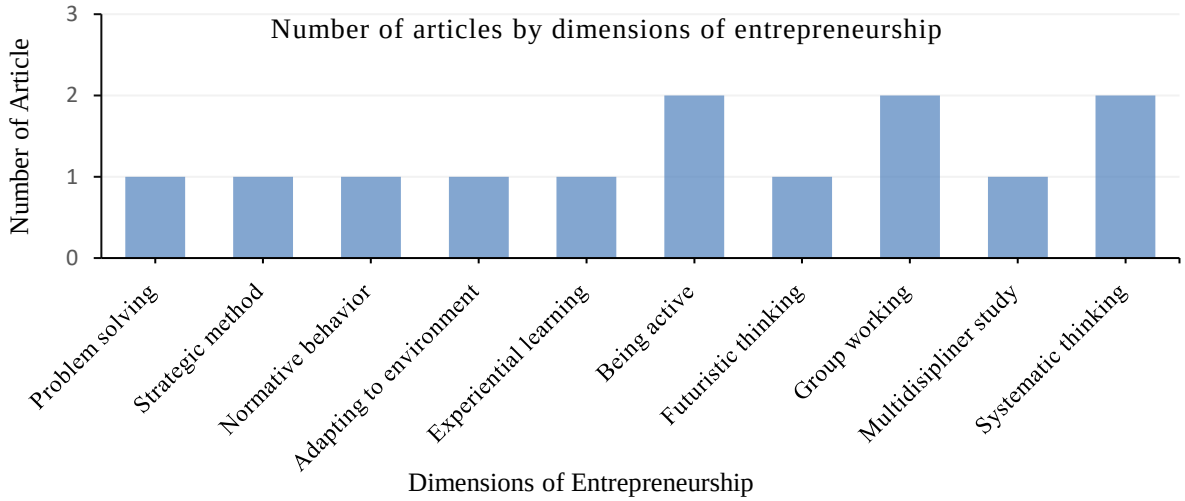

## 3.2. The Results About Effects of Lesson Study on Entrepreneurship Competences

The research question is, "What are the key findings that emerged from the lesson study and the literature on entrepreneurship competencies?" We've been looking for an answer to this question. The categories created in this context yielded the following findings:

### 3.2.1. *Dimensions of entrepreneurship*

Only three of the 17 studies included in the study focus directly on entrepreneurship. Examining the entrepreneurship dimensions of the three studies reveals the emergence of "*systematic thinking* (n = 1), *forward-looking thinking* (n = 1), *normative behavior* (n = 1), *bringing together differences and interdisciplinary diversity* (n = 1), *harmony with the environment (n = 1), taking action* (n = 2), *strategic management* (n = 1), *group work* (n = 1), *experiential learning* (n = 1), *and problem solving (n = 2)*". Figure 7 displays the findings regarding the dimensions.

**Figure 7.** *Dimensions of entrepreneurship.*



### 3.2.2. *The advantages of effectiveness of lesson study and entrepreneurial competencies*

The advantages and disadvantages of the lesson study and entrepreneurship competencies were examined. Accordingly, lesson study effect on entrepreneurship competencies development was investigated in terms of advantatges. Table 1 displays the resulting findings. In this regard, it is seen that the advantages of providing an honest, productive and constructive discussion environment (n=7), group work related to lesson study (n = 5), gaining a deeper perspective on teaching and learning processes (n = 3), and being able to take part in non-judgmental group work (n = 2) are stated.

**Table 1.** *Advantages of lesson study practice and entrepreneurship competences*.

| *Advantages* | *f* | *Sample articles* |
|---|---|---|
| Being active in non-judgmental group | *2* | Fox & Poultney, 2019 |
| Learning from experienced teachers | *1* | *Klammer & Hanfstingl, 2019* |
| Learning bu group reflections | *1* | Lee & Madden, 2019 |
| Learning achievement | *1* | Lee & Madden, 2019 |
| Planning lesson with different approaches | *1* | Lee & Madden, 2019 |
| Sharing responsibility | *1* | Lee & Madden, 2019 |
| Innovative learning approach and methods | *1* | Klammer & Hanfstingl, 2019 |
| Self-determination highly | *1* | Klammer & Hanfstingl, 2019 |
| Professional communication | *1* | Klammer & Hanfstingl, 2019 |
| Gaining a deeper perspective on teaching and learning processes | *3* | Nguyen, *et al*.., 2024 |
| Effective evaluation opportunity | *1* | Uştuk & Çomoğlu, 2019 |
| Group work | *5* | Morselli, 2024 |
| Supporting problem solving skills | *2* | Alwadi & Wilson, 2020 |
| Deepening teachers' knowledge in accordance with different levels | *2* | Darsih, Suherdi & Safrina, 2021 |
| Providing an honest, productive and constructive discussion environment | *7* | Achmad & Miolo, 2021 |

### 3.2.3. *The disadvantages of effectiveness of lesson study and entrepreneurial competencies*

Table 2 lists the disadvantages that have been identified in studies examining the effectiveness of lesson study and entrepreneurship competencies. Accordingly, the lesson study effect on entrepreneurship competencies development was investigated in terms of disadvantages. The research revealed that the inexperience of administrators and groups in observation and time management hindered their ability to implement standard-based programs, receive evaluations from others, and reflect on practices and teachers, particularly during reflection studies.

**Table 2.** Disa*dvantages of lesson study practice and entrepreneurship competences*.

| *Disadvantages* | *f* | *Sample articles* |
|---|---|---|
| Practicing standart-based curriculum | 2 | Lee & Madden, 2019 |
| Difficulty in assessment | 1 | Uştuk & Çomoğlu, 2019 |
| Lacking of reflective practices and teachers' views on reflection | 1 | Nguyen *et al.*, 2024 |
| Inexperiencing of observations of administrators and peers | 1 | Lee & Madden, 2019 |
| Difficulty in time-management | 1 | Nguyen *et al*., 2024 |

## 3.3. The Results About Lesson Study and Entrepreneurship Competencies

The research asks, "What are the main findings emerging in the literature regarding lesson study and entrepreneurial competencies of foreign language teachers?". We examined the effectiveness of foreign language teachers' lesson study and entrepreneurial competencies in 17

articles. In this regard, one article stated that lesson study and entrepreneurship competencies were ineffective for foreign language teachers, despite some articles showing no difference in effect (n = 3). 13 of the studies examined stated that lesson study was effective on entrepreneurship competencies. Table 3 presents these findings.

**Table 3.** *The results about lesson study and entrepreneurship*.

| Results | *f* | Articles |
|---|---|---|
| Lesson study is effective. | 13 | Susanto *et al*., 2020; Alwadi & Wilson, 2020; Bayram & Bıkmaz, 2021; Achmad & Miolo, 2021; Zhang & Luo, 2022; Mulualem *et al*., 2023; Mahmud *et al*., 2023; Nguyen *et al*., 2024; Fox & Poultney, 2019; Uştuk & Çomoğlu, 2019; Morselli & Gorenc, 2022; Joensuu-Solo, Viljama &Varamaki, 2022; Klammer & Hanfstingl, 2019 |
| Lesson study is not effective. | 1 | Wijnands *et al*., 2022; |
| There is no significance. | 3 | Darsih *et al*., 2021; Morselli, 2024; Lee & Madden, 2019 |

## 4. DISCUSSION and CONCLUSION

The objective of this study was to systematically analyse the impact of lesson study on the entrepreneurial skills of foreign language teachers using a systematic review approach. The analysis examined data from 17 papers that satisfied the specified criteria for inclusion in this study. Within the parameters of the initial research inquiry, we identified elements in study of lesson study and competences related to entrepreneurship. Based on the acquired results, it was analysed the distribution of research by year over the past five years. Considering the five-year period, it is seen that researchers conducted more research in 2019, 2020 and 2022. Concurrently, a thorough analysis of the research indicated that most of them (n = 13) used qualitative methodologies. The primary rationale for conducting studies using qualitative methods is that the nature of lesson study and entrepreneurship competencies is better aligned with the data collecting and analysis techniques employed in qualitative research. An analysis of the sample groups comprised in the studies indicates that they are predominantly consisting of English teachers. With the objective of investigating the entrepreneurial skills of foreign language teachers, this is a predictable result. The analysis of the nations where the research were conducted indicates a notable clustering in Indonesia, Türkiye, Italy, and the United Kingdom.

The study's findings indicate that three studies are directly relevant to the aspects of entrepreneurial abilities. The research mainly concentrates on two key entrepreneurship skills: proactive action and effective problem-solving. The elements of taking action and issue solving, especially in the context of lesson study, are thought to contribute to the cultivation of problem-solving abilities for possible challenges that may emerge in action and research courses. Zhang and Zhang (2018) argue that the variables of entrepreneurship, namely innovation, teamwork, self-awareness, and self-efficacy, are highly associated with the processes of lesson study. Furthermore, it is widely accepted that designing research lessons that are in line with collaborative work in terms of entrepreneurship competencies and the lesson study process, while also fostering creativity and teamwork among the learner group and enhancing learning skills, allows teachers to improve their problem-solving abilities. The academic research actively integrates the competency of "taking action". Lesson study, developed based on the lesson study paradigm, enable English teachers to strategically plan their own and their students' competences, as well as creatively construct their lessons by taking into account a diverse range of information and skills. In this manner, it can be confidently stated that the operational process of the lesson study model greatly improves entrepreneurship competency. Engaging in collaborative work, especially during the phases of lesson plan development, revision, and execution, allows teachers to impartially examine themselves and promotes their personal and

professional development. Furthermore, in accordance with the analyzed studies, it was reported that they encountered drawbacks in the execution of lesson study and entrepreneurship skills, particularly during observations, owing to the groups' limited experience. They also faced challenges in time management for conducting research courses and observations, as well as in assessment.

An additional research inquiry to be addressed in the study is to ascertain the efficacy of lesson study in relation to the entrepreneurial skills of foreign language teachers. The majority of the studies (n = 13) demonstrated the efficacy of lesson study in this context. Hence, the implementation of lesson study, which is a professional development paradigm, can facilitate the enhancement of teachers' entrepreneurial skills. Considering the last five years of literature indicates a lack of research that specifically investigate both lesson study and entrepreneurship skills. Nevertheless, the intended professional abilities at the conclusion of the lesson study process also include skills related to entrepreneurship. The process of entrepreneurship starts by first finding chances to provide novel value. We subsequently utilize the available resources and convert them into chances to generate profit. In the process of lesson study, educators initially analyze the accessible materials such as guides and lesson plans. They next convert current experiences into potential possibilities and proceed to design and create the research lesson. Thus, the lesson study asserts that the capacity to recognize possibilities and generate resources at the beginning of the planning process is a fundamental element of entrepreneurship.

Once the research course objectives have been established, group members proceed to design the research course by drawing upon their previous experiences, observations of their present students, and a range of available sources such as guides, reference books, and lesson plans. (Fornandez & Yoshida, 2004). Planning and management in entrepreneurship refer to the process of establishing priorities and action plans, as well as creating target objectives for the long, medium, and short term (European Commission, 2016). The above arguments elucidate that both entrepreneurship and lesson study entail a systematic planning process that commences with the establishment of objectives. Undoubtedly, the collaboration between teachers is the fundamental determinant of the success of the lesson study model (Bozkurt & Yetkin-Özdemir, 2018; Choski & Fernandez, 2004; Doig & Groves, 2011; Ermeling & Graff-Ermeling, 2014; Hj Suhaili & Khalid, 2011; Lewis & Perry, 2014; Stepanek *et al.*, 2007). The concept is structured around collaborative stages, encompassing the organization, planning, execution, and reflection of the program and objectives. In the context of entrepreneurship competencies, collaboration is characterized as the act of collectively and cooperatively working with people to generate ideas and transform them into tangible actions (European Commission, 2016).

In the realm of entrepreneurship, the ability to appraise ideas is a fundamental skill. Specifically, members of the group actively engage in listening to one another's views and assess the worth of each other's ideas by employing methods like brainstorming or empathy mapping to exchange ideas and experiences. Based on the lesson study conducted by Stepanek *et al.* (2007), it is essential for group members to have the capacity to critically and realistically evaluate their own teaching methods, work together as a cohesive unit, have trust in one another, and show respect for each other's opinions. By adopting this viewpoint, it can be inferred that the attitude of valuing the principles employed in the instruction showcases entrepreneurial competence. The learning objectives relate to the development of ethical and sustainable thinking. Entrepreneurial competency, linked to the assessment of ideas, encompasses the ethical conduct and sustainable mode of thinking. The acquisition of ethical behavior is arguably one of the key attitudes that teachers should develop during the process of studying lessons. In order to maintain a sound procedure, educators must prioritize ethical principles such as precision, integrity, accountability, and humility. Furthermore, while developing a distinctive product, they give paramount importance to the sustainability aspects that provide value and advantages to both students and teachers.

The entrepreneurial process culminates in the emergence of a distinctive product that possesses both economic and social value. An integral component of the lesson study process is fostering collaboration among teachers and cultivating the conviction that improved outcomes will ensue. Teachers will gain awareness of the professional growth they and their colleagues experience as a consequence of these practices (Grove, 2011; Lewis & Perry, 2014), and they will strategically create a value system that will be advantageous for the learners. Entrepreneurial creativity refers to the generation of novel and economically valuable concepts. As the objective of entrepreneurship is to locate and assess possibilities, generate ideas, and produce a distinctive product, creativity is a crucial component of the entrepreneurial process. According to Lewis (2002), reflection meetings, in which every teacher and teacher candidate provides information for lesson study, greatly improve the quality of the debate by enriching scholarly discourse. Thus, each research course generated in a lesson study functions as an independent library production center.

Politis (2005) emphasizes that the acquisition of knowledge in entrepreneurship is a process that involves practical experience, conscious reflection, critical thinking, and direct action. Conversely, lesson study provides teachers with the opportunity to participate in live research lessons that they create and execute independently. This method establishes a distinctive experiential learning setting for educators, which emphasizes the interests and desires of the group and incorporates their own teaching experiences (Murata, 2011).

The lesson execution phase of the lesson study involves the observation of a teacher by other teachers while they are implementing the planned lesson. This observation serves to assist the professional growth of teachers. The study conducted by Choski and Fernandez (2004) revealed that teachers displayed feelings of anxiety, enthusiasm, shyness, and timidity while being observed during their lessons. This can be associated with the self-assurance or self-efficacy of instructional professionals. Serbest's (2014) study demonstrated that teachers' outcomes outcomes from the implementation of lesson study interventions resulted in a rise in their self-assurance and self-efficacy convictions. Furthermore, self-efficacy is a crucial entrepreneurial skill.

Students that engage in lesson study are expected to develop skills such as skepticism and risk-taking, which are considered essential entrepreneurial competencies. The European Commission (2016) defines scepticism and risk as the act of making judgments in situations where the outcome of the decision is unknown, the available knowledge is incomplete or uncertain, or there is a possibility of repercussions. Lesson study indicates that teachers are progressively embracing greater risks by implementing diverse approaches in classroom practices, a phenomenon that is thought to augment their expertise in the field and pedagogical knowledge (Cajkler *et al.*, 2014). Alongside risk-taking, the ability to take initiative is a crucial skill in entrepreneurship. The learning outcomes of taking initiative, a competency in entrepreneurship, encompass the ability to tackle problems by beginning procedures that generate value, taking strategic action to accomplish objectives, and diligently carrying out planned tasks (McCallum, 2018). The utilisation of this entrepreneurial competency is observed.

Consequently, lesson study, a structured professional development process that transforms teachers' practical knowledge into actionable insights, significantly contributes to their long-term professional growth. As education systems prioritize skills like creativity, critical thinking, and innovation to prepare students for future challenges, lesson study enables teachers to align their instructional practices with these emerging priorities. The cyclical nature of lesson study, encompassing stages of planning, implementation, observation, and reflection, parallels the iterative processes fundamental to entrepreneurial thinking. Through consistent participation in this cycle, educators develop behavior and mindset linked to entrepreneurial competencies, including problem-solving, adaptability, collaboration, and ongoing learning. Over time, these competencies become established as habitual professional learning patterns, promoting

resilience, creativity, and proactive decision-making. Lesson study contributes to teachers' immediate classroom effectiveness and supports the long-term development of entrepreneurial skills, enhancing their capacity to innovate, lead, and address future challenges in their professional lives. It functions as a transformative instrument that connects conventional teaching methods with the progressive requirements of modern educational systems. Potential future research might investigate the influence of instructional designs for foreign language teachers on their entrepreneurial skills in pre-service education via using alternative Professional development models e.g. micro-teaching or school-based Profesional development models etc.

## Acknowledgments

This study was presented at 12th International Conferences on Curriculum and Instruction as an oral presentation.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

## Contribution of Authors

**Nalan Özbey**: Investigation, Resources, Visualization, Data Analysis, and Writing-original draft. **Duygu İleritürk**: Methodology, Data Analysis, Supervision, and Validation.

## Orcid

Nalan Özbey https://orcid.org/0000-0001-7849-5689
Duygu İleritürk https://orcid.org/0000-0002-0409-4281

## REFERENCES

Akalın, S., & Kılıç, A. (2022). Türkiye'de ders imecesi modeli kullanılarak yapılan lisansüstü tezler üzerine bir meta sentez çalışması [The dissertations and thesis about lesson study model in Turkey: A meta-synthesis]. *Fırat Üniversitesi Sosyal Bilimler Dergisi, 32* (3), 1013-1025

Akanbi, S.O., & Aun, I.I. (2015). Creativity and Innovation in Entrepreneurship. In D. Bögenhold, J. Bonnet, M. Dejardin and D.G.P. Lema (Eds.), *Contemporary Entrepreneurship*, (pp.66-89), Springer Publications.

Akiba, M., & Wilkinson, B. (2016). Adopting an international innovation for teachers Professional development: State and district approaches to lesson study in Florida. *Journal of Teacher Education, 67*(1), 74-93.

Ames, M., & Runco, M.A. (2005). Predicting entrepreneurship from ideation and divergent thinking creatvity. *Innovation Managament, 14*(3), 311–315

Aldemir, A.Y., & Er, K.O. (2012). A comparison of pre-service English language teachers training systems in Turkey and Japan. *Ankara University Journal of Faculty of Educational Sciences, 45*(1), 83-105.

Banjica, M. (2020). The qualities of a modern teacher. *Journal of Educational Research, 2*, 3-4.

Bozkurt, E., & Yetkin-Özdemir, İ. E. (2018). Mesleki gelişimde işbirliğine dayalı bir yaklaşım: Ders Araştırması [An approach based on cooperative in Professional development: Lesson study]. *Kastamonu Üniversitesi Kastamonu Eğitim Dergisi, 26*(1), 109-116.

Cajkler, W., Wood, P., Norton, J., & Pedder, D. (2014). Lesson study as a vehicle for collaborative teacher learning in a secondary school. *Professional Development in Education, 40*(4), 511-529.

Capella – Paris, C., Cosgrove, M.M., Salvador-Garcia, C., & Marave-Vivas, M. (2023). How to promote social entrepreneurship on urban education? An active learning implementation.

*Education and Urban Society*, *55*(7), 771-791. https://doi.org/10.1177/00131245221083551
Chokshi, S., & Fernandez, C. (2004). Challenges to importing Japanese Lesson Study: Concerns misconceptions, and nuances. *Phi Delta Kappan*, *85*(7), 520-525.
Çiçek, R., & Durna, U. (2012). Meslek idealleri ve girişimcilik niyetleri [Profession ideals and entrepreneurship intenations]. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *9*(17), 17-31. https://dergipark.org.tr/en/pub/mkusbed/issue/19553/208277
Coşkun, A. (2017). The application of lesson study in teaching English as a foreign langauge. *İnönü University Journal of the Faculty of Education*, *18*(1), 151-162. https://doi.org/10.17679/inuefd.297845.
Doig, B., & Groves, S. (2011). Japanese Lesson Study: Teacher professional development through communities of inquiry. *Mathematics Teacher Education and Development*, *13*(1), 77-93.
Dudley, P. (2015). How lesson study works and why it creates excellent learning and teaching. P. Dudley (Ed). In *Lesson study-professşonal learning for our time*. Routledge.
Elkomy, M.M., & Elkhaial, N.H. (2022). The lesson study approach to Professional development: Promoting teachers' peer mentoring and communities of practice and students' learning in Egypt. *Teaching and Teacher Education, 109*, 1-13. https://doi.org/10.1016/j.tate.2021.103538
Ermeling, B.A., & Graff-Ermeling, G. (2014). Learning to learn from teaching: A first-hand account of lesson study in Japan. *International Journal for Lesson and Learning Studies*, *3*(2), 170–191.
European Commission (2016). *EntreComp: The Entrepreneurship Competence Framework.* JRC Science For Policy Report. https://core.ac.uk/download/pdf/38632642.pdf
Fernandez, C., & Yoshida, M. (2004). *Lesson study: A Japanese approach to improving mathematics teaching and learning.* Lawrence Erlbaum Associates.
Goh, R., & Fang, Y. (2017). Improving English langauge teaching through lesson study: Case study of teachers learning in a Singapore primary school grade level team. *International Journal for Lesson and Learning Studies*, *6*(2), 135-150. https://doi.org/10.1108/IJLLS-11-2015-0037.
Grove, M.C. (2011). *Assessing the impact of lesson study on the teaching practice of middle school science teachers*. [Unpublished dissertation of doctorate]. University of California.
Gözel, E. (2017). Sınıf öğretmenlerinin problem çözmeye dayalı ders imecesi bilgilerinin gelişiminin incelenmesi [Investigation of lesson study knowledge development of primary school teachers in terms of problem solving]. *Eğitim Kuram ve Uygulama Araştırmaları Dergisi*, *3*(3), 48-62.
Günhan, C.B., & Başer, N. (2007). Geometriye yönelik öz-yeterlik ölçeğinin geliştirilmesi [Development of scale of self-efficacy towards geometry]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 2007*(33), 68-76
Heppt, B., Henschel, S., Hardy, I., Hettmannsperfer-Lippolt, R., Gabler, K., Sontag, C., Mannel, S., & Stanat, P. (2022). Professional development for language support in science classrooms: Evaluating effects for elementary school teachers. *Teaching and Teacher Education, 109*, 1-13. https://doi.org/10.1016/j.tate.2021.103518
Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., & Welch, V.A. (eds). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5 (updated August 2024). Cochrane, 2024. www.training.cochrane.org/handbook
Hisrich, R.D., Peters, M.P., & Shepherd, D.A. (2012). *Entrepreneurship* (9th Ed.). McGraw-Hill Education.
Hj Suhaili, H.A.S, & Khalid, M. (2011). Mathematics teachers' perception of lesson study as a continuous professional development programme. *Journal of Science and Mathematics Education in Southeast Asia*, *34*(1), 67-89
Hodge, B., Wright, B., & Bennett, P. (2018). The role of grit in determining engagement and academic outcomes for university students. *Research in Higher Education, 59*(4), 448-460.

İleritürk, D. (2021). *Ders araştırması modeli ile İngilizce öğretmeni özel alan yeterliklerinin geliştirilmesi [Development of English language teachers' special field proficienceies via lesson study]*. [Unpublished dissertation of doctorate]. Atatürk University.

Kanellopoulou, E.M., & Darra, M. (2019). The implementation of the lesson study in basic teacher education: a research review. *Higher Education Studies, 9*(3), 65-78

Karataş, K., Ardıç, T., & Oral, B. (2017). Öğretmenlik mesleğinin yeterlikleri ve geleceği: Metaforik bir analiz [Teachers' professional competencies and future: A metaphoric analysis]. *International Periodical for the Languages, Literature and History of Turkish or Turkic, 12*(33), 291-312. https://doi.org/*10.7827/TurkishStudies.12665*

Kozubikava, L., Dvorsky, J., Cepel, M., & Balcerzak, A.P. (2017). Important characteristics of an entrepreneur in relation to risk taking: Czech Republic case study. *Journal of International Studies, 10*(3), 220-233. https://doi.org/10.14254/2071-8330.2017/10-3/16

Lewis, C. (2002). *Lesson study: A handbook for teacher-led instructional change.* Philadelphia: Research for better schools.

Lewis, C.C., & Hurd, J. (2011). *Lesson study step by step*. Heinemann.

Lewis, C.C., & Lee, C. (2017). The global spread of lesson study. M. Akiba & G. LeTendre (Eds.), In *International Handbook of Teacher Quality and Policy* (pp.185-195). Routledge.

Lewis, C., & Perry, R. (2014). Lesson study with mathematical resources: A sustainable model for locally-led teacher professional learning. *Mathematics Teacher Education and Development, 16*(1), 22–42

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gotzsche, P.C., Ioannidis, J.P.A. Moher, D. (2009). The PRISMA statement for reporting systematic reviews and metaanalyses of studies that evaluate health care interventions: Explanation and elaboration. *BMJ, 339*, 27-32.

Liu, K. (2021). On the construction of teachers' Professional quality – oriented english practice teaching system – Exemplified with the english major of Sichuan University of arts and science. *Theory and Practice in Language Studies, 11*(4), 390-395. https://doi.org/10.17507/tpls.1104.08

McCallum E., Weicht R., McMullan L., & Price A. (2018). *EntreComp into Action: get inspired, make it happen* (M. Bacigalupo & W. O'Keeffe Eds.), Luxembourg. ISBN 978-92-79-79360- 8, https://doi.org/10.2760/574864JRC109128

Millî Eğitim Bakanlığı (MEB). (2019). 2023 eğitim vizyonu [2023 educational vision]. https://2023vizyonu.meb.gov.tr/doc/2023_EGITIM_VIZYONU.pdf

Moberg, K. (2014). *Assessing the impact of entrepreneurship education*. [Unpublished dissertation of doctorate]. Copenhagen Business School.

Mulrow, C.D. (1994). Systematic reviews: Rationale for systematic reviews. *BMJ, 309*(6954), 597–599.

Murata, A. (2011). Introduction: Conceptual overview of Lesson Study. In L. C. Hart, A. Alston & A. Murata (Eds.), *Lesson Study research and practice in mathematics education. Learning together* (pp. 1-12). Springer

Ono, Y., & Ferreira, J. (2010). A case study of continuing teachers Professional development through lesson study in South Africa. *South African Journal of Education, 30*, 59-74.

Özbey, N. (2022). *Blok tabanlı öğretim programının bilişsel olmayan girişimcilik yetkinlikleri üzerine etkililiğinin incelenmesi [ Investigation of effects of block based curriculum upon non-cognitive entrepreneurship competenciens]*.[Unpublished dissertation of doctorate]. Atatürk University.

Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., & Mulrow, C.D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ, 372*(71). https://doi.org/10.1136/bmj.n71

Politis, D. (2005). The process of entrepreneurial learning: A conceptual framework. *Entrepreneurship Theory and Practice, 29*(4), 399-424.

Raab, G., Stedham, Y., & Neuner, M. (2005). Entrepreneurial potential: An exploratory study

of business students in the US and Germany. *Journal of Business and Management, 11*(2), 71-88.https://doi.org/10.1504/JBM.2005.141135

Rahmatullah, Sariakin, Akmaluddin. (2022). Integrated socio-entreprenerurship in cultural analysis and validation semester learning plan. *Qalamuna – Jurnal Pendidikan, Social, dan Agama, 14*(2), 567-588. *https://doi.org/10.37680/qalamuna.v14i2.3440*

Saykal, A., & Sağır-Uluçınar, Ş. (2021). Türkiye'de öğretmen yeterlikleri ve teknolojik pedagojik alan bilgisi araştırmaları [The research about teachers' competencies and technological pedagogical field knowledge in Türkiye]. *Gazi Journal of Educational Science, 7*(2), 115-137.

Serbest, A. (2014). *Ders imecesinin etki alanları üzerine bir meta-sentez çalışması* [A metasyntesis on effects of lesson study] [Unpublished master thesis]. Karadeniz Technical University.

Shane, S., & Venkataraman, S. (2000). *The promise of entrepreneurship as a field of research. Academy of Management Review, 25*(1), 217–226.

Stepanek, J., Appel, G., Leong, M., Mangan, M.T., & Mitchell, M. (2007). *Leading Lesson Study*. The United States of America: Corwinn.

Swanson, L.A. (2017). *Entrepreneurship and Innovation Toolkit*. https://openlibrary-repo.ecampusontario.ca/jspui/bitstream/123456789/634/6/Entrepreneurship-and-Innovation-Toolkit-1568142286.pdf

Türkiye Yeterlilikler Çerçevesi (TYÇ). (2018). Anahtar Yetkinlikler Tavsiye Karari [Key competencies advice decision]. https://www.myk.gov.tr/images/articles/TYC/Yayinlar/Hayat_Boyu_Ogrenme_icin_

Uştuk, Ö., & Çomoğlu, İ. (2019) Lesson study for professional development of English language teachers: Key takeaways from international practices. *Journal on Efficiency and Responsibility in Education and Science, 12*(2), 41-50.

Widiastuti, A., Supriatna, N., & Nurbayani, S. (2023). Development of learning models to increase entrepreneurship competency. *Malaysian Online Journal of Educational Management, 11*(2), 1-17.

Willems, I., & Bossche, P.V. (2019). Lesson Study effectiveness for teachers' professional learning: a best evidence synthesis. *International Journal for Lesson and Learning Studies, 8*(4), 257-271

Xiao, Y., Surasin, J., & Prabjandee, D. (2020). Developing a training modüle to improve initial ELT proficiency among student-teachers in multi-ethnic community schools. *HRD Journal, 11*(1), 90-106.

Verma, M., & Chaurasia, A. (2020). Entrepreneurship and innovation. *International Journal for Multidisciplinary Research, 2*(4).

Zhang, X., & Zhang, K. (2018), The relation between creativity and entrepreneurial intention: a moderated mediating effect model. *Foreign Economies and Management, 40*, 67-78.

Zhiyong, S., Muthukrishnan, P., & Sidhu, G.K. (2020). College english language teaching reform and key factors determining EFL teachers' Professional development. *European Journal of Educational Research, 9*(4), 1393-1404. https://doi.org/10.12973/eu-jer.9.4.1393

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 240–276
https://doi.org/10.21449/ijate.1510749
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# A systematic review of the recommendations of theses on fine arts high school music departments in Türkiye

**Haşim Vapur**[1], **Ülkü Sevim Şen**[2*]

[1]Atatürk University, Institute of Educational Sciences, Department of Fine Arts Education, Department of Music Education PhD Student, Erzurum, Türkiye
[2]Atatürk University, Kazım Karabekir Faculty of Education, Department of Music Education, Erzurum, Türkiye

**ARTICLE HISTORY**
*Received: July 4, 2024*
*Accepted: Dec. 12, 2024*

**Keywords:**
Fine arts high school,
Music department,
Systematic literature review.

**Abstract:** This study aims to systematically review the recommendations of theses focusing on music departments in Turkish FAHS. The study is considered unique and significant for examining the needs of FAHS music education from a different perspective. The research employed a systematic literature review method. A systematic literature review is a structured and comprehensive synthesis of numerous studies using similar methods to identify the most robust available research evidence by experts in the field. The research data sources are the theses on music education in FAHS published in the database of the National Thesis Center of the Council of Higher Education (CoHE) between 2005 and 2023. According to the inclusion and exclusion criteria of the study, 159 theses were interpreted by content analysis. It was determined that under the education theme of the theses examined; school, course, material and equipment, student, curriculum, and teacher dimensions were presented. In the discussion section, the codes regarding these dimensions were analyzed and interpreted. When the recommendations were analyzed according to the codes, it was determined that the majority covered issues that should be resolved by the Ministry of National Education (MoNE). Additionally, there are recommendations for the principals and teachers working in FAHS. It is essential that these stakeholders should address and resolve the issues to improve the efficiency of education and increase academic success.

## 1. INTRODUCTION

Music is a field that presents sounds within an aesthetic structure through its facets of phonetic and instrumental performance, facilitating the expression of feelings/emotions. Throughout history, music has served as a vehicle for self-expression, articulating emotions such as joy, sorrow, love, and enthusiasm. Due to its impact and contribution to learning, music has also been utilized as an educational tool. “Music education is generally the process of imparting musical behaviors to individuals and modifying these behaviors” (Uçan, 2005, p. 14). Seyhan (2019) asserts that music education enriches individuals’ daily activities and fosters social cohesion (p. 9). Through music education, individuals improve their vocal performance,

*CONTACT: Ülkü Sevim ŞEN usevimsen@atauni.edu.tr Ataturk University, Kâzım Karabekir Faculty of Education, Department of Music Education, Erzurum, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

e-ISSN: 2148-7456

instrumental competencies, and listening skills in music, emerging as good listeners and academically qualified individuals.

In Türkiye, music education is organized under three main categories. These are general, amateur, and professional music education. General music education is provided within the compulsory education framework of the 4+4+4 system. "General music education aims to develop students' talents and foster a collective consciousness of music culture among them. Thus, music's national and universal characteristics are imparted to children, and intellectual individuals with an affinity for music are raised" (Tokatlı, 2020, p. 14). "Amateur music education is pursued as a hobby based on individual preferences, whether through paid or free lessons from private or public institutions. This voluntary-based education encourages individuals to utilize their leisure time effectively, engage in musical interactions, and derive spiritual satisfaction through music" (Sezen & Özerdem, 2021, p. 1696). Professional music education, on the other hand, is targeted at individuals aspiring to pursue music as a career. This type of education involves providing professional training to students selected through special talent examinations, aiming to raise experts in their field. In Türkiye, professional music education is offered in fine arts high schools (FAHS), music departments of universities' fine arts faculties, education faculties' music education departments, conservatories, and various institutions such as military bands.

Music departments in FAHS provide foundational professional music education. As of 2024, there are 105 FAHS in Türkiye. "Fine Arts High Schools hold the status of Anadolu High Schools and offer a four-year educational program. Many of these schools are boarding and offer daytime co-educational instruction. They are primarily established in locations where higher education institutions related to fine arts are situated. Instruction is conducted in Turkish, and students are taught the foreign language equivalents of art-related terms within their specialized courses. Admission to Fine Arts High Schools is based on talent examination scores and Secondary Education Achievement Scores, with placement determined by score ranking" (Ministry of National Education [MoNE], 2024). "In the talent examination, students are presented with questions that measure musical aptitude, such as the repetition of single, double, triple, and quadruple tones, chord progressions, rhythm and melody repetition, and similar assessments. Based on the examination results, students are placed into the schools they have applied to according to their ranking. The number of classes opened each year for 9th grade cannot exceed two per field. In a given academic year, the number of students admitted to each fine arts-related field cannot exceed 30" (MoNE, 2024). "Fine arts high schools, one of the most widespread institutions offering vocational music education in Türkiye, contribute to the cultivation of qualified individuals in the field by imparting fundamental knowledge and skills in fine arts to students, alongside the general objectives of secondary education" (MoNE, 2019).

According to MoNE (2009), the objectives of Anatolian FAHS are as follows:

- To provide education and training in fine arts based on students' interests and talents,
- To prepare students for higher education programs that require special talents,
- To encourage students to engage in research in their fields,
- To enable students to make interpretations and applications in line with their talents, thus nurturing them to become creative and productive individuals,
- To facilitate the understanding and interpretation of national and international works of art.

Graduates of FAHS have the opportunity to pursue further music education in the Music Education Department of Education Faculties, the Music Sciences Department of Fine Arts Faculties, Conservatories, the Military Band NCO Vocational School of the National Defense University, and Music and Performing Arts Faculties.

An examination of the literature reveals that studies focusing on fine arts high schools have investigated various areas such as performance anxiety, the utilization of music technologies, students' perceptions of Turkish folk music, and physiological disorders encountered by

students receiving instrument training (Çalışır, 2024; Gerekten & Ahmethan, 2018; Tuncay, 2019; Yalçın, 2019). Beyond these topics, there are various studies conducted using different methodologies regarding the issues experienced in vocational courses in the field of music (e.g., instrument education, elective piano, musical hearing-reading-writing) in FAHS (Apaydınlı, 2009; Derin, 2007; Kaynak-Akçaoğlu, 2021; Mustan, 2010; Soğukçam, 2007; Topalak, 2013; Uluç, 2006). Additionally, there are studies on extracurricular factors such as school management, teacher opinions, and distance education within the framework of FAHS. (Akçay & Sürmeli, 2022; Ayaydın, 2011; Umuzdaş, 2006; Umuzdaş & Umuzdaş, 2009; Yılmazçal & Küçüksüleymanoğlu, 2022).

The review revealed a lack of music research conducted using the systematic literature review method in Türkiye. Additionally, the absence of studies explicitly focusing on the recommendations sections of prior scientific research necessitated the undertaking of this study. The research problem statement was formulated as "What are the recommendations in the theses that focus on the music departments in FAHS in Türkiye?" The study sought to address the following research questions:

- What recommendations have been proposed under the "education theme" in these theses?
- What are the sub-categories within these theses?
- How are these recommendations distributed across the identified sub-categories?

### 1.1. Aim and Significance

This study aims to systematically examine the recommendations of theses on music departments of FAHS in Türkiye. The study is significant as it systematically examines the recommendations presented in theses within this field, thereby contributing to the identification of deficiencies or needs in the educational processes of FAHS.

## 2. METHOD

In this study, a systematic literature review method was employed. "Systematic review is a structured and comprehensive synthesis of numerous studies conducted with similar methods by experts in the field to determine the best available research evidence" (Karaçam, 2013, p. 26). "A systematic literature review, derived from information science, provides a standardized methodology that offers clear and replicable methods for identifying, evaluating, and synthesizing existing research literature" (Jesson *et al*., 2011, p. 46; Okoli & Schabram, 2010, p. 4). "Systematic reviews can integrate data from different research studies to produce a new integrated conclusion or combine different types of evidence to explore or explain meaning" (Snilstveit *et al*., 2012, p. 410). This method was chosen since the study aimed to systematically examine the recommendations in the theses published between 2005 and 2023 in the field of FAHS music departments. Theses published before 2005 were excluded from the study due to the gradual transition to the constructivist curriculum in Türkiye since that year. The year 2024 was also excluded from the study due to ongoing thesis submissions throughout that year.

This study used the systematic literature review method, utilizing the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist. Similar to other reporting guidelines, PRISMA enhances the quality of systematic review reports. "PRISMA also has the potential to enhance the execution of systematic reviews, as suggested by other reporting guidelines, by aiding authors in documenting the predetermined roadmap of their systematic reviews" (Moher *et al*., 2010, p. 337).

### 2.1. Data Collection Process

In the study's data collection process, a search string combination was initially formed using the English and Turkish keywords "Güzel sanatlar lise" and "Fine arts high school." Following various search combinations, the Turkish phrase "Güzel sanatlar lise" was found to be adequate and comprehensive. Therefore, this Turkish keyword group exclusively formed the basis of the study's literature search. In order to determine the relevant studies representing data sources,

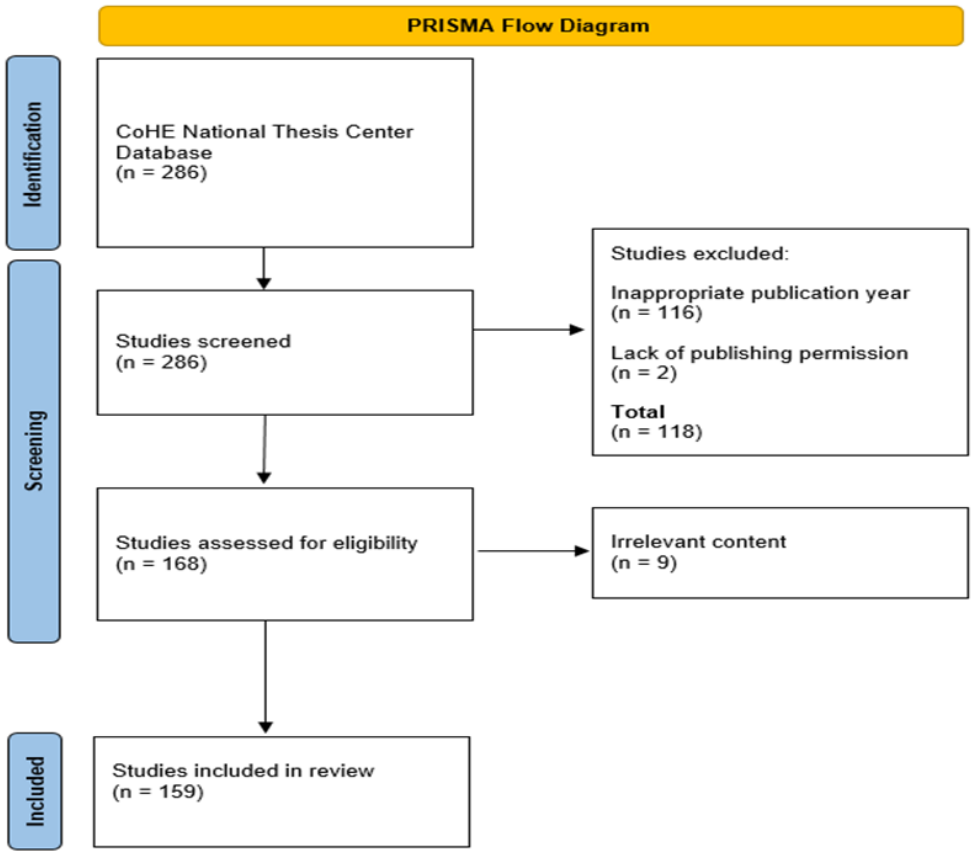
they were initially accessed by typing the word "müzik" (music) in the topic field after typing the keywords. Table 1 presents the search criteria used during the data collection process.

**Table 1.** *Inclusion and exclusion criteria.*

| Inclusion Criteria | Exclusion Criteria |
|---|---|
| ▪ Theses published in Turkish and English | ▪ Theses published in languages other than Turkish and English |
| ▪ Theses published between 2005 and 2023 | ▪ Theses published before 2005 or after 2023 |
| ▪ Theses published in the Council of Higher Education (CoHE) National Thesis Center database | ▪ Theses published outside the CoHE National Thesis Center database |
| ▪ Theses published in the field of music education in FAHS | ▪ Theses published outside the field of music education in FAHS |

The literature review and systematic reviews were reported and presented according to the PRISMA flowchart. The PRISMA flowchart allows authors to specify how many records are excluded before screening (e.g., duplicates removed or flagged as inappropriate by automated tools), indicate the number of full-text reports and how many are unavailable, detail how many studies and reports from a previous review are included in the current iteration of review, and offer the option to present the flow of records through the review process segmented by source type (e.g., bibliographic databases, websites, organizational and citation searching) (Haddaway *et al*., 2022, p. 2). Figure 1 in this study illustrates the PRISMA flow diagram depicting the data elicitation process.

**Figure 1.** *PRISMA flow diagram.*



The searches revealed 286 accessible studies published in the Council of Higher Education (CoHE) National Thesis Centre database. Initially, 116 studies were excluded since they did not meet the inclusion criteria for the publication years between 2005 and 2023. Subsequently, 170 theses were reviewed in both abstract and full-text forms, and 9 of them were excluded due to their lack of relevance in content and subject scope. Additionally, two theses were excluded from the study due to a lack of permission for publication in the database. Overall, 159 theses were included in the systematic review for data analysis.

### 2.2. Data Analysis

The qualitative analysis method was used to examine the findings related to the scope of this research. The data obtained for the research questions were analyzed using content analysis. "The fundamental process in content analysis involves grouping similar data within the framework of specific concepts and themes, then organizing and interpreting them in a manner comprehensible to the reader" (Yıldırım & Şimşek, 2006, p. 227). The final studies selected according to the inclusion and exclusion criteria were analyzed using MAXQDA software. Themes related to the research questions were examined, and corresponding codes were developed. The categorized studies were labeled as Study1-Study2-Study3...Study159 (Appendix 1). In the data analysis, multiple responses were utilized as more than one recommendation from same studies were used; thus, the total of 159 studies was considered as 100% (see Table 4).

### 2.3. Validity and Reliability

"In qualitative research, validity and reliability refer to the procedures researchers undertake to ensure that what they notice or hear can be trusted" (Fraenkel *et al*., 2012, p. 458). Expert opinions were sought for the validity and reliability measures used in the study. Feedback was obtained from two experts in the field of Music Education (Prof. Dr. from Niğde Ömer Halisdemir University and Assoc. Prof. Dr. from Atatürk University) and one expert in the field of Educational Sciences (Prof. Dr. from Kafkas University). An expert evaluation form consisting of open-ended questions was utilized to gather expert opinions. "In the expert evaluation form, questions regarding the responses to be obtained from the expert should be clearly written, and a two-option answer format of 'appropriate/valid' and 'appropriate/invalid' can be used to determine the expert's opinions on the suitability of the questions" (Büyüköztürk, 2005, p. 112). Additionally, to enhance the validity and reliability of the study's analysis, repeated searches were conducted in the CoHE National Thesis Center database at different time intervals. The inclusion and exclusion criteria specified for all synthesized studies in the research are presented in Table 1.

### 2.4. Quality Assessment Process

The studies included in the systematic literature review were evaluated according to a checklist taken from Dybå & Dingsøyr (2008, p. 839) and translated into Turkish by Önaçan (2020, p. 358) to minimize bias and achieve reliable results. The following criteria were used in the quality assessment process of the included studies:

- Is the study presented in the form of a scientific report?
- Is the research objective clearly stated?
- Is the scope of the research clearly defined?
- Does the research contribute to the field?

Based on the quality assessment criteria, each study's methodological reliability and relevance were given an overall rating of "low," "medium," or "high." These studies are presented in Appendix 1.

## 3. FINDINGS

The subcategories determined after synthesizing the recommendations of the theses as a result of the systematic literature review and content analysis are shown in Figure 2. It shows that the identified subcategories for the synthesis related to the education theme consist of school, course, material and equipment, student, curriculum, and teacher dimensions. Data related to these dimensions under the education theme are specified in the following tables.

**Figure 2.** *Subcategories determined for the synthesis related to the education theme.*

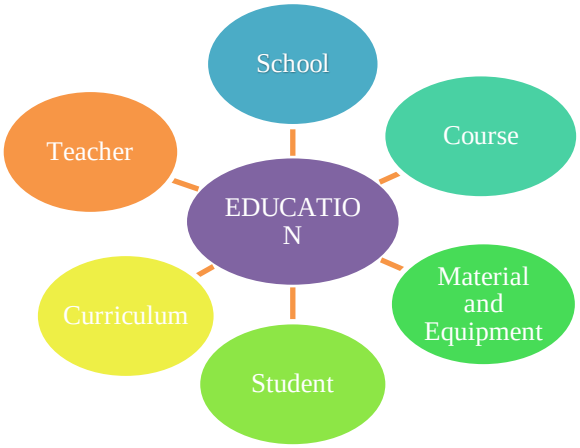



Figure 3 shows a word cloud that depicts the codes created from the subcategories identified after analyzing and synthesizing the recommendations sections of the theses. These subcategories were obtained through a systematic literature review and content analysis to obtain research data.

**Figure 3.** *Word cloud depicting the codes generated from subcategories.*

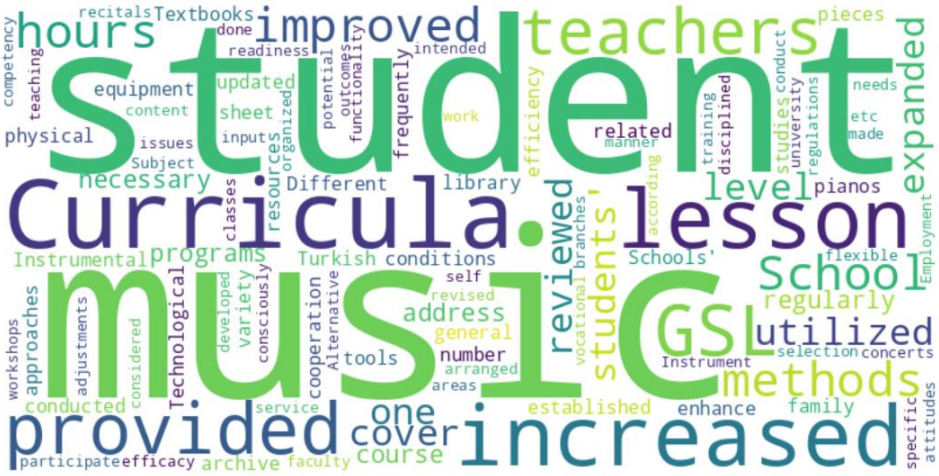



Table 2 shows the recommendation codes regarding the school dimension. According to the findings, 13.2% of the studies (21 theses) recommend that *the physical conditions of schools should be renovated*; and 8.8% of the studies (14 theses) recommend that *parent-teacher-student association should be established*.

**Table 2.** *Codes identified for the school dimension.*

| Codes | Studies | *f* | % |
|---|---|---|---|
| The physical conditions of schools should be renovated. | 2, 6, 10, 11, 14, 15, 17, 22, 27, 35, 36, 45, 55, 56, 60, 72, 78, 84, 86, 127, 159 | 21 | 13.2 |
| Parent-teacher-student association should be established. | 10, 11, 15, 19, 20, 48, 94, 95, 105, 111, 115, 116, 134, 149 | 14 | 8.8 |

Table 3 shows the recommendation codes regarding the course dimension. According to the findings, 27% of the studies (43 theses) recommend that *the number of course hours in the field of music should be increased*; 18.2% (29 theses) recommend that *different methods and approaches should be utilized to enhance the efficiency of lessons*; 13.2% (21 theses) recommend that *Turkish music studies should be increased*; 3.1% (5 theses) recommend that *the class hours should be reviewed*; and 1.8% (3 theses) recommend that i*nstrumental lessons should be conducted one-on-one.*

**Table 3.** *Codes identified for the course dimension.*

| Codes | Studies | *f* | % |
|---|---|---|---|
| The number of course hours in the field of music should be increased. | 6, 11, 14, 22, 23, 31, 32, 33, 41, 43, 45, 46, 47, 48, 49, 53, 56, 57, 60, 64, 65, 66, 67, 72, 74, 78, 84, 88, 93, 109, 111, 123, 124, 125, 127, 129, 132, 142, 147, 151, 153, 157, 159 | 43 | 27 |
| Different methods and approaches should be utilized to enhance the efficiency of lessons. | 4, 12, 13, 14, 16, 22, 26, 28, 42, 53, 59, 83, 108, 116, 117, 118, 120, 123, 129, 142, 144, 146, 147, 149, 151, 152, 153, 154, 157 | 29 | 18.2 |
| Turkish music studies should be increased. | 14, 29, 45, 49, 51, 56, 59, 64, 69, 76, 77, 83, 92, 108, 110, 118, 121, 127, 133, 141, 151 | 21 | 13.2 |
| The class hours should be reviewed. | 2, 11, 14, 15, 40 | 5 | 3.1 |
| Instrumental lessons should be conducted one-on-one. | 56, 124, 159 | 3 | 1.8 |

Table 4 shows the recommendation codes regarding the material and equipment dimensions. According to the findings, 24.5% of the studies (39 theses) recommend that s*chools should be provided with equipment and materials support*; 16.3% (26 theses) recommend that *the archive of books and sheet music should be expanded*; 10.6% (17 theses) recommend that *textbooks should be updated to cover the general level of students*; 5% (8 theses) recommend that *technological tools should be utilized*; 4.4% (7 theses) recommend that *the number of pianos should be increased*; and 2.5% (4 theses) recommend that *the variety of works given to students should be increased*.

**Table 4.** *Codes identified for material and equipment dimension.*

| Codes | Studies | *f* | % |
|---|---|---|---|
| Schools should be provided with equipment and materials support. | 1, 2, 4, 6, 11, 15, 17, 22, 25, 27, 29, 32, 35, 40, 41,64, 48, 49, 50, 51, 64, 67, 72, 78, 87, 99, 103, 105, 108, 114, 117, 127, 132, 135, 136, 137, 146, 147, 151, 157 | 39 | 24.5 |
| The archive of books and sheet music should be expanded. | 1, 2, 4, 6, 15, 22, 27, 41, 46, 48, 49, 50, 51, 64, 67, 78, 87, 99, 105, 108, 114, 132, 136, 146, 147, 151 | 26 | 16.3 |
| Textbooks should be updated to cover the general level of students. | 43, 83, 87, 90, 91, 93, 95, 99, 102, 104, 108, 110, 111, 113, 114, 158, 159 | 17 | 10.6 |
| Technological tools should be utilized. | 6, 10, 62, 103, 127, 132, 146, 153 | 8 | 5 |
| The number of pianos should be increased. | 1, 15, 22, 35, 49, 64, 127 | 7 | 4.4 |
| The variety of works given to students should be increased. | 1, 14, 49, 157 | 4 | 2.5 |

Table 5 shows the recommendation codes regarding the student dimension. According to the findings, 27% of the studies (43 theses) recommend that s*tudents should regularly participate in concerts, recitals, workshops, and similar events;* 6.9% of the studies (11 theses) recommend that *instrument selection should be made consciously*; 4.4% of the studies (7 theses) recommend that s*tudents' attitudes towards lessons, readiness levels and self-efficacy levels should be taken into consideration;* and 3.7% of the studies (6 theses) recommend that *students should study with discipline*.

**Table 5.** *Codes identified for student dimension.*

| Codes | Studies | *f* | % |
|---|---|---|---|
| Students should regularly participate in concerts, recitals, workshops, and similar events. | 1, 6, 10, 11, 14, 15, 20, 23, 27, 31, 32, 36, 37, 40, 43, 46, 49, 51, 54, 61, 64, 69, 72, 81, 85, 86, 91, 92, 93, 95, 98, 105, 107, 122, 123, 127, 131, 142, 147, 149, 151, 155, 158 | 43 | 27 |
| Instrument selection should be made consciously. | 8, 11, 31, 34, 37, 73, 91, 97, 126, 138, 146 | 11 | 6.9 |
| Students' attitudes towards lessons, readiness levels and self-efficacy levels should be taken into consideration. | 30, 123, 137, 147, 147, 150, 155 | 7 | 4.4 |
| Students should study with discipline. | 7, 100, 107, 126, 127, 155 | 6 | 3.7 |

**Table 6.** *Codes identified for curriculum dimension.*

| Codes | Studies | *f* | % |
|---|---|---|---|
| The curriculum should be improved in terms of the outcomes it covers. | 23, 27, 39, 47, 52, 57, 63, 64, 68, 81, 84, 118, 119, 123, 124, 128, 133, 136, 138, 139, 146 | 21 | 13.2 |
| The functionality of the curriculum should be regularly examined and the necessary modifications should be made. | 7, 17, 22, 41, 43, 51, 57, 65, 101, 113, 128, 129, 130, 136, 140, 143, 145, 148, 152, 157 | 20 | 12.5 |
| The curriculum should be designed with input from university faculty and FAHS teachers. | 2, 21, 23, 27, 31, 32, 48, 56, 60, 64, 67, 69, 71, 112, 127, 132, 152, 156 | 18 | 11.3 |
| The curriculum prepared for music vocational courses should address all areas of competence. | 9, 16, 79, 87, 96, 108, 115, 119, 159 | 9 | 5.6 |
| FAHS regulations should be revised according to needs. | 18, 24, 27, 32, 53, 93, 109 | 7 | 4.4 |
| Alternative programs, course content, and flexible teaching methods should be developed for potential issues in the curriculum. | 12, 13, 66, 81, 82, 124 | 6 | 3.7 |

Table 6 shows the recommendation codes regarding the curriculum dimension. According to the findings, 13.2% of the studies (21 theses) recommend that *the curriculum should be improved in terms of the outcomes it covers*; 12.5% of the studies (20 theses) recommend that *the functionality of the curriculum should be regularly examined and the necessary modifications should be made*; 11.3% of the studies (18 theses) recommend that *the curriculum should be designed with input from university faculty and FAHS teachers*; 5.6% of the studies (9 theses) recommend that *the curriculum prepared for music vocational courses should address all areas of competence*; 4.4% of the studies (7 theses) recommend that *FAHS regulations should be revised according to needs;* 3.7% of the studies (6 theses) recommend

that *alternative programs, course content, and flexible teaching methods should be developed for potential issues in the curriculum.*

Table 7 shows the recommendation codes regarding the teacher dimension. According to the findings, 22% of the studies (35 theses) recommend that *regular in-service training sessions should be organized for FAHS music teachers*; 14.4% of the studies (23 theses) recommend that *employment should be provided for the needed branch*; 12.5% of the studies (20 theses) recommend that *subject teachers should conduct the classes.*

**Table 7.** *Codes identified for teacher dimension.*

| Codes | Studies | *f* | % |
|---|---|---|---|
| Regular in-service training sessions should be organized for FAHS music teachers. | 2, 3, 4, 6, 18, 24, 25, 26, 32, 43, 44, 46, 51, 58, 62, 64, 67, 70, 75, 80, 84, 91, 93, 103, 107, 111, 127, 132, 138, 139, 143, 146, 154, 157, 159 | 35 | 22 |
| Employment should be provided for the needed branch. | 1, 2, 5, 6, 11, 14, 15, 17, 27, 37, 38, 41, 49, 55, 64, 72, 86, 91, 104, 127, 132, 158, 159 | 23 | 14.4 |
| Subject teachers should conduct the classes. | 1, 5, 14, 17, 22, 23, 29, 32, 33, 43, 50, 58, 67, 74, 89, 93, 117, 127, 132, 153 | 20 | 12.5 |

## 4. DISCUSSION and CONCLUSION

In conclusion, the recommendations under the education theme in the theses were structured around key codes encompassing school, course, material and equipment, student, curriculum, and teacher dimensions. These codes were interpreted in the light of the related literature.

• The conclusions related to the school dimension are as follows:

*Regarding the physical conditions of schools should be renovated*; environmental conditions such as structural integrity of buildings, number and layout of classrooms, lighting, heating systems, playgrounds, and green areas are among the factors that constitute the physical adequacy of schools. “An appropriate physical environment is a prerequisite for the effective implementation of educational activities” (Demirtaş & Küçük, 2014, p. 54). “School buildings serve as symbols and indicators of a city and community” (Arslan-Karaküçük, 2008, p. 318). Döş (2013) suggests that “schools should be planned suitable for social and artistic activities in terms of physical spaces, and schoolyards and classrooms should be designed according to the needs and expectations of students” (p. 278). From these perspectives, it is evident that enhancing the physical infrastructure of FAHS and new school constructions, which include study rooms, concert halls, recording studios, gymnasiums, dormitories, appropriate environmental landscaping, and security measures aligned with local contexts, are crucial factors for enhancing the functionality of education and teaching practices.

*Regarding the parent-teacher-student association*; due to the extended duration of the learning process, adjustments that positively influence the school atmosphere should be made in line with the students’ interests during this period. School, family, and student collaboration should be established to strengthen the communication network and processes that increase student achievement and support their holistic development should be integrated into learning environments through this robust partnership. Organizing this collaboration in a planned manner is crucial for monitoring the process effectively (Smith *et al*., 2019, p. 517). Aslanargun (2007) argues that “establishing effective collaboration among school, family, and students is crucial for improving students’ academic success and better preparing them for life” (p. 132). By fostering a sense of shared responsibility and solidarity, it is possible to develop solutions to address existing challenges and enhance the efficiency of education and teaching practices.

• The conclusions related to the course dimension are as follows:

*The number of course hours in the field of music should be increased* was the most emphasized suggestion in the course dimension. Since the establishment of FAHS in Türkiye, the MoNE has made numerous changes in the weekly distribution of courses. Compulsory courses in some periods have later become elective or have been completely removed from the curriculum and replaced with different courses according to needs. According to changes made by the MoNE Board of Education in 2020, the field of music was divided into three distinct departments: Music, Turkish Folk Music, and Turkish Classical Music, within the context of FAHS (MoNE, 2023). Schools can activate or deactivate these departments based on their qualifications. With the recent amendment, some music field course hours have been increased. Previously, the 9th-grade instrumental education course was scheduled for one hour weekly; however, this was extended to four hours weekly in 2020. Overall, in 2018, 9th-grade music courses were scheduled for six hours weekly, whereas the revised program now provides ten hours per week. These developments highlight the efforts of MoNE to increase the number of music course hours. Examining the opinions of education stakeholders and conducting additional scientific research can reveal the adequacy of these advancements.

*Regarding different methods and approaches should be utilized to enhance the efficiency of lessons*; in today's education system, methods and approaches that encourage students to actively discover information, apply it to their daily lives, and use it to solve problems are widely utilized. Contents structured according to cognitive, affective, and kinesthetic domains are interconnected in line with objectives and integrated into the teaching-learning process. "The teacher's teaching goal and the student's learning goal are the same, yet their expressions are different" (Ergashevich, 2024, p. 266). Therefore, it is crucial for teachers to carefully monitor the process using assessment tools and utilize different methods and approaches in deficient areas to enhance the efficiency of education.

*Regarding Turkish music studies should be increased*; the effective use of cultural elements in the learning process can be considered a prerequisite for forming a shared societal consciousness. Music undertakes a significant role in reflecting the collective emotions of society and bridging connections with other societies through sounds. In recent years, with the advancements in music education at FAHS, the intensity of Turkish music studies in instrument education textbooks has increased. Additionally, courses such as Turkish music theory and practice, Turkish Islamic music, music from the Turkic world, and the history of Turkish and Western music (MoNE, 2023) have been incorporated into the curriculum, contributing to advancements in this field. "There are also perspectives arguing that the repertoire oriented towards Turkish music in Western music instrument training courses is insufficient" (Doğru, 2022).

*Regarding the class hours should be reviewed*; it was observed in FAHS weekly course schedules that in the 9th grade, out of 40 weekly hours of standard courses, only 10 hours are music-field courses. Similarly, in the 10th grade, out of 42 weekly hours, only 10 hours are dedicated to music. Moving on to the 11th grade, 17 out of 36 weekly hours are allocated to music-related subjects, and in the 12th grade, 21 out of 36 weekly hours are dedicated to music (MoNE, 2023). Based on these ratios, it can be argued that the music-related course hours in the 9th and 10th grades are insufficient and should be increased. A. Özen's (2023) finding that the weekly course hours for instrument lessons are limited provides evidence that the planning of course hours need to be reviewed.

*Regarding the instrumental lessons should be conducted one-on-one*; individualized instruction in music education is crucial for students to showcase their performance abilities effectively and express themselves comfortably (Çaylak, 2022, p. 66). Given the acknowledged shortage of specialized music teachers in FAHS in Türkiye, it is necessary to appoint subject teachers to ensure the personalized delivery of instrument lessons.

• The conclusions related to the material and equipment dimension are as follows:

*Regarding schools should be provided with equipment and materials support*; each resource utilized during the teaching and learning process is an educational tool. "Teaching aids aim to impart lasting knowledge to students by expressing events or entities. Teachers should use appropriate tools and materials to facilitate learning" (Bayraktar, 1994, p. 24). In music education, instruments, sheet music, music stands, metronomes, music software, and similar tools are employed as teaching aids. The MoNE and school administrations must review the equipment conditions of FAHS to eliminate deficiencies and enhance the quality of education.

*Regarding the archive of books and sheet music should be expanded*; all types of written and visual materials used in education are significant elements in enhancing the diversity of information. Exposure to various stimuli strengthens students intellectually, enhances their problem-solving abilities, and fosters their creativity. Therefore, it is crucial for schools to have online and physical libraries. "School libraries serve as highly functional learning environments for students to access valid and reliable information" (Durukan, 2015, p. 108). Establishing a national online archive for books and sheet music would promote equal educational opportunities and ensure students have access to a reliable source of information.

*Regarding textbooks should be updated to cover the general level of students*; books, methods, and educational information used in music education should provide a comprehensive integration of cognitive, affective, and psychomotor skills for students. "The proportional and balanced content structure of textbooks used in education is crucial" (Akbaş *et al*., 2021, p. 3854). M. Barış's (2023) finding that instrument training textbooks are not adequately aligned with the general proficiency levels of students, leading both students and teachers to utilize supplementary resources, underscores a significant problem in this area. "The textbooks for FAHS are prepared according to the MoNE Textbooks and Educational Tools Regulation by a commission composed of various experts in the field. The input from different subject specialists and music teachers in FAHS can be considered significant in developing more high-quality and comprehensive textbooks".

*Regarding technological tools should be utilized*; the use of technology in education has become increasingly widespread. "Educational technology is the systematic application of technology products developed in the field of sciences and knowledge accumulated in behavioral sciences to educational activities" (Alpar *et al*., 2007, p. 25). "Interactive whiteboards in schools have become indispensable technological tools for modern education. There are numerous beneficial software programs available for music education, providing both teachers and students with opportunities in various study levels. Additionally, software that seamlessly interface with sound cards on computers and applications accessible via smart devices such as tablets and phones significantly augment music education" (Nart, 2016, p. 81). N. Barış (2022) determined that the technological infrastructure in schools is inadequate and noted that students experience difficulties in practical areas such as computer-assisted music.

*Regarding the number of pianos should be increased*; one pivotal factor influencing success in instrumental education is the possession of the instrument to be trained. "Piano education is one of the fundamental principles of instrument education. Due to its ability to enable polyphonic music performance, the piano holds a distinct position compared to other instrument groups" (Ertem, 2011, p. 646). Given their structural nature, the piano lessons necessitate individualized instruction. It is acknowledged that in FAHS, each teacher typically provides piano instruction to two or more students within their classrooms. Additionally, considering that many of these institutions offer boarding facilities, it is crucial to ensure that there are dedicated piano practice rooms within dormitory buildings. Hence, the oversight and remediation of deficiencies in piano availability in FAHS and their affiliated dormitories hold significant importance.

*Regarding the variety of works given to students should be increased*; the effective implementation of instrumental education requires teachers to undertake various

responsibilities. In this process, the diversity of works presented by considering students' levels of readiness is as crucial as the discipline and methods. "Music education holds the potential to contribute to individuals' physical, mental, and spiritual development. It aids in fostering behavioral changes that promote adaptability and sensitivity to the societal and cultural environment in which individuals live" (Baydağ, 2020, p. 388). Working with a diverse repertoire relevant to their fields assists students in achieving the outcomes. Additionally, teachers can enhance student development by incorporating different works based on their own experiences beyond those provided in textbooks published by MoNE.

• The conclusions related to the student dimension are as follows:

*Regarding students should regularly participate in concerts, recitals, workshops, and similar events*; it is crucial for students to engage in such activities due to the performative practical of music. "Music is an important art form with its social and unifying qualities. Educational institutions should be able to respond to students' social needs beyond their educational requirements" (Dinçer, 2009, p. 42). "Educational environments must be intertwined with real-life experiences. Making the teaching stage concrete and meaningful is possible through applications both within and outside the school. Experiments and observations that support learning processes through multiple learning environments and experiential learning activities have significant impacts on learning" (Girgin *et al*., 2011, pp. 9-10).

*Regarding instrument selection should be made consciously*; it is crucial to consider students' interests, physical attributes, and musical preferences when guiding their instrument choices. Adhering to these criteria helps students establish a deeper affinity with the instrument they choose. "The bond, which individuals aspiring to become successful instrumentalists establish with their instruments, significantly influences the course and quality of their musical lives" (Polat *et al*., 2019, p. 1722). Active participation of students in the learning process enhances motivation, rendering learning outcomes more meaningful and enduring (Virkkula, 2020, p. 2).

*Regarding students' attitudes towards lessons, readiness levels, and self-efficacy levels should be taken into consideration;* there are individual differences in understanding, comprehension, and application in music education. These distinctions necessitate careful consideration throughout the educational process. "Performance is pivotal in the advancement of music education. Demonstrating performance in music education is deemed necessary to assess the degree of learning and promptly address any deficiencies or challenges in the learning process" (Gün & Yıldız, 2014, p. 1055). The study hours dedicated to music-related studies can vary significantly among students in FAHS, depending on environmental factors such as whether they live with their families or in dormitories. Teachers should account for these multifaceted aspects to effectively plan and structure their courses according to students' individual circumstances. It is important to acknowledge that experiences fostering competence and success enhance motivation, whereas exposure to excessive challenges or feelings of inadequacy may lead individuals to perceive themselves as ineffective (Evans, 2017, p. 67).

*Regarding students should study with discipline*; the establishment of disciplined and consistent study habits significantly impacts success in music education akin to other disciplines. "Students should manage their time effectively and complete assigned tasks regularly and as required" (Özen, 2004, p. 60). "When assigning homework to students, attention should be paid to the characteristics of the instructional field, the objectives of the instruction, and the individual characteristics of the students" (Türkmen & Özal-Göncü, 2024, p. 48).

• The conclusions related to the curriculum dimension are as follows:

*Regarding the curriculum should be improved in terms of the outcomes it covers*; aligning targeted achievements with students' readiness levels and setting realistic expectations holds paramount importance. "Achievements in the teaching-learning process are the knowledge, skills, attitudes, and values expected from students through planned and organized experiences. These achievements, which have a spiral structure, are systematically conveyed to students by

the teacher according to their levels and individual characteristics" (Güven & Çelenk, 2020, p. 450). Music education is an area with a high intensity of performance practices beyond literary knowledge. Therefore, the curriculum should be carefully prepared according to students' cognitive, affective, and psychomotor domains. Additionally, it should be designed to contribute to their individual, social, and emotional characteristics and facilitate the recognition of national and universal music culture.

*Regarding the functionality of the curriculum should be regularly examined and the necessary modifications should be made*; curricula play a crucial role in the systematic and organized implementation of the teaching-learning process. "Music education should be grounded within a sufficiently robust and consistent framework in terms of conceptual foundations and theoretical models" (Uçan, 2005, p. 35). "Curricula should be developed and updated over time according to emerging needs in the field" (Akkuzu & Şen, 2017, pp. 99-100). Therefore, the units or committees responsible for curriculum development must assess the functionality of the programs and make necessary adjustments to address deficiencies.

*Regarding the curriculum should be designed with input from university faculty and FAHS teachers*; the FAHS curriculum follow a unit-based approach, and the assessment of the achievements embedded in these units should be conducted by the teachers who implement the programs. Therefore, it is crucial for MoNE to gather input from practitioners during the curriculum development phase. "In the process of developing curricula, a harmonious approach that considers the holistic developmental characteristics of individuals based on current scientific knowledge is essential" (MoNE, 2018, p. 10).

*Regarding the curriculum prepared for music vocational courses should address all areas of competence*; instructional programs should be developed that encompass cognitive, affective, and psychomotor domains in music education. "In preparing educational programs, they are organized vertically and horizontally according to target behaviors. The vertical dimension represents distant goals, general and specific objectives, while the horizontal dimension represents cognitive, affective, and psychomotor domains" (Güllü *et al*., 2011, pp. 123-124). "Learning activities should appeal to a wide range of sensory organs" (Şendurur & Akgül-Barış, 2002, p. 166).

*Regarding FAHS regulations should be revised according to needs*; the Turkish Language Association states that regulations encompass the entirety of laws, regulations, directives, and similar enactments in force in a country (TDK [TLA], 2024). The regulations pertaining to FAHS encompass statutory provisions detailing aspects such as establishment, management, program structure, teacher employment, and student admissions. In accordance with the requirements of the modern age and to address issues arising in the administration of FAHS, the regulations should be revised.

*Regarding alternative programs, course content, and flexible teaching methods should be developed for potential issues in the curriculum*; teachers should proactively seek alternative teaching methods to mitigate potential disruptions during lessons. "The methods to be used for effective learning-teaching processes, how learning will occur, and how this process will be completed in the shortest time are planned by teachers. Different methods and approaches used during lessons help to ensure students' more active participation in the classroom" (Kaybal, 2017, p. 211).

• The conclusions related to the teacher dimension are as follows:

*Regarding regular in-service training sessions should be organized for FAHS music teachers*; the engagement of music teachers in such professional development activities is crucial for enhancing their vocational competencies. "Countries that allocate substantial resources to education consistently rank at the forefront of in-service training initiatives. These countries have high productivity at individual, organizational, and national levels" (Öztürk & Sancak, 2007, p. 789). "In-service training becomes imperative for educators who encounter challenges

in adapting to the evolving demands of both society and their profession with their existing educational background. Scheduled in-service trainings serve to augment educators' confidence in their professional roles. Those educators who possess a heightened sense of confidence make more substantial contributions to their professional endeavors compared to their less assured counterparts" (Selimoğlu & Biçen-Yılmaz, 2009, pp. 5-6).

*Regarding employment should be provided for the needed branch*; in order to become a teacher in FAHS, applicants must meet the criteria of having "at least three years of teaching experience in the field and be working as a teacher in ministry positions or in other service classes as of the application date" (MoNE, 2010). Teachers meeting these criteria undergo written and oral examinations conducted by commissions established by provincial directorates of national education, and successful candidates are recruited to FAHS. In addition, temporary teaching positions or paid assignments are offered to teachers in FAHS. It is crucial for the MoNE to identify deficiencies of subject area teachers at these institutions and to ensure employment in the areas of demand for the enhancement of education and training.

*Regarding subject teachers should conduct the classes*; teachers serving in institutions affiliated with MoNE can be assigned to positions in the same or different provinces based on their branches after completing the mandatory service requirements. This policy also applies to teachers in FAHS. Particularly in less populated provinces, the replacement of transferred teachers with subject-specific teachers can pose challenges. In such instances, teachers deemed suitable by school administrations may be assigned to these classes. "However, assigning teachers to roles for which they lack suitable qualifications often leads to experiences of discomfort, as they struggle with a sense of belonging in certain non-core subjects or grade levels. These feelings of unease regarding subject matter or grade levels can detrimentally affect the school's educational environment" (Du Plessis *et al*., 2014, p. 91).

FAHS hold a vital position in Türkiye due to their service in vocational music education from their establishment to the present day and their contribution to educating numerous students in this field. According to the results of this research, which examined studies related to FAHS, it was determined that suggestions were made regarding the dimensions of school, course, material and equipment, student, curriculum, and teacher. It was observed that the majority of proposed recommendations address issues that MoNE should resolve. Additionally, there are also suggestions for the directorates of FAHS and teachers serving in these institutions. These stakeholders should address and resolve these issues to enhance the efficiency of education and increase academic success.

## Acknowledgments

This study has been developed from research prepared as part of the "seminar" course conducted by Assoc. Prof. Dr. Ülkü Sevim ŞEN in the spring semester of the 2023-2024 academic year within the scope of the Ataturk University Institute of Educational Sciences, Department of Fine Arts Education, Music Education Doctoral Program.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

## Contribution of Authors

**Haşim Vapur**: Investigation, Resources, Visualization, Software, Formal analysis, and Writing-original draft. **Ülkü Sevim Şen**: Methodology, Supervision, and Validation.

## Orcid

Haşim Vapur https://orcid.org/0000-0002-5338-7988
Ülkü Sevim Şen https://orcid.org/0000-0003-2392-2326

## REFERENCES


Akbaş, B.Z., Kaleli, Y.S., & Özdek, A. (2021). Güzel sanatlar liselerinde kullanılan bağlama ders kitaplarındaki alıştırma ve ilgili türkülerin hedef davranış bakımından incelenmesi [Investigation of exercises and related songs in baglama textbooks used in fine arts high schools in terms of target behavior]. *Journal of the Human and Social Science Researches, 10*(4), 3853-3867. https://doi.org/10.15869/itobiad.983442

Akçay, Ş.Ö., & Sürmeli, Ö. (2022). Güzel sanatlar lisesi müzik öğretmenlerinin covid-19 pandemisi sürecindeki uzaktan eğitim hakkında görüşleri [Views of fine arts high school music teachers on distance education during the covid-19 pandemic]. *MSKU Journal of Education, 9*(2), 648-668. https://doi.org/10.21666/muefd.1009210

Akkuzu, D.Y., & Şen, Ü.S. (2017). Ortaöğretim 12. sınıf müzik dersi öğretim programının işlevselliğine ilişkin öğretmen görüşleri [Teachers' views as to functionality of programmes of teaching of twelfth class music lesson in secondary education]. *Academic Sight International Refereed Online Journal,* (61), 98-116.

Alpar, D., Batdal, G., & Avcı, Y. (2007). Öğrenci merkezli eğitimde eğitim teknolojileri uygulamaları [Educational technology applications in student-centered education]. *HAYEF Journal of Education, 4*(1), 19-31.

Apaydınlı, K. (2009, 23-25 Eylül). *Anadolu güzel sanatlar lisesi müzik bölümü öğrencilerinin müziksel işitme-okuma-yazma dersinde karşılaştıkları sorunlar ve çözüm önerileri* [*Problems encountered by Anatolian fine arts high school music department students in musical hearing-reading-writing course and solution suggestions*] [Oral presentation]. 8. National Music Education Symposium, Samsun.

Arslan-Karaküçük, S. (2008). Okul öncesi eğitim kurumlarında fiziksel/mekânsal koşulların incelenmesi: Sivas ili örneği [Investigation of physical/spatial conditions in preschool education institutions: Sivas province case]. *Cumhuriyet University Faculty of Arts and Sciences Journal of Social Sciences, 32*(2), 307-320.

Aslanargun, E. (2007). Okul-aile iş birliği ve öğrenci başarısı üzerine bir tarama çalışma [A survey study on school-family cooperation and student achievement]. *Manas University Journal of Social Sciences, 9*(18), 119-135.

Ayaydın, A. (2011). Güzel sanatlar ve spor liselerinde karşılaşılan sorunlar ve çözüm önerileri: Öğretmenlerin görüşleri [Problems in fine arts and sport high school and suggestions for solutions: Opinions of teachers]. *Mustafa Kemal University Journal of Social Sciences Institute, 8*(16), 435-453.

Barış, H. (2022). *Güzel sanatlar lisesi bilişim destekli müzik dersinin işlenişine ilişkin öğretmen görüşleri* [*Teacher's opinions on the processing of the fine arts high school informatics aid music course*] (Publication No. 747052) [Master's thesis, Ataturk University]. Council of Higher Education National Thesis Center.

Barış, M. (2023). *Güzel sanatlar liselerinde verilen bağlama çalgısına yönelik derslerde uygulama açısından karşılaşılan problemler* [*Problems encountered in terms of application in bağlama instrument courses given in fine arts high schools*] (Publication No. 841454) [Master's thesis]. Council of Higher Education National Thesis Center.

Baydağ, C. (2020). Ses eğitiminde repertuvar seçiminin önemi [The importance of the repertoire choice in voice training]. *Afyon Kocatepe University Journal of Social Sciences, 22*(2), 387-398. https://doi.org/10.32709/akusosbil.546365

Bayraktar, M.F. (1994). Eğitim öğretimde araç ve gerecin önemi [The importance of tools and equipment in education]. *Journal of Suggestion, 1*(1), 24-29.

Büyüköztürk, Ş. (2005). Anket geliştirme [Questionnaire development]. *Turkish Journal of Educational Sciences, 3*(2), 133-151.

Çalışır, Ç. (2024). *Müzik performans kaygısı ile baş etmeye yönelik bir deneme: Ayvalık Sebahat-Cihan Şişman Güzel Sanatlar Lisesi örneği* [*A study on coping with music performance anxiety: The case of Ayvalık Sebahat-Cihan Şişman fine arts high school*]

(Publication No. 856968) [Master's thesis, Uludag University]. Council of Higher Education National Thesis Center.

Çaylak, D.G. (2022). *Müzik eğitimi ana bilim dallarında öğrenim gören öğrencilerin bireysel çalgı dersine yönelik akademik motivasyon ve tutumlarının incelenmesi* [*Investigation of academic motivation and attitudes of students studying in music teaching departments towards individual instrument lessons*] (Publication No. 758817) [Master's thesis, Pamukkale University]. Council of Higher Education National Thesis Center.

Demirtaş, Z., & Küçük, Ö. (2014). Okul yöneticilerinin performanslarını değerlendirme ölçütleri ve performanslarını düşüren nedenler: Nitel bir araştırma [Criteria for evaluating the performance of school administrators and the reasons for decreasing their performance: A qualitative study]. *Marmara University Ataturk Faculty of Education Journal of Educational Sciences, 40*(40), 47-67.

Derin, U.Y. (2007). *Anadolu güzel sanatlar lisesi müzik bölümü öğrencilerinin çalgı seçim yöntemi, karşılaşılan sorunlar ve çözüm önerileri* [*Method of instrument choice for students of music department of Anatolian fine arts high schools, problems encountered and suggested solutions*] (Publication No. 201578) [Doctoral thesis, Marmara University]. Council of Higher Education National Thesis Center.

Dinçer, M.G. (2009). *Müzik dersi ve ders dışı müzikal etkinliklerin ilköğretim 7. sınıf öğrencilerinin sosyalleşmelerine etkisi* [*The effect of music lessons and extracurricular musical activities to primary school 7th grade students on socialization*] (Publication No. 235083) [Master's thesis, Selçuk University]. Council of Higher Education National Thesis Center.

Doğru, L. (2022). *Güzel sanatlar lisesi klasik gitar eğitimine yönelik Türk halk ezgileri kaynaklı materyal öneri ve uygulamaları* [*Material proposals and applications of Turkish folk melodies for fine arts high school classical guitar education*] (Publication No. 757503) [Master's thesis, Burdur Mehmet Akif Ersoy University]. Council of Higher Education National Thesis Center.

Döş, İ. (2013). Mutlu okul [Happy school]. *Education and Science, 38*(170), 266-280.

Du Plessis, A.E., Gillies, R.M., & Carroll, A. (2014). Out-of-field teaching and professional development: A transnational investigation across Australia and South Africa. *International Journal of Educational Research, 66*, 90-102. https://doi.org/10.1016/j.ijer.2014.03.002

Durukan, A.Y. (2015). Dijital dünyada okul kütüphaneciliği ve 21. yüzyıl becerileri [School librarianship and 21st century skills in a digital world]. *Journal of National Education, 45*(208), 106-120.

Dybå, T., & Dingsøyr, T. (2008). Empirical studies of agile software development: A systematic review. *Information and Software Technology, 50*(9-10), 833-859.

Ergashevich, E.A. (2024). The role of bloom's taxonomy in determining pedagogical objectives in the process of educational lessons. *American Journal of Language, Literacy and Learning in STEM Education, 2*(1), 264-270.

Ertem, Ş. (2011). Orta düzey piyano eğitimi için repertuvar seçme ilkeleri [Principles of selecting repertoire for intermediate piano education]. *Kastamonu Journal of Education, 19*(2), 645-652.

Evans, P. (2015). Self-determination theory: An approach to motivation in music education. *Musicae Scientiae, 19*(1), 65-83.

Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill.

Gerekten, S.E., & Ahmethan, N.B. (2018). Güzel sanatlar lisesi müzik bölümü öğrencilerinin "Türk Halk Müziği" kavramına ilişkin algıları [Fine arts high school music students' perceptions about concept of "Turkish Folk Music"]. *Erzincan University Education Journal, 20*(2), 387-406. https://doi.org/10.17556/erziefd.334139

Girgin, M., Sever, R., & Gök, Y. (2011). Atatürk üniversitesi fen-edebiyat ve Kâzım Karabekir eğitim fakültesi coğrafya öğrencilerinin mesleki uygulama gezilerine ilişkin görüşleri

[Ataturk university faculty of science and letters and Kazım Karabekir faculty of education geography students' views on professional practice trips]. *Journal of Eastern Geography, 8*(10), 7-20.

Güllü, M., Arslan, C., Uğraş, S., & Görgüt, İ. (2011, 25-27 May). *Yeni ortaöğretim beden eğitimi ders programı kazanımlarının değerlendirilmesi* [*Evaluation of the new secondary school physical education curriculum outcomes*] [Oral presentation] National Physical Education and Sports Teaching Congress, Van.

Gün, E., & Yıldız, G. (2014). Müzik öğretmeni adaylarına yönelik piyano performansı öz yeterlik ölçeğinin geliştirilmesi [Development of a piano performance self-efficacy scale for prospective music teachers]. *Electronic Turkish Studies, 9*(5), 1053-1065.

Güven, A.K., & Çelenk, K. (2020). Müzik öğretmeni görüşleri doğrultusunda 2018 müzik dersi öğretim programı 8. sınıf kazanımları üzerine bir değerlendirme [An evaluation on the 8th grade achievements of the 2018 music course curriculum in line with the views of music teachers]. *Journal of Ataturk University Graduate School of Social Sciences, 24*(1), 447-469.

Haddaway, N.R., Page, M.J., Pritchard, C.C., & McGuinness, L.A. (2022). PRISMA2020: An R package and shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and open synthesis. *Campbell Systematic Reviews, 18*(2), 1-12. https://doi.org/10.1002/cl2.1230

Jesson, J., Matheson, L., & Lacey, F.M. (2011). Doing your literature review: Traditional and systematic techniques. *Journal of Electronic Resources in Medical Libraries, 21*(1), 46-47. https://doi.org/10.1080/15424065.2024.2319264

Karaçam, Z. (2013). Sistematik derleme metodolojisi: Sistematik derleme hazırlamak için bir rehber [Systematic review methodology: A guide to preparing systematic reviews]. *Dokuz Eylül University School of Nursing Electronic Journal, 6*(1), 26-33.

Kaybal, E. (2017). Müzikte özel öğretim yöntem ve tekniklerine ilişkin kaynaklar ve kavramlar [Resources and concepts related to special teaching methods and techniques in music]. *Online Journal of Music Sciences, 2*(3), 206-227.

Kaynak-Akçaoğlu, T. (2021). Güzel sanatlar fakültelerinde müzik teorisi ve işitme eğitimi derslerinde öğrencilerin yaşadıkları sorunlar [Problems experienced by students in music theory and hearing education courses in fine arts faculties]. *Journal of the National Academy of Education, 5*(1), 97-111. https://doi.org/10.32960/uead.888438

Millî Eğitim Bakanlığı (MEB). [Ministry of National Education]. [MNE]. (2009). *Millî Eğitim Bakanlığı güzel sanatlar ve spor liseleri yönetmeliği* [*Ministry of National Education regulation on fine arts and sports high schools*]. https://www.resmigazete.gov.tr/eskiler/2009/06/20090616-7.htm

Millî Eğitim Bakanlığı (MEB). [Ministry of National Education]. [MNE]. (2010). *Millî Eğitim Bakanlığına bağlı fen liseleri ve sosyal bilimler liselerinin öğretmenleri ile güzel sanatlar ve spor liselerinin beden eğitimi, müzik ve görsel sanatlar/resim öğretmenlerinin seçimi ve atamalarına dair yönetmelik* [*Regulation on the selection and appointment of teachers of science and social sciences high schools and physical education, music and visual arts/painting teachers of fine arts and sports high schools affiliated to the Ministry of National Education*]. https://www.resmigazete.gov.tr/eskiler/2010/12/20101219-3.htm

Millî Eğitim Bakanlığı (MEB). [Ministry of National Education]. [MNE]. (2018). *Güzel sanatlar lisesi çalgı eğitimi dersi öğretim programı* [*Fine arts high school instrument education course curriculum*]. https://mufredat.meb.gov.tr/Dosyalar/20201214235924830-%C3%87ALGI%20E%C4%9E%C4%B0T%C4%B0M%C4%B0%20G%C4%B0TAR.pdf

Millî Eğitim Bakanlığı (MEB). [Ministry of National Education]. [MNE]. (2019). *2018-2019 eğitim-öğretim yılı verileri* [*Data for the 2018-2019 academic year*]. http://ogm.meb.gov.tr/meb_iys_dosyalar/2018_09/19185304_GYzel_Sanatlar_Liseleri.pdf

Millî Eğitim Bakanlığı (MEB). [Ministry of National Education]. [MNE]. (2023). *Güzel sanatlar liseleri haftalık ders çizelgeleri* [*Weekly course schedules for fine arts high*

*schools*]. https://ttkb.meb.gov.tr/meb_iys_dosyalar/2023_09/14141220_2023-41_Guzel_Sanatlar.pdf

Millî Eğitim Bakanlığı (MEB). [Ministry of National Education]. [MNE]. (2024). *Ortaöğretim kurumlarının amaçları* [*Purposes of secondary education institutions*] https://ogm.meb.gov.tr/meb_iys_dosyalar/2023_10/11115926_10_EKYM-YENY_OKUL_TURLERY.pdf

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & The Prisma Group. (2010). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *International Journal of Surgery, 8*(5), 336-341. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Mustan, H. (2010). *Anadolu güzel sanatlar ve spor liseleri müzik bölümü 12. sınıf öğrencilerinin bireysel ses eğitimi dersinde karşılaştıkları zorluklar üzerine bir araştırma* [*A survey on difficulties faced by 12. class students in vocal education lessons at the music departments of anatolian fine art and sports high schools*] (Publication No. 278547) [Master's thesis, Selçuk University]. Council of Higher Education National Thesis Center.

Nart, S. (2016). Music software in the technology integrated music education. *Turkish Online Journal of Educational Technology*, *15*(2), 78-84.

Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research. *Communications of the Association for Information Systems,* (37), 879-910.

Önaçan, M.B.K. (2020). Türkiye'de yeşil bilişim çalışmaları: Sistematik literatür taraması [Green it studies in Turkey: A systematic literature review]. *Igdir University Journal of Social Sciences,* (21), 345-368.

Özen, N. (2004). Çalgı eğitiminde yararlanılan müzik eğitimi yöntemleri [Music education methods used in instrument education]. *Gazi University Gazi Journal of Faculty of Education, 24*(2), 57-63.

Özen, A. (2023). *Güzel sanatlar lisesi müzik bölümü keman öğrencilerinin Barok dönem keman tekniklerini uygulamalarına yönelik öğretmen görüşleri* [*Teacher's opinions on the applications of Baroque period violin techniques by violin students of fine arts high school music department*] (Publication No. 800955) [Master's thesis, Çankırı Karatekin University]. Council of Higher Education National Thesis Center.

Öztürk, M., & Sancak, S. (2007). Hizmet içi eğitim uygulamalarının çalışma hayatına etkileri [Effects of in-service training practices on working life]. *Yaşar University E-Journal, 2*(7), 761-794.

Polat, S., İlgar, K., & Öztürk, N.İ. (2019). Konservatuvar öğrencilerinin ekonomik durumlarının enstrüman seçimi ve başarı düzeyi üzerindeki etkisi [The effect of the economıc sıtuatıons of conservatory students on the ınstrument selectıon and success level]. *Social Sciences, 14*(4), 1715-1735. https://doi.org/10.29228/TurkishStudies.23450

Selimoğlu, E., & Biçen-Yılmaz, H. (2009). Hizmet içi eğitimin kurum ve çalışanlar üzerine etkileri [The effects of in-service training on the organization and employees]. *Paradox, Journal of Economics, Sociology and Politics, 5*(1), 1-12.

Seyhan, Z. (2019). *Özengen müzik eğitiminde flüt eğitimcilerinin kullandıkları yöntem ve tekniklerin incelenmesi* [*Investigation of methods and techniques used by flute educators in amateur music education*] (Publication No. 552961) [Master's thesis, Necmettin Erbakan University]. Council of Higher Education National Thesis Center.

Sezen, H., & Özerdem, D.M. (2021). Londra Trinity koleji sertifikalı müzik eğitimi programı piyano dalının özengen müzik eğitimi açısından incelenmesi [Trinity college London certified music education program examination of piano branch in terms of amator music education]. *Electronic Turkish Studies, 16*(6). 1693-1713. https://dx.doi.org/10.7827/TurkishStudies.52132

Smith, T.E., Sheridan, S.M., Kim, E.M., Park, S., & Beretvas, S.N. (2019). The effects of family-school partnership interventions on academic and social-emotional functioning: A

meta-analysis exploring what works for whom. *Educational Psychology Review, 32*, 511-544. https://doi.org/10.1007/s10648-019-09509-w

Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. *Journal of Development Effectiveness, 4*(3), 409-429. https://doi.org/10.1080/19439342.2012.710641

Soğukçam, B. (2007). *Anadolu güzel sanatlar liselerinde klarnet eğitiminde karşılaşılan sorunlar ve çözüm önerileri* [*Problems in clarinet education at Anatolian fine arts high school and suggestions to solve*] (Publication No. 240969) [Master's thesis, Trakya University]. Council of Higher Education National Thesis Center.

Şendurur, Y., & Akgül-Barış, D. (2002). Müzik eğitimi ve çocuklarda bilişsel başarı [Music education and cognitive achievement in children]. *Gazi University Gazi Journal of Faculty of Education, 22*(1), 165-174.

Tokatlı, B. (2020). *Özengen müzik eğitimi kurumlarında keman derslerini yürüten öğretmenlerin karşılaştığı problemler* [*In the institutions teaching amateur music, the problems faced by the teachers who conduct violin lessons*] (Publication No. 642006) [Master's thesis, Necmettin Erbakan University]. Council of Higher Education National Thesis Center.

Topalak, Ş. (2013). Güzel sanatlar lisesi çalgı eğitimi/öğretiminde karşılaşılan sorunların incelenmesi [Investigation of the problems encountered in fine arts high school instrument education/teaching]. *Journal of Art Education, 1*(2), 114-129.

Tuncay, T. (2019). *Güzel sanatlar liselerinde çalgı eğitimi alan öğrencilerde karşılaşılan fizyolojik rahatsızlıklar* [*Physıological disorders encountered in the education of musical instruments in fıne arts high school*] (Publication No. 589768) [Master's thesis, Ordu University]. Council of Higher Education National Thesis Center.

Türk Dil Kurumu. (TDK). (2024). "Mevzuat" maddesi ["Legislation" clause]. https://sozluk.gov.tr/

Türkmen, E.F., & Göncü, İ.Ö. (2024). Müzik derslerinde uygulanabilecek performans görevi örnekleri [Performance task models applicable to music courses]. *Turkish Journal of Educational Sciences, 22*(1), 46-64. https://doi.org/10.37217/tebd.1301019

Uçan, A. (2005). *Müzik eğitimi-temel kavramlar-ilkeler-yaklaşımlar [Music education-basic concepts-principles-approaches]* (3rd ed.). Evrensel.

Uluç, Ç. (2006). *Güzel sanatlar liselerinde keman eğitiminde karşılaşılan sorunlar ve çözüm önerileri* [*What difficulties do the students who have been trained about playing violin in fine arts high school have during the training*] (Publication No. 206793) [Master's thesis, Trakya University]. Council of Higher Education National Thesis Center.

Umuzdaş, S. (2006). *Anadolu güzel sanatlar lisesi yöneticilerinin okul yönetiminde karşılaştıkları sorunlar* [*The problems of school managers' face into school management in fine and art high schools*] (Publication No. 189156) [Master's thesis, Gaziosmanpaşa University]. Council of Higher Education National Thesis Center.

Umuzdaş, S., & Umuzdaş, M.S. (2009). Anadolu güzel sanatlar lisesi yöneticilerinin karar aşamasında karşılaştıkları sorunlar [The problems that encounted by Anatolian fine art high school (afahs) managers in decision process]. *Kastamonu Education Journal, 17*(1), 367-376.

Virkkula, E. (2020). Evaluating motivational characteristics in vocational music education within the perspective of self-determination theory. *Empirical Research in Vocational Education and Training, 12*(1), 1-15.

Yalçın, P.D. (2019). Güzel sanatlar lisesi müzik öğretmenlerinin müzik teknolojilerinden yararlanma durumlarına ilişkin görüşleri [Review/Feedback of fine arts high school music teacher's regarding their status in utilization of music technologies]. *e-Kafkas Journal of Educational Research, 6*(1), 51-62. https://doi.org/10.30900/kafkasegt.577352

Yıldırım, A., & Şimşek, H. (2006). *Sosyal bilimlerde nitel araştırma yöntemleri* [*Qualitative research methods in social sciences*] (5rd ed.). Seçkin.

Yılmazçal, M., & Küçüksüleymanoğlu, R. (2022). Güzel sanatlar lisesi müzik ve resim öğretmenlerinin uzaktan eğitim sürecinde uygulamalı derslerin etkililiği üzerine görüşleri [Views of fine arts high school music and painting teachers on the efficiency of applied lessons in the distance education process]. *International Journal of Human and Art Research, 8*(E.Ş. special issue), 39-56.

# APPENDIX

## Appendix 1. *Table showing information about the studies included in the research*

| Study number | Study title | Year | Author | Type | Methodological reliability | Relevance review |
|---|---|---|---|---|---|---|
| 1 | Anadolu güzel sanatlar liseleri, lise III müzik sınıfında okuyan yatılı ve gündüzlü öğrencilerin piyano başarıları arasındaki farklar [Differences in the success between boarding and normal daytime pupils in the music classes of the Anatolian high school of fine arts and the reasons for the differences] | 2006 | Ayça Yıkılmazoğlu | Master | High | High |
| 2 | Anadolu güzel sanatlar liselerinde verilmekte olan viyola eğitiminin değerlendirilmesi [An assessment of viola education in Anatolian high schools for fine arts] | 2005 | Hasan Hakan Okay | Master | High | High |
| 3 | Anadolu güzel sanatlar liseleri müzik bölümleri bölüm şeflerinin eğitim-öğretim etkinliklerindeki yeri ve önemi [The place and the importance of music department chiefs of Anatolian fine arts high schools in the education actıvities] | 2006 | Tuba Yokuş | Master | High | High |
| 4 | Anadolu güzel sanatlar liselerinde kullanılan başlangıç keman eğitimi metotlarının incelenmesi [Examining elementary violin training methods that are used at Anatolia fine high schools] | 2006 | Özlem Altınel | Master | High | High |
| 5 | Anadolu güzel sanatlar liselerine öğretmen ve öğrenci seçme sınavı kriterlerinin karşılaştırılması [Comparing the criteria for the exam of the selection of teachers and students for Anatolian fine arts schools] | 2006 | Ömer Mumcu | Master | High | High |
| 6 | Türkiye'deki Anadolu güzel sanatlar liseleri müzik bölümlerinde viyolonsel eğitimine yönelik öğretmen ve öğrenci görüşleri [A study on the views of violoncello teachers and students aimed at violoncello education in Anatolia fine arts high school in Turkey] | 2006 | Özhan Özgün | Master | High | High |
| 7 | Anadolu güzel sanatlar liseleri müzik bölümü öğrencilerinin "müziksel işitme okuma yazma" başarılarının incelenmesi [Examining the "musical hearing, reading and writing" achievements of Anatolian fine arts high school music department students] | 2006 | Timuçin Sezerel | Master | High | High |
| 8 | Anadolu güzel sanatlar lisesi müzik bölümü öğrencilerinin çalgı seçim yöntemi, karşılaşılan sorunlar ve çözüm önerileri [Method of instrument choice for students of music department of Anatolian fine arts high schools, problems encountered and suggested solutions] | 2007 | Uğur Yalçın Derin | Master | High | High |
| 9 | Kuzey Kıbrıs Türk Cumhuriyeti ve Türkiye Cumhuriyetindeki güzel sanatlar liselerinde okutulan piyano dersi öğretim programlarının karşılaştırılması [Comparison the cirruculums of the piano lessons lectured in Anatolia high schools for fine arts in Turkish Republic of North Cyprus and Turkey] | 2006 | Yusuf Acemoğlu | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| 10 | Anadolu güzel sanatlar liselerini tercih eden öğrenci profili [The profile of the student who prefers fine art Anatolian high school] | 2007 | Mehtap Alim | Master | High | High |
| 11 | Anadolu güzel sanatlar liselerinde verilen müzik eğitiminin öğrencilerin meslek seçimine etkisi [The effect of the music education given by Anatolian fine arts high schools on the students choice of profession] | 2007 | Erkan Sülün | Master | High | High |
| 12 | Güzel sanatlar lisesine yeni başlayan piyano öğrencilerinin biyolojik, psikolojik ve müzikal yönden incelenmesi [Examining the piano students trained in fine arts high school in bological, psychological, musical ways] | 2006 | Gamze Yılmaz | Master | High | High |
| 13 | Güzel sanatlar liselerinde keman eğitiminde karşılaşılan sorunlar ve çözüm önerileri [What difficulties do the students who have been trained about playing violin in fine arts high school have during the training] | 2006 | Çiğdem Uluç | Master | High | High |
| 14 | Anadolu güzel sanatlar liseleri müzik bölümlerinde uygulanan viyola öğretim programına dayalı üçüncü sınıf düzeyinde devinişsel hedeflere ulaşma durumları [The sate of reaching the pschomotor learning targets in the third grade based to the viola teaching programmes applied in the music departmans of Anatolian high schools of fine arts] | 2007 | Gökhan Özdemir | Master | High | High |
| 15 | Türkiye'de Anadolu güzel sanatlar liseleri (yaylı çalgılar) bireysel çalgı eğitimi dersinde piyano eşlikli çalışmalara ilişkin öğretmen görüşleri [Teachers opinionson piano accompanied studies in Turkish Anatolian high school of fine arts (strings) individual instrument education lesson)] | 2007 | Sonat Coşkuner | Master | High | High |
| 16 | Anadolu güzel sanatlar liseleri müzik bölümleri keman dersinde anlamlandırma stratejisinin kullanımı ve etkililik düzeyi [The use of elaboration strategy and its efficiency level in violin lessons given in music departments of Anatolian fine arts high schools] | 2007 | Özlem Akın | Doctoral | High | High |
| 17 | Sivas ve Tokat Anadolu güzel sanatlar liselerindeki keman öğrencilerine uygulanan keman başlangıç yöntemlerinin Ankara Anadolu güzel sanatlar lisesinde uygulanan keman başlangıç yöntemleriyle karşılaştırılması [The comparison of violin beginning methods applied to students in Tokat and Sivas Anatolian high school of fine arts with violin beginning methods applied in Ankara Anatolian high school of fine arts] | 2007 | Ahmet Hakan Baş | Master | High | High |
| 18 | Anadolu güzel sanatlar liselerinin müzik alanına öğrenci seçimi [Choosing students to music branch of Anatolian high school of fine arts] | 2007 | Gökhan Köse | Master | High | High |
| 19 | Anadolu güzel sanat tarih liselerinde okuyan müzik bölümü öğrencilerinin güzel sanatlar liselerini seçimlerinde ailenin ve sosyal çevrenin etkileri [Choosing students to music branch of Anatolian high school of fine arts] | 2007 | Senem Özyoğurtçu | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| 20 | Anadolu güzel sanatlar liseleri müzik bölümü öğrencilerinin alanları ile ilgili yükseköğretim kurumlarını tercihlerinde etkili olan faktörlerin öğretmen ve öğrenci boyutunda değerlendirilmesi [Evaluation of factors that affect the higher education institutions preferences of students who are educating in music department of high schools of fine arts with the viewpoints of both teachers and students] | 2007 | Gürkan Güleç | Master | High | High |
| 21 | Anadolu güzel sanatlar lisesi müzik bölümü öğrencilerinin dinlemeyi tercih ettikleri müzik türleri [Music types that the students at Anatolian high school of fine arts prefer to listen] | 2008 | Nihal Ulutürk | Master | High | High |
| 22 | Anadolu güzel sanatlar liseleri müzik bölümü öğrencilerinin müziksel işitme-okuma-yazma dersinde karşılaştıkları sorunlar ve çözüm önerileri [The problems and the solution of suggestions which Anatolian high school of fine arts students encounter in the musical hearing-reading and writing lesson] | 2006 | Köksal Apaydınlı | Master | High | High |
| 23 | Anadolu güzel sanatlar lisesi öğrencilerinin ses problemleri ve nedenleri [Voice problems and reasons for the students of the Anatolian high school of fine arts] | 2008 | Bilgen Özcan Coşkunsoy | Master | High | High |
| 24 | Türkiye'deki Anadolu güzel sanatlar liseleri müzik bölümleri öğrenci özel yetenek giriş sınavlarının müziksel işitme-algılama boyutunun değerlendirilmesi [The evaluation of category of musical audition-perception in the special ability entrance examinations for the students of music departments of Anatolian fine arts high schools in Turkey] | 2009 | Ufuk Yağcı | Doctoral | High | High |
| 25 | Anadolu güzel sanatlar liselerinde bilgisayar destekli müzik eğitiminin kullanılabilme durumu [The state of computer based music education that is benefited in Anatolian fine arts high schools] | 2008 | Serdar Koldemir | Master | High | High |
| 26 | Anadolu güzel sanatlar liselerinde okutulmakta olan keman eğitimi ders kitaplarında yer alan yay tekniklerinin öğrenciler tarafından kullanılırlık durumunun incelenmesi [Asessing availabity for students condition of string techniques included in violin course books of Anatolian fine arts high schools] | 2009 | Melih Günaydın | Maste | High | High |
| 27 | Anadolu güzel sanatlar liselerindeki (AGSL) Türk müziği çalgıları dersinde öğretmen ve öğrencilerin karşılaştıkları problemler ve çözüm önerileri [Problems of students and teachers enceountering in Turkish musical instrument course in Anatolian art high scohools and solition recommendations] | 2008 | Akın Uzunoğlu | Master | High | High |
| 28 | Anadolu güzel sanatlar lisesi öğrencilerinin dönemsel ses özelliklerine uygun batı müziği koro eğitimi dersi öğretiminde kullanılabilecek eserler üzerine antolojik bir çalışma [A suggestion of an antology on pieces of music that can be used at western music choral training classes suitable for | 2009 | Levent Sezgin | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | the periodical voice characteristics of the students at Anatolian fine arts high scools] | | | | | |
| 29 | Anadolu güzel sanatlar liseleri müzik bölümü öğrencilerinin müziksel işitme okuma ve yazma dersinin öğretim programında yer alan kazanımlara ulaşma durumlarının değerlendirilmesi [In Anatolian high school of fine arts to achieve the acquasitions in the musical hearing reading and writing programme] | 2009 | Ünsal Deniz | Master | High | High |
| 30 | Anadolu güzel sanatlar liseleri müzik bölümü öğrencilerinin piyano dersine yönelik tutumlarının incelenmesi [The study of the studensts at the department of music at Anatolian fine arts high schools towards piano lessons] | 2009 | Özlem Kılınçer | Master | High | High |
| 31 | Anadolu güzel sanatlar liselerinde klarnet eğitiminde karşılaşılan sorunlar ve çözüm önerileri [Problems in clarinet education at Anatolian fine arts high school and suggestions to solve] | 2007 | Barış Soğukçam | Master | High | High |
| 32 | Beşinci bölge Anadolu güzel sanatlar liselerinin müzik bölümlerinde okuyan 9, 10, 11 ve 12. sınıf öğrencilerinin beklentileri arasındaki farklar [The differences between the expectations of the 9th, 10th, 11th and the 12th graders in the fifth region's Anatolian high schools of fine arts] | 2010 | Kadir Özal | Master | High | High |
| 33 | Anadolu güzel sanatlar liseleri müzik bölümlerinde bireysel ses eğitimi dersinin öğrenci görüşlerine göre incelenmesi [Studying of individual voice education course in the music departments of Anatolia fine arts high schools] | 2008 | Özlem Üzeren | Master | High | High |
| 34 | Çalgı eğitiminde öğretmen-öğrenci iletişiminin "transaksiyonel analiz ego durumları" açısından incelenmesi: Bir Anadolu güzel sanatlar lisesi örneği [The analysis of teacher-student communication in instrumental training in terms of "transactional analysis ego states": An example of Anatolian high school of fine arts] | 2010 | Ersin Turhal | Master | High | High |
| 35 | Anadolu güzel sanatlar liselerinde piyano eğitimi alan öğrencilerin karşılaştıkları teknik problemlerin incelenmesi [The investigation of the technical problems faced by the students who take piano training in Anatolian highschools of fine arts] | 2010 | Ali Erman Tezer | Master | High | High |
| 36 | Anadolu güzel sanatlar liseleri viyolonsel eğitiminde motivasyonun yeri ve önemine ilişkin öğretmen ve öğrenci görüşleri [Opinions of teachers and students regarding the role and the importance of motivation in violoncello trainig at Anatolian fine arts high schools] | 2009 | Zeynep Özder | Master | High | High |
| 37 | Kuruluşundan günümüze kadar Anadolu güzel sanatlar liseleri öğretim sürecinin ve mesleki müzik eğitimine katkılarının farklı değişkenler açısından incelenmesi (Erzurum Anadolu güzel sanatlar lisesi örneği) [Assesment for instruction period of Anatolian fine arts high schools and contributions to the vocational music | 2013 | Mustafa Kemal Sümbüllü | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | education in terms of different variables (Erzurum Anatolian fine arts high school sample)] | | | | | |
| 38 | Güzel sanatlar liselerindeki piyanoların akort, bakım ve onarım sorunlarının değerlendirilmesi [The evaluation of the problems of tuning, maintenance and repair of pianos in fine arts high schools] | 2014 | Hazime Pakel | Master | High | High |
| 39 | Bingöl ve Sivas Anadolu güzel sanatlar lisesi öğrencilerinin öz geçmişlerinin müzik eğitimine etkisi [The effects of CV's (cirruculum vitae) of the students of Bingol and Sivas Anatolian fine arts high school to music education] | 2014 | Gökhan Aladağ | Master | High | High |
| 40 | Güzel sanatlar liseleri müzik bölümü öğrencilerinin piyano dersi başarısını etkileyen etmenlere ilişkin ders öğretmenlerinin görüşleri (Ege bölgesi örneği) [Teachers' opinions about factors that effect students' piano class academic success who are at department of music in fine art high schools (The sample of Aegean region)] | 2014 | Eylem Bayazıt Karademir | Master | High | High |
| 41 | Güzel sanatlar liselerinde uygulanan müziksel işitme-okuma-yazma derslerinin öğretim programı işlevselliğine yönelik öğretmen ve öğrenci görüşleri (Ankara ili örneği) [The opinions of the teachers and the students about the functionality of the teaching programs towards the musical auditory, reading and writing lessons which are applied in fine arts high schools (The sample of Ankara)] | 2015 | Ceren Tütüncü | Master | High | High |
| 42 | Güzel sanatlar lisesi müzik bölümü öğrencilerinin çalgıya yönelik benlik algıları ile genel benlik saygıları arasındaki ilişkinin incelenmesi [An investigation on the correlation between the self-perception for the played instrument for the students of fine arts high school and their self esteem level] | 2004 | Gülbahar Karabulut | Doctoral | High | High |
| 43 | Güzel sanatlar lisesi bireysel ses eğitimi dersi öğretim programının uygulamasında karşılaşılan sorunların öğretmen görüşleri ile incelenmesi [Analysis of the problems that are occurred in the course of individual voice training program application with teacher interviews in the fine arts high school] | 2014 | Vahide Bahar Yiğit | Doctoral | High | High |
| 44 | Güzel sanatlar liselerinde bireysel ses eğitimi dersinde kullanılan öğretim yöntemlerinin incelenmesi [Analyzing individual voice education courses in music departments of fine arts high schools] | 2014 | Zeynep Ece Öztürk | Master | High | High |
| 45 | Güzel sanatlar liseleri müzik bölümlerinde verilen Türk müziği çalgı eğitimi ile piyano dersi başarı durumları arasındaki ilişkinin incelenmesi [Examination of the relationship between the Turkish music instrument training and the piano course success situations given in the music departments of the high schools of fine arts] | 2014 | Cüneyt Öbek | Master | High | High |
| 46 | Güzel sanatlar liseleri batı müziği çalgı toplulukları dersi repertuvarına yönelik sorunların incelenmesi [The analysis of | 2015 | Pınar Dağdeviren | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | problems oriented to the repertoire of the classical music instrument ensemble classes at fine arts high schools] | | | | | |
| 47 | Güzel sanatlar liseleri müzik bölümlerinde bireysel ses eğitimi dersinin doğru güzel ve etkili konuşma becerisine etkisi [The analysis of problems oriented to the repertoire of the classical music instrument ensemble classes at fine arts high schools] | 2014 | Ayfer Sönmez | Master | High | High |
| 48 | Güzel sanatlar liseleri müzik bölümlerinde bireysel ses eğitimi derslerinde öğretilen eserlerin antolojisi [The anthology of works taught at individual vocal training in fine arts high school music departments] | 2014 | Seçkin Sardaş Çelik | Master | High | High |
| 49 | Güzel sanatlar liselerinde piyano eğitiminde karşılaşılan sorunlar ve çözüm önerileri (Karadeniz bölgesi örneği) [The problems encountered in piano education at high school of fine arts and recommendations to these problems (The sample of Black Sea region)] | 2016 | Varol Çiçek | Master | High | High |
| 50 | Güzel sanatlar liselerinde bireysel ses eğitimi dersinde uygulanan Vaccai ve Concone etütlerinin öğrencinin ses gelişimine etkisi [The effects of vaccai and concone vocal exercises on voice development of students applied in individual voice training courses at fine arts high schools] | 2015 | Gülşah Başaran Tanrıöver | Doctoral | High | High |
| 51 | Güzel sanatlar lisesi gitar ders kitaplarında yer alan Türk müziği makamlarının uygulanabilirliğine ilişkin öğretmen görüşlerinin incelenmesi [Research of teachers opinions on practicality of used makams in Turkish classical music guitar textbooks used in fine art schools] | 2015 | Şerif Gayretli | Master | High | High |
| 52 | Güzel sanatlar lisesi öğrencilerinin müzik öğretmenliği mesleğine yönelik tutumlarının incelenmesi [Research of teachers' opinions on practicality of used makams in Turkish classical music guitar textbooks used in fine art schools] | 2016 | Hüseyin Tan | Master | High | High |
| 53 | Güzel sanatlar liselerinde 9. sınıf keman ders kitabının, öğrenci ve öğretmen açısından işlevselliğinin incelenmesi [Examining functionality of 9th grader violin textbook at fine arts high schools in terms of students and teachers opinions] | 2016 | Ahmet Ulvar Tanrıöver | Master | High | High |
| 54 | Güzel sanatlar lisesindeki öğrencilerle diğer lise türlerindeki öğrencilerin yaratıcılık düzeylerinin karşılaştırılması [Comparison of creativity levels of students between in fine arts high schools and students in other types of high schools] | 2016 | Birce Doğramacıoğlu | Master | High | High |
| 55 | Trakya bölgesindeki Anadolu güzel sanatlar liselerindeki flüt eğitiminin incelenmesi [An investigation of flute education in Anatolian fine arts lyceums in thrace region] | 2016 | Merve Çokamay | Master | High | High |
| 56 | Anadolu güzel sanatlar liselerindeki bireysel ses eğitiminde Türk halk ezgilerinin kullanılma durumları [Utilization Turkish folk tunes on individual voice training in Anatolian fine arts high school] | 2016 | Banu Saraç Çetin | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| 57 | Güzel sanatlar liseleri ve güzel sanatlar fakülteleri müzik bölümü flüt öğrencilerinin müzikal gelişimini etkileyen faktörler [Factors which affects musical development of fine arts high school and flute department students in fine arts faculty] | 2016 | Ümran Ezgi Güleken | Master | High | High |
| 58 | Güzel sanatlar lisesi müziksel işitme okuma ve yazma dersi öğretmenlerinin özel alan yeterliklerine ilişkin bir tasarı örneği [A draft proposal to assess the proficiency of musical literacy educators in fine arts high schools in terms of their specialties] | 2017 | Alper Er | Doctoral | High | High |
| 59 | Güzel sanatlar liseleri başlangıç düzeyi bağlama öğretimine yönelik öğretim elemanı görüşlerinin incelenmesi [The investigation of instructors' opinions towards beginner level baglama instruction in fine arts high schools] | 2017 | Neşet Güngör | Master | High | High |
| 60 | Güzel sanatlar lisesinde görev yapan müzik branş öğretmenlerinin yaşadığı mesleki sorunlar [Professional problems of the music teachers who work at the fine arts high school] | 2017 | Ahmet Gürsu | Master | High | High |
| 61 | Kırklareli Lüleburgaz TEK güzel sanatlar lisesindeki keman eğitiminin incelenmesi [Examination of violin education in Kırklareli Luleburgaz TEK fine arts high school] | 2017 | Gökhan Yeniada | Master | High | High |
| 62 | Güzel sanatlar lisesi müzik bölümü çalgı bakım onarım dersi öğretim programının kazanımları hakkında öğretmen ve öğrenci görüşleri [The idea of teachers and students about the teaching program gains of instrument maintenance repair lesson in fine art high school] | 2017 | Kemal Karaarslan | Master | High | High |
| 63 | Güzel sanatlar liseleri batı müziği koro derslerinde koro tınısı elde etmede kullanılan yöntemlerin değerlendirilmesi [Evaluation of the methods used in producing timbre in western music choir lesson of fine arts high schools] | 2017 | Melike Bozkuş | Master | High | High |
| 64 | Güzel sanatlar liselerinde piyano öğretimi sürecinin öğretmen ve öğrenci görüşleri çerçevesinde değerlendirilmesi [The evaluation of piano education process at fine arts school in Turkey within the scope of the opinions of teachers and students] | 2017 | Pınar Şahin | Doctoral | High | High |
| 65 | Güzel sanatlar liseleri 1997 ve 2016 çoksesli koro dersi öğretim programlarının karşılaştırması [Comparison of teaching programs of polyphonic choir at fine arts high school in 1997 and 2016] | 2017 | Güngör Hatipoğlu | Master | High | High |
| 66 | Güzel sanatlar liselerinde piyano eğitiminin teknik ve pedagojik açıdan incelenmesi [Technical and pedagogic analysis of piano education in fine arts highschools] | 2018 | Yasemin Çelik Nayır | Master | High | High |
| 67 | Güzel sanatlar liselerinde flüt eğitiminde kullanılan etüt ve eserler [Etudes and works used in flute education in fine arts high schools] | 2018 | Güney Caf | Master | High | High |
| 68 | Güzel sanatlar lisesi müzik bölümü öğrencilerinin piyano dersi motivasyonlarının piyano dersi başarısı ile olan ilişkisinin incelenmesi [Examination of the relationship between the piano lesson | 2018 | Mustafa Durgun | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | motivation and piano lesson success of the fine arts high school music department students] | | | | | |
| 69 | Güzel sanatlar lisesi çello dersi öğretim programına dayalı kullanılan kitapların Türk müziği kazanımları açısından değerlendirilmesi [The assessment of books used based on the curriculum of fine arts high schools cello course in terms of their Turkish music learning outcomes] | 2018 | Funda Uğraş Bilir | Master | High | High |
| 70 | Güzel sanatlar lisesi çalgı eğitimi viyola dersi öğretmenlerinin özel alan yeterliklerinin belirlenmesine yönelik taslak önerisi [A draft suggestion on the determination of subject-specific competence of instrument training-viola lesson teachers in the fine art high school] | 2018 | Çağlar Baykal | Doctoral | High | High |
| 71 | Güzel sanatlar lisesi viyolonsel öğretim programlarının karşılaştırılması [The comparison of the cello curriculum of the fine arts high schools] | 2018 | Yiğit Alp Onat | Master | High | High |
| 72 | Doğu Anadolu bölgesinde bulunan güzel sanatlar liseleri müzik bölümlerinin mevcut durumlarının öğretmen görüşleri açısından değerlendirilmesi [The evaluation of the ongoing situation of music departments of the fine arts high schools located at Eastern Anatolia region with respect to the teachers' point of view] | 2018 | Tuğçe Elif Mutlu | Master | High | High |
| 73 | Güzel sanatlar lisesi ve diğer liselerde müzik eğitimi almış öğrencilerin sosyal ve duygusal gelişmişlik düzeylerinin belirlenmesi [Determination of social and emotional development levels of students who have been music training in fine arts high school and other high schools] | 2019 | Enes Arslan | Master | High | High |
| 74 | Güzel sanatlar liselerinde bağlama eğitimi ve öğretiminde karşılaşılan problemler ve çözüm önerileri [Solution offers for experienced in bağlama lectures for the students of fine arts high schools] | 2019 | Serdar Şen | Master | High | High |
| 75 | Güzel sanatlar lisesi müzik bölümü öğrencilerinin mükemmeliyetçilik ve akademik erteleme düzeyleri arasındaki ilişkinin incelenmesi [Investigation of the relationship between, levels of perfectionism and academic procrastination of fine arts high school music department students] | 2019 | Cemal Yücel | Master | High | High |
| 76 | TRT Türk halk müziği repertuvarına kayıtlı Giresun türkülerinin güzel sanatlar liseleri programında yer alan THM teori ve uygulaması derslerinde kullanılabilirliği üzerine bir inceleme [A review of the usability of Giresun folk songs which are registered to the trt Turkish Folk Music repertoire in the TFM theory and practise lessons in fine arts high school] | 2019 | Batuhan Tüfekçi | Master | High | High |
| 77 | Bolu yöresi türkülerinin güzel sanatlar liseleri keman eğitimi dersinde kullanılabilirliği [The availability of the Bolu region folk songs in the violinclasses at fine arts high schools] | 2019 | Sena Eyüpoğlu | Master | High | High |
| 78 | Güzel sanatlar lisesi müzik bölümleri çalgı eğitimine yönelik öğretmen ve öğrenci | 2019 | Elif Özmumcu | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | görüşlerinin çeşitli boyutlarıyla incelenmesi [Investigation of the various dimensions of teachers and students' views intended for instrument education in music departments of fine arts high school] | | | | | |
| 79 | Güzel sanatlar lisesi müzik bölümü öğrencilerinde piyano başarısı ile duygusal zekâ arasındaki ilişki [The relationship between piano success and emotional intelligence in students of fine arts high school music department] | 2019 | Ebru Karadağ | Master | High | High |
| 80 | Güzel sanatlar lisesi müzik bölümü öğrencilerinin piyano dersine yönelik tutumları ile piyano dersi başarıları arasındaki ilişki [The relationship between the attitudes and the achievement of the fine arts high school music department students on piano classes] | 2019 | Çağrı Bakıoğlu | Master | High | High |
| 81 | Güzel sanatlar liseleri müzik bölümü öğrencilerinin bireysel ses eğitimi dersine yönelik motivasyon düzeylerinin incelenmesi [Examining the motivation levels of fine arts high school music department students for individual voice training lesson] | 2019 | İrem Karslı | Master | High | High |
| 82 | Güzel sanatlar liselerinin müzik bölümlerinde flüt eğitimi alan öğrencilerin çalgılarına yönelik güdülenme düzeylerinin incelenmesi [Examination of the motivation level to their instrument of flute students in fine arts high school] | 2019 | Aygül Ülker | Master | High | High |
| 83 | Güzel sanatlar lisesi müzik alanı 9. sınıf piyano ders kitabının içeriğinin incelenmesi (Analysis of content of piano main course book used for 9th grades at fine arts high school department of music) | 2019 | Burak Baran | Master | High | High |
| 84 | Piyanoda artikülasyon tekniklerinin öğretiminde öğretmenlerin karşılaştıkları güçlükler ve bunlarla baş etme yolları: Güzel sanatlar liseleri örneği [The difficulties of teachers in the teaching of articulation techniques on piano and the ways to cope with them: The sample of fine arts high schools] | 2019 | Seçil Çoban | Master | High | High |
| 85 | Güzel sanatlar liseleri müzik bölümlerinde çalgı eğitimi viyola dersinde motivasyonun önemine ilişkin öğretmen ve öğrenci görüşlerinin değerlendirilmesi [The difficulties of teachers in the teaching of articulation techniques on piano and the ways to cope with them: The sample of fine arts high schools] | 2019 | Ozan Nabi Akın | Master | High | High |
| 86 | Anadolu güzel sanatlar liseleri 9. Sınıf piyano dersi öğretim programında yer alan kazanımlara ulaşılma durumları (Osmaniye Abdurrahman Keskiner güzel sanatlar lisesi örneği) [The statuses of achieving the gains in the education program of Anatolian fine arts high schools 9th grade piano courses (Osmaniye Abdurrahman Keskiner fine arts high school case study)] | 2019 | Burçin Korkmaz Kavuklu | Master | High | High |
| 87 | Güzel sanatlar liseleri 11. sınıf Türk halk müziği ve teori uygulama dersi öğretim programı ve ders kitabının değerlendirilmesi [The evaluation of the | 2019 | Ertuğrul Kılıç | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | curriculum and coursebook of fine arts high school 11th grade Turkish folk music and theory practice lesson] | | | | | |
| 88 | Van güzel sanatlar lisesinde görev yapan piyano öğretmenlerinin piyanoda deşifre kullanımına yönelik görüşlerinin incelenmesi [Investigation of piano teachers' perceptions on the use of transcription in piano at Van fine arts high school] | 2019 | Sultan Timur | Master | High | High |
| 89 | Güzel sanatlar liseleri batı müziği çalgı toplulukları öğretim programındaki kazanımların gerçekleşme düzeyine ilişkin öğretmen görüşleri [Teacher opinions on the achievements of the western musical instrumental communities course of the fine arts high school] | 2019 | Serhat Atalay | Master | High | High |
| 90 | Kuruluşundan günümüze güzel sanatlar liseleri başlangıç düzeyi keman eğitimi ders kitaplarının incelenmesi [The analysis of violin education textbooks in fine arts high schools beginner level since their establishment] | 2019 | Baran Gülüm | Master | High | High |
| 91 | Türkiye'deki güzel sanatlar liseleri çalgı eğitimi kanun dersi öğretim programının uygulanışına yönelik öğretmen ve öğrenci görüşleri [The opinions of teachers and students' about implementation of the instrument training and qanun lesson's curriculum at fine arts high schools in Turkey] | 2019 | Derya Avcı | Master | High | High |
| 92 | Güzel sanatlar lisesi müzik bölümü öğrencilerinin Türk müziği ve batı müziği derslerine yönelik tutumları üzerine bir inceleme [A study on the attitudes of the students of the fine arts high school music department towards Turkish music and classical music lessons] | 2019 | Merve Daştan | Master | High | High |
| 93 | Güzel sanatlar liselerindeki piyano eğitim sürecine ilişkin öğretmen ve öğrenci görüşleri (İç Anadolu Bölgesi örneği) [Opinions of teachers' and students' regarding the piano education process at high schools of fine arts (the case of Central Anatolia Region)] | 2019 | Münevver Çetindere | Master | High | High |
| 94 | Güzel sanatlar lisesi müzik bölümü öğrencilerinin mesleki olgunluk düzeyleri ve ailelerin meslek seçimlerine etkisinin incelenmesi [Professional maturity levels of fine arts high school music department students and examining the effect of families on career choices] | 2019 | Çağlar Araz | Master | High | High |
| 95 | Güzel sanatlar lisesi öğrencilerinin duygu durumlarına göre müzik tercihleri ile kişilik özellikleri arasındaki ilişkiler [The relationship between music preferences of fine arts high school students according to their personality traits and mood] | 2019 | Pınar Nizamoğlu | Master | High | High |
| 96 | Güzel sanatlar lisesi öğrencilerinin bireysel çalgı performanslarına yönelik optimal performans duygu durumlarının çeşitli değişkenler açısından incelenmesi (Orta Karadeniz bölgesi örneği) [Investigation of the flow state of fine arts high school students individual instrument performance | 2019 | Özgür Koçak | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | in terms of various varieties (Middle Black Sea region example)] | | | | | |
| 97 | Güzel sanatlar liselerinde çalgı eğitimi alan öğrencilerde karşılaşılan fizyolojik rahatsızlıklar [Physiological disorders encountered in the education of musical instruments in fine arts high school] | 2019 | Tuncay Tosun | Master | High | High |
| 98 | Güzel sanatlar lisesi müzik bölümü keman derslerinde öğrenme stratejilerinin kullanımı ve etkililik düzeyi [Usage and effectiveness level of learning strategies in violin lessons in fine arts high school music departments] | 2019 | Resulhan Bahadır Hafızoğlu | Master | High | High |
| 99 | ABRSM grade 1 ve grade 2 piyano programları ile güzel sanatlar liseleri 9. ve 10. sınıf piyano programlarının incelenmesi [Examination of ABRSM grade 1 and grade 2 piano programs and fine arts high school 9th and 10th grade piano programs] | 2019 | Nuray Esen | Master | High | High |
| 100 | Kırıkkale güzel sanatlar lisesi mezunlarına ilişkin bir inceleme 2004-2018 [A studyon graduates of Kirikkale fine arts high school 2004-2018] | 2019 | Arzu Yalçın | Master | High | High |
| 101 | Güzel sanatlar liselerindeki çalgı eğitimcilerinin mesleki doyumları ile mesleki tükenmişlik düzeylerinin belirlenmesi [Determination of job satisfaction and occupational burnout levels of musical instrument teachers in high schools of fine arts] | 2019 | Mustafa Akgül | Master | High | High |
| 102 | Halk ozanlığı geleneği ve müziğinin güzel sanatlar liseleri ve müzik eğitimi ana bilim dallarında kullanılma durumunun tespiti [The determination of the usage of minstrelsy tradition and folk music in fine arts high schools and departments of music education] | 2019 | Emin Durmuş | Master | High | High |
| 103 | Güzel sanatlar liseleri bireysel ses eğitimi derslerinde teknoloji destekli öğretimin etkililiğinin değerlendirilmesi [Assessment of the effectiveness of technology-supported teaching in individual voice education courses of fine arts high schools] | 2020 | Melike Çakan Uzunkavak | Master | High | High |
| 104 | Güzel sanatlar liselerinde tanbur eğitimine yönelik eğitmen görüşlerinin incelenmesi [Investigation of teacher opinions on tanbur education in fine arts high school] | 2020 | Hasan Koçak | Master | High | High |
| 105 | Türkiye'deki güzel sanatlar lisesi öğrencilerinin müzikal yaşantısı ve müzikal motivasyonu düzeylerinin incelenmesi [Investigation of musical experience and motivation levels of fine arts high school students in Turkey] | 2020 | Semra Atılgan Bozarslan | Master | High | High |
| 106 | Güzel sanatlar lisesi çalgı eğitimi gitar ders kitaplarındaki klasik dönem etüt ve eserlerin incelenmesi [The study of classical period etudes and its works in guitar course books in fine arts high school instrument education] | 2020 | Caner Şahin | Master | High | High |
| 107 | Güzel sanatlar lisesi viyolonsel öğrencilerinin bireysel çalgı çalışma alışkanlıklarının çalgı performansı öz yeterlik inancına etkisi [The effect of individual musical instrument study habits of fine arts high school cello students on | 2020 | Ceren Albayrak | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | musical instrument performance self-efficacy belief] | | | | | |
| 108 | Güzel sanatlar liseleri gitar dersi öğretim programının ve 9. sınıf ders kitabının incelenmesi [The analysis of secondary education fine arts high school instrument training guitar coursebook for 9th graders] | 2020 | Nur Hilal Şahin | Master | High | High |
| 109 | Güzel sanatlar lisesi keman kitaplarında yer alan eserlerin içerdiği kazanımlar ve teknik güçlükler bakımından incelenmesi [Examination of the works included in the fine arts high school violin books in terms of gains and technical difficulties] | 2020 | Ezgi Demir | Master | High | High |
| 110 | Güzel sanatlar liseleri piyano eğitiminde aksak ritimlerin kullanımı ve başlangıç düzeyi aksak ritimli etüt ve alıştırma önerileri: Van ili örneği [The use of aksak (limping) rhythms for fine arts high school piano education and beginner level aksak rhythm etudes and practice proposals: Van province sample] | 2020 | Fatih Marufoğlu | Master | High | High |
| 111 | Güzel sanatlar liseleri 10. sınıf Batı müziği teori-uygulama dersi öğretim programı ve ders kitabının değerlendirilmesi [Examination of fine arts high schools 10th class West music theory-practice course curriculum and course book] | 2021 | Ercan Akbulut | Master | High | High |
| 112 | Güzel sanatlar lisesi müzik bölümü öğrencilerinin farklı değişkenlere göre memnuniyet düzeylerinin araştırılması [Degree of satisfaction research in regard to diffirent variables in fine arts high school music department students] | 2021 | Müge Güneş | Master | High | High |
| 113 | Güzel sanatlar liselerinin çalgı eğitimi branşlarından kaval dersi öğretim programının incelenmesi [Review of the kaval lesson teaching program from the instrument education branches of fine arts high schools] | 2021 | Eren Gedik | Master | High | High |
| 114 | Türk sanat müziği eserlerinin güzel sanatlar lisesi viyola eğitiminde kullanılabilirliği [Availability of Turkish art music works in viola education in fine arts high school] | 2021 | Bilal Can Tiryaki | Master | High | High |
| 115 | Güzel sanatlar liselerindeki bireysel ses eğitimi dersinin hedeflerine ulaşma düzeyinin değerlendirilmesi (Trabzon ili örneği) [Evaluation of the level of achievement of the goals of the individual voice training course in fine arts high schools (Trabzon province example) | 2021 | Melek Yıldız | Master | High | High |
| 116 | Güzel sanatlar lisesi öğrencilerinin viyolonsel dersine ilişkin tutumlarının incelenmesi (Doğu Karadeniz örneği) [Resarch of fine arts high school students' attitudes towards cello lesson (example of Eastern Black Sea region)] | 2021 | Bahar Saraloğlu | Master | High | High |
| 117 | Müzik öğretmeni adaylarının güzel sanatlar lisesi bireysel ses eğitimi dersi kazanımlarına ulaşma düzeyleri [Music education student's achievement levels concerning fine arts high school's individual voicetraining learning outcomes] | 2021 | Ceren Tepe | Master | High | High |
| 118 | Güzel sanatlar liseleri geleneksel Türk müziği koro eğitimine yönelik yöntem önerisi ve yöntemin performans üzerine | 2021 | Nevra Güçlü Tükü | Doctoral | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | etkileri [A method recommendation for traditional Turkish music choir education at fine arts high schools and its effects on performance] | | | | | |
| 119 | Güzel sanatlar lisesi keman öğretim programında yer alan makamların geleneksel icra yöntemine ilişkin model önerisi "Hüzzam makamı örneği" [Model proposal regarding the traditional performance method of maqams in the fine arts high school violin training program "Sample of Hüzzam maqam"] | 2021 | Ahmet Hakan Baş | Doctoral | High | High |
| 120 | İlk piyano eğitimi ve güzel sanatlar lisesi müzik bölümü öğrencilerine metod, teknik çalışma ve repertuvar bağlamında öneriler [Suggestions in the context of method, technical study and repertory to the students of first piano education and fine arts high school music department] | 2021 | Nihan Sözmen | Master | High | High |
| 121 | Güzel sanatlar liseleri keman eğitiminde kullanılmak üzere oluşturulan çok sesli beş halk türküsünün icrasına yönelik model bir çalışma önerisi [A study model proposed for the performance of five polyphonic folk songs for violin education in fine arts high schools] | 2021 | Cemre Yılmaz | Master | High | High |
| 122 | Halk türkülerinin güzel sanatlar liseleri piyano eğitiminde kullanılabilirliği (Kırşehir yöresi örneği) [Folk songs in Fine Arts high school piano education availability (Example to Kırşehir region)] | 2021 | Şener Pehlevan | Master | High | High |
| 123 | Güzel sanatlar lisesinde öğrenim gören öğrencilerin piyano dersine yönelik tutumlarının çeşitli değişkenler açısından incelenmesi [Investigation of Fine Arts high school students' attitudes towards piano lessons in terms of certain variables] | 2022 | Beste Kademli Güçlü | Master | High | High |
| 124 | Güzel sanatlar liselerinde 11 ve 12. sınıf öğrencilerinin bağlama eğitimi öğretim programının incelenmesi [Investigation of the baglama education curriculum of the 11th and 12th grade students in Fine Arts higs schools] | 2022 | Umut Yaymak | Doctoral | High | High |
| 125 | Güzel sanatlar liselerindeki keman öğrenci ve öğretmenlerinin uzaktan eğitim sürecinden etkilenme durumları [The students and teachers of fine arts in fine high schools areaffected by the distance education process] | 2022 | Osman Küçükkılınç | Master | High | High |
| 126 | Güzel sanatlar lisesi öğrencilerinin bireysel çalgı eğitimi dersine yönelik tutumlarının incelenmesi [The analysis of attitudes of students of fine arts high school towards individual instrument lesson] | 2022 | Nazlısu Örs | Master | High | High |
| 127 | Güzel sanatlar liselerindeki batı müziği, Türk sanat müziği ve Türk halk müziği teori ve uygulaması derslerinin öğretmen görüşlerine göre incelenmesi [Examination of western music, Turkish classical music and Turkish folk music theory and practice courses in fine arts high schools according to teacher opinions] | 2022 | Muhammet Hanifi Kervancıoğlu | Master | High | High |
| 128 | Güzel sanatlar liseleri müzik bölümlerindeki kanun öğretimi ders kitaplarının incelenmesi, buna dair uzman | 2021 | Dilşan Aslan | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | görüşleri ve öneri etüt çalışmaları [Law on the music departments of fine arts high schools review of teaching course books, regarding this expert opinions and recommendations studies] | | | | | |
| 129 | Güzel sanatlar liselerinde Türk sanat müziği koro dersinin yeri ve önemi [The place and importance of Turkish Art Music choir lesson in Fine Arts high schools] | 2022 | Ayşe Çağlar | Master | High | High |
| 130 | Güzel sanatlar liseleri başlangıç viyolonsel eğitiminde temel davranışların öğretilmesine yönelik eğitimci görüşleri [Educator views on target and target behaviors in fine arts high schools beginning cello education] | 2022 | Seyfullah Feyat | Master | High | High |
| 131 | Güzel sanatlar lisesi müzik bölümlerinde uzaktan eğitim sürecinin öğrenci görüşleri doğrultusunda değerlendirilmesi: Konya ili örneği [Evaluation of distance education process in Fine Arts high school music departments in the line of student opinions: The case of Konya province] | 2022 | Osman Sermikli | Master | High | High |
| 132 | Güzel sanatlar lisesi bilişim destekli müzik dersinin işlenişine ilişkin öğretmen görüşleri [Teacher's opinions on the processing of the fine arts high school informatics aid music course] | 2022 | Harun Barış | Master | High | High |
| 133 | Güzel sanatlar liselerinde uygulanan viyola dersi öğretim programlarının incelenmesi ve değerlendirilmesi [Examination and evaluation of the viola teaching programs applied at fine arts high schools] | 2022 | Sevda Çetin | Master | High | High |
| 134 | Güzel sanatlar liseleri müzik bölümü tercihinde etkili olan faktörler (Karadeniz bölgesi örneği) [Factors affecting the choice of fine arts high schools music department (Black Sea region example)] | 2022 | Orkun Gönül | Master | High | High |
| 135 | Güzel sanatlar liseleri keman eğitimi öğretmenlerinin uzaktan eğitim uygulamaları ve görüşleri (Doğu Karadeniz bölgesi örneği) [Fine arts high school violin instructors' online education applications and opinions (Eastern Black Sea region sample)] | 2022 | Onur Öztürkoğlu | Master | High | High |
| 136 | Güzel sanatlar lisesi klasik gitar eğitimine yönelik Türk halk ezgileri kaynaklı materyal öneri ve uygulamaları [Material proposals and applications of Turkish folk melodies for fine arts high school classical guitar education] | 2022 | Levent Doğru | Doctoral | High | High |
| 137 | Güzel sanatlar lisesi 9. ve 12. sınıf öğrencilerinin piyano dersi tutum ve motivasyonlarının incelenmesi (Karadeniz bölgesi örneği) [Comparison of the attitudes and motivations of 9th and 12th grade students studying in fine arts high school towards piano lesson (Black Sea region example)] | 2022 | Sıla Gönül | Master | High | High |
| 138 | Güzel sanatlar liselerinde “beden farkındalığı ve müzisyen sağlığı” dersinin gerekliliğine yönelik bir ihtiyaç ve durum analizi [A need and situation analysis related to necessity of body awareness and musicians health course in fine arts high schools] | 2022 | Şevval Satıcı | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| 139 | Güzel sanatlar liselerinde şelpe tekniği kullanımına yönelik öğretmen görüşlerinin değerlendirilmesi [Evaluation of teacher's opinions on using the selpe technique in fine arts hish schools] | 2023 | Mert Balcı | Master | High | High |
| 140 | Güzel sanatlar liseleri müzik bölümü piyano derslerinde Beyer op.101 metodunun kullanımına ilişkin değerlendirme [Evaluation regarding the use of Beyer op.101 method in piano lessons of fine arts high schools music department] | 2023 | Dilay Setenay Öztürk | Master | High | High |
| 141 | Klasik gitarın güzel sanatlar liseleri çalgı toplulukları dersinde geleneksel çalgılar ile birlikte kullanılması üzerine bir inceleme [A review on the use of classic guitar with traditional instruments in fine arts high school instrument ensembles course] | 2022 | Efgan Rende | Doctoral | High | High |
| 142 | Güzel sanatlar lisesi müzik bölümü keman öğrencilerinin barok dönem keman tekniklerini uygulamalarına yönelik öğretmen görüşleri [Teacher's opinions on the applications of Baroque period violin techniques by violin students of fine arts high school music department] | 2023 | Ali Özen | Master | High | High |
| 143 | Güzel sanatlar liseleri müzik bölümü 12. sınıf öğrencilerinin lisans kariyer düşüncelerini etkileyen faktörler (Akdeniz bölgesi örneği) [Factors affecting undergraduate career thoughts of fine arts high school 12th grade music department students] | 2023 | Harun Yacan | Master | High | High |
| 144 | Güzel sanatlar liseleri müzik bölümü gitar öğrencilerinin gitar dersine yönelik tutumlarının incelenmesi [Investigation of the attitudes of fine arts high schools department of music guitar students too guitar lesson] | 2023 | Celal Gülçelik | Master | High | High |
| 145 | Güzel sanatlar lisesi akordiyon çalgı öğretimine ilişkin müzik öğretmeni görüşleri ve müfredat programı örneği [Music teacher's opinions on accordion instrument teaching in fine arts high school and sample of the curriculum] | 2023 | Faruk Mehınagıç | Master | High | High |
| 146 | Güzel sanatlar lisesi müzik öğretmenlerinin kaynaştırma öğrencileriyle yürütülen çalgı derslerine ilişkin görüşleri [Fine arts high school music teachers' opinions on instrumental lessons with inclusion students] | 2023 | Berivan Dağhan | Master | High | High |
| 147 | Güzel sanatlar liseleri müzik bölümlerindeki viyola eğitimi sürecinin incelenmesi [Investigation of viola education process in fine arts high schools music departments] | 2023 | Barkın Çalış | Master | High | High |
| 148 | Güzel sanatlar liseleri keman eğitiminde kullanılmak üzere oluşturulan keman rapsodisinin icrasına yönelik model bir çalışma önerisi [A model study proposal for the performance of the violin rapsody created to be used in fine arts high schools violin education] | 2023 | Nilüfer Nogayoğlu | Master | High | High |
| 149 | Güzel sanatlar lisesi müzik bölümü öğrencilerinin akademik güdülenme düzeylerinin çeşitli değişkenlere göre | 2023 | Bilal Kavak | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | incelenmesi [Investigation of academic motivation levels of fine arts high school music department students according to various variables] | | | | | |
| 150 | Güzel sanatlar lisesi öğrencilerinin bireysel çalgı dersi motivasyon düzeylerinin çeşitli değişkenler açısından incelenmesi [Investigation of the individual instrument lesson motivation levels of high school fine arts students in terms of various variables] | 2023 | Anıl Altaç | Master | High | High |
| 151 | Güzel sanatlar liseleri "çalgı eğitimi kanun öğretim programının" değerlendirilmesi ve taslak öğretim programı önerisi [Evaluation of high school of fine arts "Qanon teaching program" and probosal for draft teaching program] | 2023 | Özgür Yakınlar | Master | High | High |
| 152 | Güzel sanatlar liselerinde uygulanan kabak kemane dersi öğretim programlarının incelenmesi ve değerlendirilmesi [Examination and evaluation of the teaching programs of the kabak kemane lesson applied in fine arts high schools] | 2023 | Seyithan Deniz | Master | High | High |
| 153 | Güzel sanatlar liseleri batı müziği teori ve uygulamaları dersinde karşılaşılan sorunlara ilişkin öğretmen ve öğrenci görüşlerinin değerlendirilmesi [Evaluation of teacher and student opinions on the problems encountered in the Western music theory and practices course in fine arts high schools] | 2023 | Nihan Davulcu Öztürk | Master | High | High |
| 154 | Türkiye'deki güzel sanatlar liselerinde viyola eğitiminin uzaktan eğitim sürecinde sürdürülmesine ilişkin öğretmen ve öğrenci görüşleri [The teacher and student opinions on continuing viola education during distance education programme at the fine arts high school in Turkey] | 2023 | Ümit Öztürk | Master | High | High |
| 155 | Güzel sanatlar lisesi müzik bölümü öğrencilerinin bireysel ses eğitimi dersindeki özyeterlilik algılarının incelenmesi: Akdeniz bölgesindeki güzel sanatlar liseleri [Examination of the self-efficiency perceptions of students take high schools of fine arts in the Mediterranean Region individual voice education course] | 2023 | Ceren Turabik | Master | High | High |
| 156 | Güzel sanatlar liseleri müzik bölümlerinde yan flüt eğitimi alan öğrencilerin teknik çalışma programları ve teknik çalışma farkındalıkları üzerine bir inceleme [Flute in music departments of fine arts high schools technical study programs of students receiving training and a review on technical work awareness] | 2023 | Ayşe Simge Serçe | Master | High | High |
| 157 | Güzel sanatlar lisesi çoksesli koro dersi öğretim programlarının değerlendirilmesi ve program önerisi [The evaluation of the fine arts high school 1 polyvocious choruse teaching programs and program recommendation] | 2023 | Gülay Laçin | Doctoral | High | High |
| 158 | Güzel sanatlar liselerinde çalgı eğitimi (keman) dersinde eşlikli çalışmanın öğrencinin müziksel gelişimine ve keman dersine bakış açısına etkisi [The musical development of the student and the accompanying study in the lesson of instrument education (Violin) at fine arts | 2023 | Gülşah Tazegül | Master | High | High |

| | | | | | | |
|---|---|---|---|---|---|---|
| | high schools effect on the view of the violin lesson] | | | | | |
| 159 | Güzel sanatlar liselerinde verilen bağlama çalgısına yönelik derslerde uygulama açısından karşılaşılan problemler [Problems encountered in terms of application in bağlama instrument courses given in fine arts high schools] | 2023 | Mesut Barış | Master | High | High |

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 277–292
https://doi.org/10.21449/ijate.1531350
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# Teacher perceived emotional job demands scale: Validity, reliability and invariance study

**Alper Uslukaya[1*], Yılmaz Arslan[2], Füsun Yükrük[3], Oktay Ağ[3], Mehmet Akif Evcimik[3]**

[1]Çankırı Karatekin University, Faculty of Humanities and Social Sciences, Department of Educational Sciences, Çankırı, Türkiye
[2]Republic of Türkiye Ministry of National Education, Van, Türkiye
[3]Republic of Türkiye Ministry of National Education, Elazığ, Türkiye

**ARTICLE HISTORY**
*Received: Aug. 10, 2024*
*Accepted: Jan. 11, 2025*

**Keywords:**
Emotional job demands,
Validity,
Reliability,
Invariance.

**Abstract:** The aim of this study was to develop a measurement tool to determine teachers' perceived emotional job demands. The item pool created as a result of a literature review conducted by researchers was subjected to expert evaluation for content, appearance, and meaning validity, and finally, a draft scale form was created. The draft form was applied to 207 teachers, and exploratory factor analysis was conducted with the obtained data, revealing a one-dimensional scale structure consisting of seven items. To confirm this structure, data were collected from 283 teachers, and the one-factor scale structure was confirmed based on fit indices estimated through confirmatory factor analysis. Furthermore, it was observed that the scale met the conditions of convergent validity and reliability as a result of the analyses conducted. Measurement invariance of the scale was tested for gender (female-male), marital status (married-single), and age (under 40 and over 40) categories, and it was found that the scale met the requirements of configural, metric, scalar, and strict invariance. In the final stage, the scale's nomological validity was established by examining its relationship with burnout and work engagement. Data collected from 245 teachers were analyzed using structural equation modeling, and the significant relationships confirmed the scale's nomological validity. In conclusion, it was determined that the scale in question is a valid and reliable tool that can be used to measure teachers' perception of emotional job demands.

## 1. INTRODUCTION

Unfortunately, teachers who start their careers with high expectations, hopes, and ideals can often lose their positive attitudes toward their profession in a short period (Pines, 2002). This is because teaching is both a stressful profession (Borg & Riding, 1991), and teachers often find themselves faced with many challenging situations that push their limits (Chang, 2009; Skaalvik & Skaalvik, 2011). Especially in recent years, the reforms and changes in the education systems of countries (Skaalvik & Skaalvik, 2015) and the increasing professionalization of the profession have led to a steady increase in responsibilities associated

*CONTACT: Alper USLUKAYA auslukaya@gmail.com Çankırı Karatekin University, Faculty of Humanities and Social Sciences, Department of Educational Sciences, Çankırı, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

with the profession (Bartlett, 2004). In Türkiye, when the structural features of the current system are considered together with the deep reforms and rapid changes in education, the increasing responsibilities of the profession are felt much more clearly. The relatively high class sizes in Türkiye increase the obligation of teachers to respond to individual needs (Özdemir *et al*., 2024). Additionally, rising parental expectations and the added responsibilities imposed on teachers during the digital transformation of education expose teachers to emotional job demands (Uslukaya & Demirtaş, 2023). Moreover, comprehensive policy changes such as Türkiye's 2023 Education Vision Document not only emphasize academic success but also raise expectations for teachers to support students' social and emotional development (Özer, 2018).

In recent years, significant immigration from various countries (e.g., Afghanistan, Pakistan, Syria) has further diversified student demographics in Türkiye. This growing diversity requires teachers to adapt to students with different linguistic, cultural, and educational backgrounds. Efforts to support the adaptation process and academic success of these students have notably increased teachers' workload (Bozkurt, 2023). Additionally, addressing the psychological and social challenges stemming from these students' traumatic experiences necessitates greater emotional sensitivity from teachers (Küçüksüleymanoğlu & Kurt, 2020). This heightens the frequency of emotionally charged interactions in the classroom. Such expectations increasingly expose teachers to emotional demands in schools, requiring them to engage in emotionally intensive interactions. Emotional job demands, which involve experiencing and managing emotionally charged situations, interactions, and events in the workplace (Tuxford & Bradley, 2015), unfortunately, place significant psychological strain on teachers (Taris & Schreurs, 2009).

Emotional job demands constitute one dimension of emotional labor in the literature. Emotional labor is a concept that refers to the efforts of individuals to manage their emotions for a specific purpose in business life. This concept was first defined by Hochschild (1983) and later developed in the literature under two main dimensions: surface acting and deep acting. While surface acting requires the individual to pretend to express emotions that he/she does not feel, deep acting involves the individual's effort to actually feel the emotions appropriate to the job requirements (Grandey, 2002). In addition to these two approaches, more recent studies have emphasized the importance of conceptualizing emotional labor by focusing on job demands. For example, Brotheridge and Grandey (2002) conceptualized emotional labor under two main headings: employee-focused approach and job-focused approach. Employee-focused emotional labor refers to the process or experience of managing emotions and expressions to meet job demands. In contrast, job-focused emotional labor (i.e., emotional job demands) corresponds to job expectations related to expressing specific emotions (Schaubroeck & Jones, 2000) and demands related to the frequency of interactions with stakeholders (Morris & Feldman, 1997). Therefore, job-focused emotional labor indicates the level of emotional demands in a profession (Brotheridge & Grandey, 2002). Employee-focused emotional labor captures an individual's perspective, and this perspective does not align with the job-focused concept from which the concept of job demands is derived (Näring *et al*., 2012). In this sense, job-focused emotional demands correspond to challenging situations inherent in the essence of a profession.

Morris and Feldman (1997) state that these demands are directly related to the frequency and intensity of interaction in the work environment. For example, in professions where human relations are intense, such as teaching, individuals are expected to display both a positive emotional state and emotional resilience. Therefore, emotional job demands are not limited to the perceptions of the individual but are directly related to the requirements arising from the nature of the profession (Brotheridge & Grandey, 2002). Thus, emotional job demands represent the emotional standards that employees must fulfill depending on the nature of the job. These standards are often shaped by the frequency, intensity, and emotional content of professional interactions and have both psychological and physiological effects on individuals (Näring *et al.*, 2012).

In countries like Türkiye that are still developing, the responsibility for achieving globally desired success heavily relies on education and, consequently, on teachers. These responsibilities not only increase the physical and cognitive demands on teachers but also elevate the emotional demands placed upon them. As with other organizations, the success of educational institutions is partly measured by their ability to meet these emotional demands (Scheibe *et al*., 2015). Consequently, the teaching profession has evolved from its traditional molds, transforming into a service-oriented profession where addressing emotional demands is increasingly essential (Tuxford & Bradley, 2015).

For instance, teachers must maintain patience, fairness, and self-control when dealing with students who have disciplinary issues. They are also responsible for showing sensitivity to students' familial and social challenges and adopting more supportive behaviors. Moreover, attending to the high expectations of parents regarding their children's academic performance or behavior has become a fundamental duty of teachers (Bodenheimer & Shuster, 2020). Teachers must also respond to the performance pressures, change requests, and demands for implementing new practices from central administration and school management. As a result, in addition to the emotionally charged interactions and demands inherent in the profession, teachers have increasingly been exposed to these emotional demands in the context of a changing global order. Indeed, some researchers have found that employees in high-contact professions, such as teaching, are subject to increasing emotional demands over time (e.g., Peng, 2017).

Although policymakers, practitioners, and service recipients desire that teachers effectively manage emotional demands, such demands inevitably come at a cost for teachers. Emotional demands require continuous emotional effort, which can deplete an individual’s resources and make them vulnerable to job-related difficulties (Bakker & Demerouti, 2017). Some researchers have highlighted the risks posed to teachers and schools, noting that those exposed to these demands often experience increased occupational stress, along with reduced well-being, performance, and productivity (Näring *et al.*, 2012). Traditional resource and demand models have repeatedly emphasized the potential risks associated with such demands. For example, the Conservation of Resources (COR) model (Hobfoll, 2002) highlights that under challenging conditions, such as emotional demands, employees’ efforts to conserve their resources and minimize the threat of resource loss can lead to negative outcomes. Additionally, the Job Demands-Resources (JD-R) model (Demerouti *et al.*, 2001) predicts that emotionally taxing job demands can deplete an individual’s energy, thereby reducing their well-being, performance, and productivity. Numerous empirical studies testing these models’ assumptions have consistently demonstrated that emotional job demands can negatively affect employees’ occupational well-being, thereby limiting their performance and productivity (Lavelle *et al.*, 2021; Rughoobur-Seetah, 2024; Tuxford & Bradley, 2015; Van den Heuvel *et al.*, 2010).

While these international studies highlight the risks of emotional demands for employees and, consequently, for teachers, it is essential to empirically test the effects of such demands in the context of Türkiye. On the other hand, studies conducted in Türkiye have generally focused on the employee-oriented conceptualization of emotional labor, analyzing the impacts of employees' surface and deep acting behaviors on their professional lives and well-being. One significant reason for this is the nature of the measurement tools used. The scales employed in Türkiye have been developed or adapted based on the employee-oriented approach to emotional labor (e.g., Begenirbas & Yalçın, 2012). However, local studies have largely overlooked the perspective of emotional labor specific to the teaching profession, namely emotional job demands. This situation may stem from the absence of a specific measurement tool designed to assess teachers' emotional job demands in Türkiye. This study aims to fill this gap by developing the "Teachers’ Perceived Emotional Job Demands Scale (T-PEJDS)." During the development process, the structure of the scale was first determined and validated using exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). Subsequently,

reliability analyses were conducted, and the scale was subjected to measurement invariance tests to ensure its suitability for group comparisons. In the final stage, the scale's nomological validity was assessed by examining its relationship with burnout and work engagement, as theoretically linked constructs (Bakker & Demerouti, 2017). The use of this scale in empirical studies is expected to contribute to a better understanding of the emotional job demands faced by teachers in Türkiye and to inform practices aimed at improving teachers' well-being and performance.

## 2. METHOD

### 2.1. Research Sample

The data used to develop the T-PEJDS were collected from three participant groups working in the city center of Elazığ. To identify the participants of the study, cluster sampling, one of the probabilistic sampling methods, was chosen (Creswell & Creswell, 2017). This method offers a cost-effective and time-efficient way to collect data from a large and heterogeneous population while also allowing for a sample that better represents the population (Groves *et al.*, 2011). Cluster sampling is frequently employed in educational research and provides a practical solution under the field conditions of the study. In the first stage, 100 schools were randomly selected from over 200 schools located in the Elazığ city center. The administrators of these schools were contacted and informed about the purpose and scope of the study. Out of the 100 administrators, 76 agreed to allow the surveys to be conducted in their schools. After creating a form that included scale items and demographic information, the research team visited these schools to administer the forms to teachers. In the first stage, 100 schools were randomly selected from over 200 schools located in the Elazığ city center. The administrators of these schools were contacted and informed about the purpose and scope of the study. Out of the 100 administrators, 76 agreed to allow the surveys to be conducted in their schools. After creating a form that included scale items and demographic information, the research team visited these schools to administer the forms to teachers. In the first wave of data collection, usable data were obtained from 207 teachers, which were used in the EFA. Approximately three weeks after the first wave, the same schools were revisited, and data were collected from 283 teachers who had not participated in the initial wave. These data were used in the CFA. In the final stage, a form including items related to burnout, work engagement, and the scale's items, along with demographic variables, was administered approximately three weeks after the second wave of data collection at the same schools. During this stage, usable data were collected from 245 teachers, and these data were used to assess the nomological validity of the T-PEJDS. Demographic information of the participants is presented in Table 1.

**Table 1.** *Demographic information of participants.*

| | | Group 1 (N=207) | | Group 2 (N=283) | | Group 3 (N=245) | |
|---|---|---|---|---|---|---|---|
| Categories | | ***f*** | ***%*** | ***f*** | ***%*** | ***f*** | ***%*** |
| Gender | Female | 108 | 52.2 | 168 | 59.4 | 134 | 54.7 |
| | Male | 99 | 47.8 | 115 | 40.6 | 111 | 45.3 |
| Marital Status | Married | 155 | 74.9 | 203 | 71.7 | 172 | 70.2 |
| | Single | 52 | 25.1 | 80 | 28.3 | 73 | 29.8 |
| Age | 21-30 | 23 | 11.2 | 34 | 12 | 28 | 11.4 |
| | 31-40 | 68 | 32.8 | 86 | 30.4 | 72 | 29.4 |
| | 41-50 | 65 | 31.4 | 108 | 38.2 | 100 | 40.8 |
| | 51+ | 51 | 24.6 | 55 | 19.4 | 45 | 18.4 |

### 2.2. Scale

To identify items that can measure emotional job demands, researchers conducted a literature review and created a pool of 22 items. This pool was reviewed by three experts in the field of educational management. Based on their feedback, eight items were removed from the pool because they were deemed to measure the same aspects as other items in the pool. Additionally, based on the recommendations of these experts, one item was added to the pool to be considered within the scope of emotional job demands. The 15 items that emerged from the experts' evaluations were further reviewed by two measurement and assessment experts for scientific rigor. Three items that had multiple interpretations were revised. In the final stage, two experts in the Turkish language reviewed the items for appropriateness in terms of meaning and grammar. As a result of the language experts' recommendations, two items were revised due to containing unnecessary expressions, and five items were modified for clarity and comprehensibility. Considering the way the items were formulated, the scale was finalized using a five-point Likert scale (*Strongly Disagree* [1], *Disagree* [2], *Undecided* [3], *Agree* [4], and *Strongly Agree* [5]).

### 2.3. Data Analysis

For data analysis, a series of validity conditions were initially established. In this context, content validity was assessed by consulting experts in educational management, face validity by experts in measurement and assessment, and semantic validity by experts in the Turkish language. Necessary adjustments were made based on their evaluations. Subsequently, using the SPSS 22 package program, an EFA was conducted to perform the factorization process of the 15-item scale that emerged. For the adequacy of the sample size in the data set used in the EFA, the Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's Sphericity Test analyses were checked to determine the suitability of the items in the scale for EFA. A KMO coefficient above .60 indicates that the sample is sufficient, and the significance of Bartlett's Sphericity Test ($p$<.05) suggests that the scale is suitable for factor analysis (Tabachnick & Fidell, 2007).

In EFA, Principal Axis Factoring was used to identify the underlying factor structure by analyzing the shared variance among items in the data set. This method is particularly suitable for uncovering latent constructs, as it focuses on the common variance shared by variables rather than total variance (Fabrigar *et al.*, 1999). The Direct Oblimin rotation technique was applied to allow for correlations between factors, facilitating a more accurate and interpretable representation of the measured structure (Çokluk *et al.*, 2010). For the assessment of items, item factor loadings ($\lambda$) and the common factor variance explained by the items ($h^2$) were utilized, and items with load values below .32 (Tabachnick & Fidell, 2007) and items with common factor variance values below .50 were removed from the scale (Hair *et al.*, 1998).

After the factorization process, CFA was conducted using Mplus 8.3 to confirm the structure, or model, revealed by EFA. To determine the estimator to be used for CFA analysis, the data set was checked for univariate and multivariate normality assumptions. Skewness and kurtosis values were examined for univariate normality, and since all values were within the range of ± 2.0, it can be stated that the assumption of univariate normality is met (George & Mallery, 2010). However, since Mardia's skewness and kurtosis values were found to be significant ($p$ < .05), it was concluded that the data set did not meet the assumption of multivariate normality. Therefore, analyses were conducted using the Yuan-Bentler robust Maximum Likelihood (MLR) estimator, which can be used in cases where the data set is not normally distributed and can provide robust estimates, in Mplus (Şen, 2020). In model evaluation, the non-significant chi-square ($\chi^2$) value was used as a criterion. However, due to the sensitivity of $\chi^2$ to sample size, alternative fit indices were also considered. For this purpose, CFI (Comparative Fit Index), TLI (Tucker–Lewis Index), RMSEA (Root Mean Square Error of Approximation), and SRMR (Standardized Root Mean Square Residual) fit criteria were used based on the recommendations of Xu and Tracey (2017). CFI and TLI values above .95, RMSEA and SRMR values below .05

indicate excellent fit; CFI and TLI values between .90-.95, RMSEA values between .05-.08, and SRMR values between .05-1 indicate acceptable fit (Hu & Bentler, 1999).

The convergence validity and reliability tests were conducted using the values obtained from the CFA analysis. For the reliability coefficient, Cronbach's Alpha (α) and composite reliability (ω) coefficients were examined, and coefficients above .70 were considered reliable values (Hair *et al.*, 2014). For convergence validity, the Average Variance Extracted (AVE) and composite reliability (ω) values were calculated. The criterion for meeting the convergence validity condition is having a composite reliability coefficient above .70, an AVE value above .50, and the composite reliability coefficient being greater than the AVE value (ω>AVE) (Fornell & Larcker, 1981).

After testing structural validity, convergent validity, and reliability conditions through CFA, measurement invariance was examined to determine whether the scale measures the same structure in different groups (e.g., gender, marital status, age). The decision to test measurement invariance across gender, age, and marital status categories was guided by the literature and the theoretical foundations of emotional job demands in the teaching profession. Research consistently demonstrates that gender, age, and marital status are significant determinants of emotional job demands (Närin *et al.*, 2006; Näring & van Droffelaar, 2007; Näring *et al.*, 2012). These demographic factors are highly relevant to understanding differences in teachers' perceptions of emotional job demands, as supported by studies in occupational and educational psychology (Chang, 2009; Skaalvik & Skaalvik, 2015). Furthermore, focusing on these specific categories aligns with the objectives of this study, which aim to develop a reliable and valid instrument that accounts for the key demographic variables influencing teachers' emotional job demands.

Scales that do not achieve measurement invariance can lead to incorrect results and interpretations in studies involving group comparisons (Byrne, 2008). In measurement invariance, a baseline model (configural model) is first created without any invariance constraints to see if the same structure exists in different groups. After the baseline model, a series of models are tested, starting with the constrained factor loadings model (metric model), followed by the constrained factor loadings and intercepts model (scalar model), and finally the model with constrained factor loadings, intercepts, and unique factor variances (strict model) (Widaman *et al.*, 2010). In model evaluations, each model is compared to the previous model, and changes in chi-square ($\Delta\chi^2$) are analyzed. A non-significant $\Delta\chi^2$ test ($p > .05$) is considered evidence of measurement invariance for the tested model. Additionally, due to the sensitivity of the chi-square test to sample size, the threshold values for model comparisons were determined based on Chen's (2007) recommendations. Specifically, a change in the CFI of no more than .010 ($\Delta CFI \leq .01$) and a change in the RMSEA of no more than .015 ($\Delta RMSEA \leq .015$) were established as acceptable limits.

In the final stage, we conducted a latent variable path analysis to test the nomological validity of the scale. Nomological validity refers to the extent to which a measured construct fits within a network of theoretically related constructs (Kock *et al.*, 2024). In other words, a construct is considered valid and meaningful to the degree that it correlates with other variables it has been theoretically linked to. This validity is established by analyzing the relationships (e.g., positive or negative correlations) between the measured construct and different variables, demonstrating consistency with the construct's theoretical framework (Cronbach & Meehl, 1955). In the literature, exposure to emotional job demands leads to strong negative emotions and emotional effort in the workplace (Taris & Schreurs, 2009). These negative emotions and emotional efforts have the potential to deplete individuals' resources (Van den Heuvel *et al.*, 2010). Indeed, emotional demands are a significant source of strain in professions that involve intensive interpersonal relationships, such as teaching, and can adversely affect employee well-being (Sutton & Wheatley, 2003). In occupational health literature, emotional demands are

considered a cause of burnout (Maslach & Schaufeli, 2018) and are assumed to significantly undermine work engagement (Xanthopoulou *et al.*, 2013). Therefore, for the T-PEJDS scale to demonstrate nomological validity, it should positively predict burnout and negatively predict work engagement. For this purpose, the 10-item burnout scale developed by Pines (2005) and adapted into Turkish by Tümkaya *et al.* (2009), as well as the nine-item work engagement scale developed by Schaufeli *et al.* (2006), were used. The goal was to examine whether perceived emotional job demands among teachers predict burnout and work engagement.

### 2.4. Ethical

In the scope of the research, ethical approval was obtained from the Ethics Committee for Social and Human Sciences Research at Fırat University.

## 3. FINDINGS

In this section, initially, EFA was conducted for factorization of the scale. Subsequently, CFA was performed to confirm the structure revealed by EFA. Reliability and convergent validity analyses were conducted using standardized path coefficients obtained from CFA. Additionally, measurement invariance was tested using data used for CFA. Lastly, the nomological validity of the scale was examined.

### 3.1. Exploratory Factor Analysis (EFA)

The result of the EFA conducted for the factorization process yielded a KMO value of .91, and Bartlett's sphericity test ($\chi^2$ = 1518.90; *df* = 105; $p < .001$) was statistically significant. These results indicate that the sample size is sufficient, and the data set is suitable for factorization. Item-level factor loadings and common variance values were examined, and items with values below the predetermined cut-off points were gradually removed from the analysis, starting with the lowest values. As a result of this process, eight items with both factor loadings and common variance values below the specified cut-off points (ƛ< .32 and $h^2 < .50$) were excluded from the analysis. A one-factor structure consisting of seven items was predicted as a result of the EFA. The EFA results are presented in Table 2.

**Table 2.** *Results of the EFA for the T-PEJDS.*

| Items | ƛ | $h^2$ | α if items are removed |
|---|---|---|---|
| item1 | .90 | .82 | .94 |
| item4 | .85 | .72 | .94 |
| item5 | .79 | .63 | .95 |
| item6 | .84 | .71 | .94 |
| item8 | .85 | .72 | .94 |
| item11 | .86 | .75 | .94 |
| item12 | .87 | .76 | .93 |
| Eigenvalue | 5.09 | | |
| Total Variance Explained % | 72.79 | | |
| Cronbach Alpha | | | .95 |

*Note*. ƛ= Item loading; $h^2$= Common variance; α= Cronbach Alpha

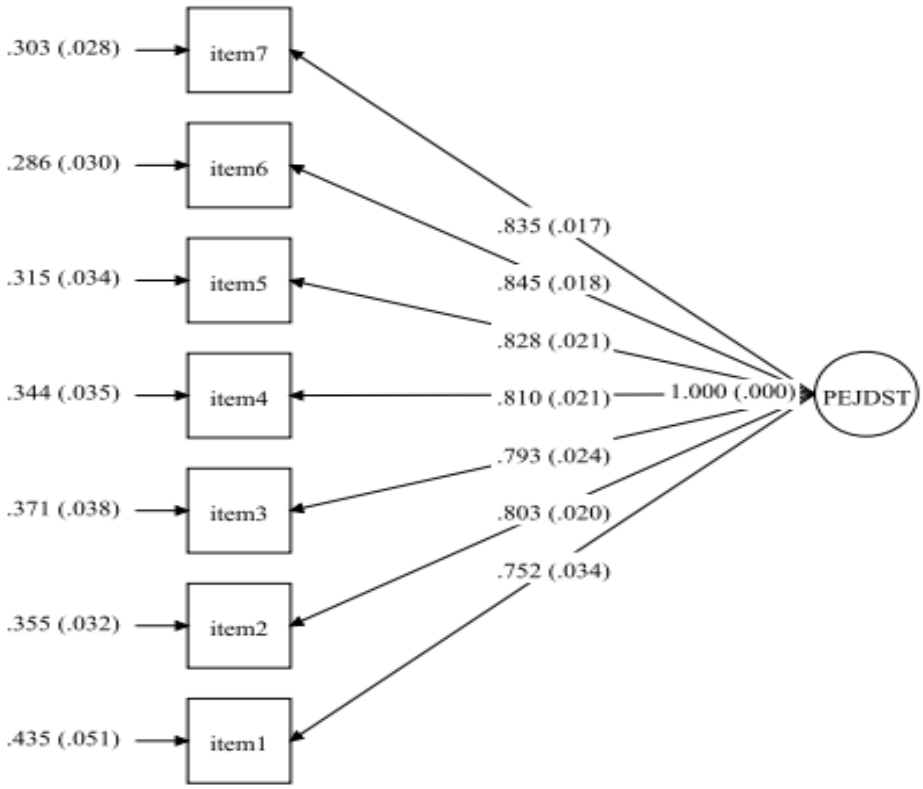
The results in Table 2 indicate that a scale structure consisting of a single factor with an eigenvalue of 5.09 has been identified. Furthermore, the single-factor structure explains 72.79% of the total variance. Item loadings (ƛ) ranged from .79 to .90, indicating strong structural relationships between these items and the single-factor structure. All items had common variance values ($h^2$) above .50, which can be considered evidence of item homogeneity (Çokluk

*et al.*, 2010). The Cronbach's Alpha coefficient for the seven-item single-factor structure was estimated at .95, indicating a high level of internal consistency.

### 3.2. Confirmatory Factor Analysis (CFA)

Confirmatory Factor Analysis (CFA) was conducted to confirm the structure of the T-PEJDS, which consists of a single factor with seven items, as a result of the EFA. For this purpose, the data collected from 283 teachers were used. Tabachnick and Fidell (2007) recommend a sample size of at least 200 for conducting multivariate statistical analyses. Additionally, Kline (2005) suggests that the ratio of cases to the number of free parameters should be 10:1 for conducting CFA. Therefore, the sample size of the study is sufficient both according to Tabachnick and Fidell's (2007) recommendation ($N > 200$) and the criterion set by Kline (2005) (7x10 = 70 < 283). The measurement model obtained as a result of the CFA analysis of the T-PEJDS, using the MLR estimator, is shown in Figure 1.

**Figure 1.** *The result of the CFA Analysis of the T-PEJDS.*




X2= 125.653 [df= 14; p= .0586], RMSEA=.054 [%90 CI = .017- .087], CFI= .989, TLI= .984, and SRMR= .020

The findings indicate that the single-factor T-PEJDS model fits the data quite well ($\chi^2$ = 125.653 [*df* = 14; *p* =.0586], RMSEA =.054 [90% CI =.017-.087], CFI =.989, TLI =.984, and SRMR = .020). It has been observed that the indicators forming the latent variable of T-PEJDS have standardized factor loadings ranging from .752 to .845. This suggests strong structural relationships between these indicators and the T-PEJDS latent variable. Additionally, using the standardized factor loadings obtained from the CFA analysis, AVE and composite reliability (ω) values were calculated. The AVE value was estimated at .656, and the composite reliability value was .930. Since the AVE value exceeds .50, the composite reliability value exceeds .70, and the composite reliability value is higher than the AVE value, it can be concluded that the convergent validity of the scale is established. Moreover, both Cronbach's Alpha and composite reliability coefficients were estimated at a high level (ω=.930, α=.927), indicating that the scale meets the reliability requirement.

### 3.3. Measurement Invariance

Measurement invariance models were tested using the data used in the CFA analysis to assess whether the scale measures the same structure across different groups (gender, marital status, and age). The results are presented in Table 3.

**Table 3.** *Measurement invariance fit indices (N=283).*

| Models | $\chi^2$ | *df* | SRMR | TLI | CFI | RMSEA | $\Delta\chi^2$ | $\Delta df$ | *p* | ΔCFI | ΔRMSEA |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Female N=168); Male (115)* | | | | | | | | | | | |
| Configural model | 53.733 | 28 | .026 | .982 | .972 | .078 | | | | | |
| Metric model | 66.158 | 34 | .059 | .977 | .972 | .079 | 12.425 | 6 | .053 | .000 | .001 |
| Scalar model | 74.375 | 40 | .059 | .975 | .974 | .078 | 8.217 | 6 | .055 | .002 | .001 |
| Strict model | 81.927 | 47 | .059 | .978 | .978 | .072 | 7.552 | 7 | .374 | .004 | .006 |
| *Married (N=203); Single (80)* | | | | | | | | | | | |
| Configural model | 40.144 | 28 | .023 | .989 | .984 | .055 | | | | | |
| Metric model | 47.587 | 34 | .051 | .988 | .985 | .053 | 7.443 | 6 | .309 | .001 | .002 |
| Scalar model | 55.144 | 40 | .060 | .987 | .986 | .052 | 7.557 | 6 | .285 | .001 | .001 |
| Strict model | 63.767 | 47 | .059 | .985 | .987 | .050 | 8.623 | 7 | .281 | .001 | .002 |
| *Early adult (N=118); Mid. adult (N=165)* * | | | | | | | | | | | |
| Configural model | 40.786 | 28 | .024 | .989 | .983 | .057 | | | | | |
| Metric model | 44.113 | 34 | .033 | .991 | .989 | .046 | 3.377 | 6 | .934 | .006 | .011 |
| Scalar model | 49.613 | 40 | .035 | .992 | .991 | .041 | 5.500 | 6 | .533 | .002 | .005 |
| Strict model | 59.188 | 47 | .049 | .989 | .991 | .043 | 9.575 | 7 | .214 | .000 | .002 |

*In this study, Levinson's (1986) life structure, the concept of the life cycle, was utilized. The author defines the age ranges as follows: 20-40 years as early adulthood; 40-60 years as middle adulthood; and 60 years and above as late adulthood. Participants who were 40 years old were categorized into the early adulthood group, while those who were 41 years and older were considered part of the middle adulthood group.

The findings in Table 3 indicate that the configural model has good fit indices for the gender variable ($\chi^2$= 53.733 (*df* = 28), RMSEA = .078, CFI = .982, TLI =.972, and SRMR =.026), confirming the structure of the configural model. In a stepwise analysis, comparisons between the metric model and the configural model ($\Delta\chi^2$=12.425, $\Delta df$= 6, $p$ >.05), the scalar model and the metric model ($\Delta\chi^2$=8.217, $\Delta df$= 6, $p$ >.05), and the strict model and the scalar model ($\Delta\chi^2$=7.552, $\Delta df$= 7, $p$ >.05) were not statistically significant. Additionally, the changes in CFI (ΔCFI ≤ .010) and RMSEA (ΔRMSEA ≤.015) did not exceed the recommended thresholds at any step of the analysis, further supporting the invariance across gender. Therefore, it can be said that the configural, metric, scalar, and strict invariance of the T-PEJDS for the gender variable are fully supported.

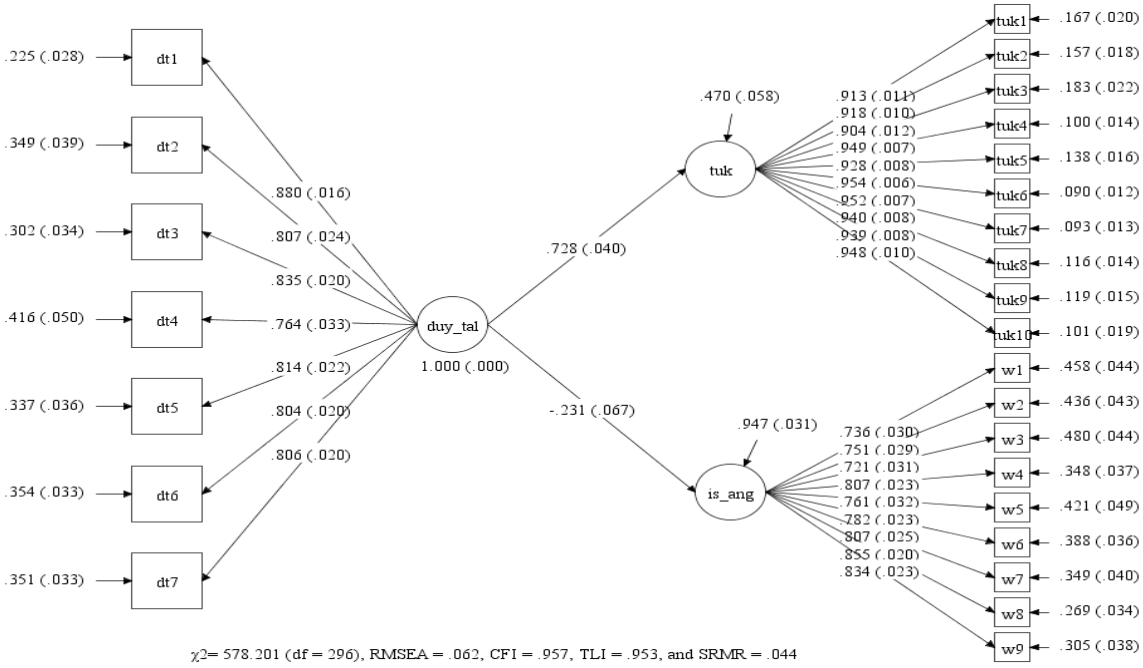
The results show that the configural model has good fit indices for the marital status variable ($\chi^2$=40.144 (*df*=28), RMSEA =.057, CFI =.989, TLI =.984, and SRMR =.023), confirming the structure of the configural model. In a stepwise analysis, comparisons between the metric model and the configural model ($\Delta\chi^2$=7.443, $\Delta df$= 6, $p$ >.05), the scalar model and the metric model ($\Delta\chi^2$=7.557, $\Delta df$= 6, $p$ >.05), and the strict model and the scalar model ($\Delta\chi^2$=8.623, $\Delta df$= 7, $p$ >.05) were not statistically significant. Similarly, the changes in CFI (ΔCFI ≤ .010) and RMSEA (ΔRMSEA ≤ .015) were within acceptable limits, confirming measurement invariance across marital status. Therefore, it can be concluded that the configural, metric, scalar, and strict invariance of the T-PEJDS for the marital status variable are fully supported.

Finally, measurement invariance was tested for the age variable. The results show that the configural model has good fit indices for the age variable ($\chi^2$=40.786 (*df* =28), RMSEA=.057, CFI=.989, TLI=.983, and SRMR=.024). Therefore, the first condition of measurement invariance was met for the age variable. In stepwise analyses for the age variable, comparisons between the metric model and the configural model ($\Delta\chi^2$=3.377, $\Delta df$= 6, $p$ >.05), the scalar model and the metric model ($\Delta\chi^2$=5.500, $\Delta df$= 6, $p$ >.05), and the strict model and the scalar model ($\Delta\chi 2$=9.575, $\Delta df$= 7, $p$ >.05) were not statistically significant. Moreover, changes in CFI ($\Delta$CFI ≤ .010) and RMSEA ($\Delta$RMSEA ≤ .015) remained below the recommended thresholds, providing additional evidence for the invariance across age. Hence, it can be concluded that the T-PEJDS, for the age variable, fully supports the configural model, as well as metric, scalar, and strict invariance.

### 3.4. Nomological Validity

In this stage, the nomological validity of the T-PEJDS scale was tested by analyzing its relationship with burnout and work engagement. To verify these associations, data collected from 245 teachers were tested using structural equation modeling, and the model fit indices along with standardized path coefficients are presented in Figure 2.

**Figure 2.** *SEM results.*




The results indicate that the model developed to test the nomological validity of the T-PEJDS scale demonstrates good fit indices ($\chi^2$ =578.201 (*df*=296), RMSEA=.062, CFI=.957, TLI=.953, and SRMR = .044). Additionally, the standardized path coefficients reveal significant relationships between emotional demands and burnout (β =.728, $p$ <.001) as well as work engagement (β=-.231, $p$<.001). These findings provide evidence that the scale possesses adequate nomological validity.

## 4. DISCUSSION

Although numerous studies have highlighted the negative impact of emotional job demands on employee well-being (e.g., Sutton & Wheatley, 2003; Taris & Schreurs, 2009; Van den Heuvel *et al.*, 2010), there is currently no measurement tool in the Turkish context specifically assessing teachers' emotional job demands. This study aimed to address this gap by developing a valid, reliable, and inter-group equivalent scale. A pool of 22 items was created through a literature review and refined through expert evaluations of content, face, and semantic validity, resulting in a 15-item draft scale. EFA with data from 207 teachers revealed a unidimensional seven-

item structure accounting for 72.78% of the total variance (see Appendix 1 for the seven-item structure). CFA with data from 283 teachers confirmed this structure, with good fit indices. The scale's convergent validity was supported by Average Variance Extracted (AVE) and Composite Reliability (ω) values, and reliability was confirmed with Cronbach's Alpha and composite reliability values exceeding .70. Measurement invariance analyses for gender, marital status, and age demonstrated configural, metric, scalar, and strict invariance, indicating that the scale reliably measures the same construct across these groups. Finally, nomological validity was established by testing the relationships between emotional job demands, burnout, and work engagement using structural equation modeling with data from 245 teachers. The significant relationships observed confirmed the scale's validity.

### 4.1. Implications

Considering the importance of simple measurements of fundamental structures in research, our seven-item scale represents a step towards facilitating further studies in educational contexts and provides a much-needed, reliable, and straightforward tool for assessing teachers' perceptions of emotional job demands. Emotional job demands are typically examined under the broader concept of emotional labor in the literature and, as such, have often been overlooked in terms of their antecedents and outcomes in educational research. This study conceptualized emotional job demands as a distinct construct based on the perspectives of previous researchers who theorized them as different from emotional labor. This approach could inspire studies aimed at uncovering the unique antecedents and outcomes of emotional job demands, thus broadening the scope of research in educational settings. Particularly in professions characterized by human interaction, such as teaching (Tuxford & Bradley, 2015), and with the growing emphasis on student-centered education in recent years, the scale could pave the way for a deeper analysis of emotional job demands (Peng, 2017). Our scale has the potential to clarify the nomological network of this construct, leading to a more comprehensive understanding of its role in educational contexts.

In developing the scale, we were guided by resource theories that assume emotionally charged and effortful interactions (i.e., emotional job demands) create challenging conditions for employees, leading to negative work outcomes (e.g., job dissatisfaction, burnout, turnover intentions). Therefore, our scale offers intriguing opportunities for studies that integrate emotional job demands to understand the mechanisms and moderating conditions underlying these negative work outcomes for teachers. For instance, researchers could use the T-PEJDS to explore under what conditions emotional job demands predict these negative work outcomes and through which mediators the effects of emotional job demands are transmitted to these outcomes. Within this framework, such studies could contribute to the development of effective interventions targeting these negative work outcomes. For example, school administrators could take targeted measures to mitigate the impact of emotional job demands on teachers with high perceptions of such demands by employing resources like organizational support that buffers these effects.

### 4.2. Limitations

This study has several limitations that present opportunities for future research. First, cross-sectional data were used to test nomological validity. Cross-sectional data capture the simultaneous relationships between variables, but these relationships may not hold over time. This lack of longitudinal validation is a significant limitation that restricts the scale's utility in assessing long-term outcomes or changes in the construct being measured. Future research could address this limitation by adopting a longitudinal design with multiple data collection points to capture time-lagged relationships between key variables. Second, the reliance on self-report measures in the study may have introduced response biases stemming from participants' desire to avoid embarrassment and project a positive image to others (Fisher, 1993). To enhance the validity of findings, future studies could incorporate multi-source and objective measures,

such as systematic observations. Third, the items developed for the scale were grounded in theoretical and empirical literature. However, teachers increasingly encounter new emotionally challenging demands. For instance, the widespread use of mass communication tools to coordinate school operations (despite one item related to this being removed during the EFA process) and the intrusion into personal life that this entails has become a notable issue in recent years. Consequently, the emergence of new items to measure emotional job demands is inevitable. This calls for regular reviews of the developed items to ensure the inclusion of the most current and relevant information. By addressing these limitations and pursuing these avenues, future research could provide a more comprehensive understanding of the impact of emotional job demands on teachers and schools.

## 5. CONCLUSION

This study aimed to develop a reliable and valid scale to measure teachers' perceptions of emotional job demands. The results show that the seven-item T-PEJDS meets the conditions of construct, convergent, and nomological validity, exhibits high reliability values, and also meets the criteria of measurement invariance. These results indicate that the T-PEJDS is a reliable and valid scale for measuring teachers' perceptions of emotional job demands. All items in the scale have been negatively coded. The scale was prepared in a five-point Likert format, with the highest possible score being 35 and the lowest score being 7. Higher scores indicate a higher level of perceived emotional job demands, while lower scores indicate a lower level of perceived emotional job demands. Therefore, it is expected that researchers using the scale pay attention to this aspect.

### Acknowledgments

The authors would like to thank all the participants.

### Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Fırat University, 05.10.2023-18829.

### Contribution of Authors

**Alper Uslukaya**: Investigation, Methodology, Software, Formal analysis, Writing-original draft, Supervision, and Validation. **Yılmaz Arslan:** Investigation, Resources, Visualization, and Writing-original draft. **Füsun Yükrük:** Data collection, Investigation, Resources, Visualization, Software, and Formal analysis. **Oktay Ağ:** Data collection, Resources, Methodology, and Validation. **Mehmet Akif Evcimik:** Data collection, Resources, Methodology, and Validation.

### Orcid

Alper Uslukaya https://orcid.org/0000-0003-1455-8438
Yılmaz Arslan https://orcid.org/0009-0004-5886-0588
Füsun Yükrük https://orcid.org/0009-0008-9690-0279
Oktay Ağ https://orcid.org/0000-0002-2745-9378
Mehmet Akif Evcimik https://orcid.org/0000-0001-5264-8018

## REFERENCES

Bakker, A.B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. *Journal of Occupational Health Psychology*, *22*(3), 273-285. http://dx.doi.org/10.1037/ocp0000056

Bartlett, L. (2004). Expanding teacher work roles: a resource for retention or a recipe for overwork?. *Journal of Education Policy*, *19*(5), 565-582. https://doi.org/10.1080/0268093042000269144

Begenirbaş, M., & Yalçın, R.C. (2012). The effects of teachers' personalities on their emotional labor display. *Çağ University Journal of Social Sciences*, *9*(1), 47-65.

Bodenheimer, G., & Shuster, S.M. (2020). Emotional labour, teaching and burnout: Investigating complex relationships. *Educational Research*, *62*(1), 63-76. https://doi.org/10.1080/00131881.2019.1705868

Borg, M.G., & Riding, R.J. (1991). Occupational stress and satisfaction in teaching. *British Educational Research Journal*, *17*(3), 263-281.https://doi.org/10.1080/0141192910170306

Bozkurt, B. (2023). Problems experienced by teachers in classrooms with high concentrations of Syrian students and solution suggestions. *MANAS Journal of Social Studies*, *12*(4), 1222-1235.

Brotheridge, C.M., & Grandey, A.A. (2002). Emotional labor and burnout: Comparing two perspectives of "people work". *Journal of Vocational Behavior*, *60*(1), 17-39. https://doi.org/10.1006/jvbe.2001.1815

Byrne, B.M. (2008). Testing for multigroup equivalence of a measuring instrument: A walk through the process. *Psicothema*, *20*(4), 872–882.

Chang, M.L. (2009). An appraisal perspective of teacher burnout: Examining the emotional work of teachers. *Educational Psychology Review*, *21*(3), 193-218. https://doi.org/10.1007/s10648-009-9106-y

Chen, F.F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464-504.

Creswell, J.W., & Creswell, J.D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests**.** *Psychological Bulletin, 52*(4), 281–302. https://doi.org/10.1037/h0040957

Çokluk, O., Şekercioğlu, G. & Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları* [*Multivariate statistics for social sciences SPSS and LISREL applications*]. Pegem-A Publications.

Demerouti, E., Bakker, A.B., Nachreiner, F., & Schaufeli, W.B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology*, *86*(3), 499-512. https://doi.org/10.1037/0021-9010.86.3.499

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299.

Fisher, R.J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of Consumer Research*, *20*(2), 303–315. https://doi.org/10.1086/209351

Fornell, C. & Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, *18*(1), 39-50. https://doi.org/10.1177/002224378101800104

George, D., & Malley, P. (2010). *SPSS for Windows step by step. A simple study guide and reference*. Pearson

Groves, R.M., Fowler Jr, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2011). *Survey methodology*. John Wiley & Sons.

Hair, J.D., Anderson, R.E., Tatham, R.L., & Black, W.C. (1998). *Multivariate date analysis*. Prentice Hall.

Hobfoll, S.E. (2002). Social and psychological resources and adaptation. *Review of General Psychology*, *6*(4), 307-324. https://doi.org/10.1037/1089-2680.6.4.307

Hochschild, A.R. (1983) *The managed heart: Commercialization of human feeling*. University of California.

Hu, L. & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. https://doi.org/10.1080/10705519909540118

Kline, R.B. (2005). *Principles and practice of structural equation modeling*. Guilford Publications.

Kock, F., Berbekova, A., Assaf, A.G., & Josiassen, A. (2024). Developing a scale is not enough: on the importance of nomological validity. *International Journal of Contemporary Hospitality Management*, *36*(10), 3233-3244.https://doi.org/10.1108/IJCHM-07-2023-1078

Küçüksüleymanoğlu, R., & Kurt, A. (2020). Suriyeli öğrenciler İçin görevlendirilen PICTES öğretmenlerinin tükenmişlik düzeyleri: Bursa örneği [Burnout levels of PICTES teachers assigned for Syrian students: Bursa case]. *Sakarya University Journal of Education Faculty*, *20*(1), 24-45.

Lavelle, J.J., Rupp, D.E., Herda, D.N., Pandey, A., & Lauck, J.R. (2021). Customer injustice and employee performance: Roles of emotional exhaustion, surface acting, and emotional demands-abilities fit. *Journal of Management*, *47*(3), 654-682. https://doi.org/10.1177/0149206319869426

Maslach, C. (1982). *Burnout: The cost of caring*. Prentice Hall.

Maslach, C., & Schaufeli, W.B. (1993). Historical and conceptual development of burn-out. In W.B. Schaufeli, C. Maslach, & T. Marck (Eds.), *Professional burn-out: Recent developments in the theory and research* (pp. 1-16). Taylor & Francis. https://doi.org/10.4324/9781315227979-1

Morris, J.A., & Feldman, D.C. (1997). Managing emotions in the workplace. *Journal of Managerial Issues*, 257-274.

Naring, G., Briet, M., & Brouwers, A. (2006). Beyond demand–control: Emotional labour and symptoms of burnout in teachers. *Work & Stress*, *20*, 303-315. https://doi.org/10.1080/02678370601065182

Naring, G., & van Droffelaar, A. (2007). Incorporation of emotional labour in the demand–control–support model: The relation with emotional exhaustion and personal accomplishment in nurses. In C.E.J. Härtel, N.M. Ashkanasy, & W.J. Zerbe (Eds.), *Research on emotion in organizations: Vol. 3* (pp. 221–236). Emerald.

Näring, G., Vlerick, P., & Van de Ven, B. (2012). Emotion work and emotional exhaustion in teachers: The job and individual perspective. *Educational Studies*, *38*(1), 63-72. https://doi.org/10.1080/03055698.2011.567026

Özdemir, F., Kaya, G., Berkil, S., & Avcı, F. (2024). Öğretmenlerin tükenmişlik düzeyinin sınıf yönetimine etkisi [The effect of teachers' burnout level on classroom management]. *Ulusal Özgün Eğitim Araştırmaları Dergisi*, *2*(1), 51-65.

Özer, M. (2018). The 2023 Education Vision and new goals in vocational and technical education. *Journal of Higher Education and Science*, *8*(3), 425-435.

Peng, K.Z. (2017). Exhaustion and emotional demands in China: A large-scale investigation across occupations. *Frontiers of Business Research in China*, *11*(1), 1-17. https://doi.org/10.1186/s11782-017-0003-9

Pines, A.M. (2002). Teacher burnout: A psychodynamic existential perspective. *Teachers and Teaching*: *Theory and Practice*, *8*(2), 121-141.

Schaubroeck, J., & Jones, J.R. (2000). Antecedents of workplace emotional labor dimensions and moderators of their effects on physical symptoms. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *21*(2), 163-183. https://doi.org/10.1002/(SICI)1099-1379(200003)21:2<163::AID-JOB37>3.0.CO;2-L

Schaufeli, W.B., Bakker, A.B., & Salanova, M. (2006). The measurement of work engagement with a short questionnaire: A cross-national study. *Educational and Psychological Measurement*, *66*(4), 701-716. https://doi.org/10.1177/0013164405282471

Scheibe, S., Stamov-Roßnagel, C., & Zacher, H. (2015). Links between emotional job demands and occupational well-being: Age differences depend on type of demand. *Work, Aging and Retirement*, *1*(3), 254-265. https://doi.org/10.1093/workar/wav007

Skaalvik, E.M., & Skaalvik, S. (2011). Teacher job satisfaction and motivation to leave the teaching profession: Relations with school context, feeling of belonging, and emotional exhaustion. *Teaching and Teacher Education*, *27*(6), 1029-1038. https://doi.org/10.1016/j.tate.2011.04.001

Skaalvik, E.M., & Skaalvik, S. (2015). Job satisfaction, stress and coping strategies in the teaching profession-what do teachers say?. *International Education Studies*, *8*(3), 181-192. https://doi.org/10.5539/ies.v8n3p181.

Sutton, R.E., & Wheatley, K.F. (2003). Teachers' emotions and teaching: A review of the literature and directions for future research. *Educational Psychology Review*, *15*(4), 327-358. https://doi.org/10.1023/A:1026131715856

Şen, S. (2020). *Mplus ile yapısal eşitlik modellemesi uygulamaları* [*Structural equation modeling applications with Mplus*]. Nobel Pablications.

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics*. Allyn & Bacon Publishing.

Taris, T.W., & Schreurs, P.J. (2009). Explaining worker strain and learning: How important are emotional job demands?. *Anxiety, Stress & Coping*, *22*(3), 245-262. https://doi.org/10.1080/10615800802460401

Tumkaya, S., Cam, S., & Cavusoglu, I. (2009). Turkish adaptation of the burnout snydrome inventory short version. *CU Social Sciences Institute J*, *18*, 387-98.

Tuxford, L.M., & Bradley, G.L. (2015). Emotional job demands and emotional exhaustion in teachers. *Educational Psychology*, *35*(8), 1006-1024. https://doi.org/10.1080/01443410.2014.912260

Uslukaya, A., & Demirtas, Z. (2023). Can a supportive supervisor be a treasure for teachers? A multilevel analysis of the relationship between perceived supervisor support, work engagement, and burnout. *Psychology in the Schools*, *60*(11), 4667-4685. https://doi.org/10.1002/pits.23006

Van den Heuvel, S.G., Geuskens, G.A., Hooftman, W.E., Koppes, L.L., & Van den Bossche, S.N. (2010). Productivity loss at work; health-related and work-related factors. *Journal of Occupational Rehabilitation*, *20*(3), 331-339. https://doi.org/10.1007/s10926-009-9219-7

Widaman, K.F., Ferrer, E., & Conger, R.D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*(1), 10-18. https://doi.org/10.1111/j.1750-8606.2009.00110.x

Xu, H., & Tracey, T.J. (2017). Use of multi-group confirmatory factor analysis in examining measurement invariance in counseling psychology research. *The European Journal of Counselling Psychology*, *6*(1), 75-82. https://doi.org/10.5964/ejcop.v6i1.120

## APPENDIX

**Appendix 1.** *Teacher perceived emotional job demands scale - Turkish version*

| Maddeler | Öğretmenlerin Algıladıkları Duygusal İş Talepleri Ölçeği | Kesinlikle Katılmıyorum | Katılmıyorum | Kararsızım | Katılıyorum | Kesinlikle Katılıyorum |
|---|---|---|---|---|---|---|
| 1 | Öğretmenlik, duygusal olarak yoğun çaba gerektiren bir meslektir. | | | | | |
| 2 | Okul yöneticilerinin baskılarına veya nezaketsiz davranışlarına rağmen onlara karşı iyi davranmak zorunda hissederim | | | | | |
| 3 | İsteksiz olmama rağmen, davranış sorunları olan öğrencilerle ilgilenmek durumunda kalırım. | | | | | |
| 4 | Öğrenciler için elimden geleni yapmama rağmen, velilerden şikayet alırım. | | | | | |
| 5 | Uyum sağlayamadığım öğretmen arkadaşlarımla çalışmak beni duygusal olarak zorlar. | | | | | |
| 6 | Milli Eğitim Bakanlığı'nın standartlarını karşılamaya çalışmak beni duygusal olarak yıpratır. | | | | | |
| 7 | İl veya İlçe Milli Eğitim Müdürlüğü'nün onaylamadığım talepleriyle uğraşmak zorunda kalırım. | | | | | |

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 293–316
https://doi.org/10.21449/ijate.1467609
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# Preschool teachers' awareness and views of children's self-regulation scales: A professional development program's effect

**Asiye İvrendi**[1*]

[1]Pamukkale University, Faculty of Education, Department of Primary Education, Preschool Education, Türkiye

**ARTICLE HISTORY**

*Received: Apr. 12, 2024*

*Accepted: Mar. 6, 2025*

**Keywords:**

Preschool teachers' awareness and views,
Self-regulation skills,
Professional development.

**Abstract:** There has been increased evidence about the importance of supporting children's self-regulation skills in early childhood and the role of teachers in this process. This current research consisted of two studies. Study 1 aimed to develop two scales: one to assess preschool teachers' awareness of children's self-regulation and the other to determine their views of children's self-regulation skills. To examine the scales' construct validity, Exploratory Factor Analysis (EFA- $n = 201$) and Confirmatory Factor Analysis (CFA- $n = 123$) were conducted for both the awareness and views scales. EFA yielded two subscales for the awareness scale: self-regulation (7 items- $\omega = .81$) and relation to learning (3 items- $\omega = .76$). CFA results showed that the awareness scale's 10-item and two-factor structure had good fit indices. A single factor emerged in EFA results for the teachers' views of children's self-regulation skills scale (8 items- Cronbach's $\alpha = .92$). CFA results demonstrated that the views scale's 8-item and single factor structure exhibited good fit indices. Based on the results, these two scales can be utilized validly and reliably to determine teachers' cognizance of self-regulation and opinions about children's self-regulation. Using these two scales in a pretest-posttest design, study 2 examined a professional development program's effect on preschool teachers' ($n = 21$) awareness and views about children's self-regulation. The Wilcoxon Signed Rank Test was utilized to analyze the data. The results showed a statistically significant difference between the participant teachers' pretest and posttest awareness and views scores on children's self-regulation. Implications for future research and teacher professional development programs were discussed.

## 1. INTRODUCTION

The promising research findings on the impact of children's strong self-regulation skills have encouraged educators, policymakers, and researchers to focus on this topic. The first years of life are essential for children's development of self-regulation (SR), which is defined in numerous ways in the literature (Berger *et al*., 2007; Kopp, 1982; Murray *et al*., 2016). According to Berger *et al*. (2007). SR is about monitoring and adjusting cognition, emotion, and behavior to reach a goal and comply with the cognitive and social requirements in a context. Another definition focuses on regulating thoughts and emotions to facilitate goal-oriented actions, including behavior organization, impulse control, and constructive problem-solving

*CONTACT: Asiye İVRENDİ aivrendi@pau.edu.tr Pamukkale University, Faculty of Education, Department of Primary Education, Preschool Education, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

(Murray *et al*., 2016). For McClelland and Cameron (2012), SR is "the capability of controlling or directing one's attention, thoughts, emotions, and actions" (p. 136). Evident in these definitions, self-regulation consists of cognitive, emotional, and behavioral domains (Jahromi & Stifter, 2008; Murray *et al*., 2016), reflecting a threefold model. Still, self-regulation is also approached as emotionality and regulation and as a one-dimensional construct (Raffaelli *et al*., 2005). The threefold model investigates these structures independently, yet these three processes interact (Raffaelli *et al*., 2005). As Murray *et al*. (2016) stated, the interaction between cognition and emotion procures a basis for behavioral regulation.

Cognitive self-regulation encompasses a broader range of executive functions, extending beyond their traditional conceptualization (Roebers, 2017). Essential executive functions are working memory (keeping and manipulating information in mind), inhibitory control (intentional control of attention, behavior, thoughts, and/or emotions to prevent predominant responses) (Diamond, 2012), and attentional flexibility (altering perspective/thinking, adjusting to change) (Blair & Diamond, 2008; Diamond, 2012). Cognitive SR enables children to utilize and enhance cognitive processes essential for academic learning and problem-solving (Bodrova & Leong, 2006). Emotional regulation refers to managing, controlling, and modulating feelings (Murray *et al*., 2016), while behavioral SR is about the conscious application of essential executive functions to behaviors in social contexts (McClelland *et al*., 2007). In a classroom context, executive functions help children learn how to consciously and effectively manage their behavior: play with other children, follow directions, calm down when upset, remember directions, and persist when challenged (McClelland & Tominey, 2014). On the contrary, low SR skills are associated with difficulty paying attention, calming down when needed, acting appropriately, and coordinating information from different sources during learning (Bautista *et al*., 2024).

Studies have been increasingly investigating the relationship between early SR and socio-demographic factors (Lenes *et al*., 2020), early math and reading skills (Ivrendi, 2011; Welsh *et al*., 2010), its longitudinal influence on academic skills (Ribner *et al*., 2017), and the impact of intervention programs (McClelland *et al*., 2019). Many studies have provided evidence about self-regulation's relation to academic skills, such as early mathematics, language, and literacy skills (Howse *et al*., 2003; Korucu *et al*., 2022; Sezgin & Ulus, 2020; Uyanık *et al*., 2021; Welsh *et al*., 2010). In their longitudinal study, Ribner *et al*. (2017) found that children's executive function levels at the age of five predict academic skills in the fifth grade. Children with strong executive function skills but low early math skills can catch up to those with strong academic capacities. Furthermore, research on the intervention programs' effect in supporting SR skills in early childhood environments is promising (Dias & Seabra, 2015; Schmitt *et al*., 2015).

Other studies focus more on enhancing teachers' SR skills while paying less attention to teachers' gains concerning the conceptual foundations of children's SR (Perels *et al*., 2009). Evidence supports this claim, demonstrating a lack of focus on teachers' knowledge and effective practices to enhance children's SR skills in the classroom (Bautista *et al*., 2024; Hamre *et al*., 2017; Montoya *et al*., 2023; Willis *et al*., 2014). In an interview study with 71 kindergarten teachers, Bautista *et al*. (2024) found that teachers had a limited understanding of SR and lacked knowledge of strategies to strengthen these skills. Given the importance of SR in children's learning as well as such skills' impact on later education levels, it would seem essential to support these skills from the first years of life (Howse *et al*., 2003; Ribner *et al*., 2017), and teachers are the key facilitators in this process (Ackerman & Friedman-Krauss, 2017; Silkenbeumer *et al*., 2018).

### 1.1. Place of Self-Regulation in Early Childhood Programs/Curriculum Frameworks

Researchers who view self-regulation as a developmental achievement question how much early childhood curriculum frameworks include SR (Størksen *et al*., 2015; Vasseleu *et al*., 2022), how to foster these skills in early childhood, and how to implement intervention

programs accordingly (Schmitt *et al*., 2015). Early childhood programs/frameworks appear to not fully reflect the critical role of SR in children's well-being (Lenes *et al*., 2020; Vasseleu *et al*., 2022). For instance, Lenes *et al*. (2020) reported that the Norwegian Framework Plan for the Content and Tasks of Kindergartens does not mention children's SR. Similarly, the Turkish Ministry of National Education-Preschool Education Program (MoNE-PEP, 2013) does not explicitly emphasize SR. However, MoNE-PEP (2013) is a child-centered and play-based program, and some of its principles relate to this construct, such as fostering autonomous behaviors and self-control which is a feature of inhibitory control (Diamond, 2014). As Lenes *et al*. (2020) state, child-centered and play-based programs may foster children's SR skills. Meanwhile, the literature emphasizes the importance of intentionally planned support for children's self-regulation skills (Vasseleu *et al*., 2022). For this to happen, teachers need to be cognizant of the SR construct. Yet, the topic is relatively new for some preschool teachers, who need to become more familiar with its theoretical foundations and means of systematically supporting such skills in the classroom (Vasseleu *et al*., 2022; Willis *et al*., 2014).

### 1.2. Teacher Professional Development on Children's Self-Regulation

If provided with training and support, preschool teachers can enhance children's executive function skills (Diamond & Lee, 2011). The necessity of professional development (PD) on children's SR skills is tied to the teacher's central role in improving such skills in early childhood (Silkenbeumer *et al*., 2018; Venitz & Perels, 2018). The PD activities have the potential to assist teachers in acquiring the necessary knowledge, skills, and dispositions to promote children's development and learning (Atmaca & Ertürk Kara, 2023; Snyder *et al*., 2012). This is because providing teachers with a theoretical and practical component is crucial in building a strong foundation for children's SR skills (Willis *et al*., 2014). Iriogbe-Efionayi (2020) conducted a study with 97 pre-service and in-service early childhood teachers, finding a positive relationship between the teachers' level of SR knowledge, years of teaching experience, and the amount of professional development (PD). Another study involved conducting SR teacher training with 35 kindergarten teachers and 97 children. The training included trainers acting as a model for teachers, relating content on SR with kindergarten activities, opportunities for being proactive in learning, such as sharing experiences, fostering the transfer of training to the classroom through assignments, and keeping diaries about self-regulation. The results showed positive effects of teacher training on both teachers' knowledge about SR and children's self-regulated learning (Perels *et al*., 2009).

The PD content on SR seems essential to adopting a broader perspective, including teachers' beliefs, classroom interactions, guidance, and structure (Diamond & Lee, 2011; Rosanbalm & Murray, 2017; Silkenbeumer *et al*., 2018). These are considered key aspects of the teacher's knowledge and instruction of SR skills in the classroom. Literature calls this perspective an indirect intervention to improve such skills (Venitz & Perels, 2018). Direct interventions specifically train children on executive function skills (Mattera *et al*., 2021), such as storytelling and games in which they will have to inhibit their behavior and strengthen memory through repetitive songs (Mann *et al*., 2016). Supporting indirect intervention, Diamond and Lee (2011) emphasize that attending to socio-emotional and physical development rather than solely focusing on dimensions of executive function may increase the effect of SR training on teachers. This type of intervention aims to establish a learning environment that promotes the development of executive function skills (Mattera *et al*., 2021). Evidence reinforcing the effectiveness of direct intervention (Schmitt *et al*., 2015) and indirect intervention was found in research showing that close teacher-child interactions and high classroom quality were related to children's high SR skills (Cadima *et al*., 2016). In a study, Dias and Seabra (2015) provided preschool teachers with six hours of training to implement the Intervention Program for Self-Regulation. The training's content consisted of knowledge, discussions on research results, teacher roles, modeling, and intervention program practice. Children in the trained teachers' classrooms showed better attention and inhibition abilities performance than their

counterparts. The authors concluded with a remark on the program's positive effect in supporting children's executive function skills.

Existing research highlights a complex challenge in supporting preschool children's self-regulation, revealing nuanced and sometimes conflicting perspectives on teachers' roles. The first issue is that preschool teachers lack or have limited knowledge about child SR, as demonstrated by research conducted in different cultures (Bautista *et al*., 2024; Vasseleu *et al*., 2022; Willis *et al*., 2014). For instance, a study with 115 early childhood teachers found that teachers lack knowledge and instruction about SR skills. Over half were unaware of this construct, and over two-thirds did not engage in PD activities to support these skills (Willis *et al*., 2014). A recent study showed that Australian teachers mainly focus on examples of dysregulation when discussing SR (Vasseleu *et al*., 2022). Therefore, it appears that teachers need PD activities related to children's SR.

Another issue is that the effectiveness of a few PDs is evaluated in the literature (Popova *et al*., 2022). When PD is effective, a change could occur in the form of enhancements in teachers' knowledge, instructional practice, and student learning outcomes (Wei *et al*., 2009). Determining how much PD activities attain their goals (Markussen-Brown *et al*., 2017) is essential and approached differently (Guskey, 2003; Schachter, 2015). Schachter (2015) analyzed 73 PD studies regarding the design, delivery, and measurement of the PD's effectiveness. The results showed that half of them investigated how teacher practice changed, 18% evaluated how teachers' knowledge changed, 40% measured how children's learning changed, and 11% evaluated how children's behavior changed. That is, less attention is paid to the change in teacher knowledge, which is also evident in the current literature indicating that there is little evidence on what teachers know about SR and how well they support these skills (Bautista *et al*., 2024; Montoya *et al*., 2023). One reason is linked to the scarcity of measurement tools for assessing what teachers know and/or have learned about SR (Vasseleu *et al*., 2024).

The effectiveness of PD programs can be determined by considering teachers' self-reports of PD features that enhance their knowledge and transform instructional practices (Guskey, 2003). Although limited in number, some instruments specifically measure preschool teachers' knowledge and practices of SR (Adagideli *et al*., 2015; Perels *et al*., 2009; Vasseleu *et al*., 2021; Willis *et al*., 2014). For example, Adagideli *et al*. (2015) developed an instrument to measure preschool teachers' practices to foster self-regulated learning in young children. The scale consisted of five subscales: "emotional and motivational regulation," "metacognitive regulation during the task," "metacognitive knowledge of the task and strategy," "metacognitive regulation after the task," and "metacognitive knowledge of the person" (p. 431). Perels *et al*. (2009) constructed a scale to measure teachers' knowledge about SR before and after an intervention program to determine the effectiveness of the PD program. This scale included items about the kindergarten teachers' use of support, encouragement, verbal rewards, and modeling for the preschoolers. Another study developed an instrument to measure early childhood teachers' knowledge and instruction of SR. This measure had three factors: teacher attitudes and beliefs, classroom management techniques, and child behavior (Willis *et al*., 2014). Moreover, a scale developed by Vasseleu *et al*. (2021) measures early childhood educators' beliefs about fostering children's self-regulation. It had the subscales of perceived knowledge, attitudes, and self-efficacy, which are considered cognitive beliefs.

The current research encompasses two studies that address the issue of less emphasis on assessing teachers' knowledge (Study 1) and the lack of SR knowledge among teachers (Study 2). In Study 1, two brief teacher self-report scales were developed by examining the literature on the conceptual basis of self-regulation (e.g., Berger *et al*., 2007; Kopp, 1982; Murray *et al*., 2016). The first scale assesses preschool teachers' awareness of children's SR. The second scale measures preschool teachers' views of children's SR skills. Previous studies provide proper

instruments for assessing SR from different perspectives, such as classroom management skills and modeling for preschoolers (Perels *et al*., 2009; Willis *et al*., 2014), practices for fostering self-regulated learning (Adagideli *et al*., 2015), and cognitive beliefs (Vasseleu *et al*., 2021). The scales developed in this present study differ from the previous instruments (Adagideli *et al*., 2015; Perels *et al*., 2009; Vasseleu *et al*., 2021; Willis *et al*., 2014) in terms of focus. The awareness scale explicitly targets whether preschool teachers know the conceptual basis of SR's cognitive dimension and self-regulation's relation to learning. The views scale focuses on the indicators of children's self-regulatory behaviors observed in a typical classroom environment. Using these two measurement tools, Study 2 investigated the effect of a PD program on preschool teachers' awareness and views about children's SR skills. The research questions were:

1. What are the psychometric properties of preschool teachers' awareness and views scales regarding children's self-regulation (Study 1)?

2. How does the PD program affect the preschool teachers' awareness and views about children's self-regulation skills (Study 2)?

## 2. STUDY-1

### 2.1. Method

Study 1 included developing two scales using the survey method, which is approached as a proper research model in studies aiming to identify certain group features (Christensen *et al*., 2015). The characteristics determined through the developed scales in this study were preschool teachers' awareness and views about children's self-regulation.

#### 2.1.1. *Study group*

A convenience sampling method was used to determine the participants of Study 1. The scales were filled by preschool teachers (exploratory factor analysis (EFA)) $n$ = 201 and confirmatory factor analysis (CFA), $n$ = 123, for a total of 324 participants). Based on EFA data, the average teaching experience was 14.43 years (*SD*=5.71). Considering the educational status of the teachers in the EFA study group, 180 (89.6%) had an undergraduate level of education, and 18 (9%) had a graduate level of education. Some teachers did not report their educational level.

The average teaching experience of the CFA study group was 9.50 years (*SD*=6.06). The educational backgrounds of CFA study group teachers were as follows: Three (2.4%) had associate degrees, 103 (83.7%) had bachelor's degrees, and 17 (13.8%) had master's degrees.

#### 2.1.2. *Scale development*

The development process of the two teacher self-report scales, one measuring awareness about children's SR and the other assessing views on children's SR skills, involved a literature review and pursuing field experts along with preschool teachers' opinions. First, a literature review focusing on the conceptual explanations of SR (behavioral SR, cognitive SR, including dimensions of executive functions; working memory, attention flexibility, and inhibitory control) and related research was conducted (e.g., Berger *et al*., 2007; Bronson, 2000; Diamond, 2014; Diamond, 2012; Kopp, 1982; McClelland *et al*., 2019; McClelland & Tominey, 2015; McClelland & Cameron, 2012; Murray *et al*., 2016; Raffaelli *et al*., 2005; Savina, 2021; Tominey *et al*., 2018). The awareness scale was structured to measure the extent to which preschool teachers know the conceptual foundations of SR and its relation to learning. Therefore, only the theoretical explanations of SR and information on its relationship with learning were considered in determining the items of the awareness scale. This procedure resulted in the creation of an item pool and the determination of a draft of 24 items for the awareness scale. The views scale aimed to measure the extent to which teachers identify indicators related to children's SR skills in the context of preschools (e.g., raising a hand before talking, following rules). Hence, the scale included only items related to children's SR skills

that are more likely to occur in the preschool environment. This process resulted in the creation of a draft of the views scale, which included 12 items.

Four field experts and two preschool teachers examined the scope and face validity of both scales' draft items. The field experts primarily offered revisions to improve the clarity of the items, such as "Self-regulation is effective in learning academic skills (e.g., mathematics and language)." After adjusting both scales' items in line with experts' feedback, teachers were asked to fill in the scales by examining the clarity and comprehensibility of the items. Teachers generally suggested word choice (e.g., regulating instead of adjusting) and ambiguous statements. The scales' items were revised according to the teachers' feedback and were re-examined by a field expert and the researcher.

This procedure resulted in the construction of the final versions of the two scales: the awareness scale with 24 items and the views scale with 12 items. Both scales were teacher self-report measures that asked preschool teachers to rate their level of agreement about the items on a five-point Likert-type scale (1-Strongly Disagree-5-Totally Agree). The scales were delivered to preschool teachers via an online survey platform and face-to-face. Concerning the data collection process[†], first, data for EFA was collected to investigate the structure of the two scales. Then, data for CFA was collected to confirm the structure of the awareness scale obtained from EFA.

### 2.1.3. *Data analysis*

Data analysis consisted of EFA and CFA to explore the scales' construct validity, the McDonald's Omega (ω) and Cronbach's Alpha (α) coefficients to determine internal consistency reliability, and the Pearson correlation coefficient to examine the relationship between the awareness and views scales. The data was examined as part of the pre-analysis to determine missing values and outliers. There were no missing or extreme values in the EFA data. The data for CFA contained missing values. Six cells were left blank in the data set. These missing values were addressed with the series mean method. The views scale data contained 13 outliers identified by examining the box plot and *z* scores +-3 (Çokluk *et al*., 2012), which were removed from the data set.

By examining initial item-total correlations, items with a correlation under .30 were excluded from the scales to strengthen the structure (Büyüköztürk, 2009; Çokluk *et al*., 2012). EFA was conducted for both scales on data from 201 preschool teachers. The varimax method was preferred to obtain a more interpretable structure since it minimizes the complexity of the factors by maximizing the variance of the loadings on each factor (Tabachnick & Fidell, 2015). The Kaiser-Meyer-Olkin (KMO) and Bartlett Test of Sphericity were inquired to decide whether the data was appropriate for factor analysis. Factors with an eigenvalue greater than 1 and the scree plots were utilized to determine the number of factors in the scales (Henson & Roberts, 2006). Decisions regarding inclusion and exclusion of items were made based on item loadings, cross-loadings, and a minimum of three salient loadings in a factor (Carpenter, 2018). Hence, the distribution of each item to the factors and the factor loadings were examined. Items with factor loadings between .30 and .59 are considered moderate for the correlation between the item and the factor (Çokluk *et al*., 2012). Based on these value ranges, items with .45 or higher factor loadings were retained in the scales, and items loading more than one factor were excluded. The EFA data's normality was examined using kurtosis and skewness values. The awareness scale's total and subscales' skewness values changed from -0.17 to .38, and the kurtosis values were between 0.52 and 0.87. The views scale's skewness value was -0.48, and

[†] EFA data of Study 1 and Study 2 were funded by the Scientific and Technological Research Council of Türkiye. Only CFA data were collected after the project was completed.

the kurtosis value was 1.45. The normality assumption was met since these values were between -1.5 and +1.5 (Tabachnick & Fidell, 2015).

CFA was conducted on data obtained from a different sample of 123 preschool teachers to verify the structure determined by EFA. For CFA data, the multivariate normal distribution of variables was tested with Mardia's skewness and kurtosis coefficients. Results showed that the awareness data ($\chi^2$ =753; $p$ <.001) and views scale data ($\chi^2$ =788; $p$ <.001) did not meet the multivariate normality assumption. Therefore, Robust Maximum Likelihood was utilized as the estimation method. In CFA, the following acceptable fit values were evaluated by considering the value ranges specified by Çokluk *et al*., (2012): Chi-square ratio = $\chi^2/df \leq 2.5$; RMSEA = ≤ .08; SRMR = ≤ .08; CFI = ≥ .90; TLI ≥ .90).

The Jamovi 2.3.28 program was used for CFA only, and SPSS version 23 was used for the other analyses.

### 2.2. Results

The findings obtained from the validity and reliability analyses of the awareness and views scales are presented below.

#### 2.2.1. *Validity and reliability results for the Teachers' Awareness of Self-regulation in Children (TASRIC) Scale*

The construct validity of the TASRiC scale was examined with EFA. Initial corrected item-total correlations showed that eight items having low correlations with other items (under .30) were excluded from the scale (Büyüköztürk, 2009; Çokluk *et al*., 2012). EFA was conducted with the remaining items. Based on the examination of item factor loadings and cross-loadings, six items that had under .45 factor loadings and entered more than one factor in the exploratory factor analysis were removed from the scale for optimizing scale structure (Büyüköztürk, 2009). The corrected item-total correlations of TASRiC's with the remaining 10 items were presented in Table 1.

**Table 1.** *Item-total statistics of TASRiC.*

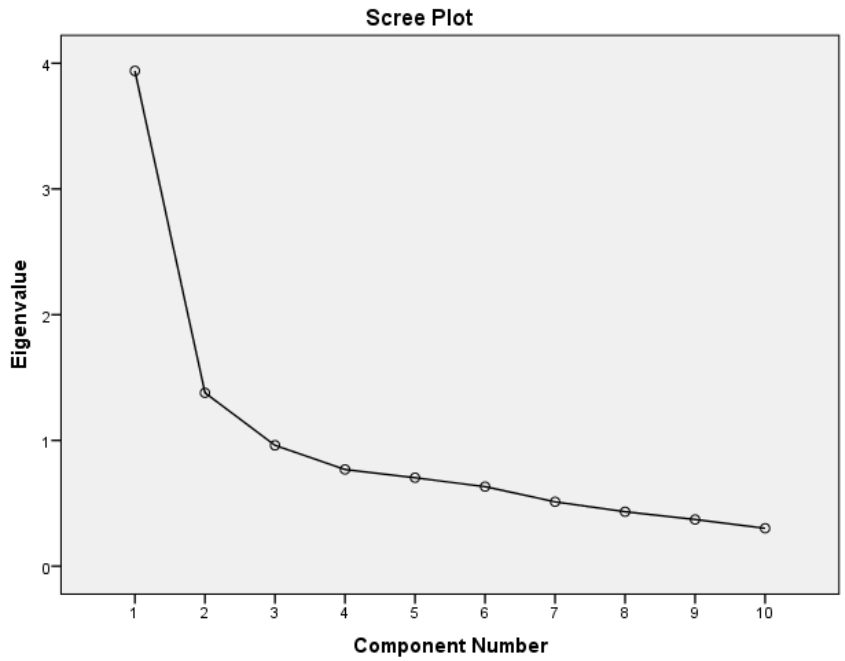
| Item | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|---|---|
| 1 | 34.6169 | 15.158 | .553 | .800 |
| 2 | 34.8060 | 14.447 | .507 | .805 |
| 3 | 34.9055 | 14.346 | .544 | .800 |
| 4 | 34.4527 | 15.209 | .581 | .798 |
| 5 | 34.5323 | 15.020 | .575 | .797 |
| 6 | 34.7711 | 14.877 | .436 | .813 |
| 7 | 34.3532 | 15.370 | .505 | .804 |
| 8 | 34.4428 | 15.848 | .405 | .814 |
| 9 | 34.4129 | 15.454 | .555 | .801 |
| 10 | 34.6020 | 15.351 | .439 | .811 |

Table 1 shows that the lowest item-total correlation value is .41, and the highest value is .58. The values higher than .30 demonstrate that each item sufficiently correlates with the other items in the scale measuring preschool teachers' awareness of cognitive self-regulation.

The Kaiser-Mayer-Olkin (KMO=.817) and Bartlett Test of Sphericity ($\chi^2$ = 611.050; $p$ < .05) values were found to be suitable for factor analysis of the data. Factors with an eigenvalue greater than 1 and the scree plot were used to determine the number of factors (Büyüköztürk, 2009). The scree plot is presented in Figure 1. The plot shows that the scale does not exhibit precise lines after the second factor, indicating the presence of two factors in the scale. The first

factor, named self-regulation (eigenvalue = 3.94; variance (%) = 31.62), loaded seven items that related to cognitive self-regulation's aspect of executive functions: working memory, attention flexibility, and inhibitory control. "Relation to Learning", the second factor, loaded three items reflecting the associations between self-regulation and learning (eigenvalue = 1.38; variance (%) = 21.56). The cumulative explanation rate was 53.18%.

**Figure 1.** *The scree plot for TASRiC.*




The item loadings are presented in Table 2. According to Table 2, the factor loadings for the self-regulation subscale were between .58 and .74, and for the relation to learning subscale, they were between .67 and .90. The mean (*M*), standard deviation (*SD*), and reliability coefficients of the TASRiC's subscales were as follows: *M* = 3.79; *SD* =0.479; ω =.81, α=.80 for self-regulation, *M* =3.94; *SD*=0.519; ω=.76, α=.74 for relation to learning, and *M* =3.87; *SD*=0.423; and ω = .83, α= .82 for the total scale.

**Table 2.** *The TASRiC's factor loadings.*

| | | Factor Loadings | |
|---|---|---|---|
| | Items | 1 | 2 |
| Self-regulation | 1. | .737 | .105 |
| | 2. | .704 | .102 |
| | 3. | .687 | .066 |
| | 4. | .659 | .282 |
| | 5. | .644 | .306 |
| | 6. | .593 | .247 |
| | 7. | .575 | .102 |
| Relation to learning | 8. | .038 | .896 |
| | 9. | .271 | .796 |
| | 10. | .216 | .669 |

CFA was performed to confirm the two-factor structure of the TASRiC scale obtained from EFA. Acceptable fit values were not reached in the first CFA results for the scale (Chi-square $\chi^2/df$ = 95.3/34 = 2.802; RMSEA = .107; SRMR = .066; CFI = .897; TLI = .864). In other words, RMSEA, CFI, and TLI showed poor model-data fit since the RMSEA value was greater

than .08, and CFI and TLI values were lower than .90 (Çokluk *et al*., 2012). Therefore, three modifications (1 to 4, 4 to 7, and 6 to 7) were made by examining the modification index. All these items were within the subscale of self-regulation. Moreover, modifications between these items theoretically make sense since they all reflect the executive functions aspect of cognitive self-regulation, including attention flexibility, working memory, and inhibitory control (Roebers, 2017). For example, item 1 (using/assessing sudden opportunities) is related to attentional flexibility, and item 7 (behaving by the rules in the absence of a teacher or adult) is associated with working memory and inhibitory control. After these modifications, CFA results were as follows: Chi-square $\chi^2/df$ =63.5/31 =2.048; RMSEA=.074; SRMR =.058; CFI = .934; TLI =.904. These indices showed acceptable model-data fit. As a result, the scale's 10-item two-factor structure showed an acceptable fit, which verifies the structure obtained in EFA.

### 2.2.2. *Validity and reliability results for the Teachers' Views on Children's Self-Regulation Skills (TVoCSRS) Scale*

Exploratory factor analysis was used to investigate the views scale's construct validity. The initial corrected item-total correlation values showed that two items had low correlations with other items (under .30), so they were removed from the scale. The corrected item-total correlations of the remaining 10 items were given in Table 3.

**Table 3.** *Item-total statistics of TVoCSRS.*

| Item | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|---|---|
| 1 | 34.7910 | 24.016 | .824 | .895 |
| 2 | 34.8408 | 23.605 | .785 | .896 |
| 3 | 34.9005 | 24.330 | .729 | .900 |
| 4 | 34.8955 | 23.754 | .725 | .899 |
| 5 | 34.9353 | 23.911 | .733 | .899 |
| 6 | 34.9303 | 23.925 | .708 | .900 |
| 7 | 34.9154 | 24.458 | .634 | .905 |
| 8 | 35.1841 | 23.181 | .629 | .907 |
| 9 | 34.9154 | 24.288 | .630 | .905 |
| 10 | 35.0796 | 24.764 | .477 | .916 |

Table 3 shows that the lowest item-total correlation value is .48, and the highest value is .82. The corrected item-total correlations are higher than .30, which indicates that each item adequately correlates with the other items in the scale measuring preschool teachers' views of children's SR skills. EFA was conducted using the remaining 10 items. The Kaiser-Mayer-Olkin (KMO = .906 - excellent) and Bartlett Test of Sphericity ($\chi^2$ = 1282.731; $p$ < .05) values show that the data is suitable for factor analysis (Büyüköztürk, 2009; Çokluk *et al*., 2012). EFA revealed a single-factor structure with 10 items (eigenvalue = 5.780; variance (%) = 57.80). The explanation rate was 58%. The scree plot is presented in Figure 2. It demonstrates that very precise lines did not separate the scale after the first factor, which reveals a single-factor structure. The TVoCSRS's factor loadings were presented in Table 4. It demonstrates that the item loadings of the TVoCSRS ranged from .55 to .88. The mean ($M$ = 3.88) and standard deviation values ($SD$ = 0.541) were computed. The internal consistency coefficient (Cronbach's α) was .91.

CFA was performed to verify TVoCSRS's structure. The first CFA results for the scale didn't indicate model-data fit (Chi-square $\chi^2/df$ = 236/35 = 6.74; RMSEA =.216; SRMR =.092; CFI =.773; TLI =.709). The RMSEA and SRMR values were greater than .08, and CFI and TLI values were lower than .90 (Çokluk *et al*., 2012). Parameter estimates revealed that standardized

regression coefficients were between .169 and .616, standardized beta coefficients ranged from .321 to .935, and explained variance-$R^2$ were between .103 to .875. Specifically, these values were low for items 8 (Standardized regression coefficient =.169, standardized beta coefficient = .321, $R^2$ =.103) and 10 (Standardized regression coefficient =.320, standardized beta coefficient = .466, explained variance-$R^2$ =.217), and thus they were removed from the model.

**Figure 2.** *The scree plot for TVoCSRS.*

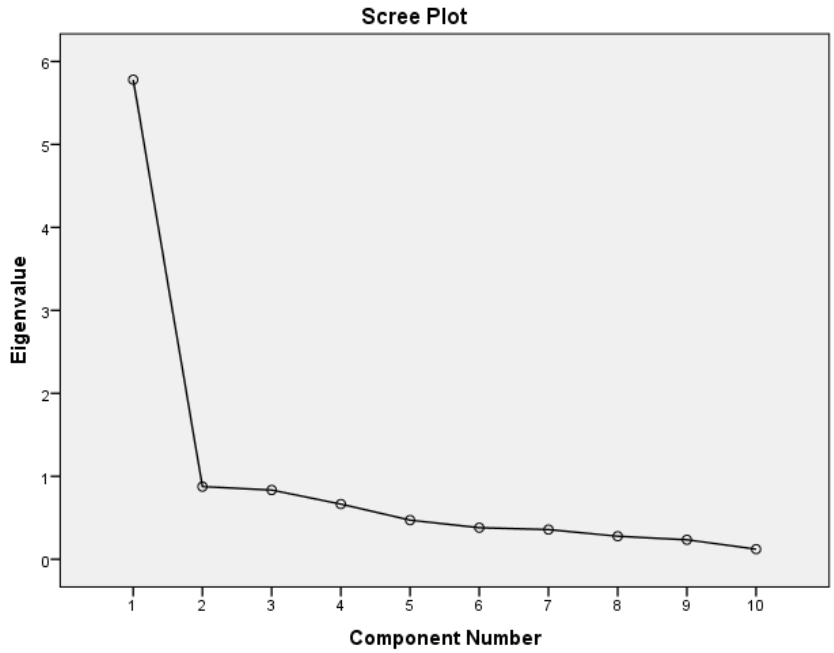



**Table 4.** *The TVoCSRS's factor loadings.*

| Item no | Factor Loadings |
|---|---|
| 1 | .884 |
| 2 | .856 |
| 3 | .802 |
| 4 | .801 |
| 5 | .797 |
| 6 | .776 |
| 7 | .704 |
| 8 | .690 |
| 9 | .688 |
| 10 | .546 |

After removing items 8 and 10 from the model and conducting modifications between items (6 and 7, 2 and 4), the goodness of fit values improved (Chi-square $\chi^2/df$ = 29.6/18 = 1.64 (Excellent); RMSEA=.028 (Excellent); SRMR=.031 (Good); CFI=.998 (Good); TLI=.997 (Good)). Theoretically, items 6 and 7 measure working memory, and items 2 and 4 are about working memory and inhibitory control. Therefore, it was thought that modifications could be made to these items since they measure similar skills related to self-regulation. With modifications, the remaining 8 items' standardized regression coefficients were between 0.321 and 0.655, standardized beta coefficients ranged from .579 to .964, and the explained variance was between .336 and .929. As a result of this procedure, the model data fit was obtained. After removing items 8 and 10 based on parameter estimates, EFA was repeated to examine the

construct validity of the remaining 8 items. The corrected item-total correlations of the items and factor loadings were presented in Table 5.

**Table 5.** *Item-total statistics and factor loadings of TVoCSRS.*

| Item | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted | Factor Loadings |
|---|---|---|---|---|---|
| 1 | 27.4129 | 14.704 | .855 | .894 | .904 |
| 2 | 27.4627 | 14.360 | .815 | .896 | .877 |
| 3 | 27.5224 | 14.991 | .746 | .902 | .816 |
| 4 | 27.5174 | 14.561 | .734 | .903 | .812 |
| 5 | 27.5572 | 14.798 | .719 | .904 | .793 |
| 6 | 27.5522 | 14.679 | .720 | .904 | .787 |
| 7 | 27.5373 | 15.170 | .629 | .912 | .706 |
| 8 | 27.5373 | 15.260 | .581 | .916 | .663 |

According to Table 5, the lowest item-total correlation value is .58, and the highest value is .86. The KMO (.894 - excellent) and Bartlett Test of Sphericity ($\chi^2$ = 1107,306; $p$ <.05) values indicate that the data is appropriate for factor analysis (Büyüköztürk, 2009; Çokluk *et al*., 2012). The results of EFA showed a single-factor structure (eigenvalue = 5.098; variance (%) = 63.72), and item loadings were between .663 and .904 with an explanation rate of 63.72%. The mean ($M$ = 3.93) and standard deviation ($SD$ = 0.546) were computed. The internal consistency coefficient (Cronbach's α) was .92.

Additionally, the relationship between TASRiC and TVoCSRS scales was calculated using the Pearson correlation coefficient. The findings are presented in Table 6. As seen in Table 6, there is a high level of positive correlation between the total TASRiC score and the self-regulation subscale ($r$ = .84; $p$ < .01) and in relation to learning ($r$ = .86; $p$ < .01). TVoCSRS positively correlates at a moderate level with TASRiC's self-regulation subscale and TASRiC's total score ($r$ = .649; $p$ < .01, $r$ = .545; $p$ < .01, respectively) and positively at a low level with TASRiC's relation to learning subscale ($r$ = .29; $p$ <.01).

**Table 6.** *Correlations between TASRiC and TVoCSRS (N=201).*

| | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1. TASRiC- Self-regulation | - | | | |
| 2. TASRiC – Relation to learning | .439** | - | | |
| 3. TASRiC - Total | .835** | .861** | - | |
| 4. TVoCSRS - | .649** | .289** | .545** | - |

**$p$<.01

## 2.3. Discussion and Conclusion

Study 1 sought to construct two scales, one to assess preschool teachers' awareness of self-regulation and the other to measure their views on children's self-regulation skills. As discussed previously, preschool teachers' knowledge and skill levels on this issue have not been investigated sufficiently (Bautista *et al*., 2024; Montoya *et al*., 2023). Relatedly, the limited number of measurement tools on the subject also creates a methodological constraint; that is, measuring the construct quantitatively becomes difficult (Vasseleu *et al*., 2024).

EFA, conducted to examine the construct validity of TASRiC, revealed a two-factor structure: self-regulation and relation to learning. The scale's total explanation rate was 53.18%, which is sufficient based on the recommendation that the explained variance should be between 40% and 60% (Tabachnick & Fidell, 2015). Testing the TASRiC's internal consistency demonstrated that the McDonald's Omega coefficient of the self-regulation subscale and total scale is good.

The internal consistency of the relation to learning subscale is acceptable (Nunnally, 1978, cited in Streiner, 2003). It can be concluded that the scale's items are consistent and measure the construct of teachers' awareness of children's self-regulation. Moreover, CFA showed that the fit indices of the two-factor structure related to teachers' awareness of self-regulation were sufficient. The second scale developed in this study is TVoCSRS. The results of EFA indicated that the scale has a single-factor structure with an explanation rate of 63.72%. The scale had high internal consistency measured with Cronbach's alpha, demonstrating that its items are closely related as a single factor measuring preschool teachers' views on children's self-regulation skills. CFA results demonstrated that the fit indices obtained were adequate.

Further, the Pearson correlation coefficient was calculated to determine the relationship between TASRiC and TVoCSRS. The results showed a high level of positive correlation between the total TASRiC score and the self-regulation subscale and in relation to learning subscale. TVoCSRS had a moderately positive relationship with TASRiC's self-regulation subscale and TASRiC's total score. It also had a lowly positive relationship with TASRiC's relation to learning subscale. In conclusion, the correlations between the two scales were positive and ranged between low and moderate.

A moderate positive correlation between TASRiC and TVoCSRS implies that as the teachers' awareness about children's SR increases, their views on identifying children's self-regulatory skills in the classroom environment increase. It is intriguing that there is a positive but low correlation between TVoCSRS and TASRiC's relation to learning subscale. This suggests that teachers' understanding of the relationship between SR skills and learning, such as their effectiveness in academic learning and their learnability, weakly associates with their opinions about the indicators of these skills. Although adults may identify children's difficulties in following rules and ignoring distractions (Ackerman & Friedman-Krauss, 2017), they may not relate these difficulties to the SR construct or to their influence on learning, as seen in the correlations between TASRiC and TVoCSRS. As Blair and Razza (2007) proved, inhibitory control was an eminent correlate of early math and reading ability.

Although the findings of this study provide evidence for the construct validity and reliability of awareness and views scales, they should be interpreted with caution. One of the limitations of this study is related to the modified model obtained after conducting modifications in CFA. Despite the modified model's good fit with the CFA data, Çokluk *et al*. (2012) recommend testing it on a new sample to ensure its generalizability. Further research with a different study group is necessary in this case to evaluate the modified model fit, potentially providing further evidence about the construct validity of the awareness and views scales. Another limitation of this study relates to the study group's features. Both the EFA study group, with an average teaching experience of 14.43 years (*SD* = 5.71), and the CFA study group, with an average teaching experience of 9.50 years (*SD* = 6.06), seem to comprise experienced teachers. Subsequent studies have the potential to re-examine the structure of these measurement tools, incorporating study groups with demographic characteristics such as those new to their profession and those working in rural areas.

Along with these limitations, this study contributes to the literature by providing two scales to measure preschool teachers' awareness of SR and their views on SR skills in a valid and reliable way. English and Turkish versions of the scales are presented in the Appendix. As emphasized earlier, there is less focus on determining teacher knowledge, partly because of a lack of measurement tools (Vasseleu *et al*., 2024). Teachers' in-service training needs regarding this issue can be determined using these scales. Similarly, these scales can be useful in cultures where practicing teachers are unfamiliar with the concept of children's SR and in preschool education programs that do not explicitly address this construct. Furthermore, gathering data from teachers via scales is becoming increasingly challenging due to their intense workload. The TASRiC and TVoCSRS are short scales. They are not time-consuming for teachers to fill

out. Future research may benefit from using these brief scales to identify teachers' awareness and views of SR and how they correlate with children's SR skills.

## 3. STUDY-2

### 3.1. Method

Study 2 examined the effect of the PD program on the preschool teachers' awareness of self-regulation and their views about children's self-regulation skills, using a pretest and posttest design in an experimental group without a control group.

#### 3.1.1. *Study group*

The participants were determined by using a criterion sampling technique. These criteria included teaching children in the 4-5 age group, volunteering, and not having previously participated in any training related to children's SR. Public preschool teachers working in a mid-sized city in the southwestern part of Türkiye applied to this PD program by completing an online application form. Then, 23 teachers were determined among the applicants by drawing lots online. Due to missing data, the analysis did not include data from two participants. Participants' average teaching experience was 14.10 years (*SD* = 3.12). Most were (81.0%) college graduates, and only four (19.0%) had a graduate degree. The average number of children in their classrooms was 18.71 (*SD* = 4.38). Nine of them (42.9%) worked with five-year-olds, nine participants (42.9%) taught four-year-olds, one (4.8%) worked with 3-4-year-olds, and two of them (9.5%) worked with 5-6-year-old children.

Two questions in the demographic information form determined the participants' ratings of their level of knowledge and practical skills about SR (see below). According to the participants' self-report, eight (38.1%) had either no knowledge or a low level, 13 (61.9%) had a moderate level, and none had a good/advanced level of knowledge about children's SR skills. Regarding the practical skills needed to support children's SR skills, approximately half of the participants had either no skills or a low level (47.6%), 10 participants (47.6%) had moderate skills, and 1 participant (4.8%) had advanced skills.

#### 3.1.2. *Data collection tools*

Study 2 employed the demographic information form, the awareness (TASRiC), and views (TVoCSRS) scales created in Study 1 as data collection tools.

*Demographic information form*: An information form was used to ascertain the characteristics of the participants, including teaching experience, the number of children in their classrooms, educational status, and items for self-assessment of their knowledge and practical skills about children's SR. The items for self-assessment asked teachers to rate their knowledge (1 = no knowledge to 5 = advanced level of knowledge) and skill levels (1 = I do not have any to 5 = I have advanced skills) regarding children's self-regulation using a five-point rating system.

*Teachers' Awareness of Self-Regulation in Children (TASRiC)*: TASRiC aims to assess preschool teachers' awareness of children's self-regulation. The instrument has 10 items with two subscales: self-regulation (7 items—e.g., self-regulation is about keeping in mind what is said/described) and relation to learning (3 items—e.g., self-regulation skills make it easier to learn a knowledge/skill). Teachers rate their level of agreement on a five-point Likert-type scale (1-Strongly Disagree-5-Totally Agree). The scale's explanation rate was 53.18%. The reliability coefficients of the subscales were ω =.81, α =.80 for self-regulation, ω =.76, α =.74 for relation to learning, and ω =.83, α =.82 for the total scale.

*Teachers' Views on Children's Self-Regulation Skills (TVoCSRS)*: This scale assesses preschool teachers' views on the indicators of children's self-regulation skills. It has a single-factor structure with 8 items (e.g., raising a hand and waiting a turn at an activity). Preschool teachers rate their level of agreement on a five-point Likert-type scale (1—Strongly Disagree—5—

Totally Agree). The scale has a 63.72% explanation rate with Cronbach's alpha coefficient of .92.

### 3.1.3. *Data collection and analysis*

The participant teachers completed the TASRiC and TVoCSRS scales as pretests at the beginning of the PD program. After teachers participated in the PD program for five days, the same instruments were implemented as posttests. Then, the pretest and posttest data were entered into the SPSS program. In data analysis, descriptive statistics about the pretest and posttest were computed. Due to the low number of participants in Study 2 ($n$ = 21), the PD program's effect on teachers' awareness of children's SR and views on children's SR skills was examined using a nonparametric technique known as the Wilcoxon Signed Rank Test, and calculating the effect size by using the formula given in Equation 1 (Fritz *et al*., 2012). For effect size, 0-0.29 means small, and 0.30-0.49 indicates medium effect. Values 0.50 and above are considered large effects (Cohen, 1988).

$$r = \frac{z}{\sqrt{N}} \quad (1)$$

### 3.1.4. *Procedure: Teacher PD program on children's self-regulation*

The preschool teachers' PD program lasted five days and approximately four to six hours daily. Six early childhood faculty members were responsible for constructing and conducting the PD program sessions (see Ivrendi *et al*., 2022) for details about the theoretical framework used for this PD). The first purpose of the program was to raise preschool teachers' awareness about the concept of self-regulation, its domains, and the different ways of supporting it on a typical preschool day. The second purpose was to foster teachers' understanding and skills to implement the Red Light Purple Light Self-Regulation (RLPL) Intervention program in their classrooms. The RLPL is a game, music, and movement-based program developed by Tominey *et al*. (2018) (See details about the permission to use RLPL and its adaptation to Turkish in Ivrendi *et al*., [in press]). The sessions were constructed by considering the two purposes of the PD and roughly had the following structure: a theoretical introduction, questions, an exchange of experiences, practice examples, and small-large group discussions. A summary of the PD content is given below:

Day 1: The sessions centered on the theoretical explanations of SR, including its definition, development, and impact on various skills; the sharing of teachers' previous experiences with SR; and strategies for promoting such skills in a typical classroom setting.

Day 2: The sessions focused on providing parental guidance for SR, introducing the RLPL, watching sample videos, discussing the RLPL's sessions, and singing songs from the RLPL.

Day 3: The sessions aimed to help the participants experience the RLPL games as if they were children. The researcher acted as a teacher, modeling the implementation of the games in a child-friendly manner. The focus of the aftermath discussions was on the participants' observations and thoughts about the games and their potential implementation in preschool classrooms.

Day 4: The role-play of RLPL games was conducted in which the participants acted as teachers and children while the researcher offered guidance when needed. The aftermath discussions centered on the participants' feelings about implementing the games, as well as their thoughts on potential problems and solutions that could arise when implementing them in the classroom.

Day 5: The sessions were about what teachers need to be careful about when implementing the RLPL in their classrooms, embedding the dimensions of cognitive SR (e.g., working memory, inhibitory control) in other traditional games and music, and in the Ministry of National Education Preschool Education Program (e.g., classroom organization, visual clues, giving choices) and evaluating the PD program.

This way of organizing the content and implementing the PD program aimed to assist the participants in progressing step by step, developing an understanding of the theoretical foundations of SR and related practical skills. Documentations that emerged from group activities and reflective journals were kept, but they are not included in this study. Although the participants received follow-up support through an online meeting, classroom visits, and online platforms after the completion of the PD, these supports were not ongoing.

### 3.2. Results

Descriptive statistics for participating teachers' pretest and posttest scores of TASRiC were as follows: Self-regulation subscale ((Pretest) = 4.00, *SD* = 0.39; (Posttest) = 4.75, *SD* = 0.28), relation to learning subscale ((Pretest) = 4.19, *SD* = 0.45; (Posttest) = 4.92, *SD* =0.18), and total of TASRiC ((Pretest) = 4.05, *SD* =0.35; (Posttest) = 4.80, *SD* =0.21). The pretest average mean score of the self-regulation subscale is lower than the relation to the learning subscale. All the posttest scores reflect the increase in teachers' agreement about whether the items related to children's self-regulation skills (5 = Totally Agree). The descriptive values of the TVoCSRS (pretest mean = 4.07, *SD* = 0.56; posttest mean = 4.82, *SD* = 0.25) demonstrate an increase in the participants' views about identifying the indicators of self-regulation skills exhibited in classrooms. The effect of PD on the participants' pretest-posttest scores regarding their awareness of children's SR was compared using the Wilcoxon Signed-Rank Test results (Table 7).

**Table 7.** *Wilcoxon Signed Ranks Test results about the pretest-posttest scores of the TASRiC.*

| Posttest-Pretest | | *N* | Mean rank | Sum of ranks | *Z* | *p* |
|---|---|---|---|---|---|---|
| Self-regulation posttest-pretest scores | Negative ranks* | 1 | 1.00 | 1.00 | | |
| | Positive ranks** | 19 | 11.50 | 209.00 | | |
| | Ties | 1 | | | -3.890 | .001 |
| Relation to learning posttest-pretest scores | Negative ranks | 0 | 0.00 | 0.00 | | |
| | Positive ranks | 17 | 9.00 | 153.00 | | |
| | Ties | 4 | | | -3.662 | .001 |
| TASRiC's total posttest-pretest scores | Negative ranks | 0 | 0.00 | 0.00 | | |
| | Positive ranks | 21 | 11.00 | 231.00 | | |
| | Ties | 0 | | | -4.015 | .001 |

*Based on negative ranks; **Based on positive ranks

According to Table 7, there is a statistically significant difference between the participants' pretest and posttest ($p$ <.001) scores in self-regulation, relation to learning, and total scale. These differences favor the posttest. The calculated effect size shows a large difference between the scores ($r$ = .88). Hence, the PD program seems to increase preschool teachers' awareness about children's self-regulation. Using the Wilcoxon Signed Ranks Test, the participants' pretest-posttest scores about their views on children's SR skills were investigated, and the results are presented in Table 8.

**Table 8.** *Wilcoxon Signed Ranks Test results about the pretest-posttest scores of the TVoCSRS*

| Posttest-Pretest | | *N* | Mean rank | Sum of ranks | *Z* | *p* |
|---|---|---|---|---|---|---|
| Teachers' Views on Children's Self-Regulation Skills | Negative ranks* | 1 | 2.50 | 2.50 | | |
| | Positive ranks** | 18 | 10.42 | 187.50 | | |
| | Ties | 2 | | | -3.735 | .001 |

*Based on negative ranks; **Based on positive ranks

Table 8 shows a statistically significant difference between the participants' pretest and posttest scores ($p$ < .001), which favors the posttest. The calculated effect size demonstrates a large

difference between the scores ($r$ = 82). According to these findings, the PD program increased teachers' views on children's self-regulation.

### 3.3. Discussion and Conclusion

Study 2 investigated the PD program's effect on the participant preschool teachers' awareness and views about children's SR. One of the findings of this study was that the statistically significant difference between the participants' pretest and posttest awareness scores measured by TASRiC was in favor of the posttest. This finding implies that after attending the PD program, participants' understandings of SR's cognitive domain and its facilitation of children's learning increased significantly compared to their pretest scores. They could relate features, such as retaining what is said in mind, responding appropriately to instructions, mentally associating parts of an event, and intentional control of attention, behavior, thoughts, and/or emotions, to children's SR. Possible explanations for this finding might be that the participants volunteered to participate in this PD program. Also, the descriptive statistics about the participants' level of knowledge and practical skills for supporting self-regulation indicated that SR was a novel notion for them. Their responses demonstrated that 38% of the participants either never heard of the term self-regulation or had limited knowledge about it. Considering these factors, the PD activities of this study appeared to match the participant teachers' need for knowledge about children's self-regulation, which coincides with the argument that promoting teacher learning and classroom practices reflects features of high-quality PD (Garet *et al*., 2001). This PD program provided participants with theoretical information about the definitions of self-regulation, its dimensions, and how it facilitates learning. In addition, the participants gained insight into how a broader perspective, which considers classroom interactions, guidance, and structure, can support SR (Diamond & Lee, 2011). They learned about the intervention program, RLPL, specifically designed to promote SR. Moreover, the participants had opportunities to share their thoughts, practice, and discuss new learning with their colleagues and research team. As emphasized in the literature, knowledge sharing (Lee *et al*., 2023) and PD activities that allow teachers to think and reflect on their practice (Vujičić & Tambolaš, 2017) are essential in encouraging them to implement appropriate classroom practices. The study's finding, which showed a difference in the participants' pretest and posttest awareness scores, aligns with research findings that link teacher knowledge to the PD activities they participate in (Atmaca & Ertürk Kara, 2023; Iriogbe-Efionayi, 2020). Additional support for this comes from current research, which found that educators' confidence in their knowledge of SR improved after participating in a program promoting children's SR (Vasseleu *et al*., 2024).

The results further showed that the participants' posttest scores regarding their views of children's SR skills were higher than their pretest scores. This indicates that the participants, after participating in this PD, were able to relate indicators such as raising hands, waiting for their turn, and complying with the rules regarding learning centers with SR. Without appropriate theoretical knowledge and explanations, teachers may have difficulty connecting such indicators to SR, and they may be unable to distinguish between children with well-developed and underdeveloped self-regulatory skills. As Vasseleu *et al*. (2024) point out, children with well-developed SR can sustain attention by ignoring distractions, interacting positively with others, and not giving up easily on challenging tasks. The descriptives in the study group section reveal that nearly half of the participants in this study either lacked or had low skills to support children's SR in the classroom environment. This finding is consistent with other research, indicating that teachers did not know about children's SR skills (Willis *et al*., 2014), lacked strategies to foster these skills (Bautista *et al*., 2024), and perceived this construct differently from the literature (Vasseleu *et al*., 2022). Teachers' shortcomings in conceptual and practical skills may stem from the lack of emphasis on children's SR skills in preservice teacher education and in-service training programs. Hence, pre-service teacher education and in-service training programs could be restructured to support the knowledge base about developing and promoting children's SR in early childhood settings.

Based on the results discussed above, the structure of this current study's PD effectively increased preschool teachers' cognizance and opinions about child SR. This study's PD program structure worked well in the context where SR is relatively new for early childhood teachers. However, PD providers may directly focus on practical skills and intervention programs in cultures where there is no need for theoretical foundations, which can be determined using measurement tools such as TASRiC and TVoCSRS. This approach increases the likelihood of creating PDs that are more purposeful and tailored to teachers' needs.

Although this current study found that PD activities effectively promoted teachers' awareness and views on SR, it also has limitations. One limitation is that the study employed a pretest-posttest design to examine the impact of PD activities on teachers' awareness of SR and their opinions of children's SR skills. More specifically, this study did not include a control group and a persistence test. Planning the research with a control group could aid in comparing the knowledge and skill levels of teachers regarding this concept between those who participated in PD activities and those who did not receive such training. Simultaneously, the PD activities can more clearly reveal the gains on this issue. Including a persistence test in the research may allow future research to determine the level of retained knowledge by using teacher self-report measures, such as TASRiC and TVoCSRS. Implementing newly learned knowledge is necessary for retaining PD knowledge (Liu & Phelps, 2020). Hence, other researchers may conduct follow-up studies that measure whether teachers retain gained knowledge and to what extent their knowledge influences children's skills. The small size of the study group ($n$ = 21) is another limitation of this study. The study's findings indicated that PD activities effectively enhanced teachers' SR knowledge, suggesting the possibility of offering these activities to larger sample groups as part of preschool teachers' in-service training.

### 3.3.1. *Implications*

The present study underscores the need to empower preschool teachers by supporting them through targeted PD activities about children's self-regulation. Results indicated that teachers who experienced the PD program increased their awareness and views about children's SR. Therefore, given the significance of SR in children's lives and the lack of knowledge and strategies among teachers to promote such skills in the classroom (Bautista *et al*., 2024; Willis *et al*., 2014), it is necessary to implement teacher training programs that primarily focus on knowledge gain and practical skills. Evidence from research supports this, demonstrating a correlation between teachers' self-regulation knowledge and PD (Iriogbe-Efionayi, 2020). Also, research reports that only a small percentage of teachers participate in PD programs, but providing support on how to transfer theory to the classroom environment may increase their participation (Willis *et al*., 2014). Consequently, it is essential to provide PD programs that meet teachers' needs, incorporate differentiating features such as content that combines theoretical knowledge and practical skills, and methods that offer opportunities for small group reflection on the newly learned knowledge and past experiences.

## Acknowledgments

This research was supported by The Scientific and Technological Research Council of Türkiye, grant 121B351. The opinions, views, and ascertains expressed are those of the author and do not represent the views of the Council.

## Declaration of Conflicting Interests and Ethics

The author declares no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author. **Ethics Committee Number**: Pamukkale University, Protocol No:10.131.1.47 430 Dated: 2021, and 10.131.1.35 Dated:2023.

## Orcid

Asiye İvrendi https://orcid.org/0000-0002-0555-9247

## REFERENCES


Ackerman, D.J., & Friedman-Krauss, A.H. (2017). Preschoolers' executive function: Importance, contributors, research needs and assessment options. *Policy Information Report and ETS Research Report Series*, 1-24. https://doi.org/10.1002/ets2.12148

Adagideli, F.H., Saraç, S., & Ader, E. (2015). Assessing preschool teachers' practices to promote self-regulated learning. *International Electronic Journal of Elementary Education*, *7*(3), 423-440.

Atmaca, B., & Ertürk Kara, H.G. (2023). Impact of distance professional development course on early childhood teacher's knowledge, skills, and beliefs about self-regulation. *Journal of Learning and Teaching in Digital Age*, *9*(2), 1-11. https://doi.org/10.53850/joltida.1301612

Bautista, A., Williams, K.E., Lee, K., & Ng, S.P. (2024). Early self-regulation: Kindergarten teachers' understandings, estimates, indicators, and intervention strategies. *Journal of Early Childhood Teacher Education*, *45*(3), 290-311. https://doi.org/10.1080/10901027.2024.2327424

Berger, A., Kofman, O., Livneh, U., & Henik, A. (2007). Multidisciplinary perspectives on attention and the development of self-regulation. *Progress in neurobiology*, *82*(5), 256-286.

Blair, C. & Diamond, A. (2008). Biological processes in prevention and intervention: The promotion of self-regulation as a means of preventing school failure. *Development and Psychopathology*, *20*(3), 899–911. https://doi.org/10.1017/S0954579408000436

Blair, C., & Razza, R.P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647-663.

Bodrova, E., & Leong, D.J. (2006). Self-regulation as key to school readiness: How early childhood teachers promote this critical competency. In M. Zaslow & I. Martinez-Beck (Eds.), *Critical issues in early childhood professional development* (pp. 203–224). Paul H. Brookes Publishing.

Bronson, M.B. (2000). Recognizing and supporting the development of self-regulation in young children. *Young Children*, *55*(2), 32-37.

Büyüköztürk, Ş. (2009). *Sosyal bilimler için veri analizi el kitabı [Data analysis handbook for social sciences]*. PEGEM.

Cadima, J., Verschueren, K., Leal, T., & Guedes, C. (2016). Classroom interactions, dyadic teacher–child relationships, and self–regulation in socially disadvantaged young children. *Journal of Abnormal Child Psychology*, *44*, 7-17. https://doi.org/10.1007/s10802-015-0060-5

Carpenter, S. (2018). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures*, *12*(1), 25 44. https://doi.org/10.1080/19312458.2017.1396583

Christensen, L.B., Johnson, B., & Turner, L.A. (2015). *Research methods, design, and analysis*. Pearson.

Clark, L.A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319.

Cohen, J. (1988). *Statistical Power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları [Multivariate statistics for social sciences: SPSS and LISREL applications]*. PEGEM.

Diamond, A. (2014). Want to optimize executive functions and academic outcomes? Simple, just nourish the human spirit. *Minnesota Symposia on Child Psychology*, *37*, 205–232.

Diamond, A. (2012). Activities and programs that improve children's executive functions. *Current Directions in Psychological Science*, *21*(5), 335-341. https://doi.org/10.1177/0963721412453722

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4-12 years old. *Science, 19(*333-6045), 959±964. https://doi.org/10.1126/science.1204529

Dias, N.M., & Seabra, A.G. (2015). Is it possible to promote executive functions in preschoolers? A case study in Brazil. *International Journal of Child Care and Education Policy*, *9*(6), 1-18. https://doi.org/10.1186/s40723-015-0010-2

Garet, M.S., Porter, A.C., Desimone, L., Birman, B.F., & Yoon, K.S. (2001). What makes professional development effective? Results from a national sample of teachers. *American Educational Research Journal*, *38*(4), 915-945. https://doi.org/10.3102/00028312038004915

Guskey, T.R. (2003). What makes professional development effective? *Phi Delta Kappan*, *84*(10), 748-750. https://doi.org/10.1177/003172170308401007

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage.

Fritz, C.O., Morris, P.E., & Richler, J.J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2-18.

Hamre, B.K., Partee, A., & Mulcahy, C., 2017. Enhancing the impact of professional development in the context of preschool expansion. *AERA Open*, *3*(4). https://doi.org/10.1177/2332858417733686

Henson, R.K., & Roberts, J.K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement*, *66*(3), 393-416.

Howse, R.B., Lange, G., Farran, D.C., & Boyles, C.D. (2003). Motivation and self-regulation as predictors of achievement in economically disadvantaged young children. *Journal of Experimental Education*, *71*(2), 151–174. https://doi.org/10.1080/00220970309602061

Iriogbe-Efionayi, S. (2020). Promoting self-regulation in early childhood education: teachers' knowledge of self-regulation. *Urban Education Research and Policy Annuals- Tennessee State University Special Edition*, *7*(1), 5-21.

İvrendi, A., Atan, A., & Erol, A. (in press). Promoting Children's Self-Regulation and Early Numeracy Skills through an Intervention Program. *Pamukkale University Journal of Education*.

Ivrendi, A., Cevher Kalburan, N., & Şimşek, Z.C. (2022). Experiences of preschool teachers in the professional development program of children's self-regulation skills. *Journal of Teacher Education and Educators*, *11*(3), 309-330.

Ivrendi, A. (2011). Influence of Self-Regulation on the Development of Children's Number Sense. *Early Childhood Education Journal*, *39*(4), 239-247. https://doi.org/10.1007/s10643-011-0462-0

Jahromi, L.B., & Stifter, C.A. (2008). Individual differences in preschoolers' self-regulation and theory of mind. *Merrill-Palmer Quarterly*, *54*(1), 125-150. https://doi.org/10.1353/mpq.2008.0007

Kopp, C.B. (1982). Antecedents of self-regulation: A developmental perspective. *Developmental Psychology*, *18*(2) 199-214.

Korucu, I., Ayturk, E., Finders, J.K., Schnur, G., Bailey, C.S., Tominey, S.L., & Schmitt, S.A. (2022). Self-regulation in preschool: Examining Its factor structure and associations with pre-academic skills and social-emotional competence. *Frontiers in Psychology*, *12*, 1–14. https://doi.org/10.3389/fpsyg.2021.717317

Lenes, R., Gonzales, C.R., Størksen, I., & McClelland, M.M. (2020). Children's self-regulation in Norway and the United States: The role of mother's education and child gender across cultural contexts. *Front. Psychol*, *11*(566208). https://doi.org/10.3389/fpsyg.2020.566208

Lee, K-C., Chang, I-H., Wang, I-L., & Chen, R-S. (2023). Effects of knowledge sharing on sustainable happiness of preschool teachers: The mediating roles of self-efficacy and helping behavior. *Current Psychology*, *42*, 13648-13657. https://doi.org/10.1007/s12144-022-02725-8

Liu, S., & Phelps, G. (2020). Does teacher learning last? Understanding how much teachers retain their knowledge after professional development. *Journal of Teacher Education*, *71*(5) 537–550. https://doi.org/10.1177/0022487119886290

Mann, T.D., Hund, A.M., Hesson-McInnis, M.S., & Roman, Z.J. (2016). Pathways to school readiness: Executive functioning predicts academic and social–emotional aspects of school readiness. *Mind, Brain, and Education*, *11*(1), 21-31. https://doi.org/10.1111/mbe.12134

Markussen-Brown, J., Juhl, C.B., Piastad, S.B., Blesese, D., Højene, A., & Justice, L.M. (2017). The effects of language- and literacy-focused professional development on early educators and children: A best-evidence meta-analysis. *Early Childhood Research Quarterly*, *38*, 97–115. https://doi.org/10.1016/j.ecresq.2016.07.002

Mattera, S., Rojas, N.M., Morris, P.A. & Bierman, K. (2021). Promoting EF with preschool interventions: Lessons learned from 15 years of conducting large-scale studies. *Frontiers in Psychology*, *12* (640702). https://doi.org/10.3389/fpsyg.2021.640702

McClelland, M.M., Tominey, S.L., Schmitt, S.A., Hatfield, B.E., Purpura, D.J., Gonzales, C.R., & Tracy, A.N. (2019). Red Light, Purple Light! Results of an intervention to promote school readiness for children from low-income backgrounds. *Frontiers in Psychology*, *10*(2365), 1-15. https://doi.org/10.3389/fpsyg.2019.02365

McClelland, M.M., & Tominey, S.L. (2014). The development of self-regulation and executive function in young children. *Zero to Three*, *35*(2), 2-8.

McClelland, M.M., & Tominey, S.L. (2015). *Stop, think, act*. Taylor and Francis.

McClelland, M.M., & Cameron, C.E. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives*, *6*(2), 136–142. https://doi.org/10.1111/j.1750-8606.2011.00191.x

McClelland, M.M., Cameron, C.E., Wanless, S., & Murray, A. (2007). Executive function, behavioral self-regulation, and social-emotional competence: Links to school readiness. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on social learning in early childhood education*. (p. 113–137). Information Age Publishing.

Montoya, M.F., Susperreguy, M.I., & Morrison, F.J. (2023). Self-regulation scaffolding behaviors of teachers in Chilean preschool classrooms. E*arly Education and Development*, *34*(6), 1305-1324. https://doi.org/10.1080/10409289.2022.2135867

MoNE-PEP (2013). *36-72 Aylık Çocuklar İçin Okul Öncesi Eğitim Programı [Preschool Education Curriculum for 36-72 Months Old Children]*. *Ministry of National Education.*

Murray, D.W., Rosanbalm, K., & Christopoulos, C. (2016). *Self-regulation and toxic stress report 3: A Comprehensive review of self-regulation interventions*. OPRE Report # 2016-34, Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Perels, F., Merget-Kullmann, M., Wende, M., Schmitz, B., & Buchbinder, C. (2009). Improving self-regulated learning of preschool children: Evaluation of training for kindergarten teachers. *British Journal of Education Psycholgy*, *79*(2), 311-327. https://doi.org/10.1348/000709908X322875

Popova, A., Evans, D.K., Breeding, M.E., & Arancibia, V. (2022). Teacher professional development around the world: The Gap between evidence and practice. *The World Bank Research Observer*, *37*(1), 107–136. https://doi.org/10.1093/wbro/lkab006

Raffaelli, M., Crockett, L.J., & Shen, Y.L. (2005). Developmental stability and change in self-regulation from childhood to adolescence. *The Journal of Genetic Psychology: Research and Theory on Human Development*, *166*(1) 54-75.

Ribner, A.D., Willoughby, M.T., Blair, C.B., & The Family Life Project Key Investigators (2017). Executive function buffers the association between early math and later academic skills. *Frontier Psychology*, *8*(869). https://doi.org/10.3389/fpsyg.2017.00869

Roebers, C.M. (2017). Executive function and metacognition: Towards a unifying framework of cognitive self-regulation. *Developmental Review*, *45*, 31–51.

Rosanbalm, K.D., & Murray, D.W. (2017). *Promoting self-regulation in early childhood: A practice brief*. OPRE Brief #2017-79. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, US. Department of Health and Human Services.

Savina, E. (2021). Self-regulation in preschool and early elementary classrooms: Why it is important and how to promote it. *Early Childhood Education Journal*, *49*(3), 493-501. https://doi.org/10.1007/s10643-020-01094-w

Sezgin, E., & Ulus, L. (2020). An examination of self-regulation and higher-order cognitive skills as predictors of preschool children's early academic skills. *International Education Studies*, *13*(7), 65-87. https://doi.org/10.5539/ies.v13n7p65

Schachter, R.E. (2015). An analytic study of the professional development research in early childhood education. *Early Education and Development*, *26*(8), 1057-1085. https://doi.org/10.1080/10409289.2015.1009335

Schmitt, S.A., Geldhof, G.J., Purpura, D.J., Duncan, R., & McClelland, M.M. (2017). Examining the relations between executive function, math, and literacy during the transition to kindergarten: A multi-analytic approach. *Journal of Educational Psychology*, *109*(8), 1120-1140. https://doi.org/10.1037/edu0000193

Schmitt, S.A., McClelland, M.M., Tominey, S.L., & Acock, A.C. (2015). Strengthening school readiness for Head Start children: Evaluation of a self-regulation intervention. *Early Childhood Research Quarterly*, *30*, 20–31. https://doi.org/10.1016/j.ecresq.2014.08.001

Silkenbeumer, J.R., Schiller, E., & Kärtner, J. (2018). Co- and self-regulation of emotions in the preschool setting. *Early Childhood Research Quarterly*, *44*, 72-81. https://doi.org/10.1016/j.ecresq.2018.02.014

Snyder, P., Meeker, K.A., & Pasia, C. (2012). Characterizing key features of the early childhood professional development literature. *Infants & Young Children*, *25*(3) 188-212. https://doi.org/10.1097/IYC.0b013e31825a1ebf

Streiner, D.L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99-103. https://doi.org/10.1207/S15327752JPA8001_18

Størksen, I., Ellingsen, I.Y., Wanless, S.B., & McClelland, M.M. (2015). The influence of parental socioeconomic background and gender on self-regulation among 5-year-old children in Norway. *Early Education and Development*, *26*(5-6), 663-684, https://doi.org/10.1080/10409289.2014.932238

Tabachnick, B.G., & Fidell, L.S. (2015). *Çok değişkenli istatistiklerin kullanımı [Using Multivariate Statistics]* (M. Baloğlu, Trans.; 7th ed.). Nobel. (Original work published 2013)

Tominey, S., McClelland, M., & Tracy, A. (2018). *Red Light, Purple Light a self-regulation intervention*. Oregon State University.

Uyanık, G., Bayındır, D., & Değirmenci, Ş. (2021). Akademik ve sosyal duygusal okula hazır bulunuşluk ile öz düzenleme becerileri arasındaki ilişkinin incelenmesi [An investigation of the relationship between academic and socio-emotional school readiness and self-regulation skills]. *Elektronik Sosyal Bilimler Dergisi*, *20*(79), 1090-1104. https://doi.org/10.17755/esosder.811871

Vasseleu, E., Neilsen-Hewett, C., & Howard, S.J. (2024). An early start to self-regulation: Evaluating the effects of an early childhood self-regulation intervention on educator beliefs, knowledge, and practice. *Journal of Research in Childhood Education*, 1-21. https://doi.org/10.1080/02568543.2024.2396907

Vasseleu, E., Neilsen-Hewett, C., Clif, K., & Howard, S.J. (2022). How educators in high quality preschool services understand and support early self-regulation: a qualitative study of knowledge and practice. *The Australian Educational Researcher*, *49*, 915-941. https://doi.org/10.1007/s13384-021-00466-4

Vasseleu, E., Neilsen-Hewett, C., Ehrich, J., Cliff, K., & Howard, S.J. (2021). Educator beliefs around supporting early self-regulation: Development and evaluation of the self-regulation

knowledge, attitudes and self-efficacy scale. *Frontiers in Education*, *6*, 6. https://doi.org/10.3389/feduc.2021.621320

Venitz, L., & Perels, F. (2018). Promoting self-regulated learning of preschoolers through indirect intervention: a two-level approach. *Early Child Development and Care*, *189*(13), 2057-2070. https://doi.org/10.1080/03004430.2018.1434518

Vujičić, L., & Tambolaš, A.C. (2017). Professional development of preschool teachers and changing the culture of the institution of early education. *Early Child Development and Care*, *187*(10), 1583-1595. https://doi.org/10.1080/03004430.2017.1317763

Wei, R.C., Darling-Hammond, L., Andree, A., Richardson, N., & Orphanos, S. (2009). *Professional learning in the learning profession: A status report on teacher development in the United States and abroad*. National Staff Development Council.

Welsh, J. A., Nix, R.L., Blair, C., Bierman, K.L., & Nelson, K.E. (2010). The development of cognitive skills and gains in academic school readiness for children from low-income families. *Journal of Educational Psychology*, *102*(1), 43-53. https://doi.org/10.1037/a0016738

Willis, E., Dinehart, L., & Bliss, L. (2014). Teachers don't always do what they think they should: A preliminary validation of the early childhood educators' knowledge of self-regulation skills questionnaire. *Journal of Early Childhood Teacher Education*, *35*(2), 168-184. https://doi.org/10.1080/10901027.2014.905806

## APPENDIX

**Appendix 1.** *Turkish and English versions of the scales.*

### Çocuklarda Öz-Düzenlemeye ilişkin Öğretmenlerin Farkındalığı

| Öz-düzenleme | Kesinlikle Katılmıyorum | Katılmıyorum | Kısmen Katılıyorum | Katılıyorum | Tamamen Katılıyorum |
|---|---|---|---|---|---|
| 1. Ani fırsatları kullanabilmesi/değerlendirebilmesi ile ilgilidir. | | | | | |
| 2. Öğretmenin yönergelerine uygun bir şekilde tepki vermesi ile ilgilidir. | | | | | |
| 3. Söylenenleri/anlatılanları aklında tutması ile ilgilidir. | | | | | |
| 4. Farklı ortamlarda/etkinliklerde davranışlarının yoğunluğunu, sıklığını, süresini düzenlemesi ile ilgilidir. | | | | | |
| 5. Bir olayın/durumun parçalarını zihinsel olarak ilişkilendirmesi ile ilgilidir. | | | | | |
| 6. Bir görevi tamamlamama isteğine karşı koyabilmesi ile ilgilidir. | | | | | |
| 7. Öğretmen ya da yetişkin yokluğunda kurallara uygun davranması ile ilgilidir. | | | | | |
| 8. Akademik becerileri (matematik, dil gibi) öğrenmesinde etkilidir. | | | | | |
| 9. Bir bilgiyi/beceriyi öğrenmesini kolaylaştıran becerilerdir. | | | | | |
| 10. Öğrenebildiği becerilerdir. | | | | | |

### Teachers' Awareness of Self-Regulation in Children

| Self-regulation is about | Strongly Disagree | Disagree | Partially Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1. Being able to use/assess sudden opportunities. | | | | | |
| 2. Responding appropriately to the teacher's instructions. | | | | | |
| 3. Keeping in mind what is said/told. | | | | | |
| 4. Regulating behaviors' intensity, frequency, and duration in different environments/activities. | | | | | |
| 5. Mentally associating the parts of an event/situation. | | | | | |
| 6. Being able to resist the urge not to complete a task. | | | | | |
| 7. Behaving by the rules in the absence of a teacher or adult. | | | | | |
| 8. Self-regulation is effective in learning academic skills (such as mathematics and language). | | | | | |
| 9. Self-regulation skills make it easier to learn a knowledge/skill. | | | | | |
| 10. Self-regulation skills can be learned. | | | | | |

**Çocukların Öz-Düzenleme Becerileri ile ilgili Öğretmen Görüşleri**

| Öz-düzenleme becerileri | Kesinlikle Katılmıyorum | Katılmıyorum | Kısmen Katılıyorum | Katılıyorum | Tamamen Katılıyorum |
|---|---|---|---|---|---|
| 1. Dikkatini, bitirilen bir etkinlikten sonra yeni başlanılan etkinliğe verebilmeyle ilgilidir. | | | | | |
| 2. Diğer etkinliğe başlamadan önce yaptığı etkinlik ile ilgili materyalleri yerine koymayla ilgilidir. | | | | | |
| 3. Yönergeleri takip ederek diğer etkinliklerden büyük grup zamanına ve büyük grup zamanından da diğer etkinliklere geçiş yapmayla ilgilidir. | | | | | |
| 4. Güvenlik kurallarını takip etmeyle (örneğin, binaya girerken yürümek, kaydırakta bir seferde bir çocuğun kayması gibi) ilgilidir. | | | | | |
| 5. Öğrenme merkezleri ile ilgili kurallara uymayla ilgilidir. | | | | | |
| 6. Hikâyeye dikkat ettiğini gösteren sorular sorma ile ilgilidir. | | | | | |
| 7. Öğretmenlerin ya da akranların sorularına cevap vermekle ilgilidir. | | | | | |
| 8. Bir etkinlikte parmak kaldırmak ve sırasını beklemek ile ilgilidir. | | | | | |

**Teacher's Views on Children's Self-Regulation Skills**

| Self-regulation skills are about | Strongly Disagree | Disagree | Partially Agree | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1. Being able to pay attention to a new activity after completing an activity. | | | | | |
| 2. Putting the materials related to the activity he/she is doing before starting the other activity. | | | | | |
| 3. Transitioning from other activities to large group time and from large group time to other activities, following directions. | | | | | |
| 4. Following safety rules (e.g., walking when entering the building, going down the slide one child at a time, etc.). | | | | | |
| 5. Complying with the rules regarding learning centers. | | | | | |
| 6. Asking questions that show he/she is paying attention to the story. | | | | | |
| 7. Answering questions from teachers or peers. | | | | | |
| 8. Raising a hand and waiting a turn at an activity. | | | | | |

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 317–340
https://doi.org/10.21449/ijate.1602294
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# A review of automatic item generation techniques leveraging large language models

**Bin Tan**[1*], **Nour Armoush**[1], **Elisabetta Mazzullo**[1], **Okan Bulut**[2], **Mark J. Gierl**[2]

[1]University of Alberta, Faculty of Education, Measurement, Evaluation, and Data Science, Edmonton, Canada
[2]University of Alberta, Faculty of Education, Centre for Research in Applied Measurement and Evaluation, Edmonton, Canada

**ARTICLE HISTORY**

*Received: Dec. 16, 2024*

*Accepted: Apr. 28, 2025*

**Keywords:**

Automatic item generation,
Automatic question generation,
Large language models,
Generative artificial intelligence,
Educational technology.

**Abstract:** This study reviews existing research on the use of large language models (LLMs) for automatic item generation (AIG). We performed a comprehensive literature search across seven research databases, selected studies based on predefined criteria, and summarized 60 relevant studies that employed LLMs in the AIG process. We identified the most commonly used LLMs in current AIG literature, their specific applications in the AIG process, and the characteristics of the generated items. We found that LLMs are flexible and effective in generating various types of items across different languages and subject domains. However, many studies have overlooked the quality of the generated items, indicating a lack of a solid educational foundation. Therefore, we share two suggestions to enhance the educational foundation for leveraging LLMs in AIG, advocating for interdisciplinary collaborations to exploit the utility and potential of LLMs.

## 1. INTRODUCTION

As a foundational component of the modern education system, assessments serve multiple purposes such as facilitating student learning, evaluating teaching outcomes, and informing educational policy (Black, 1998; Lee *et al*., 2020). Advancements in educational technology, such as e-learning platforms (Granić *et al*., 2022), adaptive testing (Yen *et al*., 2012), and formative assessments (Dalby & Swan, 2019; Spector *et al*., 2016), have increased the demand for high-quality assessment items. However, traditional item development methods are struggling to keep pace with this demand. For example, the cost of developing a single item for high-stakes assessments can reach up to $2,000 (Rudner, 2010). Additionally, items for low-stakes assessments often suffer from low-quality content, thus failing to provide comprehensive or valuable feedback to students (Lim, 2019; Wylie & Lyon, 2015).

In response to the growing demand for high-quality items, educational measurement researchers first turned their attention toward a solution called automatic item generation (AIG). The goal of AIG is to generate large banks of high-quality items while reducing overall costs

*CONTACT: Bin TAN ✉ btan4@ualberta.ca Measurement, Evaluation, and Data Science, Faculty of Education, University of Alberta, 6-110 Education Centre North, 11210 87 Ave NW, Edmonton, AB T6G 2G5 CANADA

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

(Gierl & Lai, 2012; Lai *et al*., 2009). The template-based method proposed by Gierl *et al*. (2012) is the most widely used method for AIG, which involves a three-step process that requires the construction of cognitive models and item models (Gierl & Lai, 2012, 2016). This method has been successfully implemented in various educational practices (e.g., medical licensing examinations), demonstrating its viability and cost-effectiveness in creating high-quality items (Kosh *et al*., 2018; Pugh *et al*., 2020). Coincidentally, there is a related line of research in the domain of computer science, specifically focusing on automatic question generation (AQG) for educational purposes. Originally rooted in natural language processing (NLP) research for developing chatbots and conversational agents, AQG for educational purposes has evolved to share a similar goal with AIG: the automatic generation of a large number of assessment items using computer algorithms. Despite this shared goal, key terminological differences persist between the two fields. In educational measurement, the term "items" can be presented in diversified types to measure students' knowledge and skills, while "questions" typically have a much narrower format, usually referring to interrogative sentences ending with a question mark (Gierl *et al*., 2021). Although an item may include a question, not all questions qualify as complete items. Despite this distinction, researchers of AQG for educational purposes often use the term "questions" broadly, even when referring to non-interrogative item formats.

The work conducted in AQG research for educational purposes has been summarized in two recent review studies (Ch & Saha, 2018; Kurdi *et al*., 2020). Regrettably, in both reviews, there were no keywords of "automatic item generation" or even word variants of "items" in the search terms used. As a result, many significant publications in the AIG literature were not included in their review. Furthermore, educational measurement researchers have been less familiar with AQG due to differences in research interests, technical skill sets, and limited participation in NLP conferences. This gap is evident in the minimal overlap between the references and citations used in the two research domains. For example, nearly all references presented in the review paper of Kurdi *et al*. (2020) were published in computer science venues, while the references used in a review paper of AIG (Falcao *et al*., 2022) were mostly in education, psychology, and testing journals. As a result, historically, researchers in both domains have had limited knowledge of each other's work or were not even aware of each other's existence. More importantly, the use of different terminology (e.g., "items" vs. "questions") to describe and address similar problems has created communication barriers between the AIG and AQG research domains, hindering their advancement.

Marked as recent breakthroughs in generative artificial intelligence research, pre-trained language models (PLMs) and large language models (LLMs) have overcome the limitations of earlier autoregressive algorithms (e.g., rule-based systems) which relied on fixed architectures for rigid, sequential text prediction. Unlike traditional autoregressive algorithms, modern language models offer the distinct advantage of being retrainable and fine-tunable, enabling them to develop more sophisticated language comprehension and produce more contextually appropriate assessment items (e.g., Vu & Van Nguyen, 2022). These language models focus specifically on text generation but have also demonstrated their flexibility and extraordinary performance in understanding the context of language and in various language processing tasks (Ackerman & Balyan, 2023). The versatility and potential application of the models has have significantly sparked the interest of both educational measurement researchers and applied NLP researchers in employing them as an innovative methodology for automating the generation of a large number of items. Technically, LLMs are essentially scaled-up versions of PLMs, characterized by an increased number of parameters and larger training data sizes. However, there is no strict cutoff value that differentiates the number of parameters of PLMs from those of LLMs (Zhao *et al*., 2023). Therefore, in the remaining sections of this review, the term "LLMs" is used to refer to both PLMs and LLMs, as it is a more commonly recognized term.

As AIG and AQG researchers are increasingly sharing the same goal of generating questions for educational purposes and are also adopting similar methodologies such as the use of LLMs,

there is substantial decrease in the distinctions between them. Therefore, it has become crucial to raise awareness of both AIG and AQG research and map the literature of both fields to gain a comprehensive understanding of the utility and potential of LLMs. The purpose of this review is not only to identify what has been achieved in the literature regarding the use of LLMs for AIG but also to explore the knowledge gaps between the two research domains. Additionally, we aim to provide suggestions for future research to advance automated item generation practices in both fields. To ensure consistent terminology, we will use "AIG" to refer to both AIG and AQG for educational purposes in this paper, as the term "item" is more inclusive than "question" (Gierl *et al.*, 2021). However, it is important to note that these terms have distinct historical origins and development trends, but they are used interchangeably here due to their similar research objectives and the focus on leveraging LLMs in this review. To map the current state of research on leveraging LLMs for AIG, our review is guided by the following research questions (RQs):

RQ1: What are the most used LLMs for AIG, and how do their underlying architectures and features differ?

RQ2: In what specific ways are LLMs most frequently used, and how do the LLMs' features influence their use in the AIG process?

RQ3: What types of items were generated in the reviewed studies in terms of item type, subject domain, and language? Can LLMs generate valid, reliable, and high-quality items?

## 2. METHOD

### 2.1. Study Search and Selection

The study search and selection process are displayed in Figure 1. Keyword searches were conducted in seven research databases, which were selected for their comprehensive coverage of literature on AIG (Alsubait, 2015) or relevance to LLMs. The searches used two keyword components: AIG and LLMs, connected by an "AND" operator, as detailed in Table 1. Our search targeted full texts, with no restrictions on publication types in the search strategy, and was limited to publications after 2018, marking the emergence of pre-trained language models (Zhao *et al.*, 2023). The last search was conducted on August 1st, 2023, which yielded 831 unique study records from these research databases.

After the study search, two reviewers independently screened the titles, abstracts, and keywords of the 831 identified studies, using predefined inclusion criteria for selection. We only retained studies with full texts in English and focused exclusively on empirical studies, excluding editorials, commentaries, and position papers. Given that question generation has applications in other domains like chatbots and customer service, we confined our selection to studies in the education field. This focus was achieved by identifying education-related words or phrases in the title, abstracts, or keywords. Additionally, we selected publications that focused on item generation, excluding other forms of text generation, such as the creation of reading passages without items generated. After the initial screening, we noted 47 disagreements in selection decisions, resulting in a high agreement rate of 93.44%. The disagreements primarily resulted from vague mentions of education-related phrases in abstracts from papers published in computer science venues. These disagreements were resolved through further discussion and consultation of the full texts of the studies, resulting in the selection of 65 studies for full-text review and information extraction. After a detailed review, 5 studies were excluded based on our criteria, leaving 60 studies in this review.

**Figure 1.** *The flow diagram of search process.*

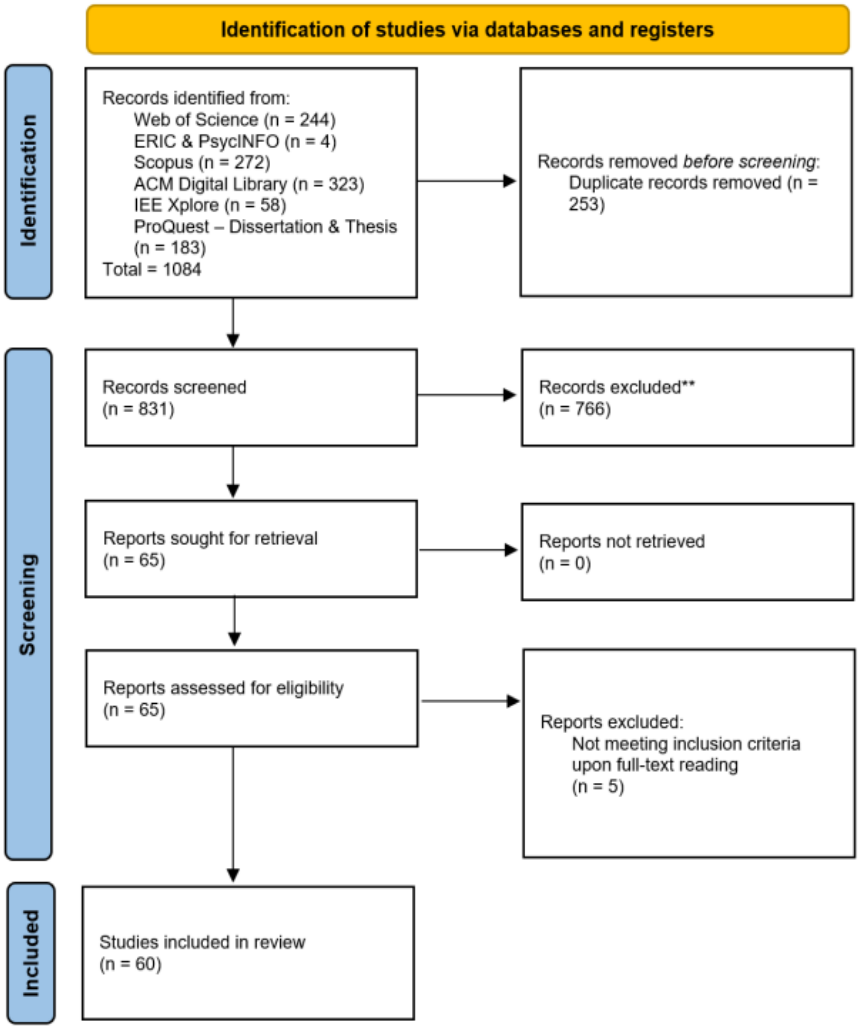



**Table 1.** *Keywords used for study search.*

| Keywords related to AIG | Keywords related to LLMs |
|---|---|
| Item generation*,<br>AIG,<br>Question generation,<br>Distractor generation,<br>Test development,<br>Item development,<br>Generat* Item*,<br>Generat* Question* | Large language model, Pre-trained language model, PLM, LLM, language model, BERT, GPT*, ChatGPT, T5, mT5, PanGu, T0, CodeGen, Tk-Instruct, UL2, OPT, NLLB, GLM, Flan-T5, BLOOM, mT0, Galactia, BLOOMZ, OPT-IML, LLaMA, CodeGeeX, Pythia, GShard, Codex, ERNIE, Jurassic-1, HyperCLOVA, FLAN, Yuan, Anthropic, WebGPT, Gopher, GlaM, LaMDA, MT-NLG, AlphaCode, InstructGPT, Chinchilla, PaLM, AlexaTM, Sparrow, WeLM, U-PaLM, Flan-U-PaLM, (Transformer AND Model) |

*Note:* We focused our search on publications written in English and published in or after 2018. We conducted searches using titles, abstracts, and keywords. "*" represents variants of the word, such as plurals.

## 2.2. Information Extraction and Analysis

To address the proposed research questions, a coding framework was developed through an iterative process to extract information from the included studies. Initially, one reviewer proposed a preliminary coding framework and piloted it by extracting information from 20 of the included studies. Refinements to the coding framework, such as adding new coding

subsections, removing redundant or irrelevant ones, and defining coding categories within each subsection, were conducted. Corresponding to the RQs, the finalized coding framework comprised three key aspects: (1) the specific LLMs employed in each study, (2) their implementation methods categorized by role in the AIG process, and (3) the characteristics of generated items. Subsequently, with the coding framework, two reviewers independently coded all the studies after training. During this process, several ambiguities emerged, such as the distinctions between LLM applications (e.g., item filtering versus quality ranking for quality control purposes) and classification of LLM roles in multi-stage AIG processes (as LLMs might be employed at multiple stages). To resolve these ambiguities and any coding disagreements, the reviewers first discussed their interpretations, with a third reviewer consulted when consensus could not be reached. In the final analysis phase, one reviewer systematically quantified the studies according to each coding category and synthesized the findings. For instance, LLM applications were categorized into three distinct AIG process stages: pre-generation stage, item generation stage, and post-generation stage.

While conducting the information extraction, we realized some information was missing in the current literature for us to answer our research questions. Particularly, many studies did not report the measurement properties of the generated items. Therefore, we employed a simple search strategy to identify the severity of this issue – searching for keywords that are expected to appear in the reviewed studies. We developed a set of basic keywords that are related to the measurement properties of items as well as individuals who are often in charge of evaluating item quality such as content specialists and SMEs. Then, we built a simple information extractor using the Python programming language and used it to summarize the patterns of occurrences of keywords across the reviewed studies.

## 3. RESULTS

Our review indicated that there were two studies published in 2019, followed by an increase to five in 2020, 9 in 2021, a peak of 33 in 2022, and finally, eleven studies as of August 2023. The included studies can be found in Appendix A. Over half of these studies ($n$ = 31) were papers published at conferences, such as the *International Conference on Artificial Intelligence in Education* and the *European Conference on Technology Enhanced Learning*. Fourteen of the reviewed studies were journal articles, which span two broad research domains, with educational technology and assessment journals like *Education and Information Technologies* and *International Journal of Assessment Tools in Education*, and applied artificial intelligence journals such as *Frontiers in Artificial Intelligence* and *IEEE Access*. Moreover, there were non-refereed studies, including two master's theses, one doctorate dissertation, and 12 preprints. These avenues imply the multidisciplinary and developing nature of studies in this field.

### 3.1. RQ1: What Are the Most Used LLMs for AIG, and How Do Their Underlying Architectures and Features Differ?

We identified Google's Text-to-Text Transfer Transformer (T5; $n$ = 32), Bidirectional Encoder Representations from Transformers (BERT; $n$ = 26), and OpenAI's Generative Pre-Trained Transformer (GPT; $n$ = 19) as three major base types of LLMs commonly used in AIG. In addition to these most common LLMs, we also found models that were only used once or a couple of times such as BART and PEGASUS. Because there were cases where more than one LLM was used in a study, we summarized the frequency of use by instance. That is, each study could be counted multiple times depending on how many LLMs they used.

Though all follow the transformer-learning paradigm, each type of LLM employs a distinct approach to language modeling, encompassing unique features and strengths for specific tasks. Due to the transfer learning paradigm, these LLMs often undergo further training with additional text data to enhance the base model to the desired task. This results in a variety of

variants and series. For a better understanding of their use in the AIG stages, the features and variants of each base type of LLM are outlined in the following subsections.

### 3.1.1. *T5 variants*

Introduced by Raffel *et al*. in 2020, the T5 model treats all language processing tasks as text-to-text conversions. That is, the model receives text input and directly generates text output. The dataset used to train this LLM was the C4 dataset which contains approximately 750GB of English texts sourced from the public Common Crawl web scrape (Raffel *et al*., 2020). The T5 model is available in several sizes, such as T5-small, T5-base, and T5-large, each differing in the number of parameters. The different sizes of these models enable their application in varied contexts, accommodating diverse requirements in terms of computational resources and performance capabilities.

### 3.1.2. *BERT variants*

BERT (Devlin *et al*., 2018) employs a bidirectional approach to understanding the context of words within a sentence. That is, the model analyzes texts from both left to right and right to left, which helps it gain a more comprehensive understanding of sentence contexts. BERT was originally trained using 3.3 billion words sourced from Wikipedia and BooksCorpus, based on two key training strategies: masked language model and next-sentence prediction. BERT can also be further trained with additional text data or training strategies, allowing it to exhibit differential performance in specific contexts. For example, multilingual BERT (Devlin *et al*., 2019) are trained with additional text data from multiple languages to improve performance in multilingual understanding. The size of training datasets, training architecture, and the number of parameters can also lead to variants such as ALBERT (BERT with reduced parameters and sentence order prediction; Lan *et al*., 2019), DistilBERT (BERT with distillation; Sanh *et al*., 2019), RoBERTa (BERT without NSP but dynamic masking; Liu *et al*., 2019), and XLNet (BERT with all tokens masked but in random order; Yang *et al*., 2019).

### 3.1.3. *GPT series*

GPT (OpenAI, 2018) employs the next-word prediction strategy to generate words sequentially. That is, it predicts the next word in a sentence based on probabilities conditioned on both the original input it receives (i.e., the prompt) and the words it has previously generated. The series of GPT models has seen remarkable advancements over the last few years, with each iteration increasing in parameters and enhancing capabilities and complexity. In this series, GPT-1 contained 117 million parameters, GPT-2 expanded to 1.5 billion parameters, while GPT-3 further increased the scale to 175 billion parameters. At the time of this review, GPT-4 (Open AI, 2023) stood as the latest iteration, reflecting substantial improvements in language processing, reasoning, and multimodal functionality. Subsequent models have continued to advance the field, focusing not only on scale but also on real-world usability by optimizing for faster inference speeds, reduced computational costs, and greater affordability for end users. Furthermore, these models have become increasingly accessible, with intuitive platforms enabling seamless deployment for both developers and non-technical users.

Building upon the core models of such as GPT-4 and GPT-3, Open AI has developed specialized derivatives tailored for specific applications. For example, the well-known ChatGPT (Open AI, 2022) is specifically fine-tuned to engage in interactive conversational tasks. This fine-tuning involves encoding given human language inputs (prompts) into rich, contextual-embedded textual representations that the machine system can understand. Subsequently, it decodes these textual representations sequentially, conditioned on the original prompts and the previously generated words, to produce coherent and contextually appropriate responses in human languages. Similarly, Codex, built based on GPT-3, is trained to specialize in understanding and generating codes of programming languages (Open AI, 2021).

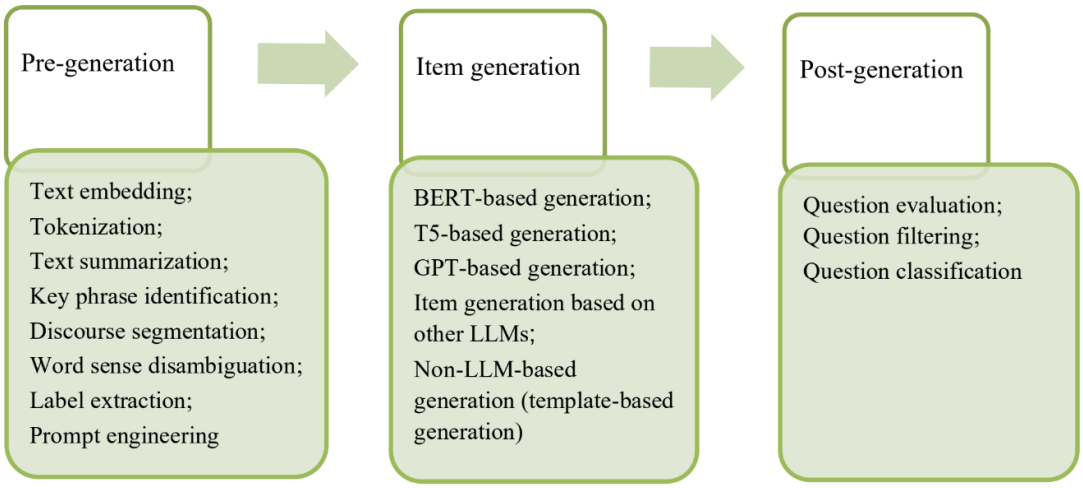
### 3.1.4. *Summary of findings for RQ1*

The most used base types of LLMs are T5, BERT, and GPT. These models exhibit differences in their training data sources, the size of their training sets, their training architectural frameworks, and their unique features.

## 3.2. RQ2: In What Specific Ways Are LLMs Most Frequently Used, and How Do the LLMs' Features Influence Their Use in the AIG Process?

As depicted in Figure 2, we summarized and synthesized the specific uses of LLMs into three stages: pre-generation, item generation, and post-generation. In the pre-generation stage, LLMs are used for preparation tasks before generating items such as cleaning, structuring, and understanding the original input text data which can be lengthy and covers various topics. This preparation is essential for the subsequent item generation stage to produce items that are contextually relevant, accurate, and coherent. In the item generation stage, items are created either directly by LLMs or through traditional methods like template-based approaches. Studies using template-based approaches were included because they employed LLMs in either the pre- or post-generation stages. After item generation is the post-generation stage, where items are evaluated and selected based on certain criteria such as quality and difficulty. With LLMs, this stage can be automated, resulting in filtered items as the final output of the AIG process. The majority of these studies (n = 44) employed LLMs in only one of the three AIG stages, 13 studies used LLMs in two stages, and 3 studies used LLMs across all three stages.

**Figure 2.** *The specific use of LLMs in the AIG process.*




### 3.2.1. *Pre-generation stage*

In the pre-generation stage, BERT was used 20 times, T5 was used 13 times, and GPT was used 9 times. We identified and categorized the following specific uses of LLMs: (1) text embedding ($n$ = 8), (2) key phrase identification ($n$ = 8), (3) text summarization ($n$ = 6), (4) word tokenization ($n$ = 3), (5) prompt engineering ($n$ = 2), (6) word sense disambiguation ($n$ = 2), (7) label extraction ($n$ = 1), and (8) discourse segmentation ($n$ = 1). These specific uses make up a toolbox for processing texts for subsequent tasks of generating desired items. The specific uses of LLMs were categorized into three themes, as described below.

The first theme is data transformation, which converts original texts into computationally tractable and manageable units. For example, word tokenization is the process of transforming texts into smaller, more manageable tokens. This task is often necessary before many subsequent preprocessing NLP tasks. LLM tokenizers have several advantages over traditional rule-based tokenizers, such as understanding context (Singh *et al.*, 2019). In addition, text embedding transforms text into numerical vectors, known as textual representations. These

numerical vectors can capture the semantic meaning of the original texts; texts with similar meanings are positioned closely in the embedding space. These numerical vectors are used and processed in later language processing tasks. For instance, they can be utilized for subsequent NLP tasks such as calculating the cosine similarity metric between items or distractors of multiple-choice questions to evaluate their semantic similarity (e.g., Min *et al*., 2021).

The second theme is related to semantic processing, which involves understanding the context of the text data or specific sentences. For example, key phrase identification involves extracting important and relevant phrases from texts. This step can help to frame the focus of the item generation. For example, Tsai *et al*. (2021) employed BERT to extract keywords from an input textbook, which were then used to construct important complete sentences as preparation for the item generation stage. In addition, word sense disambiguation is the process of determining the actual meaning of a word with multiple meanings in a given sentence and context. This step is crucial for computers to interpret words correctly in context, ensuring that later generated items are contextually appropriate and unambiguous. Moreover, we found that discourse segmentation was performed in two studies. This task divides texts into coherent segments such as sentences, paragraphs, or topics, which helps to structure the text and generate items in a way that reflects the logical and semantic composition of the input texts. Furthermore, label extraction assigns labels to texts according to categories like reading difficulty, content domain, or question type. The extracted labels serve as control labels for creating targeted and relevant items. For example, Zhao *et al*. (2022) trained an LLM to extract one of seven question types from input texts, which was then used to generate questions matching the intended type. Lastly, text summarization aims to convey the main points of the original text while significantly reducing its length, thereby improving the efficiency of subsequent item-generation tasks (Malhar *et al*., 2022).

The last theme is prompt engineering, which involves constructing commands or instructions that effectively communicate a task to LLMs. Some generative artificial intelligence such as GPT operate based on the prompt it receives. In AIG, the characteristics and quality of the prompt determine how LLMs generate items that meet specific assessment criteria, cater to various contexts, or assess across different domains. Therefore, researchers have used LLMs to generate effective prompts for subsequent tasks. For instance, in the study by Ghanem *et al*. (2022), T5 was trained on how to formulate questions and consider relevant aspects for subsequent item generation.

### 3.1.2. *Item generation stage*

We identified that 53 studies used LLMs during the item generation stage. Some studies used multiple LLMs, either to complement each other in different tasks or to compare and identify the best-performing models in the same task. Despite BERT and GPT-2 being introduced earlier than T5, the latter emerged as the most frequently employed LLM for item generation, being used 33 times. For example, Akyön *et al*. (2022) trained a variant of T5 (mT5) to first extract answers and then used these extracted answers as text input to generate questions as text output. GPT also comes as a popular option, having been used 15 times. GPT-variants predict the next word in a text sequence, enabling prompt-based generation. For instance, Wang *et al*. (2022) compared the performance of GPT-3 in item generation when using different prompts. BERT was used in 9 instances. Given BERT's characteristics, BERT variants were employed to predict masked tokens, creating specific item types such as fill-in-the-blank and cloze questions (e.g., Matsumori *et al*., 2023). In addition, we identified instances where other LLMs, such as BART and PEGASUS, were used.

### 3.1.3. *Post-generation stage*

To ensure the quality and relevance of the generated items for a given educational context, LLMs can be used in the post-generation stage to evaluate, filter, or classify the previously generated items. Among them, evaluation was conducted by calculating specific criteria or

metrics. For instance, Jiao *et al*. (2023) utilized GPT-2 to calculate perplexity values, reflecting the fluency of the generated items. They also employed BERT-large for coherence evaluation and BertScore for assessing creativity by calculating semantic differences among items. Having the evaluation metrics calculated, filtering involves removing low-quality or irrelevant questions or distractors of multiple-choice questions that were generated by the LLM. For example, Offerijns *et al*. (2020) used GPT-2 to generate question-answer pairs and then used BERT to remove the questions that could not be answered or did not make sense. Lastly, classification refers to categorizing questions by type or topic. For example, Nguyen *et al*. (2022) first employed T5 to generate questions and then used GPT-3 to classify these questions based on their utility in learning specific topics.

#### 3.1.4. *Summary of findings for RQ2*

We revealed distinct usage patterns of BERT, GPT, and T5 across the three AIG stages, reflecting the inherent features and strengths of these LLMs. As BERT excels at language understanding, it was predominantly utilized in the pre-generation stage. On the other hand, GPT was primarily used in the item-generation stage, highlighting its strength in text generation. T5 demonstrated its flexibility by being used comparably in all three AIG stages.

### 3.3. RQ3: What Types of Items Were Generated in the Reviewed Studies in Terms of Item Type, Subject Domain, and Language? Can LLMs Generate Valid, Reliable, and High-Quality Items?

#### 3.3.1. *Item type*

We found 39 studies that generated constructed-response items; 29 studies generated selected-response items. Eight studies generated both types. The constructed-response items can be further categorized as Wh-questions, cloze questions, and Fill-in-the-blank questions, whereas selected-response items included True-False and multiple-choice questions, including their distractor generation.

#### 3.3.2. *Subject domain*

Our review revealed that the majority of items generated primarily focus on two subject domains: language learning ($n$ = 24) and general knowledge acquisition ($n$ = 17). General knowledge items are often developed using general datasets, such as the Stanford Question Answering Dataset (SQuAD), which contains passages from Wikipedia and thus does not focus on a specific subject domain. Following closely behind are science-related disciplines ($n$ = 9) such as agronomy, biology, chemistry, physics, and science history. In addition, computer science and mathematics education, had five and four studies on item generation, respectively. We also found studies addressing other subject domains, including social science, medicine, and even literary experiences such as fairy tales.

#### 3.3.3. *Language*

Overall, we identified a total of 12 different languages in the items generated. This finding suggests the potential for generalizability and the wide linguistic context in which LLMs can be utilized for AIG. In most instances, items were generated in English ($n$ = 51), but there were also instances of generation in other languages, including Arabic, Chinese, French, German, Hindi, Indonesian, Korean, Lao, Marathi, Spanish, Swedish, Turkish, and Vietnamese.

#### 3.3.4. *Data source*

LLMs are typically developed for general language processing purposes and often require additional training to effectively perform specific tasks, such as generating context-relevant questions. Such additional training involves training LLMs on new datasets to adapt them to different tasks or content domains. In the reviewed studies, the most commonly utilized datasets for additional training have included SQuAD ($n$ = 24), which encompasses a collection of passages with corresponding reading comprehension questions, and the ReAding

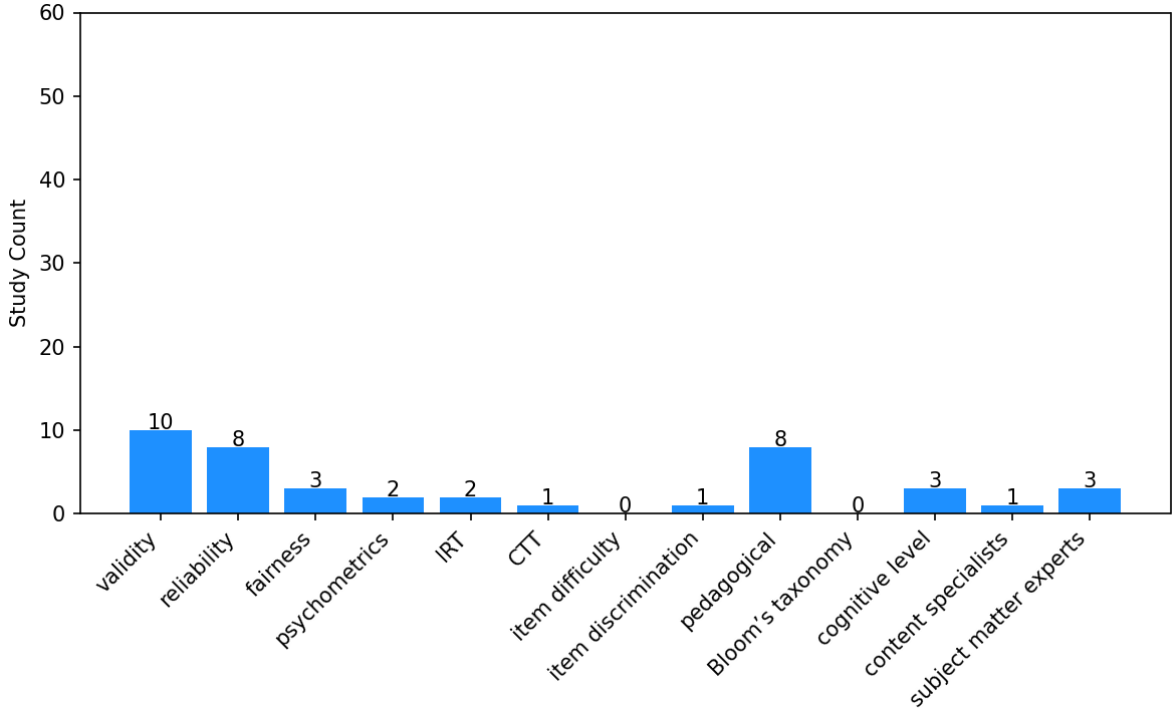
Comprehension dataset from Examinations (RACE) dataset ($n$ = 10), which focuses on English language exams. Additionally, some researchers have crafted their self-collected datasets by aggregating educational materials. These materials consisted of open-access journal articles (e.g., von Davier 2019), questions extracted from online learning platforms or communities (e.g., Stack Overflow; Tsai *et al*., 2021), LLM-generated texts (Bulut & Yildirim-Erbasli, 2022), teacher-created questions (Matsumori *et al*., 2023), stories and fairy tales (e.g., Ghanem *et al*. 2022), knowledge maps (Aigo *et al*., 2021), slides (Chughtai *et al*., 2022), textbooks (e.g., Steuer *et al*. 2020), and other course materials (e.g., Gopal, 2022).

### 3.3.5. *Item properties*

As noted in the methodology section, we did not find many studies reporting the measurement properties of the generated items. Therefore, instead, we searched for keywords pertinent to measurement properties across the 60 reviewed studies. Figure 3 depicts the number of studies containing each keyword. Notably, only 10 out of the 60 studies mentioned "validity". This was followed by "reliability" and "pedagogical", which found their places in eight studies. Other keywords were used even less frequently in the reviewed studies. The infrequent occurrences of these keywords signal a concerning issue: the majority of the reviewed studies seem to neglect measurement properties of items when generating items for educational purposes, which potentially impacts the validity and reliability of the assessment results. Moreover, this could result in generating questions that are too simple and do not require higher cognitive thinking to answer, failing to meet the measurement or pedagogical purposes (e.g., delivering feedback or evaluating achievement).

We further extracted sentences containing the keywords. However, we found that many occurrences of these keywords were not related to the measurement properties of the generated items. For instance, most descriptions of "reliability" were in contexts other than the reliability of items or assessment results. They commonly referred to terms like "inter-annotator reliability" and "the reliability of the data collection". Similarly, the instances of "fairness" were exclusively related to the fairness of experiments comparing model performance, rather than to the fairness of the assessment items themselves. Therefore, the issue of lacking sufficient consideration for the measurement properties of items may be even more severe than it appears in Figure 3.

**Figure 3.** *Number of studies containing each keyword.*



#### 3.3.6. *Summary of findings for RQ3*

We found that LLMs can be an effective and flexible solution to generating a large number of items, with few constraints on item type, language, subject domain, or the data source used for training LLMs to create items. However, we did not find many studies reporting the measurement properties of the generated items.

## 4. DISCUSSION and CONCLUSION

### 4.1. The Current State of Research on AIG

As a summary of the findings from this review, we identified the most commonly used LLMs in the current AIG literature, such as T5, BERT, GPT, and their variants. We described the characteristics and features of each base type of LLM, linking them to their specific uses in the AIG process. LLMs used in the pre-generation stage focus on preparing, processing, and understanding texts for subsequent item generation to ensure high quality. In the post-generation stage, LLMs are primarily used to filter out low-quality items (i.e., mostly focusing on the correctness of grammar and syntax as well as the semantic relevancy and similarity) or determine the usefulness of the generated items.

After reviewing the existing studies, we conclude that LLMs prove useful in generating large banks of items. Additionally, we revealed that LLMs offer a highly flexible solution for AIG, as they have virtually no constraints in terms of item type, language, the subject domain of the items to be generated, or the data source used for further training of LLMs.

### 4.2. Current Research Gaps in the Literature

While the reviewed studies often show that LLMs are effective and flexible in creating a large number of items, we found that many studies applying LLMs in AIG often lacked a solid educational foundation. This might be because many of the authors were NLP researchers who possessed limited recognition and knowledge of learning or measurement theories. Alternatively, it could be because creating high-quality items that are readily usable for educational contexts was not their primary interest or research focus. Accordingly, many of those items are generated without deep consideration of their measurement purposes and item properties, which are essential to meet the requirements of educational assessment.

In traditional item development or template-based AIG, item generation starts with a clear definition of what to measure (i.e., identifying the construct to be measured by considering the expected learner outcomes and instructional objectives), why to measure (i.e., the purpose of assessment), and how to measure (i.e., assessment design and item format), while only a few studies leveraging LLMs for AIG considered these important aspects. For example, many of those generated items in the existing studies do not attempt to evaluate the higher-level cognitive processes specified in Bloom's taxonomy, such as applying, analyzing, evaluating, or creating. This is because they are mostly created by researchers from the AQG area whose original purpose was to develop chatbots and conversational agents. Therefore, their item generation predominantly focuses on the levels of remembering or understanding, similar to the goal of conversational agents, which does not always meet or fit the measurement purposes and goals that are to evaluate the complex learning progress and outcomes of human students.

Moreover, while some studies invited human participants to evaluate the quality of the items after being generated, only a few involved SMEs or measurement specialists in the AIG process. However, human experts are crucial for guiding item development, ensuring quality, and potentially enhancing students' learning outcomes through formative use (e.g., Gierl & Lai, 2012; Lu *et al*., 2021). The absence of expert guidance has led many existing AIG studies using LLMs to overlook important measurement properties such as item difficulty and item discrimination. Thus, the current literature mostly provides evidence that LLMs can be leveraged to generate a large number of items, but little is known about whether these items possess the high quality necessary for educational purposes such as pedagogical teaching and

assessment. Considering the goal of AIG, which is to generate large banks of high-quality items while reducing overall costs (Gierl & Lai, 2012; Lai *et al*., 2009), we argue that a thorough item evaluation after generation is missing in the current literature. That is, AIG does not end up generating a large number of items but ensuring that these items are of high quality and can fulfill their demands for educational purposes and contexts. Thus, we argue that AIG is still a developing and promising research domain, and its further development and application will depend on the involvement of human experts and the integration of learning and measurement theories.

### 4.3. Suggestions for Future Research

The current research gaps, particularly the lack of educational foundations in the reviewed AIG studies, led us to advocate for two suggestions for future research in this area.

First, future research should prioritize clarifying the assessment context and measurement goals in AIG applications. While AIG benefits from advancements in NLP, such as LLMs, it differs from general text generation by creating items that meet the specific purposes of assessments in real educational contexts. Newton (2007) distinguished 18 different educational assessment applications, such as formative, diagnostic, qualification certification, and comparative purposes. For example, formative assessments take place during students' learning processes and support learning by providing effective feedback, while summative assessments aim to collect information about students' learning outcomes after the learning process has occurred. As the assessment purposes differ, the desired characteristics of the generated items also differ. When generating items for a specific educational and assessment context, researchers must identify the nature and desired characteristics of these items. For instance, in the case of formative assessments aimed at aiding students' learning, items should prioritize pedagogical value by providing feedback to help identify misconceptions. To assess students' learning outcomes following a teaching program, items should be balanced and cover a broad range of key concepts taught, rather than disproportionately focusing on concepts within a narrow content area. Furthermore, regardless of whether items are formative or summative or created for other assessment purposes, they must demonstrate high quality, with strong reliability and validity, to be effective, because they directly impact the feedback provided to students or the ability to draw inferences about students' learning progress and outcomes.

Second, we recommend evaluating both the measurement properties and pedagogical soundness of generated items as an essential step in AIG. From a practical standpoint, item development does not end with item generation; the evaluation of item quality is crucial to ensure usability in educational contexts. For example, measurement properties of items (e.g., difficulty and discrimination parameters) and tests (e.g., reliability and validity) can be examined using measurement theories such as classical test theory and item response theory. Finally, after items are generated and evaluated for their measurement properties, the focus should shift back to meeting the intended measurement purposes or pedagogical value of assessments. Educators and SMEs can assess the educational or pedagogical value of automatically generated items. It is important to ensure that the generated items can effectively serve their intended purpose. If not, it is necessary to revisit previous item development stages for revisions to create items that better align with the learning objectives and measurement goals.

The three-stage AIG framework emerging from our analysis (i.e., pre-generation stage, item generation stage, post-generation stage) offers a structured approach for integrating these two recommendations into practice. This framework not only categorizes LLM applications but also has a potential for seamlessly integrating human expertise for a human-in-the-loop approach to strengthen educational foundations of AIG practices. In the pre-generation stage, which focuses on preparation tasks before item generation, AIG practitioners can enhance outcomes by incorporating human expertise. For example, they may define assessment goals and formats or

contribute to model fine-tuning by providing context-relevant educational materials (e.g., item banks). Such practices have been shown to improve both the model's contextual awareness and task-specific performance (Ghanem *et al*., 2022; Ratcheva *et al*., 2022). During the item generation stage, human experts (e.g., educators) can refine prompts to elicit more appropriate items aligned with specific learning or assessment goals. Finally, in the post-generation stage, subject-matter experts, students (as target test-takers), and even LLMs themselves may participate in evaluating the generated items.

While not explicitly framed in these terms, Sayin and Gierl's (2024) study demonstrates strong alignment with this three-stage approach when analyzed through our conceptual framework. Their methodology began with what we would characterize as pre-generation preparation, involving the development of three foundational models: a cognitive model to define target knowledge and skills, an item model to establish construction guidelines, and a specialized text analysis model to ensure passage coherence and plausible distractors. These models then informed precise prompt engineering for GPT-3.5, enabling systematic generation of 12,500 grade-level items for high-stakes examinations - a process corresponding to our item generation stage. Finally, their comprehensive quality assurance procedures, including expert reviews and empirical field testing to assess critical psychometric properties like item difficulty and discrimination indices, exemplify what our framework identifies as post-generation validation.

### 4.4. Limitations

We did not conduct a formal assessment of the methodological quality of the included studies, which is generally considered a limitation of scoping reviews. This is because the primary purpose of this study, and of scoping reviews in general, is to map the existing literature on a particular topic to explore the range, nature, and extent of research activities related to the topic, rather than to evaluate the quality of the studies (Arksey & O'Malley, 2005). Additionally, we acknowledge that the field of leveraging LLMs is rapidly evolving. As we completed our keyword search, new LLMs have been introduced, boasting significantly larger parameter sizes and enhanced functionalities. This rapid development has not only expanded the utility and potential of LLMs but has also made advanced technology more accessible to users with varying levels of programming skills. Consequently, the use of LLMs in education continues to evolve, and understanding how these models can best serve the AIG process remains an ongoing journey.

### 4.5. Conclusion

As a category of generative artificial intelligence technology, LLMs focus on language understanding and text generation, which has sparked researchers' interest in using them to automatically generate questions for educational purposes. Through the mapping and synthesis of the reviewed studies on leveraging LLMs for AIG, we identified that the most commonly used LLMs are T5, BERT, GPT, and their variants. We have also categorized the current applications of LLMs into three stages of the AIG process: pre-generation, item generation, and post-generation. The findings reveal that using LLMs to generate items is an effective and flexible solution, with few constraints on item type, language, subject domain, or the data source used for training LLMs to create items. Due to the exceptional language understanding abilities of LLMs, the generated items are typically free from grammar errors and contextually relevant to the desired content domain. However, we also noted a lack of a solid educational foundation in many of the reviewed studies, as they did not incorporate learning and measurement theories into the item generation process. We attribute this issue to the absence of involvement of human experts such as SMEs and measurement specialists. Although one part of the goal of AIG was to reduce financial costs and human burdens, it was never meant to exclude human involvement from the item development process. Importantly, AIG is still considered an augmented intelligence approach, which means it requires both human expertise and the capabilities of a computer. Considering the goal of AIG, we not only want to generate a large number of items

but also care more about the item quality and characteristics. Hence, future researchers should consider enhancing the educational foundation in the AIG process. The three-stage framework proposed in this review provides a structured approach to integrating human expertise at each stage—pre-generation, item generation, and post-generation—to ensure high-quality item generation. This way, LLMs can be leveraged to produce items that are not only grammatically correct and contextually relevant but also reliable, valid, and pedagogically sound.

## Acknowledgments

This work has been previously presented at the 2024 National Council on Measurement in Education (NCME) Annual Meeting, Philadelphia, Pennsylvania.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

## Contribution of Authors

**Bin Tan:** Conceptualization, methodology, literature search, article screening, formal analysis and investigation, writing – original draft preparation, writing – review and editing. **Nour Armoush:** Conceptualization, methodology, literature search, article screening, formal analysis and investigation, writing – original draft preparation. **Elisabetta Mazzullo:** Formal analysis and investigation, writing – original draft preparation. **Okan Bulut:** Conceptualization, methodology, writing – review and editing, supervision. **Mark Gierl:** Conceptualization, methodology, writing – review and editing, supervision.

## Orcid

Bin Tan https://orcid.org/0000-0001-6717-5620
Nour Armoush https://orcid.org/0009-0008-2310-5098
Elisabetta Mazzullo https://orcid.org/0009-0008-4847-9934
Okan Bulut https://orcid.org/0000-0001-5853-1267
Mark J. Gierl https://orcid.org/0000-0001-6717-5620

## REFERENCES

Ackerman, R., & Balyan, R. (2023). Automatic multilingual question generation for health data using LLMs. In International *Conference on AI-generated Content* (pp. 1-11). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-99-7587-7_1

Agrawal, A., & Shukla, P. (2023). Context aware automatic subjective and objective question generation using Fast Text to text transfer learning. *International Journal of Advanced Computer Science and Applications*, *14*(4), 456-463.

Aigo, K., Tsunakawa, T., Nishida, M., & Nishimura, M. (2021). Question generation using knowledge graphs with the T5 language model and masked self-attention. In *2021 IEEE 10th Global Conference on Consumer Electronics* (pp. 85-87). IEEE. https://doi.org/10.1109/GCCE53005.2021.9621874

Akyön, F.Ç., Cavusoglu, A.D.E., Cengiz, C., Altinuç, S.O., & Temizel, A. (2022). Automated question generation and question answering from Turkish texts. *Turkish Journal of Electrical Engineering and Computer Sciences*, *30*(5), 1931-1940. https://doi.org/10.55730/1300-0632.3914

Alsubait, T., Parsia, B., & Sattler, U. (2016). Ontology-based multiple choice question generation. *KI-Künstliche Intelligenz*, *30*, 183-188. https://doi.org/10.1007/s13218-015-0405-9

Alves, C.B., Gierl, M.J., & Lai, H. (2010, April). *Using automated item generation to promote test design and development* [Paper presentation]. American Educational Research Association Annual Meeting, Denver, CO, United States.

Arksey, H., & O'malley, L. (2005). Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology*, *8*(1), 19-32. https://doi.org/10.1080/1364557032000119616

Attali, Y., Runge, A., LaFlair, G.T., Yancey, K., Goodwin, S., Park, Y., & Von Davier, A.A. (2022). The interactive reading task: Transformer-based automatic item generation. Frontiers *in Artificial Intelligence*, *5*, 903077. https://doi.org/10.3389/frai.2022.903077

Berger, G., Rischewski, T., Chiruzzo, L., & Rosá, A. (2022). Generation of English question answer exercises from texts using transformers-based models. In *2022 IEEE Latin American Conference on Computational Intelligence* (pp. 1-5). IEEE. https://doi.org/10.1109/LA-CCI54402.2022.9981171

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7-74. https://doi.org/10.1080/0969595980050102

Bulathwela, S., Muse, H., & Yilmaz, E. (2023). Scalable educational question generation with pre-trained language models. In International *Conference on Artificial Intelligence in Education* (pp. 327-339). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_27

Bulut, O., & Yildirim-Erbasli, S.N. (2022). Automatic story and item generation for reading comprehension assessments with transformers. *International Journal of Assessment Tools in Education*, *9*(Special Issue), 72-87. https://doi.org/10.21449/ijate.1124382

Bulut, O., Gorgun, G., Yildirim-Erbasli, S.N., Wongvorachan, T., Daniels, L.M., Gao, Y., ... & Shin, J. (2023). Standing on the shoulders of giants: Online formative assessments as the foundation for predictive learning analytics models. *British Journal of Educational Technology*, *54*(1), 19-39. https://doi.org/10.1111/bjet.13276

Ch, D.R., & Saha, S.K. (2018). Automatic multiple choice question generation from text: A survey. *IEEE Transactions on Learning Technologies*, *13*(1), 14-25. https://doi.org/10.1109/TLT.2018.2889100

Chiang, S.H., Wang, S.C., & Fan, Y.C. (2024). Cdgp: Automatic cloze distractor generation based on pre-trained language model. *arXiv preprint arXiv:2403.10326*. https://doi.org/10.18653/v1/2022.findings-emnlp.429

Chughtai, R., Azam, F., Anwar, M.W., But, W.H., & Farooq, M.U. (2022). A lecture centric automated distractor generation for post-graduate software engineering courses. In *2022 International Conference on Frontiers of* Information *Technology (FIT)* (pp. 100-105). IEEE. https://doi.org/10.1109/FIT57066.2022.00028

Chung, H.L., Chan, Y.H., & Fan, Y.C. (2020). A BERT-based distractor generation scheme with multi-tasking and negative answer training strategies. *arXiv preprint arXiv:2010.05384*. https://arxiv.org/abs/2010.05384

Dalby, D., & Swan, M. (2019). Using digital technology to enhance formative assessment in mathematics classrooms. *British Journal of Educational Technology*, *50*(2), 832-845. https://doi.org/10.1111/bjet.12606

Dembitzer, L., Zelikovitz, S., & Kettler, R.J. (2017). Designing computer-based assessments: Multidisciplinary findings and student perspectives. *International Journal of Educational Technology*, *4*(3), 20-31. https://educationaltechnology.net/ijet/index.php/ijet/article/view/47

Desai, T. (2021). *Discourse parsing and its application to question generation* [Unpublished dissertation]. The University of Texas at Dallas.

Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://doi.org/10.48550/arXiv.1810.04805

Dijkstra, R., Genç, Z., Kayal, S., & Kamps, J. (2022). Reading comprehension quiz generation using generative pre-trained transformers. In S. Sosnovsky, P. Brusilovsky, & A. Lan (Eds.), *Proceedings of the Fourth International Workshop on Intelligent Textbooks 2022* (pp. 4–7). CEUR-WS. http://ceur-ws.org/Vol-3192/

Drori, I., Zhang, S., Shuttleworth, R., Tang, L., Lu, A., Ke, E., ... & Strang, G. (2022). A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. *Proceedings of the National Academy of Sciences*, *119*(32), e2123433119. https://doi.org/10.1073/pnas.2123433119

Falcão, F., Costa, P., & Pêgo, J.M. (2022). Feasibility assurance: a review of automatic item generation in medical assessment. *Advances in Health Sciences Education, 27*(2), 405-425. https://doi.org/10.1007/s10459-022-10092-z

Femi, J.G., & Nayak, A.K. (2022). EQGTL: An Ensemble Model for Relevant Question Generation using Transfer Learning. In *2022 International Conference on Machine Learning, Computer Systems and Security* (pp. 253-258). IEEE. https://doi.org/10.1109/MLCSS57186.2022.00054

Fuadi, M., & Wibawa, A.D. (2022). Automatic question generation from indonesian texts using text-to-text transformers. In *2022 International Conference on Electrical and Information Technology (IEIT)* (pp. 84-89). IEEE. https://doi.org/10.1109/IEIT56384.2022.9967858

Fung, Y.C., Kwok, J.C.W., Lee, L.K., Chui, K.T., & U, L.H. (2020). Automatic question generation system for english reading comprehension. In *Technology in Education. Innovations for Online Teaching and Learning: 5th International Conference, ICTE 2020, Macau, China, August 19-22, 2020, Revised Selected Papers 5* (pp. 136-146). Springer Singapore. https://doi.org/10.1007/978-981-33-4594-2_12

Fung, Y.C., Lee, L.K., & Chui, K.T. (2023). An automatic question generator for Chinese comprehension. *Inventions*, *8*(1), 31. https://doi.org/10.3390/inventions8010031

Ghanem, B., Coleman, L.L., Dexter, J.R., von der Ohe, S.M., & Fyshe, A. (2022). Question generation for reading comprehension assessment by modeling how and what to ask. *arXiv preprint arXiv:2204.02908*. https://doi.org/10.48550/arXiv.2204.02908

Gierl, M.J., & Lai, H. (2012). The role of item models in automatic item generation. International *Journal of* Testing, *12*(3), 273-298. https://doi.org/10.1080/15305058.2011.635830

Gierl, M.J., & Lai, H. (2015). Automatic item generation. In *Handbook of test development* (pp. 410-429). Routledge.

Gierl, M.J., & Lai, H. (2016). A process for reviewing and evaluating generated test items. *Educational Measurement:* Issues *and Practice, 35*(4), 6-20. https://doi.org/10.1111/emip.12129

Gierl, M.J., Lai, H., & Tanygin, V. (2021). *Advanced methods in automatic item generation*. Routledge.

Godslove, J.F., & Nayak, A.K. (2023). Generative model for formulating relevant questions and answers using transfer learning. In *AIP Conference Proceedings* (Vol. 2819, No. 1). AIP Publishing. https://doi.org/10.1063/5.0136892

Gopal, A. (2022). Automatic question generation for Hindi and Marathi. In *2022 International Conference on Advanced Learning Technologies (ICALT)* (pp. 19-21). IEEE. https://doi.org/10.1109/ICALT55010.2022.00012

Goyal, R., Kumar, P., & Singh, V.P. (2023). Automated question and answer generation from texts using text-to-text transformers. *Arabian Journal for Science and Engineering*, 1-15. https://doi.org/10.1007/s13369-023-07840-7

Granić, A. (2022). Educational technology adoption: A systematic review. *Education and Information Technologies, 27*(7), 9725-9744. https://doi.org/10.1007/s10639-022-10951-7

Grover, K., Kaur, K., Tiwari, K., Rupali, & Kumar, P. (2021). Deep learning based question generation using t5 transformer. In *Advanced Computing: 10th International Conference, IACC 2020, Panaji, Goa, India, December 5–6, 2020, Revised Selected Papers, Part I 10* (pp. 243-255). Springer Singapore. https://doi.org/10.1007/978-981-16-0401-0_18

Han, Z. (2022). *Unsupervised multilingual distractor generation for fill-in-the-blank questions* [Unpublished thesis]. Uppsala University.

Jiao, Y., Shridhar, K., Cui, P., Zhou, W., & Sachan, M. (2023). Automatic educational question generation with difficulty level controls. In *International Conference on Artificial Intelligence in Education* (pp. 476-488). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36272-9_39

Kalpakchi, D., & Boye, J. (2021). BERT-based distractor generation for Swedish reading comprehension questions using a small-scale dataset. *arXiv preprint arXiv:2108.03973*. https://doi.org/10.48550/arXiv.2108.03973

Kasakowskij, R., Kasakowskij, T., & Seidel, N. (2022). Generation of multiple true false questions. 20. Fachtagung Bildungstechnologien. https://doi.org/10.18420/delfi2022-026

Khandait, K., Bhura, S., & Asole, S.S. (2022). Automatic question generation through word vector synchronization using lamma. *Indian Journal of Computer Science and Engineering*, *13*(4), 1083-1095. https://doi.org/10.21817/indjcse/2022/v13i4/221304046

Kosh, A.E., Simpson, M.A., Bickel, L., Kellogg, M., & Sanford-Moore, E. (2019). A cost–benefit analysis of automatic item generation. *Educational Measurement: Issues and Practice*, *38*(1), 48-53. https://doi.org/10.1111/emip.12237

Kumar, A., Kharadi, A., Singh, D., & Kumari, M. (2021). Automatic question-answer pair generation using deep learning. In *2021 Third International Conference on Inventive Research in Computing Applications* (pp. 794-799). IEEE. https://doi.org/10.1109/ICIRCA51532.2021.9544654

Kumar, N.S., Mali, R., Ratnam, A., Kurpad, V., & Magapu, H. (2022). Identification and addressal of knowledge gaps in students. In *2022 3rd International Conference for Emerging Technology* (pp. 1-6). IEEE. https://doi.org/10.1109/INCET54531.2022.9824483

Kumar, S., Chauhan, A., & Kumar C, P. (2022). Learning enhancement using question-answer generation for e-book using contrastive fine-tuned T5. In *International Conference on Big Data Analytics* (pp. 68-87). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-24094-2_5

Kumari, V., Keshari, S., Sharma, Y., & Goel, L. (2022). Context-based question answering system with suggested questions. In *2022 12th International Conference on Cloud Computing, Data Science & Engineering* (pp. 368-373). IEEE. https://doi.org/10.1109/Confluence52989.2022.9734207

Kuo, C.Y., & Wu, H.K. (2013). Toward an integrated model for designing assessment systems: An analysis of the current status of computer-based assessments in science. *Computers & Education, 68*, 388-403. https://doi.org/10.1016/j.compedu.2013.06.002

Kurdi, G., Leo, J., Parsia, B., Sattler, U., & Al-Emari, S. (2020). A systematic review of automatic question generation for educational purposes. *International Journal of Artificial Intelligence in Education*, *30*, 121-204. https://doi.org/10.1007/s40593-019-00186-y

Lai, H., Alves, C., & Gierl, M.J. (2009). Using automatic item generation to address item demands for CAT. In D.J. Weiss (Ed.), Proceedings of the *2009 GMAC Conference on Computerized Adaptive Testing*. www.psych.umn.edu/psylabs/CATCentral/

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2019). Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942*. https://doi.org/10.48550/arXiv.1909.11942

Lee, H., Chung, H.Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, *33*(2), 124-140. https://doi.org/10.1080/08957347.2020.1732383

Lim, Y.S. (2019). Students' perception of formative assessment as an instructional tool in medical education. *Medical Science Educator*, *29*(1), 255-263. https://doi.org/10.1007/s40670-018-00687-w

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*. https://doi.org/10.48550/arXiv.1907.11692

Lu, O.H.T., Huang, A.Y.Q., Tsai, D.C.L., & Yang, S.J.H. (2021). Expert-authored and machine-generated short- answer questions for assessing students' learning performance. *Educational Technology & Society*, *24*(3), 159–173. https://www.jstor.org/stable/27032863

Maheen, F., Asif, M., Ahmad, H., Ahmad, S., Alturise, F., Asiry, O., & Ghadi, Y.Y. (2022). Automatic computer science domain multiple-choice questions generation based on informative sentences. *PeerJ Computer* Science, *8*, e1010. https://doi.org/10.7717/peerj-cs.1010

Malhar, A., Sawant, P., Chhadva, Y., & Kurhade, S. (2022). Deep learning-based Answering Questions using T5 and Structured Question Generation System. In *2022 6th International Conference on Intelligent Computing and Control Systems* (pp. 1544-1549). IEEE. https://doi.org/10.1109/ICICCS53718.2022.9788264

Mathur, A., & Suchithra, M. (2022). Application of abstractive summarization in multiple choice question generation. In *2022 International Conference on Computational Intelligence and Sustainable Engineering Solutions* (pp. 409-413). IEEE. https://doi.org/10.1109/CISES54857.2022.9844396

Matsumori, S., Okuoka, K., Shibata, R., Inoue, M., Fukuchi, Y., & Imai, M. (2023). Mask and cloze: Automatic open cloze question generation using a masked language model. *IEEE Access*, *11*, 9835-9850. https://doi.org/10.1109/ACCESS.2023.3239005

Maurya, K.K., & Desarkar, M.S. (2020). Learning to distract: A hierarchical multi-decoder network for automated generation of long distractors for multiple-choice questions for reading comprehension. In *Proceedings of* the *29th ACM international conference on information & knowledge management* (pp. 1115-1124). https://doi.org/10.1145/3340531.3411997

Mazzullo, E., Bulut, O., Wongvorachan, T., & Tan, B. (2023). Learning Analytics in the Era of Large Language Models. *Analytics*, *2*(4), 877-898. https://doi.org/10.3390/analytics2040046

Min, B., Ross, H., Sulem, E., Veyseh, A.P.B., Nguyen, T.H., Sainz, O., ... & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, *56*(2), 1-40. https://doi.org/10.1145/3605943

Muse, H., Bulathwela, S., & Yilmaz, E. (2022). Pre-training with scientific text improves educational question generation. *arXiv preprint arXiv:2212.03869*. https://doi.org/10.48550/arXiv.2212.03869

Newton, P.E. (2007). Clarifying the purposes of educational assessment. *Assessment in education*, *14*(2), 149-170. https://doi.org/10.1080/09695940701478321

Nguyen, H.A., Bhat, S., Moore, S., Bier, N., & Stamper, J. (2022). Towards generalized methods for automatic question generation in educational domains. In *European conference on technology enhanced learning* (pp. 272-284). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-16290-9_20

Nittala, S., Agarwal, P., Vishnu, R., & Shanbhag, S. (2022). Speaker Diarization and BERT-Based Model for Question Set Generation from Video Lectures. In *Information and Communication Technology for Competitive Strategies ICT: Applications and Social Interfaces* (pp. 441-452). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-19-0095-2_42

Offerijns, J., Verberne, S., & Verhoef, T. (2020). Better distractions: Transformer-based distractor generation and multiple-choice question filtering. *arXiv preprint arXiv:2010.09598*. https://doi.org/10.48550/arXiv.2010.09598

Pochiraju, D., Chakilam, A., Betham, P., Chimulla, P., & Rao, S.G. (2023). Extractive summarization and multiple-choice question generation using XLNet. In *2023 7th International Conference on Intelligent Computing and Control Systems* (pp. 1001-1005). IEEE. https://doi.org/10.1109/ICICCS56967.2023.10142220

Pugh, D., De Champlain, A., Gierl, M., Lai, H., & Touchie, C. (2020). Can automated item generation be used to develop high quality MCQs that assess application of knowledge?.

*Research and Practice in Technology* Enhanced *Learning*, *15*, 1-13. https://doi.org/10.1186/s41039-020-00134-8

Qiu, X., Xue, H., Liang, L., Xie, Z., Liao, S., & Shi, G. (2021). Automatic generation of multiple-choice cloze-test questions for lao language learning. In *2021 International Conference on Asian Language Processing* (pp. 125-130). IEEE. https://doi.org/10.1109/IALP54817.2021.9675153

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, *21*(140), 1-67. https://doi.org/10.48550/arXiv.1910.10683

Raina, V., & Gales, M. (2022). Multiple-choice question generation: Towards an automated assessment framework. *arXiv preprint arXiv:2209.11830*. https://doi.org/10.48550/arXiv.2209.11830

Ratcheva, M.G., Navale, R., & Desai, B.C. (2022). An online MCQ sub-system for CrsMgr. In *Proceedings of the 26th International Database Engineered Applications Symposium* (pp. 128-133). https://doi.org/10.1145/3548785.3548789

Rodriguez-Torrealba, R., Garcia-Lopez, E., & Garcia-Cabot, A. (2022). End-to-end generation of multiple-xhoice questions using text-to-text transfer transformer models. *Expert Systems with Applications*, *208*, 118258. https://doi.org/10.1016/j.eswa.2022.118258

Rudner, L.M. (2009). Implementing the graduate management admission test computerized adaptive test. In *Elements of adaptive testing* (pp. 151-165). Springer New York. https://doi.org/10.1007/978-0-387-85461-8_8

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*. https://doi.org/10.48550/arXiv.1910.01108

Sayin, A., & Gierl, M. (2024). Using OpenAI GPT to generate reading comprehension items. *Educational Measurement: Issues and Practice*, *43*(1), 5-18. https://doi.org/10.1111/emip.12590

Shan, J., Nishihara, Y., Maeda, A., & Yamanishi, R. (2022). Question generation for reading comprehension test complying with types of question. *Journal of Information Science & Engineering*, *38*(3). https://doi.org/10.6688/JISE.202205_38(3).0005

Shan, J., Nishihara, Y., Yamanishi, R., & Maeda, A. (2019). Question generation for reading comprehension of language learning test: A method using Seq2Seq approach with transformer model. In *2019 International Conference on Technologies and Applications of Artificial Intelligence* (pp. 1-6). IEEE. https://doi.org/10.1109/TAAI48200.2019.8959903

Shridhar, K., Macina, J., El-Assady, M., Sinha, T., Kapur, M., & Sachan, M. (2022). Automatic generation of socratic subquestions for teaching math word problems. *arXiv preprint arXiv:2211.12835*. https://doi.org/10.48550/arXiv.2211.12835

Singh, J., McCann, B., Socher, R., & Xiong, C. (2019). BERT is not an interlingua and the bias of tokenization. In *Proceedings of the 2nd Workshop on Deep Learning Approaches for Low-Resource NLP (DeepLo 2019)* (pp. 47-55). https://doi.org/10.18653/v1/D19-6106

Spector, J.M., Ifenthaler, D., Samspon, D., Yang, L., Mukama, E., Warusavitarana, A., … Gibson, D.C. (2016). Technology enhanced formative assessment for 21st century learning. *Educational Technology & Society, 19*(3), 58-71. https://www.jstor.org/stable/jeductechsoci.19.3.58

Srivastava, M., & Goodman, N. (2021). Question generation for adaptive education. *arXiv preprint arXiv:2106.*04262. https://doi.org/10.48550/arXiv.2106.04262

Steuer, T., Filighera, A., & Rensing, C. (2020). Exploring artificial jabbering for automatic text comprehension question generation. In *Addressing Global Challenges and Quality Education: 15th European Conference on Technology Enhanced Learning, EC-TEL 2020, Heidelberg, Germany, September 14–18, 2020, Proceedings 15* (pp. 1-14). Springer International Publishing. https://doi.org/10.1007/978-3-030-57717-9_1

Tsai, D.C., Chang, W., & Yang, S. (2021). Short answer questions generation by Fine-Tuning BERT and GPT-2. In *Proceedings of the 29th International Conference on Computers in Education Conference* (Vol. 64). https://icce2021.apsce.net/wp-content/uploads/2021/12/ICCE2021-Vol.II-PP.-508-514.pdf

von Davier, M. (2019). Training Optimus prime, MD: Generating medical certification items by fine-tuning OpenAI's gpt2 transformer model. *arXiv preprint arXiv:1908.08594*. https://doi.org/10.48550/arXiv.1908.08594

Vu, N., & Van Nguyen, K. (2022). Enhancing Vietnamese question generation with reinforcement learning. In *Asian Conference on Intelligent Information and Database Systems* (pp. 559-570). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-21743-2_45

Wang, B., Yao, T., Chen, W., Xu, J., & Wang, X. (2021). Multi-lingual question generation with language agnostic language model. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (pp. 2262-2272). https://aclanthology.org/2021.findings-acl.199.pdf

Wang, H.C., Maslim, M., & Kan, C.H. (2023). A question–answer generation system for an asynchronous distance learning platform. *Education and Information Technologies*, *28*(9), 12059-12088. https://doi.org/10.1007/s10639-023-11675-y

Wang, Z., Valdez, J., Basu Mallick, D., & Baraniuk, R.G. (2022). Towards human-like educational question generation with large language models. In *International conference on artificial intelligence in education* (pp. 153-166). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-11644-5_13

Wylie, E.C., & Lyon, C.J. (2015). The fidelity of formative assessment implementation: Issues of breadth and quality. *Assessment in Education: Principles, Policy & Practice*, *22*(1), 140-160. https://doi.org/10.1080/0969594X.2014.990416

Xie, J., Peng, N., Cai, Y., Wang, T., & Huang, Q. (2021). Diverse distractor generation for constructing high-quality multiple choice questions. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *30*, 280-291. https://doi.org/10.1109/TASLP.2021.3138706

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R.R., & Le, Q.V. (2019). Xlnet: Generalized autoregressive pretraining for language understanding. *Advances in neural information processing systems*, *32*. https://dl.acm.org/doi/10.5555/3454287.3454804

Yen, Y.-C., Ho, R.-G., Liao, W.-W., & Chen, L.-J. (2012). Reducing the impact of inappropriate items on reviewable computerized adaptive testing. *Educational Technology & Society*, *15*(2), 231–243. https://www.jstor.org/stable/jeductechsoci.15.2.231

Zhang, C. (2023). Automatic Generation of Multiple-Choice Questions. *arXiv preprint arXiv: 2303.*14576v1. https://doi.org/10.48550/arXiv.2303.14576

Zhao, W.X., Zhou, K., Li, J., Tang, T., Wang, X., Hou, Y., ... & Wen, J.R. (2023). A survey of large language models. *arXiv preprint arXiv:2303.18223*. https://doi.org/10.48550/arXiv.2303.18223

Zhao, Z., Hou, Y., Wang, D., Yu, M., Liu, C., & Ma, X. (2022). Educational question generation of children storybooks via question type distribution learning and event-centric summarization. *arXiv preprint arXiv:2203.14187*. https://doi.org/10.48550/arXiv.2203.14187

## APPENDIX A

*Summary of the included studies*

| Authors | Venue | Description of the use of LLMs |
|---|---|---|
| Agrawal & Shukla, 2023 | International Journal of Advanced Computer Science and Applications | T5 was trained to generate questions using texts and answers together as input. |
| Aigo *et al*., 2021 | Global Conference on Consumer Electronics | T5 was used for question generation based on knowledge graphs. |
| Akyön *et al*., 2022 | Turkish Journal of Electrical Engineering and Computer Sciences | mT5 was trained to extract answers first, then use these as input for generating questions. Other models were used for comparisons. |
| Attali *et al*., 2022 | Frontiers in Artificial Intelligence | GPT-3 was first used to generate question generation instructions, which were then used to generate associated questions and answers. |
| Berger *et al*., 2022 | Latin American Conference on Computational Intelligence | T5-small was used to generate questions with text-answer pairs as input. |
| Bulathwela *et al*., 2023 | arxiv | T5 was used to generate questions with raw question-answer pair texts as input |
| Bulut & Yildirim-Erbasli, 2022 | International Journal of Assessment Tools in Education | GPT was used to predict the next word for generating reading stories with two prompts, while T5 was utilized for text-to-text generation with texts as input to generate questions. |
| Chiang *et al*., 2022 | Empirical Methods in Natural Language Processing (Conference) | BERT was used with a QA pair as input and a distractor candidate as output, which were then filtered by a text embedding model. |
| Chughtai *et al*., 2022 | International Conference on Frontiers of Information Technology | T5 was used to tokenize and summarize texts, and to generate question-answer pairs using contexts and paragraphs as input. |
| Chung *et al*., 2020 | arxiv | BERT was employed to iteratively predict the next token of distractors based on the context, question, correct answer, and previously predicted distractor tokens. GPT was used as a baseline model. |
| Desai, 2021 | Doctoral dissertation | BERT was used for discourse segmentation, textual representation, and discourse parsing (identifying connections of tokens). However, the question generation process still relies on template-based methods with masked tokens/words. |
| Dijkstra *et al*., 2022 | International Workshop on Intelligent Textbooks | Two prompts were compared for GPT to generate questions: (1) Questions were generated based on the context first and then combined with the generated answers to create distractors. (2) GPT directly generated the quiz based on the context alone. |
| Drori *et al*., 2022 | arxiv | The LLMs were used to generate questions using two prompting strategies: one shot vs. few shots examples. |
| Femi & Nayak, 2022 | International Conference on Machine Learning, Computer Systems and Security | BERT was initially used to make sense of the keywords generated by WordNet and to put the keywords in context, followed by T5 generating questions, answers, and distractors. |
| Fuadi & Wibawa, 2022 | International Conference on Electrical and Information Technology | T5 was firstly used to generate answers based on contexts, and these answers, along with the contexts, were then used to generate questions. |

| | | |
|---|---|---|
| Fung *et al*., 2020 | International Conference on Technology in Education | T5 was trained to generate questions using a source sentence with an answer phrase as input. |
| Fung *et al*., 2023 | Inventions | T5 was used for question generation with contexts and answers as input. The generated questions were compared with existing questions for feedback, which fed into the next iteration of question generation. |
| Ghanem *et al*., 2022 | arxiv | T5 was trained to teach the model how to ask questions (a skill label), then used to generate questions with a story as input and QA pairs as output. |
| Godslove & Nayak, 2023 | International Conference on Applied Mathematics in Science and Engineering | BertWSD was used for word sense disambiguation, while T5 was utilized to generate questions, answers, and distractors with processed texts as input (including text summarization and extraction). |
| Gopal, 2022 | International Conference on Advanced Learning Technologies | T5 and GPT-2 were used to generate texts with input texts. |
| Goyal *et al*., 2023 | Arabian Journal for Science and Engineering | T5 was used to identify answers and was trained to generate answers with texts as input, which were then combined with these answers to generate questions. |
| Grover *et al*., 2021 | International Advanced Computing Conference | T5 was trained to generate multiple questions by providing context paragraphs. |
| Han, 2022 | Master's thesis | BERT variants and BART variants were used to generate foreign language distractors following the masked token prediction method. The LLMs were also used to generate distractors. |
| Jiao *et al*., 2023 | International Conference on Artificial Intelligence in Education | T5, BERT, and GPT were compared for AIG tasks. In addition, GPT-2 was used to calculate perplexity values, indicating the fluency of the generated items. BERT-large was also used to evaluate coherence. BertScore was employed to calculate semantic differences in order to assess creativity. |
| Kalpakchi & Boye, 2021 | arxiv | BERT was used to generate questions with question, context, and correct answers as input in two modes: sequential and non-ordered. |
| Kasakowskij *et al*., 2022 | Fachtagung Bildungstechnologien (DELFI) | GPT-2 was used to generate false statements for true/false questions based on the extracted correct statements from texts. |
| Khandait *et al*., 2022 | Indian Journal of Computer Science and Engineering | T5 models received input paragraphs to generate questions. Other LLMs were used as baseline models without detailed information for implementation. |
| Kumar *et al*., 2021 | International Conference on Inventive Research in Computing Applications | First, T5 was employed to extract answers from given passages. Next, T5 was used to generate questions with passages and answers as inputs. |
| Kumar *et al*., 2022 | International Conference for Emerging Technology | T5 was used to summarize texts and to generate single-line questions with the candidate sentences as input and QA pairs as outputs. GPT-2 was used to generate true/false questions. Alternative sentences were firstly generated by GPT-2, followed by using Sentence BERT to filter out alternative sentences that are too similar to the candidate sentences. Sentence BERT was also used to calculate the similarity scores of distractors for question filtering. |
| Kumar *et al*., | International Conference on | T5 was used to summarize the raw texts and generate question-answer pairs. BERT is used to rank the top |

| | | |
|---|---|---|
| 2022 | Big Data Analytics | question-answer pairs. |
| Kumari *et al.*, 2022 | International Conference on Cloud Computing, Data Science & Engineering | T5 was used to detect answers from texts and then use these answers to generate questions. |
| Maheen *et al.*, 2022 | PeerJ Computer Science | BERT was used solely for text embeddings. |
| Malhar *et al.*, 2022 | International Conference on Intelligent Computing and Control Systems | The T5 model was used to tokenize sentences, generate question-answer pair output from paragraph input, and then use the question-answer output to generate distractors for multiple-choice questions. BERT was used for question filtering and evaluation, as well as for text summarization to identify the most important sentences. |
| Mathur & Suchithra, 2022 | International Conference on Computational Intelligence and Sustainable Engineering Solutions | BERT was used for text embeddings to extract key phrases (serving as answers), which were then used to generate questions and distractors by another NLP model (not a large language model). |
| Matsumori *et al.*, 2023 | IEEE Access | BERT variants are used to predict masked tokens to create cloze questions. |
| Maurya & Desarkar, 2020 | International Conference on Information & Knowledge Management | BERT was used to embed the text of input triplets in the format of <article, question, correct answer>. The embedded tokens were then passed to a sequence-to-sequence model to generate distractors. |
| Muse *et al.*, 2022 | arxiv | T5 was utilized to generate questions with either pre-training or no pre-training. |
| Nguyen *et al.*, 2022 | European Conference on Technology Enhanced Learning | T5 was used to generate questions with texts processed by different methods. GPT-3 was then used to evaluate the soundness of the generated questions. |
| Nittala *et al.*, 2022 | Information and Communication Technology for Competitive Strategies | BERT was used for tokenization and word embedding. SCIBERT served as the generator, using the tokens of contexts and answers as input and also creating embeddings for these inputs. |
| Offerijns *et al.*, 2020 | arxiv | In the first phase, GPT-2 generated questions using context and answers as input, and in the second phase, it generated distractors using context, answers, and questions as input. BERT was used to filter questions for answerability and coherence (question filtering). |
| Pochiraju & al., 2023 | International Conference on Intelligent Computing and Control Systems | XLNet and BERT were used as text summarizers, with the output passed to other general NLP tools for question generation. |
| Qiu *et al.*, 2021 | International Conference on Asian Language Processing | BERT was utilized for text embedding and question filtering. The distractor generation was still based on traditional methods such as random selection, cosine similarity, and Lowenstein distance. |
| Raina & Gales, 2022 | arxiv | GPT-3 was used in a zero-shot manner, while T5 was used with context texts (without answers or key phrases) as input. |
| Ratcheva *et al.*, 2022 | International Database Engineered Applications Symposium | T5 models were fine-tuned to generate questions with texts as input and with texts and answers combined as input. |
| Rodriguez-Torrealba *et al.*, 2022 | Expert Systems with Applications | First, T5-small was used to extract answers based on input texts. Next, the sentence-answer pairs were fed into a T5-base model to generate questions. |

| | | |
|---|---|---|
| Shan *et al*., 2019 | International Conference on Technologies and Applications of Artificial Intelligence | BERT was first used for sentence embedding of articles, enabling their entry into a transformer model to extract sentences corresponding to a given question and the answer (key phrase identification). Next, the question generation task is conducted using a sequence-to-sequence model with the extracted sentences as input and questions as output. |
| Shan *et al*., 2022 | Journal of Information Science and Engineering | BERT was used for text embedding, followed by question generation by a transformer model using the embedded tokens as input. |
| Shridhar *et al*., 2022 | arxiv | T5 was used with questions and contexts as input. |
| Srivastava & Goodman, 2021 | arxiv | GPT-2 was used to generate new questions based on student performance and desired difficulty for the next item. |
| Steuer *et al*., 2020 | European Conference on Technology Enhanced Learning | Key phrase extraction was conducted before training GPT-2 to generate questions based on discussing the determined key phrases (a type of template-based generation). |
| Tsai *et al*., 2021 | International Conference on Computers in Education Conference | BERT was used to extract keywords from input textbooks, which were then used to construct important complete sentences. Subsequently, GPT-2 iteratively predicted the next word for question generation, using these complete sentences as input. |
| von Davier, 2019 | arxiv | GPT-2 was utilized to generate distractors with QA inputs like “Q: What was A?” and “A: A was Y”, and to create question passages based on prompts. |
| Vu & Van Nguyen, 2022 | Asian Conference on Intelligent Information and Database Systems | LLMs were trained to generate texts by employing reinforcement learning for attention selection in their model architecture. |
| Wang *et al*. 2021 | Annual Meeting of the Association for Computational Linguistics | BERT served as a baseline model in their study, with no further details of implementation provided. |
| Wang *et al*. 2022 | International Conference on Artificial Intelligence in Education | GPT-3 was employed using various prompting strategies for comparison. The first strategy involved using context and answer as input, while the second also utilized context and answer as input. |
| Wang *et al*. 2023 | Education and Information Technologies | SentenceBERT was used to filter texts from slides and speech recognition, using the results as input for training T5, which was then used to generate answers. These answers were combined with the input texts to generate questions. |
| Wu, 2022 | Master’s thesis | Prompts with masked tokens: “Question for answer A: [MASK] for context C”. |
| Xie *et al*., 2021 | IEEE/ACM Transactions on Audio, Speech, and Language Processing | T5 was used to generate questions with the filtered key sentences as input. |
| Zhang, 2023 | arxiv | T5 was trained using texts and answers as input. |
| Zhao *et al*., 2022 | Annual Meeting of the Association for Computational Linguistics | First, BERT was used to extract the question type information. Second, this information was used as a signal for the BART model to summarize texts, which then passed to another BART model to generate questions. BERT was used solely for text embeddings. |

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 341–352
https://doi.org/10.21449/ijate.1581217
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# Exploring number of response categories in factor analysis: Implications for sample size

**Fatih Orçan**[1*]

[1]Kahramanmaraş Sütçü İmam University, Faculty of Education, Department of Educational Sciences, Türkiye

**ARTICLE HISTORY**
*Received: Nov. 8, 2024*
*Accepted: Jan. 30, 2025*

**Keywords:**
Factor analysis,
Sample size,
Number of response categories,
Number of items.

**Abstract:** Factor analysis is a statistical method to explore the relationships among observed variables and identify latent structures. It is crucial in scale development and validity analysis. Key factors affecting the accuracy of factor analysis results include the type of data, sample size, and the number of response categories. While some studies suggest that reliability improves with more response categories, others find no significant relationship between the number of response categories and reliability. A key consideration is that increasing the number of response categories can introduce measurement errors, especially when there are too many categories for participants to respond accurately. The study examines how different numbers of response categories affect sample size requirements in factor analysis, particularly under misspecified and correctly specified models. MonteCarloSEM package in R was used to simulate data sets based on sample size, number of response categories, model specification, and test length. Results show that a higher number of categories helps reduce bias and improve model fit, especially in smaller samples. However, when sample sizes are small or when fewer categories are used, increasing the number of items or the number of categories can improve parameter estimation. The findings suggest that for optimal results, researchers should carefully balance sample size, number of items, and response categories, particularly in studies with categorical data.

## 1. INTRODUCTION

Factor analysis is a multivariate statistical technique that aims to explore the relationships among observed variables and uncover the underlying latent structures of these variables. This technique, which is widely used in the fields of social sciences and psychometrics, plays a crucial role in scale development and validity analysis. The type of data used in factor analysis and the sample size are significant factors affecting the accuracy of the results obtained (Kyriazos, 2018). In factor analysis, the number of response categories used in items can influence the quality and reliability of the results (Bandalos & Enders, 1996; Lozano *et al*., 2008). While researchers often propose different recommendations, there is no clear consensus on the best number of response categories (Simms *et al*., 2019; Wakita *et al*., 2012; Yoon, 2024). Abulela and Khalaf (2024) note that while some studies recommend between 4 and 7

*CONTACT: Fatih ORÇAN ✉ fatihorcan@ksu.edu.tr Kahramanmaraş Sütçü İmam University, Faculty of Education, Department of Educational Sciences, Kahramanmaraş, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

response categories, others focus on the range between 5 and 7, however; most scales use 4 to 5 categories.

Several studies investigate the relationship between the number of response categories and the reliability of measurement (Abulela & Khalaf, 2024; Lozano *et al*., 2008; Matell & Jacoby, 1971; Wakita *et al*., 2012; Yoon, 2024). While many argue that reliability improves with more categories, some studies contradict this statement. That is, some researchers suggest that increasing the number of categories improves reliability, while other studies show no significant relationship between reliability and the number of response categories (Abulela & Khalaf, 2024; Yoon, 2024). Specifically, "both reliability and validity are independent of the number of response categories" (Matell & Jacoby, 1971). The number of response categories does not significantly affect descriptive statistics or Cronbach's alpha (Wakita *et al*., 2012). Having more categories does not necessarily guarantee higher reliability (Abulela & Khalaf, 2024) nor does reducing longer scales to two-point or three-point scales may compromise the reliability or validity of the results (Matell & Jacoby, 1971).

Increasing the number of categories can also lead to drawbacks. Komorita and Graham (1965, as cited in Abulela & Khalaf, 2024) suggest that using more than seven categories may exceed participants' capacity to respond accurately, which increases measurement error. The complexity of a 7-point scale is unwarranted; four- or five-point scales provide adequate differentiation for most measurement needs (Wakita *et al*., 2012). Participants tend to avoid selecting extreme responses on a 7-point scale, as "an increase in the number of options biases responders against answering the strongest expressions" (Wakita *et al*., 2012, p. 543). For instance, Abulela and Khalaf (2024) point out that using more than seven categories can make labeling each option more challenging. Similarly, Simms *et al*. (2019) argues that there is no psychometric advantage (e.g., in terms of alpha) to using more than six categories. In fact, presenting respondents with many similar options (e.g., "strongly disagree," "disagree," "slightly disagree") can make it more difficult to differentiate between them, potentially leading to confusion.

Increasing the number of categories can also increase the time required for respondents to complete the questionnaire (Preston & Colman, 2000). According to Preston and Colman (2000), 5-point scales are the easiest to use, but they may be insufficient for expressing emotions or thoughts compared to scales with more categories. However, items with fewer categories are faster to answer compared to those with more categories, and thus scales with as few as 3 categories can be used depending on the study's purpose (Preston & Colman, 2000). Additionally, the number of categories may depend on the respondents' age group. For example, 3 to 4 categories may be appropriate for children, while adults may benefit from scales with 5 or more categories (Abulela & Khalaf, 2024). This is similar to the increase in response options for multiple-choice questions that accompany an increase in test-takers' age. Therefore, if participants find it too difficult or too easy to express their attitudes or feelings using the scale, their motivation to respond may decrease, leading to lower-quality data (Preston & Colman, 2000). Moreover, if the number of response categories is not chosen carefully when drafting items, it may result in significant measurement errors during parameter estimation (Abulela & Khalaf, 2024; Bandalos & Enders, 1996).

From another point of view, increasing the number of response categories leads to better model fit, regardless of whether the model is correctly specified or misspecified (Maydeu-Olivares *et al*., 2017). It was suggested that scales with more than 5 categories may help detect misspecified models more effectively. However, it was also pointed out that reducing the number of response alternatives may also decrease the likelihood of rejecting a misspecified model (Maydeu-Olivares *et al*., 2017). Specifically, decreasing the number of categories can improve model fit while introducing parameter bias, thus reducing the likelihood of rejecting a model that should be rejected (Abdelsamea, 2020). Moreover, the Root Mean Square Error of Approximation

(RMSEA) tends to increase with the number of response categories, whereas the impact of the number of categories on SRMR is less pronounced (Maydeu-Olivares *et al*., 2017).

The sample size is also a critical factor in factor analysis regarding model fit and the accuracy of parameter estimates (Kline, 2011). In particular, the sample size plays a crucial role in the effectiveness of factor analysis (MacCallum *et al*., 1999). Rhemtulla *et al*. (2012) defined studies with sample sizes ranging from 100 to 200 as small, and those with up to 600 participants were categorized as medium-sized. However, researchers often debate the ideal sample size, as this can vary depending on factors such as the complexity of the model, the type of data, and the analytical approach used. Kyriazos (2018) provides a comprehensive overview of the various factors that influence sample size in factor analysis. Specifically,

> "In CFA, being a SEM category, sample size depends on a number of features like study design (e.g. cross-sectional vs. longitudinal); the number of relationships among indicators; indicator reliability, the data scaling (e.g., categorical versus continuous) and the estimator type (e.g., ML, robust ML etc.), the missing data level and pattern and model complexity" (p. 2208).

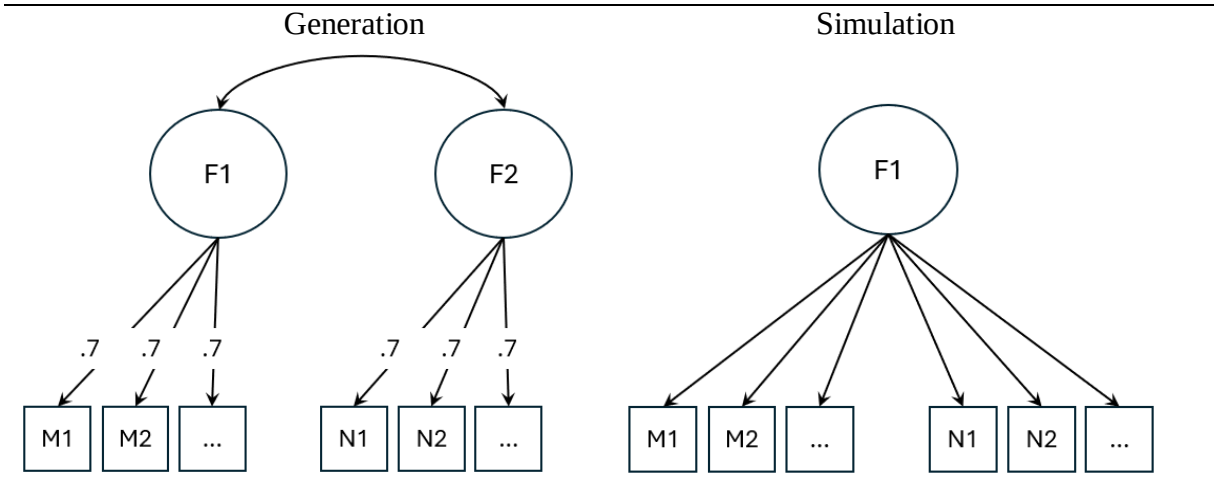
While the distinction between categorical and continuous data is often emphasized in discussions of sample size, the number of categories within categorical data also plays a critical role. The relationship between the number of categories and sample size remains underexplored in the literature. The purpose of this study is to examine the effects of the number of response categories in scales on sample size requirements in factor analysis.

## 2. METHOD

### 2.1. Data Generation Procedure

The data generation process was conducted using the MonteCarloSEM package (Orçan, 2021) within the R-CRAN environment (R Core Team, 2020). This package facilitates the simulation and analysis of data sets under various simulation conditions, including sample size and normality, for a specified model (Orçan, 2021). The data were generated based on two-factor models as shown in Figure 1. Based on the simulation conditions, the data generation model was modified. Various design factors were considered in this simulation study. After the data were generated, a one-factor model, as shown in Figure (right panel), was used to analyze the data.

**Figure 1.** *Models for data generation and simulation.*



### 2.1.1. *Design factors for data generation*

Four design factors were considered for the study.

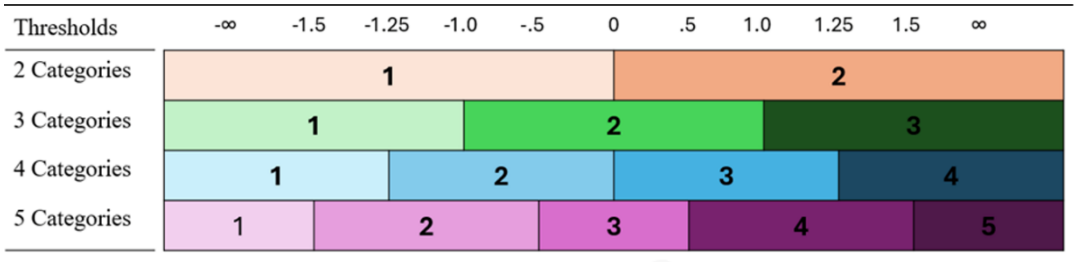
- Sample Sizes: Five different sample sizes were considered: 100, 200, 300, 500 and 1000. These sample sizes were deliberately chosen to cover a broad spectrum, ranging from small to large, within the context of Structural Equation Modeling (SEM) analysis (Orcan & Yang, 2016).
- Test Length: Three different test lengths were considered: 6, 10, 20. The number of items was equally distributed across factors. For example, when there are 6 items in total, three of those were assigned to the first factor and the rest assigned to the second factor. The factor loadings were all set at .7 within the study. That is, the factor loading was not a design factor for this study.
- Number of Response Categories: Four different categories were used within this study: 2 (C2), 3 (C3), 4 (C4), and 5 (C5). Items with two categories represent responses such as yes/no or pass/fail. Similarly, five categories represent five-point Likert-type items such as strongly disagree, disagree, neutral, agree, strongly agree. Categories larger than five were not considered in this study since scales with four or five points are sufficient (Wakita *et al*., 2012).
- Model Specifications: Three different models were used in this study: Strong misspecified model ($r$ = .5), moderate misspecified model ($r$ = .8), and correctly specified model ($r$ = 1). Under the strongly misspecified model (MS_S), the correlation between the factors was set to be .5. However, the data analyzed with one-factor model as if the correlation between the factors was assumed to be 1. Under the moderately misspecified model (MS_M), the correlation was .8 and in the correctly specified model, the correlation was set to 1. That is, under the correctly specified (CS) model, the data generation and simulation model were the same.

Consequently, the data were generated based on a total of 180 distinct conditions using these four design factors: 5 sample sizes x 3 Number of items x 4 number of categories x 3 model specifications.

## 2.2. Data Analysis

A total of 1000 data sets were generated using the MonteCarloSEM package for each of the conditions. The data generation process began by generating normally distributed data sets. Later, using per-given threshold values, the normally distributed data sets were transformed into categorical data sets. The threshold values which were used are given in Figure 2. For example, to create two categories, the threshold value was set to 0 (zero). The simulated values lower than 0 were set to 1 and larger than 0 were set to 2 to create two categories (Kılıç, 2022). The CFA modes were estimated using the Weighted Least Squares with Mean and Variance (WLSMV) estimation method in the lavaan packages (Rosseel, 2012). The WLSMV is considered "the best available categorical estimator" and is recommended for data sets with variables containing fewer than five categories (Rhemtulla *et al*., 2012, p. 354). Moreover, the WLSMV method could produce model solutions with sample sizes as small as 100 (Flora & Curran, 2004), indicating its robustness even in smaller datasets.

**Figure 2.** *Threshold values used for categorical data generation.*



Model data fits were estimated using the p-value from the chi-square test. In addition to the chi-square test, supplementary fit indices such as the Comparative Fit Index (CFI) and RMSEA should be examined to obtain a comprehensive evaluation of model fit (Flora & Curran, 2004). Therefore, CFI, RMSEA, and the standardized root mean square residual (SRMR) values were also assessed for model-data fit. For the evaluation of model data fit, Hu and Bentler's (1999) criteria were used. Also, parameter estimates for the first factor loadings were examined in detail. For this purpose, the relative biases were calculated using Equation 1.

$$Bias = \frac{Abs(Estimated\ Value - True\ Value)}{True\ Value} \quad (1)$$

Although Flora and Curran (2004) indicated that biases below 5% are considered "trivial", the critical bias threshold was set at 5% in order to be more conservative. The factor loadings for the models were all set at .7 within the study. Therefore, the true value in this study was .7.

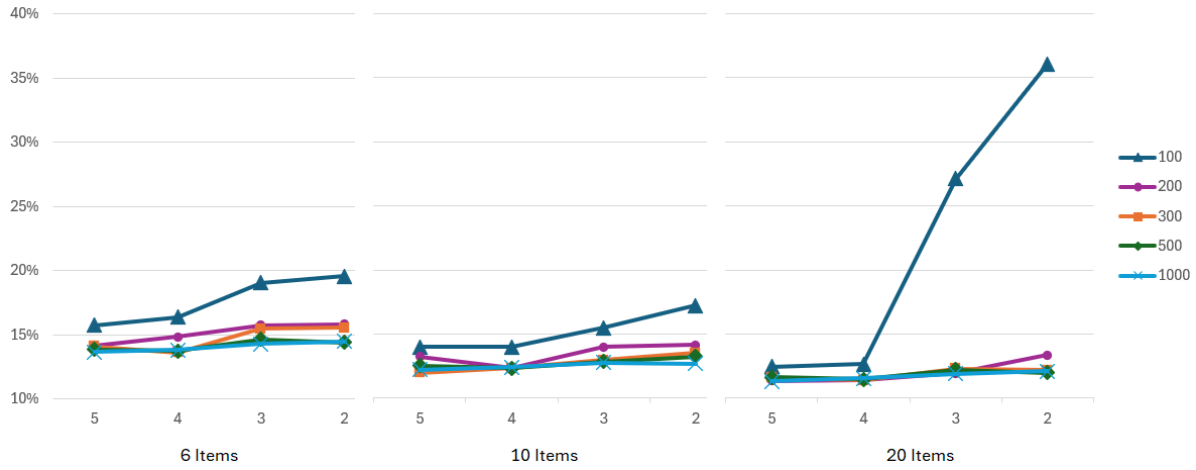
## 3. RESULTS

All the replications across the models converged to a solution. That is, non-converge was not an issue for the simulation conditions.

### 3.1. Results Based on Mis-specified Models

Two misspecified models were examined in this study. Results based on the bias calculations are presented in Figure 3 and Figure 4. When the correlation between the factors was .5, all the bias estimates exceeded the 5% critical value (see Figure 3). When examined in detail, as the sample size increased, the percentage of bias decreased, regardless of test length. Also, when the number of items increased, the percentage of bias decreased slightly. However, the decreased bias values were still much higher than the 5% critical value.

**Figure 3.** *Bias of the parameter estimates for MS_S.*




In short, under the MS_S (larger miss specification), having fewer categories for an item does not require more items or larger sample sizes, except when the sample size is too small, such as 100. That is, when the sample size is larger than 100, the estimated bases were almost identical, regardless of the number of items or categories. However, when the sample size was 100, confusing results were obtained.

When the misspecification level was decreased, i.e., when the correlation between the factors was set to .8 (MS_M), the levels of bias decreased for all conditions. Although the bias values decreased, similar patterns can be observed across the figures. Specifically, as the sample size increased, the percentage of bias decreased. Under the small sample size conditions, the increasing number of items affected the percentage of bias. For instance, when the sample size was 200 and the number of categories was 2, the bias was 10% for 6 items; however, increasing the number of items to 10 reduced the bias to 8%. Under small sample sizes, it appears that the

number of categories influences the percentage of bias. Holding everything constant but increasing the number of response categories from 2 to 4 reduces the bias estimates. Namely, under the MS_M model, similar to the results of MS_S, to achieve smaller item parameter bias, it is better to have larger sample sizes or more response categories for items. When the number of categories is low, increasing the sample size becomes "a necessity" according to the simulation results. It is also important to note that all estimates of the bias were still larger than the 5% critical value.

Besides biases in parameter estimation, model-data fit indices were also examined under misspecified models (MS_S and MS_M). Specifically, the p-values of the chi-square test, CFI, RMSEA, and SRMR values were checked.

**Figure 4.** *Bias of the parameter estimates for MS_M.*

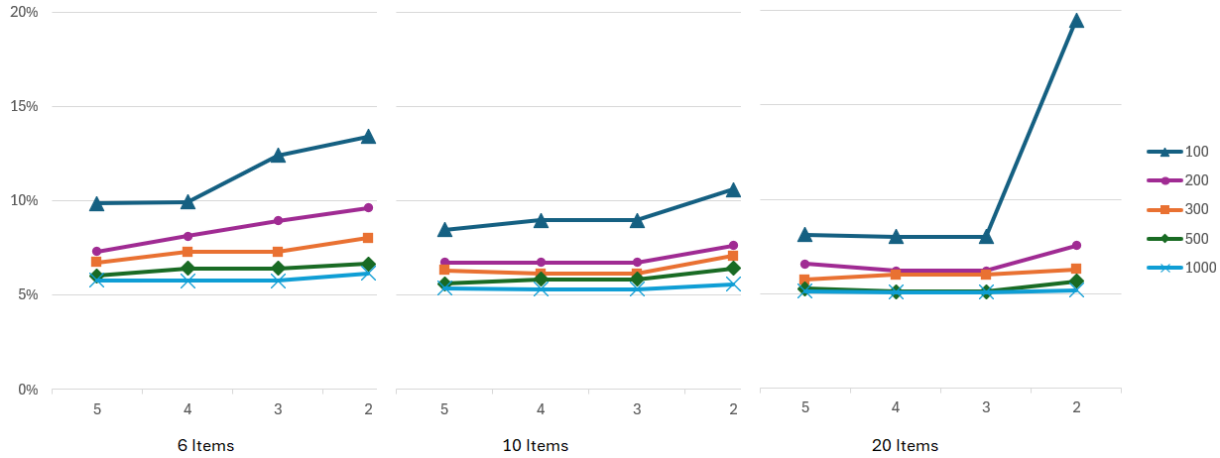



Table 1 shows the percentage of model fits for each CFI, RMSEA, and SRMR. The models reported in Table 1 are misspecified: The correlations between the factors were .5 and .8, respectively. Therefore, the values in the table were expected to be small. As the numbers increased, detecting misspecified models became less likely. For example, with 6 items, a sample size of 100, and a .5 correlation, the CFI value was .28. This means 28% of the data fit the model, even though the model was misspecified. The corresponding value for the .8 correlation (MS_M) was .98. Under the moderate misspecification with a .8 correlation, the percentage values increased. Most of the values were 1, indicating 100%. That is, 100% of the data showed fit even though it was a misspecified model. Based on the results, the values differed for the strong and moderate misspecification. It seems that when the misspecification was more evident (MS_S), as the number of categories decreased, the percentages of not rejecting the misspecified model increased based on the CFI and RMSEA values. However, as the sample size increases, up to 1000, this pattern disappears. For SRMR, the values remained almost the same regardless of size of response categories. However, under the small number of items, SRMR did not work properly, indicating biases larger than 5%, especially for the small sample sizes. Moreover, as the number of items increased, the values all became 0%. The pattern was different for the MS_M. The values were all much larger than the critical threshold for expected biases.

The results for the p-value of chi-square tests are given in Table 2. Similar to Table 1, the smaller the values, the better the misspecified model is detected. For example, with 6 items, sample size of 100, and a correlation of .5, the *p*-value was .10. This means 10% of the data fits the model according to the chi-square test, even though the model is misspecified. The corresponding value for a correlation of .8 (MS_M) was .89. When the corresponding values of MS_S and MS_M were compared, MS_M indicated larger values, in general. That is, a moderate level of misspecification is less likely to detect the problem. Based on the results, as the category decreased under the small item and sample size, detecting misspecification became less likely. As the number of items and/or sample size increased, the effect diminished.

**Table 1.** *Percent of model-data fit for misspecified models.*

| Index | Number of Items | SS | MS_S | | | | MS_M | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | C5 | C4 | C3 | C2 | C5 | C4 | C3 | C2 |
| CFI | 6 | 100 | .28 | .33 | .40 | .46 | .98 | .99 | .98 | .96 |
| | | 200 | .14 | .13 | .22 | .27 | 1.00 | 1.00 | 1.00 | .99 |
| | | 300 | .06 | .06 | .11 | .18 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 500 | .02 | .01 | .04 | .07 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 1000 | .00 | .00 | .00 | .01 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 10 | 100 | .26 | .29 | .41 | .44 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 200 | .10 | .12 | .17 | .22 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 300 | .03 | .05 | .08 | .12 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 500 | .01 | .01 | .02 | .03 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 1000 | .00 | .00 | .00 | .00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 20 | 100 | .24 | .29 | .37 | .44 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 200 | .10 | .10 | .14 | .20 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 300 | .03 | .04 | .04 | .09 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 500 | .00 | .01 | .01 | .02 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 1000 | .00 | .00 | .00 | .00 | 1.00 | 1.00 | 1.00 | 1.00 |
| RMSEA | 6 | 100 | .03 | .05 | .20 | .31 | .70 | .76 | .87 | .87 |
| | | 200 | .00 | .00 | .03 | .11 | .67 | .72 | .85 | .88 |
| | | 300 | .00 | .00 | .01 | .03 | .60 | .66 | .83 | .91 |
| | | 500 | .00 | .00 | .00 | .01 | .48 | .58 | .84 | .92 |
| | | 1000 | .00 | .00 | .00 | .00 | .31 | .49 | .86 | .96 |
| | 10 | 100 | .01 | .03 | .18 | .29 | .85 | .90 | .95 | .96 |
| | | 200 | .00 | .00 | .02 | .08 | .81 | .88 | .97 | .98 |
| | | 300 | .00 | .00 | .00 | .02 | .80 | .87 | .99 | 1.00 |
| | | 500 | .00 | .00 | .00 | .00 | .76 | .85 | .98 | 1.00 |
| | | 1000 | .00 | .00 | .00 | .00 | .68 | .84 | 1.00 | 1.00 |
| | 20 | 100 | .01 | .01 | .12 | .26 | .95 | .98 | 1.00 | 1.00 |
| | | 200 | .00 | .00 | .00 | .06 | .96 | .99 | 1.00 | 1.00 |
| | | 300 | .00 | .00 | .00 | .01 | 95 | .99 | 1.00 | 1.00 |
| | | 500 | .00 | .00 | .00 | .00 | .94 | .99 | 1.00 | 1.00 |
| | | 1000 | .00 | .00 | .00 | .00 | .93 | .99 | 1.00 | 1.00 |
| SRMR | 6 | 100 | .07 | .08 | .07 | 08 | .90 | .87 | .65 | .51 |
| | | 200 | .07 | .06 | .06 | .07 | .99 | .98 | .94 | .86 |
| | | 300 | .05 | .04 | .04 | .05 | 1.00 | 1.00 | .98 | .95 |
| | | 500 | .02 | .01 | .04 | .04 | 1.00 | 1.00 | 1.00 | .99 |
| | | 1000 | .00 | .00 | .01 | .01 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 10 | 100 | .00 | .00 | .00 | .00 | .84 | .74 | .32 | .13 |
| | | 200 | .00 | .00 | .00 | .00 | .99 | .98 | .87 | .69 |
| | | 300 | .00 | .00 | .00 | .00 | 1.00 | 1.00 | .99 | .92 |
| | | 500 | .00 | .00 | .00 | .00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 1000 | .00 | .00 | .00 | .00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | 20 | 100 | .00 | .00 | .00 | .00 | .69 | .50 | .03 | .00 |
| | | 200 | .00 | .00 | .00 | .00 | .99 | .98 | .80 | .37 |
| | | 300 | .00 | .00 | .00 | .00 | 1.00 | 1.00 | .99 | .88 |
| | | 500 | .00 | .00 | .00 | .00 | 1.00 | 1.00 | 1.00 | 1.00 |
| | | 1000 | .00 | .00 | .00 | .00 | 1.00 | 1.00 | 1.00 | 1.00 |

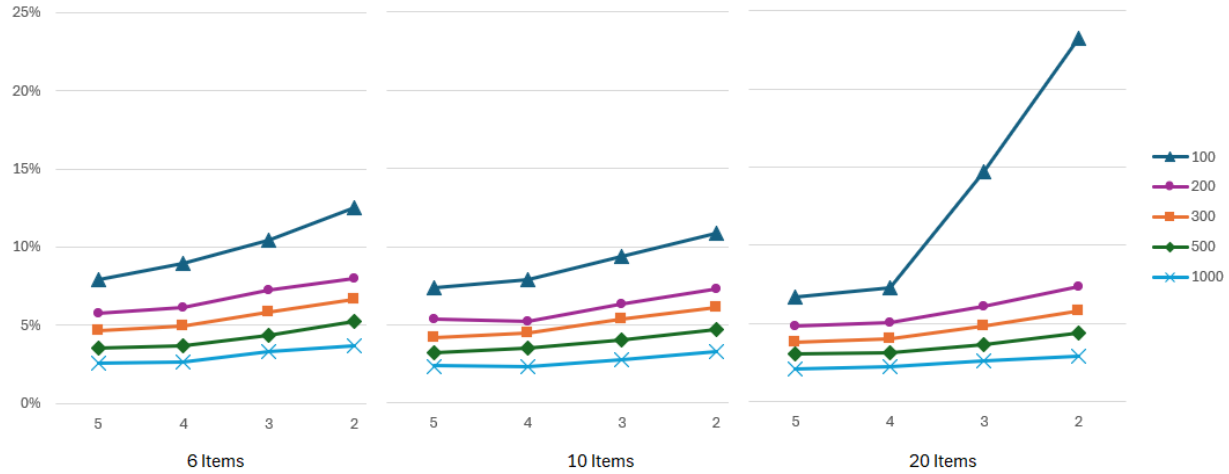
**Table 2.** *Percent of model-data fit based on the p-value for misspecified models.*

| Number of Items | SS | MS_S | | | | MS_M | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | C5 | C4 | C3 | C2 | C5 | C4 | C3 | C2 |
| 6 | 100 | .10 | .19 | .40 | .55 | .89 | .92 | .96 | .96 |
| | 200 | .00 | .01 | .04 | .14 | .72 | .78 | .88 | .92 |
| | 300 | .00 | .00 | .00 | .02 | .04 | .09 | .09 | .48 |
| | 500 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .12 |
| | 1000 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| 10 | 100 | 01 | .04 | .20 | .34 | .22 | .30 | .30 | .88 |
| | 200 | .00 | .00 | .01 | .02 | .00 | .00 | .00 | .57 |
| | 300 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .26 |
| | 500 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .01 |
| | 1000 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| 20 | 100 | .00 | .00 | .04 | .10 | .02 | .03 | .03 | .84 |
| | 200 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .41 |
| | 300 | .00 | .00 | .00 | .00 | .00 | .00 | 00 | .05 |
| | 500 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |
| | 1000 | .00 | .00 | .00 | .00 | .00 | .00 | .00 | .00 |

## 3.2. Results Based on Correctly Specified Models

Percentages of bias for the correctly specified model (CS) were given in Figure 5. Under the CS model, the data were generated with a correlation of 1 between factors (see Figure 1). As expected, as the sample size increases the percentage bias decreases, regardless of the test length or number of categories. Although the effect is limited, increasing the number of items contributes to a reduction in bias. For example, with a sample size was 200 and five response categories, biases were 5.8, 5.4, and 4.8 for 6, 10, and 20 item models respectively. It is worth noting that as the number of items reached 20, the bias dropped below the 5% critical value. Additionally, the number of response categories influenced bias estimates. Specifically, as the number of categories decreased, bias increased, regardless of the item and sample sizes.

**Figure 5.** *Bias of the parameter estimates for CS.*




Increasing sample sizes or the number of items eventually reduces the percent bias values below the critical value. Furthermore, increasing the number response options also reduces bias. Percentages of model data fit for the correctly specified model were reported in Table 3. The values in the table were expected to be small, indicating the percentages of non-fit for the model. For example, with 6 items and a sample size of 100, the RMSEA value was .02. This means that only 2% of the data does not fit the model. However, when viewed comprehensively, most

of the values in Table 3 were zero. Non-zero values were visible only for the sample size of 100. Based on the results, only SRMR values distinguish model-data fit. CFI and RMSEA values were almost all zero, while non-zero values were smaller than .05. Also, the *p*-value of the Chi-square test was all zero. Therefore, Chi-square, CFI, and RMSEA do not provide any feedback about model-data fit. However, SRMR indicates some non-zero values.

**Table 3.** *Percent of model-data fit for correctly specified models.*

| Index | Number of Items | SS | C5 | C4 | C3 | C2 |
|---|---|---|---|---|---|---|
| CFI | 6 | 100 | 0 | 0 | 0 | 0 |
| | | 200 | 0 | 0 | 0 | 0 |
| | 10 | 100 | 0 | 0 | 0 | 0 |
| | | 200 | 0 | 0 | 0 | 0 |
| | 20 | 100 | 0 | 0 | 0 | 0 |
| | | 200 | 0 | 0 | 0 | 0 |
| RMSEA | 6 | 100 | .02 | .01 | .01 | .03 |
| | | 200 | 0 | 0 | 0 | 0 |
| | 10 | 100 | 0 | 0 | 0 | 0 |
| | | 200 | 0 | 0 | 0 | 0 |
| | 20 | 100 | 0 | 0 | 0 | 0 |
| | | 200 | 0 | 0 | 0 | 0 |
| SRMR | 6 | 100 | 0 | 0 | .06 | .18 |
| | | 200 | 0 | 0 | 0 | 0 |
| | 10 | 100 | 0 | 0 | .16 | .52 |
| | | 200 | 0 | 0 | 0 | 0 |
| | 20 | 100 | 0 | 0 | .43 | .95 |
| | | 200 | 0 | 0 | 0 | 0 |

*Note:* Since the values were zero (0) for sample sizes of 300, 500, and 1000, these cases were omitted from the table for clarity.

All these non-zero values corresponded to samples with sizes below 100, with the numbers of categories being three and two. That is if all three supplementary fit indices were examined together, under the small sample sizes and a smaller number of response categories, model data fit points to a problem. Therefore, if the sample size is small, increasing the number of response categories can solve model data fit issues. From another point of view, as long as an adequate sample size is given (larger than 200), the number of response categories does not have a direct impact on model-data fit. Furthermore, as the model complexity (*df*) increases, the effect on model-data fit increases as well. For instance, in a structure with six items and two response categories, model-data fit was not achieved 18% of the time, whereas, with 10 and 20 items, this rate rises to 52% and 95%, respectively.

## 4. DISCUSSION and CONCLUSION

The effects of the number of categories on factor analysis were examined in this study. For this purpose, the Monte Carlo simulation technique was used. Sample size, test length, response categories, and model specifications were used as the design factor of the simulation. Under a misspecified model, considering that the required sample size is expected to increase as the number of items grows, a sample of 100 for 20 items may be regarded as rather low. Furthermore, in such a small sample, reducing the number of response categories appears to

have significantly impacted the estimation of model parameters. Consequently, having fewer categories changes the results considerably until the sample size reaches a certain threshold. To reduce bias, it would be beneficial either to increase the test length or the number of categories, in cases where an adequate sample size is not feasible due to the study population. More broadly, regardless of sample size, increasing the number of categories can at the very least prevent an increase in bias under a misspecified model. On the other hand, when the number of categories is low, increasing either the sample size or the number of items, even slightly, can help reduce bias. Therefore, the number of categories is linked to both the sample and test length requirements.

For model-data fits, it is better to evaluate MS_S and MS_M separately. Moderate levels of misspecification were likely more challenging for the model to identify, as they more closely resembled the true structure. In other words, when the correlation is as high as .8, distinguishing a single-factor structure in the predictive model becomes more challenging. Thus, in assessing misspecification, it would be more appropriate to focus on the result of MS_S.

As with supplementary fit indices such as SRMR, the number of response categories also impacts chi-square values (Shi *et al*., 2021). Specifically, when sample size and test length are low, reducing the number of categories makes it harder to correctly identify a misspecified model (Maydeu-Olivares *et al*., 2017). In this context, in terms of chi-square values, increasing the number of categories can make a positive contribution to the identification of a misspecified model (Abulela & Khalaf, 2024), especially when sample size and test length are limited.

In sum, when estimating a model that is not correctly specified (such as MS_S), it becomes essential to increase either the number of items in the model or the sample size if the number of categories is low. Put differently, if a small number of response options is to be used when designing items, it is important to ensure a larger number of items or a substantial sample size. Therefore, in cases where item development or access to samples is challenging, increasing the number of categories becomes crucial to accurately estimating model-data fit in misspecified models.

Similar to misspecified models, the results showed that, for the correctly specified model, increasing the sample size (MacCallum *et al*., 1999) or the number of items (Simms *et al*., 2019) reduces bias values. Furthermore, results indicate that the number of response categories used in items also helps lower estimated bias values (Shi *et al*., 2021). Consequently, in situations where sample access is limited or the number of items is restricted, increasing the number of response options in categorically defined items serves as a compensating factor. Conversely, if the number of response options is reduced, achieving consistent results would require either more items or a larger sample size. In a nutshell, when the sample size and test length are limited, choosing more response options can help lower the bias estimates.

In a correctly specified model, having a sample size above 200 can help prevent potential issues related to model-data fit. However, when the sample is insufficient or challenging to obtain, using items with only two or three response categories can lead to model-data misfit, at least as indicated by the SRMR value. Considering that supplementary fit indices are generally evaluated together, a low number of response categories may falsely suggest problems with model-data fit.

In conclusion, findings show that in misspecified models, low sample sizes and fewer response categories increase bias and lower the accuracies of parameter estimation (Lozano *et al*., 2008, Shi *et al*., 2021). To mitigate these effects, larger samples or more response categories are recommended. Particularly in models with small sample sizes and/or item counts, the number of response categories must be increased to reduce bias. Correctly specified models also benefit from additional categories, reducing bias and supporting model-data fit. A sample size of over 200 is ideal for avoiding model misfit, as low category counts with limited samples may inaccurately indicate poor fit. Finally, the study demonstrates that sample size requirements are

influenced not only by factors such as study design and estimator type (Kyriazos, 2018) but also by the number of response categories in the items. Specifically, in studies employing categorical data, using a larger number of response categories (e.g., 5) rather than fewer (e.g., 2 or 3) positively affects the adequacy of the sample size. This suggests that employing a greater number of response options can help meet sample size requirements more effectively.

Future research could examine scenarios involving studies with more than five categories, as this study is limited to a maximum of five. Furthermore, it would be beneficial to analyze how outcomes might vary in the context of more complex models. Finally, evaluating the use of a different estimation method could yield valuable insights.

## Declaration of Conflicting Interests and Ethics

The author declares no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the author.

## Orcid

Fatih Orçan https://orcid.org/0000-0003-1727-0456

## REFERENCES

Abulela, M.A.A., & Khalaf, M.A. (2024). Does the number of response categories impact validity evidence in self-report measures? A scoping review. *Sage Open, 14*(1), 1-16. https://doi.org/10.1177/21582440241230363

Abdelsamea, M. (2020). The effect of the number of response categories on the assumptions and outputs of item exploratory and confirmatory factor analyses of measurement instruments in psychological research. *Journal of Education Sohag UNV, 76*, 1153-1222. https://doi.org/10.21608/edusohag.2020.103373

Bandalos, D.L., & Enders, C.K. (1996). The effects of nonnormality and number of response categories on reliability. *Applied Measurement in Education, 9*(2), 151-160. https://doi.org/10.1207/s15324818ame0902_4

Flora, D.B., & Curran, P.J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466-491. https://doi.org/10.1037/1082-989X.9.4.466

Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kline, R.B. (2011). *Principles and Practice of Structural Equation Modeling*. Guilford Press.

Komorita, S.S., & Graham, W.K. (1965). Number of scale points and the reliability of scales. *Educational and Psychological Measurement, 25*(4), 987-995. https://doi.org/10.1177/001316446502500404

Kılıç, A.F. (2022). The effect of categories and distribution of variables on correlation coefficients. *Ege Eğitim Dergisi, 23*(1), 50-80. https://doi.org/10.12984/egeefd.890104

Kyriazos, T.A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology, 9*(8), 2207-2230. https://doi.org/10.4236/psych.2018.98126

Lozano, L., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology 4*(2). 73-79. https://doi.org/10.1027/1614-2241.4.2.73

MacCallum, R.C., Widaman, K.F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Matell, M.S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? I. Reliability and validity. *Educational and Psychological Measurement, 31*, 657-674. https://doi.org/10.1177/001316447103100307

Maydeu-Olivares, A., Fairchild, A.J., & Hall, A.G. (2017). Goodness of fit in item factor analysis: Effect of the number of response alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *24*(4), 495-505. https://doi.org/10.1080/10705511.2017.1289816

Orçan, F. (2021). MonteCarloSEM: An R package to simulate data for SEM. *International Journal of Assessment Tools in Education*, *8*(3), 704-713. https://doi.org/10.21449/ijate.804203

Orçan, F., & Yanyun, Y. (2016). A note on the use of item parceling in structural equation modeling with missing data. *Journal of Measurement and Evaluation in Education and Psychology*, *7*(1), 59-72. https://doi.org/10.21031/epod.88204

Preston, C.C., & Colman, A.M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1–15. https://doi.org/10.1016/S0001-6918(99)00050-5

R Core Team. (2020). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.R-project.org/

Rhemtulla, M., Brosseau-Liard, P.É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354-373. https://doi.org/10.1037/a0029315

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Shi, D., Siceloff, E.R., Castellanos, R.E., Bridges, R.M., Jiang, Z., Flory, K., & Benson, K. (2021). Revisiting the effect of varying the number of response alternatives in clinical assessment: Evidence from measuring ADHD symptoms. *Assessment*, *28*(5), 1287-1300. https://doi.org/10.1177/1073191120952885

Simms, L.J., Zelazny, K., Williams, T.F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, *31*(4), 557-566. http://dx.doi.org/10.1037/pas0000648

Yoon, G. (2024) No one optimal way to measure people's attitudes? Preferred length of scales in advertising research. *Journal of Current Issues & Research in Advertising*, *45*(1), 43-70, https://doi.org/10.1080/10641734.2023.2246049

Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological distance between categories in the Likert scale: Comparing different numbers of options. *Educational and Psychological Measurement*, *72*(4), 533-546. https://doi.org/10.1177/0013164411431162

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 353–366
https://doi.org/10.21449/ijate.1457123
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# Project guidance self-efficacy scale: Validity and reliability study

**Perihan Güneş**[1*], **Esra Çakırlar Altuntaş**[2], **Miraç Yılmaz**[2]

[1]Aksaray University, Faculty of Education, Department of Science Education, Aksaray, Türkiye
[2]Hacettepe University, Faculty of Education, Department of Biology Education, Ankara, Türkiye

**ARTICLE HISTORY**
*Received: Mar. 22, 2024*
*Accepted: Jan. 17, 2025*

**Keywords:**
Project-based learning,
Self-efficacy,
Scale development,
Teacher self-efficacy.

**Abstract:** Self-Efficacy scale for teachers. The items of the scale were prepared as a result of examining the literature on self-efficacy and project-making. In the first stage, the scale, which contains 40 items, was applied to a total of 578 teachers working in different branches in 7 geographical regions of Türkiye. The collected data were used to perform exploratory factor analysis ($N$=199) and confirmatory factor analysis ($N$=379). As a result of the exploratory factor analysis, it was determined that the scale, which was reduced to 19 items, had a 3-factor structure (project topic selection guidance, project implementation guidance, and project reporting guidance). The confirmatory factor analysis showed a good fit of the 19-item three-factor structure. The results revealed Cronbach's Alpha coefficient (α=.96) and McDonald's Omega (ω=.96) in the exploratory factor analysis data set, indicating a good internal consistency for the overall scale; and Cronbach's Alpha coefficient (α=.95.) and McDonald's Omega (ω=.95) in the confirmatory factor analysis data set, reflecting an excellent internal consistency. The results of the study show that the developed project guidance self-efficacy scale has good psychometric properties and reliability to measure teachers' self-efficacy for project guidance.

## 1. INTRODUCTION

The methods of accessing, processing and sharing information, which have changed dramatically in recent years, have further affected the ways we communicate and connect with each other. In order to participate in a world of rapidly expanding global networks and changing economic challenges, and to have the human skills/competencies needed, school curricula also need to be carefully developed. Advocates of 21st-century learning argue that schools have a responsibility to provide students with the opportunity to be intellectually creative, innovative, collaborative and often to think beyond boundaries (Bernhardt, 2015). Likewise, UNESCO points out that a student-centered learning approach should be adopted in education instead of solely teacher-lectured content (International Commission on the Futures of Education, 2021). These developments in perspectives have led to the prominence of student-centered approaches.

Student-centered approaches often positively affect students' knowledge and skills in various fields (Chen & Yang, 2019; Iwamoto *et al*., 2016). Kokotsaki *et al*. (2016) argued that Project-Based Learning (PBL), which is one of the course delivery methods that ensures active

*CONTACT: Perihan GÜNEŞ perihanguness@gmail.com Aksaray University, Faculty of Education, Department of Science Education, Aksaray, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

participation of students and improves their interaction and communication, stands out among all student-centered pedagogical approaches. PBL, which is an inquiry-based learning style that provides a meaningful learning experience along with a learning context with questions or problems involving real-world applications, is characterized as a teaching style that supports active and autonomous students, enables them to do creative research, set goals, collaborate, communicate and experience real-world applications (Kokotsaki *et al*., 2016).

In recent years, it has gained importance in curricula to provide students with the competence to develop project production skills and to use these skills effectively. In our country, organizations such as MoNE, Development Agency, European Union, and TUBITAK prepare local, national and international projects, and organize competitions and regional exhibitions in order to encourage students to work in basic, social and applied science fields, to guide their studies and to contribute to the development of their existing scientific studies. Especially in projects where global and social problems come to the forefront, students are expected to offer solutions by focusing on real-life problems. In the project development process, the competence of the project advisor constitutes one of the main components in terms of the quality of the project. Therefore, it has become critical to understand teachers' perceived self-efficacy in guiding student projects. Teachers' PBL guidance self-efficacy is an important construct that has not yet been sufficiently studied in education. In order to develop a comprehensive understanding of teachers' PBL guidance self-efficacy, teacher beliefs in this area need to be measured.

Although PBL is increasingly emphasized in contemporary school curricula, measurement tools that provide information about teacher self-efficacy are still limited. When the self-efficacy scales developed for PBL in the literature were examined, only two studies were found (Mutlu & Yıldız Fidan, 2018; Yaşar & Oral, 2024); in these studies, it was determined that the scale items did not reflect self-efficacy beliefs in a way to express overcoming a challenge.

Given the limitations of the existing scale, the purpose of this study is to develop a valid and reliable scale that measures teachers' self-efficacy for guiding project-based instruction. Such a scale could be valuable in helping all teachers worldwide assess their self-efficacy in guiding PBL, thus facilitating and enhancing the effectiveness of teachers' preparation and professional development programs.

### 1.1. Project Based Learning

Project-based learning (PBL) is a method of learning that encourages students to participate in authentic activities and deal with real-world problems like professionals (Kokotsaki *et al*., 2016). Specifically, it allows students to learn by asking questions, discussing ideas, designing plans, communicating with others and seeking solutions (Choi *et al*., 2019). Beginning the project with a leading question, exploring the leading question by engaging in authentic and situational inquiry as part of a collaborative activity, learning experience supported by teachers, addressing the learning objectives and a series of products or artifacts created by the students to answer the original question are referred to as the common features of the PBL approach (Krajcik & Blumenfeld, 2006). PBL is an approach that supports active participation, independent learning (Condliffe *et al*., 2017) and collaboration (Condliffe *et al*., 2017, Chu *et al*., 2017) in the learning process. PBL is also reported to be enabling the student to deepen and integrate knowledge (Iwamoto *et al*., 2016, Maros *et al*., 2023). PBL teaches students to apply knowledge to the real-world cases and to use it to solve problems, answer complex questions, and create high-quality products (Larmer & Mergendoller, 2015). PBL is an effective method for developing 21st-century skills such as critical thinking, problem solving, communication (Chu *et al*., 2017) and creativity (Duchovicova *et al*., 2018). It is also stated that PBL positively affects students' learning motivation (Chiang & Lee, 2016; Mahasneh & Alwan, 2018).

Despite the benefits of PBL, teachers in Türkiye prioritise traditional assessment tools (Gelbal & Kelecioğlu, 2007; Önel *et al*., 2020; Türkben, 2022) and teacher-centred teaching practices

(Çoban *et al*., 2021; Şahin & Ulucan, 2023; Yaylak, 2020). As the new Turkish Curriculum Guidelines aim to trigger a shift from a teacher-focused pedagogy to a student-focused pedagogy, improving teacher qualifications and, in this context, determining the relevant self-efficacy of teachers regarding teaching with PBL is significant for future policy interventions and studies on the development of teacher beliefs.

### 1.2. Measuring Project-Based Learning and Teacher Self-Efficacy

Bandura (1977, 1994) defined self-efficacy as an individual's belief in his or her capacity to execute behaviors necessary to produce specific performance attainments. It is known that self-efficacy belief is necessary to perform a certain behavior (Bandura 1977, 1994). Bandura (1977) states that self-efficacy plays an important role in how one approaches goals, tasks and challenges. It is reported that individuals with high self-efficacy beliefs tend to remain resilient when faced with difficulties, set high goals for themselves, spend more effort on their goals, and are strongly attached to their goals (Bandura, 2012). Bandura (1997) states that individuals with low self-efficacy beliefs exhibit behaviours such as not being able to do a job voluntarily, giving up quickly and often limiting their options because they doubt that they will be successful. In addition, individuals with high self-efficacy attribute their failures to correctable situations such as lack of effort and inappropriate strategies, while those with low self-efficacy attribute their failures to a lack of ability.

Self-efficacy is related to a certain field. In other words, individuals with high self-efficacy beliefs in one field may have low self-efficacy beliefs in another field (Cassidy & Eachus, 2002). Individuals develop self-efficacy with information collected from four sources: Mastery Experiences (Performance Outcomes), vicarious experiences, verbal persuasion and physiological feedback. According to Pajares (1997), self-efficacy mostly emerges in relation to specific areas or specific tasks. It would not be wrong to mention that one of the most important of these specific areas is teacher self-efficacy. Teacher self-efficacy, which is considered for long as one of the most important factors in teachers' functioning, is defined as teachers' judgment of their ability to effectively organize and execute specific courses of action (Skaalvik & Skaalvik, 2010).

Although PBL is emphasised more in contemporary school curricula, it was determined that the measurement tools that provide information about teacher self-efficacy related to PBL guidance are limited. As a result of the literature review, only two scale development studies on teachers' PBL guidance self-efficacy were found. In the study conducted by Mutlu and Yıldız Fidan (2018), a project-based teaching self-efficacy scale was developed for science and classroom teachers and applied to only 256 science and classroom teachers. The scale, which was found to include different dimensions, was applied to a limited number and disciplines of teachers and only exploratory factor analysis was used for analysis. It was determined that many sub-dimensions related to project guidance were found together in the scale and addressed with a limited number of items. When the scale developed by Yaşar and Oral (2024) is examined, it is noteworthy that the items do not reflect the self-efficacy structure to overcome a difficulty or challenge. In summary, the inadequacy of the studies in the literature to develop a valid and reliable measurement tool for teacher self-efficacy related to PBL necessitated the development of a new scale on this subject.

This study aims to develop a valid and reliable scale that measures teachers' project guidance self-efficacy.

## 2. METHOD

### 2.1. Study Design

This is a general survey and screening study conducted with the aim of developing a scale. As the data are collected from a wide population, it is possible to generalize the research results (Fraenkel *et al.*, 2012).

## 2.2. Study Group

In the study, teachers working in all public (primary, middle and high school) schools in seven regions of Türkiye were randomly selected and a total of 578 people who agreed to participate in the study were included in the sample. Teachers from every province and every school level were included to increase the widespread impact of the study. Two separate study groups were used for the purpose of this research. The first group, from which exploratory factor analysis (EFA) data were collected, consisted of 199 individuals 56.3% of whom were women and 43.7% were men; the second group, from which confirmatory factor analysis (CFA) data were collected, consisted of 379 individuals 58.3% of whom were women and 41.7% of whom were men (Table 1). Disciplines concerning the participants included in the exploratory and confirmatory factor analysis and the teachers whose project guidance self-efficacy were examined are exhibited in Table 1. Accordingly, the participants in the first study group from which EFA data were collected consisted of 199 individuals, 46.2% of whom were Science-Mathematics teachers, 46.7% were social sciences teachers and 11.1% were sports-art-language teachers; the second group from which CFA data were collected consisted of 379 individuals, 46.7% of whom were Science-Mathematics teachers, 38.3% were social sciences teachers and 15.0% of whom were sports-art-language teachers (Table 1).

**Table 1.** *Distributions of three participant groups by gender and disciplines.*

| | | EFA | CFA |
|---|---|---|---|
| Qualifications | | *f* (%) | *f* (%) |
| Gender | Male | 87 (43.7) | 158 (41.7) |
| | Female | 112 (56.3) | 221 (58.3) |
| Discipline | Science-Mathematics | 92 (46.2) | 177 (46.7) |
| | Social Sciences | 85 (42.7) | 145 (38.3) |
| | Sports-Art-Language | 22 (11.1) | 57 (15.0) |
| | Total | 199 (100) | 379 (100) |

## 2.3. Data Analysis

In order to determine whether the data obtained from the developed scale meets the requirements of factor analysis, the normality of the distribution and extreme values were examined. In addition, before the analyses, missing data were examined with Little's MCAR (Missing Completely at Random) test. However, since there were no missing values in the data set, the estimated means value could not be calculated. The data set was divided into two parts and EFA and CFA analyses were performed. Exploratory factor analysis (EFA) was performed to reveal the status of the data structure and reduce the factor whereas confirmatory factor analysis was performed to test the resulting structure. EFA was performed using Promax rotation for the construct validity of the measurements. Since it is a widely used method in the field of Social Sciences, the Principal Component was used as the factor extraction method in this study (Açıksöz *et al*., 2024). The suitability of these data for EFA was evaluated with the Kaiser criterion and Bartlett's Sphericity Test. In addition, in determining the number of factors, the eigenvalues greater than 1 were taken into account and their compatibility with the prepared draft scale was checked. In addition, the scree plot was checked. The item-total statistics of the measurement tool were examined. The averages of the scores in the lower and upper 27% groups of the total scores obtained from the scale were compared with the *t*-test.

The internal consistency of the measurement tool in reliability analyses was calculated using Cronbach's Alpha (α) and McDonald's Omega (ω) coefficients. CFA fit indices were compared with excellent and acceptable values reported in the literature. In CFA, the perfect fit of the measurement models to the data (CFI=.99; SRMR=.03; NFI=.98; NNFI=.98; IFI=.99) and acceptable fit (RMSEA=.04; PNFI=.86) were evaluated as evidence of construct validity

(Schermelleh-Engel *et al*., 2003). IBM SPSS Statistics 15 and JASP programs were used in the study.

### 2.4. The Steps of Scale Development

This study was designed in accordance with the scale development process (item generation, questionnaire administration, item reduction and CFA) proposed by Wang (2003). In the preparation phase of the developed Project Guidance Self-Efficacy Scale, the literature related to self-efficacy, project making and project process (Schwarzer & Jerusalem, 1995; Schmitz & Schwarzer, 2000; Mutlu & Yıldız Fidan, 2018) was examined. After the examinations, a total of 40 items were prepared within the scope of project preparation (13 items), project implementation (13 items) and project reporting dimensions (14 items). The scale was examined for content validity by three field educators, two measurement and evaluation experts, and a Turkish language expert. Content validity is the primary validity analysis that provides information about whether a scale measures the desired feature (Cronbach, 1970). In line with the expert opinions, the finalized draft scale included 40 items. In the scale, the participants were asked to give graded responses as “Totally Agree (5), Mostly Agree (4), Somewhat Agree (3), Somewhat Agree (2), Strongly Disagree (1)”. An online survey was conducted to collect scale development data. The questionnaire was created on Google Forms and the link was sent to the teachers. Afterwards, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were conducted to reveal the construct validity of the scale

In order to reduce bias in the data collection process, attention was paid to having fewer items in the scale. In addition, care was taken to keep the items short so that the scale could be answered in a short time. In addition, care was taken to use correct and clear expressions so that the scale items only tested one behavior and could be understood in the same way by each reader. Finally, negative statements were added to the scale and underlined to draw the reader's attention. Afterwards, expert opinion was obtained, and a pilot study was conducted. Before the expert opinion, the scale included 14 items in each of the three dimensions (project preparation, project implementation and project reporting dimensions). In line with the opinions of the experts, one item was removed from each of the project preparation and project implementation dimensions because the items covered each other. The validity, comprehensibility and expression problems of the items in the scale were not detected. A pilot study was conducted with 17 teachers from different branches and no changes were made after the pilot study. As a result of the pilot study, no negative situation was detected regarding the comprehensibility of the questions and their effect on the participants. No ethical problems were encountered during the data collection process. In addition, the time required to complete the study was sufficient for teachers to answer the questions without getting bored.

## 3. FINDINGS

### 3.1. Findings Regarding Exploratory Factor Analysis

The appropriateness of the dataset for factor analysis was confirmed with the Kaiser Meyer Olkin (KMO) value and Bartlett's Test of Sphericity. The analysis indicated that the KMO value was .951 and the result of Bartlett's Test of Sphericity was significant. ($\chi^2$ = 3439.075; $p$=.000).

The principal components extraction method was used to determine the construct validity of the scale and to size the items by determining their factor loadings. Items loaded on more than one factor were removed from the analysis (m6, m7, m8, m9, m10, m11, m12, m13, m23, m24, m25, m26, m27, m33, m34, m35, m36, m37, m38, m39, m40). The remaining items were grouped under three factors (guidance for choosing the project topic, guidance for project implementation, and guidance for project reporting). The total Variance Explained by the Scale was calculated at 69.208%. Eigenvalues and the variance explained by each factor are presented in Table 2.

**Table 2.** *Eigenvalues for the scale and explained variances.*

| Factor | Eigenvalue | Variance % | Total Variance % |
|---|---|---|---|
| 1. Guidance for Project Implementation Self-efficacy | 11.516 | 60.608 | 60.608 |
| 2. Guidance for Project Reporting Self-efficacy | 1.420 | 7.475 | 68.083 |
| 3. Guidance for Choosing the Project Topic Self-efficacy | 1.085 | 5.711 | 73.794 |

Three factors created in line with the meanings of the items were named Guidance for Choosing the Project Topic Self-efficacy (Factor 3), Guidance for Project Implementation Self-efficacy (Factor 1) and Guidance for Project Reporting Self-efficacy (Factor 2). Promax rotation was preferred for the factors revealed by the Principal components method. Items collected under three factors are in Table 2 and their common variances are presented in Table 3.

**Table 3**. *Factor loadings.*

| | | Post Rotation | | |
|---|---|---|---|---|
| Item No | Common Factor Variance | Choosing the Project Topic (Factor 3) | Project Implementation (Factor 1) | Project Reporting (Factor 2) |
| M1 | .642 | .888 | | |
| M2 | .703 | .713 | | |
| M3 | .792 | .878 | | |
| M4 | .782 | .890 | | |
| M5 | .699 | .628 | | |
| M14 | .666 | | .820 | |
| M15 | .571 | | .608 | |
| M16 | .761 | | .905 | |
| M17 | .685 | | .874 | |
| M18 | .720 | | .888 | |
| M19 | .690 | | .706 | |
| M20 | .722 | | .666 | |
| M21 | .761 | | .781 | |
| M22 | .698 | | .733 | |
| M28 | .797 | | | .738 |
| M29 | .825 | | | .711 |
| M30 | .867 | | | .987 |
| M31 | .883 | | | .917 |
| M32 | .755 | | | .819 |

Table 3 reveals that there are 5 items under the Guidance on Choosing the Project Topic factor. There are 9 items under the Guidance on Project Implementation factor and 5 items under the Guidance on Project Reporting factor. Inter-factor correlation values are given in Table 4.

**Table 4**. *Inter-factor correlation values.*

| Factors | *r* |
|---|---|
| Guidance for Choosing the Project Topic Self-efficacy-Guidance for Project Implementation Self-efficacy | .70* |
| Guidance for Choosing the Project Topic Self-efficacy-Guidance for Project Reporting Self-efficacy | .72* |
| Guidance for Project Implementation Self-efficacy - Guidance for Project Reporting Self-efficacy | .77* |

*p*<.01

### 3.2. Results Regarding the Reliability of the Scale

In order to determine the item discrimination levels, the item discrimination of the upper and lower groups of 27% was analysed. For this purpose; internal consistency of the total scale and sub-dimensions, in other words, Cronbach's Alpha, Stratified Alpha and McDonald's Omega values, which determine how closely the items are related to each other were analysed. The results concerning the reliability of the scale are presented in Table 5.

**Table 5.** *Results concerning the reliability of the scale.*

| Factor | Item No | *t* Lower %27-Upper %27 (*n*=54) | Cronbach's Alpha (α) | McDonald's Omega (ω) |
|---|---|---|---|---|
| Guidance for Choosing the Project Topic Self-efficacy | M1 | -9.452* | .900 | .902 |
| | M2 | -10.359* | | |
| | M3 | -10.663* | | |
| | M4 | -10.474* | | |
| | M5 | -11.142* | | |
| Guidance for Project Implementation Self-efficacy | M14 | -10.167* | .941 | .941 |
| | M15 | -10.551* | | |
| | M16 | -12.983* | | |
| | M17 | -12.226* | | |
| | M18 | -11.065* | | |
| | M19 | -11.088 | | |
| | M20 | -10.628* | | |
| | M21 | -11.517* | | |
| | M22 | -11.298* | | |
| Guidance for Project Reporting Self-efficacy | M28 | -11.837* | .943 | .943 |
| | M29 | -13.715* | | |
| | M30 | -11.088* | | |
| | M31 | -10.900* | | |
| | M32 | -12.334* | | |
| Overall Project Guidance Self-Efficacy Scale (EFA)** | | | .960 | .960 |

*$p$<.001,** For the reliability results of the scale's CFA, see the results and discussion sections.

The items of the 19-item scale are included in Appendix. It was further observed that the difference in item score averages between the upper and lower 27% groups was significant ($p$ <.001). Considering these values, it is possible to conclude that the reliability of the items of the scale is high.

Project Guidance Self-Efficacy Scale's total reliability was confirmed with the values Cronbach's Alpha α=.960 and McDonald's Omega ω=.960 (Table 5). This value indicates a high degree of reliability (Can, 2022). Cronbach's Alpha was calculated as α=.900 for the Guidance for Choosing the Project Topic Self-efficacy, α=.941 for Guidance for Project Implementation Self-efficacy and α=.943 for Guidance for Project Reporting Self-efficacy. The stratified alpha of the entire scale is .97. McDonald's Omega values in the sub-dimensions were calculated as ω=.902, ω=.941 and ω=.943, respectively.

### 3.3. Findings Regarding Confirmatory Factor Analysis (CFA)

Confirmatory factor analysis (CFA) was applied to confirm the construct validity of the Project Guidance Self-Efficacy Scale. Cronbach's Alpha value of the CFA data set for the overall scale was calculated as α=.953; McDonald's Omega value was calculated as ω=.952. Cronbach's

Alpha and McDonald's Omega values for the three factors were α=.90 and ω=.90 for Guidance for Choosing the Project Topic Self-efficacy, α=.94 and ω=.94 for Guidance for Project Implementation Self-efficacy and α=.94, ω=.94 for Guidance for Project Reporting Self-efficacy.

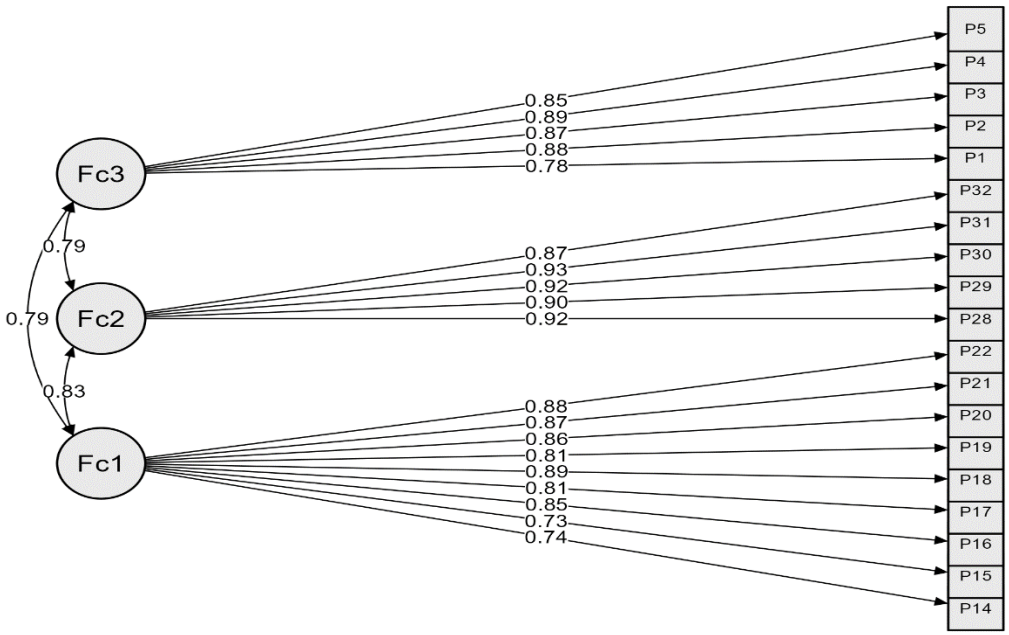
Then, the Chi-Square Fit Test was performed to demonstrate the adequacy of the model tested in CFA and it was reported that values less than 2 for $\chi^2/df$ indicate perfect fit and values between 2 and 3 indicate acceptable fit (Schumacker & Lomax, 1996; Schermelleh-Engel *et al*., 2003). Fit indices calculated with CFA were obtained as RMSEA=.049, CFI=.99. The 90% confidence intervals for the RMSEA of the scale were reported as .041 and .058. Ideally, the lower value of the 90% confidence interval includes or is very close to zero (Demir, 2022). It was observed that the statistics of $\chi^2$=285.82 ($df$=149) were significant ($p$<.01) and $\chi^2/df$ =1.91. The fit indices obtained by applying CFA are presented in Table 6.

**Table 6.** *Fit indices derived by applying CFA.*

| Analysed Fit Indices | Perfect Fit Criteria | Acceptable Compliance Criteria | Derived Fit Indices | Results |
|---|---|---|---|---|
| $\chi^2/df$ | $0 \leq \chi^2/df \leq 2$ | $2 \leq \chi^2/df \leq 3$ | 1.91 | Perfect Fit |
| CFI | .95 ≤CFI≤ 1.00 | .90 ≤CFI≤ .95 | .99 | Perfect Fit |
| NFI | .95 ≤ NFI ≤ 1.00 | 90 ≤ NFI ≤ .95 | .99 | Perfect Fit |
| NNFI | .95 ≤ NNFI ≤ 1.00 | .90 ≤ NNFI ≤ .95 | .99 | Perfect Fit |
| IFI | .95 ≤ IFI ≤ 1.00 | .90 ≤ IFI ≤ .95 | .99 | Perfect Fit |
| RMSEA | .00 ≤ RMSEA ≤ .05 | .05 ≤ RMSEA ≤.08 | .04 | Perfect Fit |
| SRMR | .00 ≤ SRMR ≤ .05 | .05 ≤ SRMR ≤ .10 | .03 | Perfect Fit |
| PNFI | .95 ≤ PNFI ≤ 1.00 | .50 ≤ PNFI ≤ .95 | .86 | Acceptable Fit |

Comparative Fit Index (CFI), Standardized Root Mean Square Residual (SRMR), Normed Fit Index (NFI), Non-normed Fit Index (NNFI), Incremental Fit Index (IFI) values of the model were observed to be at the perfect fit level (CFI=.99; SRMR=.03; NFI=.98; NNFI=.98; IFI=.99); Absolute Fit Index (RMSEA) and Parsimonious Normed-Fit Indices (PNFI) values were found to be within acceptable values ( RMSEA=.04; PNFI=.86). The fit index values derived as a result of the analysis revealed that the model had a good fit. Figure 1 shows the DFA results (Parameter Estimates-Standardized) for the Project Guid-ance Self-Efficacy Scale.

**Figure 1.** *Model plot (parameter estimates-standardized).*



## 4. DISCUSSION and CONCLUSION

The main purpose of this study was to examine teachers' project guidance self-efficacy levels and to develop a valid and reliable scale that measures teachers' project guidance self-efficacy. In this context, the Project Guidance Self-Efficacy Scale (PGSS) developed by the authors of the study was administered to teachers from different disciplines. The results of the analysis indicate that the psychometric properties of the constructed scale are acceptable and that it can be used to determine the project guidance characteristics of teachers from all disciplines.

It was found that two scales were developed in the literature to measure project guidance self-efficacy (Mutlu & Yıldız-Fidan, 2018; Yaşar & Oral, 2024). The scale developed by Mutlu and Yıldız-Fidan (2018) was applied to a limited number of teachers in a limited number of disciplines. Although the scale developed by Yaşar and Oral (2024) was applied to a sufficient number of teachers from many different disciplines, the data were obtained from a limited number of provinces. In the scale development process, we tried to ensure the generalizability of the scale by reaching teachers from different regions, provinces, branches, professional experience levels and school types (public, private, rural, urban).

When the psychometric properties of the existing scales in the literature were examined, it was determined that the scale developed by Yaşar and Oral (2024) showed sufficient psychometric values. However, it is noteworthy that the scale items do not reflect the self-efficacy structure to overcome a difficulty or challenge. This situation weakens the construct validity of the scale. In the scale developed by Mutlu and Yıldız-Fidan (2018), it was determined that many sub-dimensions related to project guidance were found together and addressed with a limited number of items. In addition, only exploratory factor analysis was conducted to determine the construct validity of the scale. This situation indicates that the scale has weak psychometric properties. Due to these problems in the developed scales, it was thought that it would be more appropriate to evaluate the project guidance self-efficacy potential holistically. For this purpose, it was decided to develop a valid and reliable project guidance self-efficacy scale. Our project guidance self-efficacy scale is based on the preparation, implementation and finalization of a project.

In the study, there were 5 items under the Guidance on Choosing the Project Topic Self-Efficacy factor (Items 1-5.), 9 items under the Guidance on Project Implementation Self-Efficacy factor (Items 6-14.) and 5 items under the Guidance on Project Reporting Self-Efficacy factor (Items 15-19). In the literature, there are two studies addressing project guidance self-efficacy in different factors (Mutlu & Yıldız-Fidan, 2018; Yaşar & Oral, 2024). When these studies are examined, it is seen that in the scale developed by Mutlu and Yıldız-Fidan (2018), self-efficacy consists of five factors: “dominance of the project process, guidance”, “planning, preparation and reflection”, “implementation and evaluation”, “giving feedback, alternative evaluation” and “group process and high-level learning”. In the scale developed by Yaşar and Oral (2024), it is seen that self-efficacy is handled in four factors “implementation process”, “supervisor self-efficacy, ‘guidance’ and ‘ethics and responsibility’. In these studies, there is no scale that combines the factors of project topic selection guidance, project implementation guidance and project reporting guidance. However, there is a parallelism between the dimensions of “implementation and evaluation” and “implementation process” in the scales developed by Mutlu and Yıldız-Fidan (2018) and Yaşar and Oral (2024) and the dimensions obtained in this study.

In conclusion, the constructed Project Guidance Self-Efficacy Scale has a total of 19 items and 3 factors, and the scores that can be obtained from the scale vary between 19 and 95. The total variance explained as a result of the scale was calculated as 69.208%. Higher scores obtained in the Project Guidance Self-Efficacy Scale indicate that the teacher has a firm belief in successfully performing the behaviors related to guiding a project. There were no reverse-coded items in the scale. Participants indicate how much they agree or disagree with the items included

in the Project Guidance Self-Efficacy Scale by selecting the answers between "Strongly Agree (5), Agree (4), Neither Agree nor Disagree (3), Disagree (2), Strongly Disagree (1)". Cronbach's Alpha and McDonald's Omega reliability values calculated with the CFA data set of the Project Guidance Self-Efficacy Scale indicated perfect internal consistency (Overall PGSS α=.95, ω=.95; Guidance for Choosing the Project Topic Self-efficacy α=.90, ω=.90; Guidance for Project Implementation Self-efficacy α=.94, ω=.94; Guidance for Project Reporting Self-efficacy α=.94, ω=.94) (Özdamar, 2002; George & Mallery, 2003). These values mean that we can keep each dimension of the scale separate.

The results of the study further indicated that the psychometric properties and reliability of the developed Project Guidance Self-Efficacy Scale were appropriate to measure the teachers' level of self-efficacy belief that they can perform behaviors related to guiding students' projects. PGSS is concerned with the individual's belief in successfully engaging in behaviors related to getting a project done. In this respect, it is compatible with Bandura's (1977) definition of self-efficacy belief in being able to do certain tasks.

## 5. SUGGESTIONS

### 5.1. Implications for Education and Teaching

The results of the study suggest that the Project Guidance Self-Efficacy Scale may help to measure the self-efficacy levels of teachers from disciplines such as science, social, arts etc. with regard to their ability to lead and guide their students in their projects and to develop further practices to improve the teachers' competencies in this regard. As it was concluded in this study that teachers can successfully guide their students in their projects, these experiences can be reflected more frequently in the learning-teaching environment to improve project guidance experiences with students. PGSS, which will be used in detailed studies on teachers, can help in the design of new project-based learning-teaching studies. With the help of this scale, further detailed studies examining the level and sources of project guidance self-efficacy beliefs of wider groups of teachers can be supported. Thus, by using the PGSS constructed herein, further research may be conducted to improve the quality of teachers' project guidance at schools, a higher level of benefit may be gained from project-based learning, which is one of the most important learning approaches that serve student skills and learning to contribute to the development of the 21st-century education system in Türkiye/Turkish-speaking countries.

### 5.2. Limitation

This study has the following limitations:

- Teachers working in different countries, regions or education systems may have different experiences. This may limit the applicability of the scale.
- Teachers' socio-economic or cultural backgrounds may cause them to respond differently to the scale items.
- Some dimensions or items may differ according to teachers' experience and branches.
- It may not be applied in the same way to teachers from different branches and experience levels (more or less project experience).
- Different reasons (heavy workload, length of the scale, fatigue, etc.) may cause teachers to leave the scale without completing it or to respond quickly and carelessly.
- Teachers may be concerned about the confidentiality of their responses and may not give honest answers.

### Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Aksaray University, 2022/04-53.

## Contribution of Authors

**Perihan Güneş:** Conception, Design, Investigation, Literature review, Methodology, Data collection, Data interpretation, and Writing-original draft. **Esra Çakırlar Altuntaş**: Methodology, Formal analysis, Finding. **Miraç Yılmaz:** Conception, Design, Supervision, Critical review, and Writing-original draft.

## Orcid

Perihan Güneş https://orcid.org/0000-0003-4551-9327
Esra Çakirlar Altuntaş https://orcid.org/0000-0002-3566-8655
Miraç Yilmaz https://orcid.org/0000-0003-3200-2767

## REFERENCES

Açıksöz, A., Dökme, İ., & Önen, E. (2024). Development and validation of STEM motivation scale for middle school students. *International Journal of Assessment Tools in Education, 11*(4), 699-720. https://doi.org/10.21449/ijate.1401339

Bandura, A. (1977) Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1994). *Self-efficacy*. In V.S. Ramachaudran (Ed.), Encyclopedia of human behavior (Vol. 4, pp. 71-81). Academic Press

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management, 38*(1), 9-44.

Bernhardt, P.E. (2015). 21st century learning: Professional development in practice. *The Qualitative Report, 20*(1), 1-19. https://doi.org/10.46743/2160-3715/2015.1419

Can, A. (2022). *SPSS ile bilimsel araştırma sürecinde nicel veri analizi* (10. Baskı) [Quantitative data analysis in the scientific research process with SPSS (10th ed.)]. Pegem Akademi.

Cassidy, S., & Eachus, P. (2002). Developing the computer user self-efficacy (CUSE) scale: Investigating the relationship between computer self-efficacy, gender and experience with computers. *Journal of Educational Computing Research*, *26*(2), 133-153. https://doi.org/10.2190/JGJR-0KVL-HRF7-GCNV

Chen, C.H., & Yang, Y.C. (2019). Revisiting the effects of project-based learning on students' academic achievement: A meta-analysis investigating moderators. *Educational Research Review*, *26*(1), 71-81. https://doi.org/10.1016/j.edurev.2018.11.001

Chiang, C.L., & Lee, H. (2016). The effect of project-based learning on learning motivation and problem-solving ability of vocational high school students. *International Journal of Information and Education Technology, 6*(9), 709-712. https://doi.org/10.7763/IJIET.2016.V6.779

Choi, J., Lee, J.-H., & Kim, B. (2019). How does learner-centered education affect teachers' self-efficacy? The case of project-based learning in South Korea. *Teaching and Teacher Education, 85*, 45–57. https://doi.org/10.1016/j.tate.2019.05.005

Chu, S.K.W., Reynolds, R.B., Tavares, N.J., Notari, M., & Lee, C.W.Y. (2017). *21st century skills development through inquiry-based learning.* Springer.

Condliffe, B., Quint, J., Visher, M., Bangser, M.R., Drohojowska, S., Saco, L., & Nelson, E. (2017). *Project-based learning: A literature review.* https://files.eric.ed.gov/fulltext/ED578933.pdf

Çoban, Ö.K., Çelik, A.Y., & Kılıç, Z. (2021). Kimya öğretmenlerinin öğretim stratejileri ve bu stratejilere etki eden faktörler [Chemistry teachers' instructional strategies and the factors that affect these strategies]. *Eğitim ve Toplum Araştırmaları Dergisi, 8*(2), 345-361. https://doi.org/10.51725/etad.1030346

Cronbach, L.J. (1970). *Essentials of psychological testing*. Harper & Row.

Demir, E. (2022). Ölçek geliştirme sürecinde doğrulayıcı faktör analizi [Confirmatory factor analysis in the scale development process]. M. Acar-Güvendir, & Y. Özer-Özkan (Ed.), *Tüm yönleriyle ölçek geliştirme süreci [Scale development process in all aspects]*. Pegem Akademi.

Duchovicova, J., Sabo, A., Petrova, G., & Hosova, D. (2018). Stimulation of creativity as a prerequisite of permanent sustainability form personality development of gifted learners. *Ad Alta - Journal of Interdisciplinary Research, 8*(2), 54-60. https://www.magnanimitas.cz/ADALTA/0802/papers/A_duchovicova.pdf

Fraenkel, J.R., Wallen, N., & Hyun, H. (2012). *How to design and evaluate research in education* (8th ed.). McGraw-Hill

Gelbal, S., & Kelecioğlu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar [Teachers' perceived competence in measurement and evaluation methods and the challenges they face]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 33*(33), 135-145. https://dergipark.org.tr/en/download/article-file/87640

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Allyn & Bacon.

International Commission on the Futures of Education (2021). *Progress update of the international commission on the futures of education.* UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000375746/

Iwamoto, D.H., Hargis, J., & Vuong, K. (2016). The effect of project-based learning on student performance: An action research study. *International Journal for the Scholarship of Technology Enhanced Learning, 1*(1), 24-42. https://cdn.fs.teachablecdn.com/CcQpxRwPSs6yJyK7D6pP

Kokotsaki, D., Menzies, V., & Wiggins, A. (2016). Project-based learning: A review of the literature. *Improving Schools, 19*(3), 267-277. https://doi.org/10.1177/1365480216659733

Krajcik, J., & Blumenfeld, P. (2006). Project-Based learning. In the Cambridge handbook of the learning sciences (pp. 317-333). Cambridge University Press. https://doi.org/10.1017/CBO9780511816833.020

Larmer, J., & Mergendoller, J. (2015). *Gold standard PBL: Project based teaching practices*. Buck Institute for Education.

Mahasneh, A.M., & Alwan, A.F. (2018). The effect of project-based learning on student teacher self-efficacy and achievement. *International Journal of Instruction, 11*(3), 511–524.

Maros, M., Korenkova, M., Fila, M., Levicky, M., & Schoberova, M. (2023). Project-based learning and its effectiveness: evidence from Slovakia. *Interactive Learning Environments, 31*(7), 4147-4155. https://doi.org/10.1080/10494820.2021.1954036

Mutlu, F., & Yıldız Fidan, N. (2018). Fen bilimleri ve sınıf öğretmenlerinin proje tabanlı öğretim yapabilme özyeterlilikleri üzerine bir ölçek geliştirme çalışması [A study on developing a scale for science and primary teacher's self efficacy regarding providing project-based teaching]. *Adnan Menderes Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 5*(2), 191-203. https://doi.org/10.30803/adusobed.387109

Önel, F., Dalkılınç, F., Özel, N., Deniz, Ş., Balkaya, T., & Birel, G. (2020). Ortaokul matematik öğretmenleri ölçme-değerlendirmeyi nasıl yapıyor? Bir durum çalışması [How middle school mathematics teachers do measurement and assessment? A case study]. *Kastamonu Eğitim Dergisi, 28*(3), 1448-1459. https://doi.org/10.24106/kefdergi.4113

Özdamar K. (2002). *Paket Programlarla İstatistiksel Veri Analizi-1 (4. Baskı) [Statistical Data Analysis with Software Packages-1(4th ed.)]*. Kaan Kitabevi.

Pajares, F. (1997). Current directions in self-efficacy research. *Advances in Motivation and Achievement, 10*(149), 1-49. https://www.dynaread.com/current-directions-in-self-efficacy-research

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods*

*of Psychological Research Online*, *8*(2), 23-74. https://www.stats.ox.ac.uk/~snijders/mpr_Schermelleh.pdf

Schmitz, G.S., & Schwarzer, R. (2000) Selbstwirksamkeitserwartung von Lehrern: Längsschnitt-befunde mit einem neuen Instrument [Perceived self-efficacy of teachers: Longitudinal findings with a new instrument]. *Zeitschrift für Pädagogische Psychologie, 14*(1), 12-25. https://doi.org/10.1024//1010-0652.14.1.12

Schumacker, E.R., & Lomax, G.R. (1996). *A beginner's guide to structural equation modeling.* Erlbaum.

Schwarzer, R., & Jerusalem, M. (1995). *General Self-Efficacy Scale (GSE)* [Database record]. APA PsycTests. https://doi.org/10.1037/t00393-000

Skaalvik, E.M., & Skaalvik, S. (2010). Teacher self-efficacy and teacher burnout: A study of relations. *Teaching and Teacher Education, 26*(4), 1059-1069. https://doi.org/10.1016/j.tate.2009.11.001

Şahin, Y., & Ulucan, P. (2023). Sınıf öğretmenlerinin çeşitli öğretim yöntem ve tekniklerini kullanma durumları. [Using conditions of various teaching methods and techniques of primary school teachers] *Ordu Üniversitesi Sosyal Bilimler Enstitüsü Sosyal Bilimler Araştırmaları Dergisi, 13*(1), 995-1030. https://doi.org/10.48146/odusobiad.1099218

Türkben, T. (2022). Türkçe öğretmenlerinin sınıf içi öğrenme becerilerini ölçme ve değerlendirme yeterlikleri [Turkish language teachers' competencies in assessing and evaluating classroom learning skills]. *Eğitim ve İnsani Bilimler Dergisi: Teori ve Uygulama, 13*(25), 45-72. https://dergipark.org.tr/tr/pub/eibd/issue/70603/1088012

Wang, Y.-S. (2003). Assessment of learner satisfaction with asynchronous electronic learning systems. *Information & Management, 41*(1), 75–86.

Yaylak, E. (2020). Sosyal bilgiler öğretmenlerinin öğretim yöntem ve uygulamalarının değerlendirilmesi [Evaluation of teaching methods and practices of social studies teachers]. *Ahi Evran Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 6*(1), 187-205. https://doi.org/10.31592/aeusbed.621664.

Yaşar, E., & Oral, B. (2024). Öğretmenlerin proje danışmanlık yeterlik düzeyleri: Ölçek geliştirme çalışması [Competency levels of teachers in project consultancy: Scale development study]. *Türk Eğitim Bilimleri Dergisi*, *22*(1), 464-486. https://doi.org/10.37217/tebd.1363368

## APPENDIX

**Appendix 1.** *Items included in the Project Guidance Self-Efficacy Scale.*

1. I can enable students to express salient modern-day problems awaiting solutions.
2. I can enable students to conduct literature research to choose project topics regarding salient problems
3. I can support students in developing creative project ideas to solve modern-day problems.
4. I can enable students to construct a project goal for solving problems, even if they have not done any projects before.
5. I can guide students to develop a research hypothesis suitable for the project's goal.
6. Even if my experience in project guidance is limited, I can support to work collaboratively of the project team.
7. I believe I can support my students to reach out to field experts for help.
8. I can enable my students to be in good communication with teachers, friends and others during the project process.
9. I can ensure everyone in the project team to fulfil their responsibilities.
10. I can ensure students in conducting their experiments/fieldworks in the implementation phase according to plan.
11. I can assist students to correctly use laboratory/field tools and equipment.
12. I believe I can guide students to manage any problem that may emerge during implementation.
13. I can enable my students to appropriately conduct their experiments/fieldworks during implementation.
14. I believe I can guide them to use time efficiently, in accordance with the project plan.
15. I believe I can support my students to interpret project results, bringing in their own commentary.
16. I can support students to make appropriate pertinent inferences regarding the obtained results.
17. I can to oversee my students as they defend their inferences using proof from the literature.
18. I can support them in explaining their results with suitable reasoning.
19. I am able to guide the writing of project results in accordance with scientific methodology (objective, methodology, findings, discussion).

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 367–386
https://doi.org/10.21449/ijate.1518335
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# The perception scale for curriculum leadership: Validity and reliability study

**Cemile Keski**[1], **İlkay Doğan Taş**[1*]

[1]Kırıkkale University, Faculty of Education, Department of Curriculum and Instruction, Kırıkkale, Türkiye

**ARTICLE HISTORY**
*Received: July 18, 2024*
*Accepted: Mar. 24, 2025*

**Keywords:**
Curriculum leadership,
Branch teacher,
Middle school,
Leadership,
Scale development.

**Abstract:** The purpose of this study was to create a measurement instrument that would be both valid and reliable for assessing middle school branch teachers' perceptions of curriculum leadership. A straightforward random sample technique was used to choose the participants. 343 middle school branch teachers made up the study's sample. The researchers created the "Curriculum Leadership Perception Scale," with 21 items to collect quantitative data. The scale's reliability was investigated utilizing McDonald's omega coefficient, item discrimination, and split-half reliability analyses, while its validity was evaluated through face, content, and construct validity. For face and content validity, expert opinion and a pilot study were carried out; EFA and CFA were used for construct validity. While the 213 participants' data were used for EFA, the 130 participants' data were used for CFA. Eighteen items were eliminated following the EFA, and a CFA was carried out. The following fit indices were obtained for the scale by the analysis: RMSEA =.062, GF =.982, NNFI/TL =.929, CFI=.940, and IFI=.941. The reliability coefficient for McDonald's omega was .932. Based on the analyses, a valid and reliable scale with 21 items and four components was created to determine middle school branch teachers' perception of curriculum leadership.

## 1. INTRODUCTION

The curriculum plays a fundamental role in shaping the educational experience, serving as the foundation for the teaching and learning process. It outlines the knowledge, skills, and values learners must acquire. Beyond merely organizing content, curricula reflect the goals of education systems, societies' priorities, and learners' needs. A well-designed curriculum conveys academic knowledge and nurtures critical thinking, creativity, and ethical awareness, empowering individuals to navigate and contribute to an ever-evolving world. In this context, curricula are important in education as a concept that changes and expands over time in the historical process.

Although curriculum began to be used in education as "the path followed," it has had many different definitions up to the present day. Until the 20th century, the definition and scope of the curriculum, regarded as a list of subjects, have changed over time (Demirel, 2021). This transformation in the definition of curriculum parallels the change in educational philosophies. In particular, the prominence of teaching approaches and individual learning processes has led

*CONTACT: İlkay Doğan TAŞ ilkaydogantas@kku.edu.tr Kırıkkale University, Faculty of Education, Department of Curriculum and Instruction, Kırıkkale, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

to the reshaping of the curriculum. It is not only limited to course content but also covers students' entire school experience. In this context, different views have enriched the definition and scope of the curriculum. Ertürk (2013) defined curriculum as "the totality of planned learning experiences designed to educate a specific cohort of students within a defined temporal framework." According to Taba (1962, p.11), the curriculum consists of goals, objective behaviors, content selection and organization, the learning-teaching process, and the evaluation of goals. Pinar *et al.* (1995), who view curriculum as school experiences, state that the curriculum includes not only course subjects but also all other activities in the school, such as meals, games, and social interactions. Curriculum is the general and most important framework that guides teachers and includes the knowledge, skills, and attitudes students need to acquire (Ministry of Education [MEB], 2017). According to definitions and literature, the curriculum is one of the main components of the educational system. The school's education and teaching processes demonstrate how students, teachers, and the curriculum interact. The effectiveness of the interaction between the curriculum and its stakeholders significantly impacts the overall efficiency of the educational system. Therefore, curriculum is among the first concepts addressed to solve problems related to education (Akpınar, 2017). Many factors, such as teachers, students, equipment, school facilities, and environmental conditions, play a role in adequately implementing the curriculum and achieving its objectives. Among these, the teacher, who plays a key role in implementing the curriculum, is more prominent than others.

In this context, teachers who are curriculum implementers have significant responsibilities. An official curriculum loses meaning unless teachers translate it into their own curriculum. Posner (1995) compares a teacher giving life to a curriculum to a director bringing a script to life and states that there are many factors a teacher must consider when implementing the curriculum. Dewey asserts that all teachers must recognize the dignity of their responsibilities and understand their role as social servants in upholding social order and promoting proper social growth (Dewey, 1897, cited in Akbaş & Keskin, 2021, p.302). Teachers should actively lead change rather than passively receive it (Harris *et al.*, 2020). Many studies in the field of leadership and management suggest that teachers should transition from their traditional role as curriculum implementers to curriculum leaders (Ho, 2010). As curriculum implementers, teachers can only accompany these dynamic processes as curriculum leaders in the changing and transforming world (Wiles, 2016).

Curriculum leadership is a perspective on using problem-solving applications involving curriculum, teaching, and leadership issues (Akbaş & Keskin, 2021, p. 310). According to Henderson *et al*. (2018), educators should consider the following ten fundamental principles on their path to becoming curriculum leaders:

**Professional Responsibility.** Measurability and accountability in schools often confuse responsibility, a crucial concept for the teaching profession. Measurability emphasizes test results, and accountability highlights the vertical hierarchical structure in schools. Accountability causes teachers to focus on their weaknesses rather than students, considering rewards or punishments. True professional responsibility can be described as a teacher carefully directing their caregiving skills and competencies toward students (Noddings, 2013).

**Critical Pragmatism.** Curriculum leadership is considered a problem-solving process, requiring teachers to act as critical thinkers. Critical pragmatism is based on visions of what is beautiful, good, and right rather than fixed, structured moral or objective certainties (Cherryholmes, 1988). Teachers with a critical perspective actively participate in solving problems in education rather than ignoring them. The critical perspective greatly contributes to the development of both the teacher and the school (Henderson *et al.*, 2018).

**Continuous Valuation of Values.** Valuing values enables the implementation of critical pragmatism. People need to question and review their values in decision-making processes, as it is impossible to think of them separately from their values. Teachers who guide society should

continuously review their values and curriculum values. Teachers should work by integrating democratic values and realities into solving daily problems (Henderson *et al*., 2018).

**Awareness.** Awareness begins with the individual's search for self-affirmation. It is argued that teachers, as curriculum leaders, should be historically conscious representatives of democratic values, free of narrow-mindedness and fanaticism. Only in this way can it be known that each student has their learning process; individual differences can be respected and it can be accepted that all individuals are unique and indispensable by prioritizing the individual (Henderson *et al*., 2018).

**Non-Linear Process.** Multi-Stage Interaction: Curriculum leadership is not a linear process that achieves results through strict adherence to rules. On the contrary, it encompasses a multi-stage and broad interaction framework. Curriculum leadership requires being dynamic rather than routine (Akbaş & Keskin, 2021; Wiles, 2016).

**Integrating Curriculum, Teaching, and Leadership Eclectically.** According to Schwab (1971), the eclectic approach is a fundamental feature of dynamic curriculum studies because curriculum is a practical field. The connection between curriculum, teaching, and leadership should be well understood and integrated into practice for a holistic educational understanding. To integrate curriculum, teaching, and leadership, academic, political, and managerial problems must be resolved, and teachers should be supported in this regard (Giroux, 1991; Henderson *et al*., 2018).

**The Ethics of Practical Wisdom.** According to Aristotle, virtues are ethical ways of knowing that inspire actions. He asserts that only virtuous wisdom can bring about pleasure and happiness. Practical wisdom means people perform their actions by filtering them through reason, morality, virtue, and value rather than following their impulses. Schools need teachers who want to do what is right with their moral will (Henderson *et al*., 2018).

**Defense Against Crude Pragmatism.** The four-stage process is a disciplined, open-minded inquiry focused on continuous learning. The first stage, professional awakening, questions whether you are experiencing a challenging professional awakening. In the second stage, holistic teaching, it is important to experience the art of interactive teaching that accepts the democratic self and social learning in students. The third stage, productive leadership, involves reorganizing a unique culture that fosters capacity development in teaching. Participatory evaluation, the final stage, aims to answer how this pedagogical journey's outcomes and social impacts can be democratically reviewed (Henderson *et al*., 2018).

**Democratic Learning Leadership.** This structure, which expresses leading learning, is collaborative and community-based. The leader integrates the organization's core values, vision, mission, and philosophy. Transforming schools into learning structures requires fundamental changes in school culture and philosophy. Learning leaders' lifelong learning responsibility is one of their most important strengths (Çelik, 2012).

**Inspiring Professional Development.** Learning leaders should inspire those who follow them (Akbaş & Keskin, 2021). Curriculum leaders need to understand the change process in schools clearly. The power of a curriculum leader comes not from their title but from their ability to direct and persuade those who follow them to seek help (Wiles, 2016). In addition to Henderson *et al*.'s (2018) ten points, Wilesstates that school curriculum leadership consists of four basic tasks. These tasks are defined as follows: defining the school's curriculum, fostering collaboration among all members, providing a working method or path that others can follow, and coordinating activities to achieve the desired curriculum (Wiles, 2016, p. 22).

Defining the curriculum for the school: The curriculum leader's first task is to define the curriculum by identifying the school's goals. The first step in the curriculum definition process is to identify the common views supported by stakeholders such as teachers, parents, administrators, and students. The aim here is to help all stakeholders discover how they want

the curriculum to be (Jacobs, 2004). Once the curriculum leader identifies standard views, they will provide criteria to guide the change, clarify goals, define all school members' roles, and select classroom strategies and tactics (Wiles, 2016).

Fostering collaboration among all school members: Stakeholders must be involved in the planning process to guarantee the successful adoption of the defined curriculum by all. In this context, to develop collaboration among stakeholders, it is necessary to avoid professional jargon and educational clichés that everyone cannot understand; use different tools such as email, written messages, newsletters, and face-to-face meetings to communicate; conduct group work to overcome individuality and ensure partnership; prepare appropriate environments for exchanging ideas and collaboration; and value the diversity (gender, race, language, religion, ability, etc.) that emerges in collaborative work.

Providing a working method or path others can follow: The curriculum leader should envision how the process will progress. Additionally, all stakeholders should understand and follow this vision. Accordingly, the curriculum leader should clarify their ideas, be predictable, regularize all processes, set goals and standards, and continuously receive stakeholder feedback. Coordinating activities to achieve the desired curriculum: At every stage, the curriculum leader should plan, organize, and prepare. They should report every development in the process in a way that all participants can understand using visual tools such as charts and graphs, and develop effective strategies to ensure change in the school. They should relate all actions taken and efforts made in the process to the goals of the curriculum and outline the path to achieving the goal. Given all the principles and tasks emphasized, Henderson *et al*. (2018) and Wiles (2016) present four key dimensions of curriculum leadership. These dimensions include being receptive to the role of a curriculum leader, mastering the curriculum, structuring the curriculum to align with the needs and values of all stakeholders, and advising colleagues on curriculum-related matters while persuading them to collaborate.

Curriculum leadership requires meticulous handling of the content, progression, evaluation, and pedagogy, which are the essence of teaching. This underscores the importance and necessity of curriculum leadership in school and system improvement (Harris *et al*., 2020). Curriculum leadership plays a pivotal role in shaping educational practices and ensuring the effective implementation of curricula. At this point, determining teachers' perceptions and competencies about curriculum leadership is essential. While several scales exist to measure curriculum leadership perception, their applicability and relevance to the Turkish educational context and teachers remain questionable. Four primary scales have been identified in the literature that measure various aspects of curriculum leadership: (1) the School Principals' Curriculum Leadership Scale (SPCLS) developed by Bayirli and Balcı (2021), (2) the Curriculum Leadership Indicators Scale for Elementary School Deans by Chuan-Hsing and Mei-Ju (2016), (3) the Curriculum Leadership Competency Scale by Bolat and Baş (2023), and (4) the Teacher Curriculum Leadership Scale by Chen *et al*. (2021).

The SPCLS developed by Bayirli and Balcı (2021) targets school principals, emphasizing their instructional leadership role rather than assessing teachers' perceptions and competencies in curriculum leadership. Similarly, Chuan-Hsing and Mei-Ju (2016) designed a scale that focuses on the curriculum leadership role of elementary school deans. This administrative-centric approach does not adequately reflect the distributed leadership model that modern educational frameworks advocate. This is because teachers directly responsible for implementing and adapting the curriculum are overlooked. Therefore, a teacher-focused scale is needed to assess how educators perceive and enact curriculum leadership within their professional practice. The other two scales in the literature are geared towards teachers rather than administrators. However, their appropriateness also needs to be discussed.

Bolat and Baş (2023) consider curriculum leadership a sub-dimension, along with organizational leadership, instructional leadership, and assessment and evaluation leadership.

However, curriculum leadership should be considered a main topic, not a sub-dimension. While this perspective provides valuable insights into teachers' understanding of curriculum content, it does not comprehensively capture leadership behaviors, strategies, and decision-making processes. The scale developed by Chen *et al*. (2021) was based on data collected from teachers in China, analyzing curriculum leadership through the lens of field dynamic theory. While informative, this study does not account for the Turkish education system's unique sociopolitical, educational, and cultural dynamics. Türkiye's policies, teacher training programs, and curriculum implementation practices differ significantly from those in China, necessitating a scale that reflects the local educational context.

The existing curriculum leadership scales provide valuable insights into various aspects of leadership in educational settings. However, their focus on administrative roles, their treatment of curriculum leadership as a secondary construct, and their lack of cultural adaptability highlight the need for a new scale tailored to Türkiye. By addressing these gaps, the proposed new Curriculum Leadership Perception Scale (CLPS) will offer a more precise and contextually relevant measurement tool to assess teachers' curriculum leadership perceptions and competencies, ultimately contributing to educational improvement in Türkiye.

This study aims to create a reliable and valid measurement tool to determine middle school branch teachers' perceptions of curriculum leadership by referencing existing literature. The following questions were sought to achieve this aim:

a. Is the scale developed to measure middle school branch teachers' perceptions of curriculum leadership valid?
b. Is the scale developed to measure middle school branch teachers' perceptions of curriculum leadership reliable?
c. Based on a solid theoretical foundation and empirical evidence, what are the dimensions of a reliable and valid curriculum leadership perception scale for middle school branch teachers?

## 2. METHOD

### 2.1. Research Design

A mixed-method exploratory sequential design was employed to develop the CLS for middle school branch teachers. Qualitative and quantitative data collection and analyses were conducted (Creswell & Guetterman, 2018). The scale development process (DeVellis & Thorpe, 2021) was administered to develop CLS for middle school branch teachers in its development phase. In the validation phase, Exploratory Factor Analyses (EFA) were performed using data obtained from Sample 1. Confirmatory Factor Analysis (CFA) was performed using data obtained from Sample 2 to test the validity and reliability characteristics of the scores obtained from CLS.

### 2.2. Development of the CLS for High School Students

#### 2.2.1. *Determine clearly what it is you want to measure*

This study aims to develop a scale to measure middle school branch teachers' perceptions of curriculum leadership. CLS was constructed in four dimensions: curriculum organization, mastery, collaboration with colleagues, and openness to being a curriculum leader. However, these dimensions were determined not at the beginning of the study but by considering the items under the factors as a result of exploratory and confirmatory factor analysis.

#### 2.2.2. *Generate an item pool*

Each item used in the scale should be related to curriculum structure and the concept of curriculum leadership and be directed toward teachers (DeVellis & Thorpe, 2021). For this reason, an item pool was created by considering the relevant literature (Akbaş & Keskin, 2021; Akbaş *et al*., 2021; Glatthorn *et al*., 2005; Harris *et al*., 2020; Henderson *et al*., 2018; Ho, 2010;

Wiles, 2016) and the items in the scale and sub-dimensions used in studies measuring curriculum leadership (Bayirli & Balcı, 2021; Bolat & Baş, 2023; Chen *et al*., 2021; Chuan-Hsing & Mei-Ju, 2016). The items prepared in this direction were presented to experts for their opinions. As a result of the literature review, a 45-item pool was written by the researchers to determine teachers' perceptions of curriculum leadership.

### 2.2.3. *Determine the format for measurement*

Likert-type measurement is a widely used and effective form of obtaining attitudes, beliefs, or opinions (DeVellis & Thorpe, 2021). This study used a four-point Likert scale. According to Johns (2006), the wisdom of offering a midpoint varies depending on the respondent's tendencies. If midpoint respondents are primarily indifferent, omitting the midpoint may overstate minority opinions. Conversely, when respondents use the midpoint to conceal socially undesirable attitudes, offering it can understate those perspectives. Given these considerations, Krosnick (1991) argues that respondents often seek easy escape options when cognitive demands increase. Meanwhile, Berinsky (2004) suggests that non-response is a strategy to avoid cognitive effort and social embarrassment. In line with these insights, this study employs a 4-point Likert scale to minimize central tendency bias and encourage participants to take a clear stance. By removing the neutral option, the scale reduces the likelihood of respondents selecting an easy escape response rather than engaging thoughtfully with the items. The response categories-1 (Does Not Reflect Me at All), 2 (Does Not Reflect Me), 3 (Reflects Me), and 4 (Completely Reflects Me)-were carefully selected to ensure clarity and facilitate meaningful interpretation. Additionally, this format aligns with findings suggesting that forced-choice response structures help mitigate social desirability bias while maintaining ease of response (Krosnick & Presser, 2010). Given the study's focus on teachers, eliminating a neutral option was deemed particularly important to obtain more engaged and authentic responses while reducing cognitive load.

### 2.2.4. *Have initial item pool reviewed by experts*

Two expert groups provided opinions to examine the appropriateness of the CLS's questions and response options. First, four curriculum development and instruction experts, one educational measurement and evaluation expert, and one Turkish language expert reviewed the items in the pool for face and content validity, language, and writing rules. Based on the feedback of these experts, five items were relocated, seven items were removed, and three items were shortened and edited for clarity.

### 2.2.5. *Cognitive interviewing*

The participants' comprehensibility of the items was examined as evidence of the scale's construct validity. The scale was presented to 10 branch teachers to gather their feedback on comprehensibility and ease of response. By the teachers' opinions, no items were removed from the scale, and only minor corrections were made to improve comprehensibility. The scale, revised to 38 items, was ready for application.

### 2.2.6. *Consider inclusion of validation items*

No validation items were used to keep the CLS clear, simple, and concise.

### 2.2.7. *Administer items to sample*

Evidence derived from response processes can be utilized as a form of validity evidence. The internal structure of the responses was investigated to obtain validity evidence of the CLS. Since this method relies on response processes, items were administered to two samples. A simple random sampling technique was used to select the study participants. The target population was middle school branch teachers working in various public schools. Specifically, the sample was drawn from middle schools in Ankara, Türkiye, with official permission from the Ministry of National Education (MoNE). Data was collected through official school WhatsApp groups

managed by school administrators. The scale link was shared within these groups, ensuring only authorized teachers participated voluntarily. This approach enabled access to a wide range of teachers from multiple schools, significantly enhancing the representativeness and generalizability of the findings. The participants' teaching experience ranged from 1 to 30 years. Responses from the first sample were used to explore the internal structure of responses via exploratory factor analysis procedures. A total of 358 middle school branch teachers participated in the study, with 222 in the first and 136 in the second. Participants responded to the scale via Google Forms. The scale link was shared through the WhatsApp groups of the schools through the administrators of the schools that received permission from the Ministry of National Education. Before the scale was administered, participants were given an informed consent form detailing the study's aims, confidentiality measures, and the condition of voluntary participation. Participants who accepted the conditions continued to fill out the scale. After examining the data, we excluded nine participants from the first round who either did not approve the informed consent form (1) or disrupted the normality of the distribution (8) and six from the second round who disrupted the normality of the distribution (6) from the analysis. In the normality of the distribution, kurtosis-skewness values and Q-Q plots were considered, and outliers were excluded from the data set. Therefore, the form consisting of 38 items was applied to the first group of 213 (36.6% men & 63.4 women) teachers in the first round. EFA and internal consistency coefficients were calculated to determine the scale's reliability and number of factors. In the second round, the form consisting of 21 items was applied to the second group of 130 (29.2% men & 70.8% women) teachers. CFA was conducted to confirm the structure determined by the EFA. Ho (2006) states that the sample size for factor analysis should be more than five times the number of items. Similar to this criterion, Bentler and Chou (1987) suggested that the sample size should be at least 5 times more than the number of estimated parameters.

Additionally, Kline (2014) emphasizes that a sample size of 200 is usually sufficient to extract reliable factors. This number can be reduced to 100 in cases where the factor structure is clear. Based on these standards, the sample size reached in the present study is sufficient.

### 2.2.8. *Evaluate the items*

In this study, Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were conducted to evaluate the scale's construct validity. EFA was performed to explore the factor structure and determine which items were loaded onto each factor. CFA was then conducted to confirm the identified factor structure and assess model fit. The internal consistency of the scale was examined using McDonald's omega coefficient. Item discrimination analysis included comparisons of the top and bottom 27% groups. Additionally, a split-half reliability analysis was conducted to evaluate the reliability of the scale scores.

A two-step approach was administered to understand the validity and reliability properties of the CLS scores obtained. The first step involved Exploratory Factor Analysis (EFA) to understand the internal structure of the data obtained from the first sample. The number of dimensions in the data and the items highly related to each dimension are determined. The K1 method, which suggests that the eigenvalue should be greater than 1, and the parallel analysis method developed by Horn (1965) were used for the representation of the sub-dimensions in the scale. Furthermore, the eigenvalue difference among consequent dimensions was investigated. We assumed that no new dimension emerged when the slope of the scree plot became close to flat (Figure 1).

The suitability of the data for factor analysis was tested to perform EFA. A Kaiser-Meyer-Olkin (KMO) value greater than .60 and a significant result from Bartlett's test of sphericity ($p$ =<.05) suggest that the data are suitable for factor analysis (Frankel & Wallen, 2008; Pedhazur & Schmelkin, 1991). According to the analysis, the KMO value was .883, and Bartlett's test of sphericity was 2077.447 ($p$ =.001), yielding a significant result. Hutcheson and Sofroniou

(2006) indicate that KMO values between .8 and .9 indicate excellent suitability, and values above .90 indicate the best fit. These results show that the data are suitable for analysis.

Mardia's multivariate normality test was conducted to assess whether the assumption of multivariate normality was met (Mardia, 1970).

**Table 1.** *Mardia's Test of Multivariate Normality.*

| Measure | Value | Statistic | *df* | *p* |
|---|---|---|---|---|
| Skewness | 143.522 | 5095.036 | 1771 | < .001 |
| Small Sample Skewness | 143.522 | 5173.390 | 1771 | < .001 |
| Kurtosis | 686.387 | 47.752 | | < .001 |

As seen in Table 1, the results indicated that the data significantly deviated from multivariate normality ($p$<.001). Exploratory factor analysis is used to preserve the total variance contained in the measured variables and transform it into a component with fewer variables (Park *et al*., 2002). In cases where the multivariate normality assumption is violated, one appropriate approach for exploratory factor analysis (EFA) is Principal Axis Factoring (PAF). This method is robust against violations of normality assumptions and produces more accurate and stable results by focusing on common variance (Fabrigar *et al*., 1999). Field (2009) recommends that these values be higher than .40. The common variance values of the items in the scale vary between .480 and .844. Accordingly, it was accepted that the common variance values of the items were appropriate for the scale.

**Figure 1.** *A line graph of scale items' eigenvalues.*

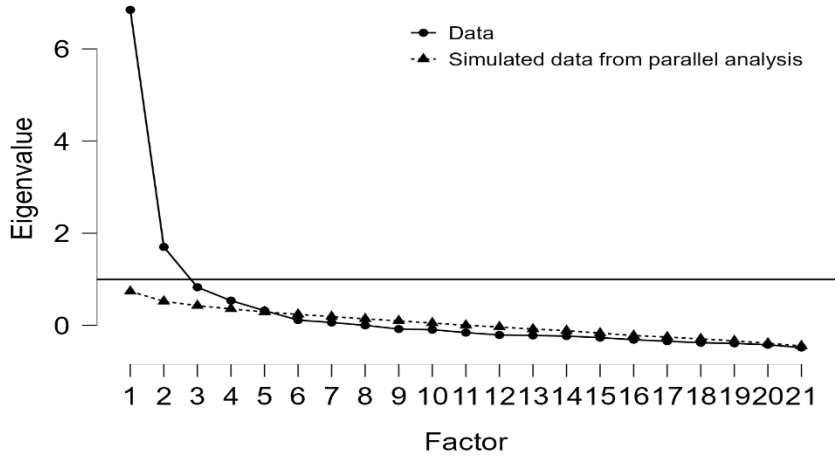



Figure 1 illustrates the four cut-off points that limit the number of factors to four. Table 2 presents the four-factor scale's eigenvalues and variance percentages.

**Table 2.** *Eigenvalues and variance percentages of scale items.*

| Factors | Eigenvalue | % of variance | % of total variance |
|---|---|---|---|
| Factor 1 | 7.503 | 18.4 | 18.4 |
| Factor 2 | 2.393 | 15 | 33.3 |
| Factor 3 | 1.553 | 9.7 | 43.0 |
| Factor 4 | 1.173 | 8.3 | 51.4 |

As seen in Table 2, the first factor contributes 18.4% to the total variance (eigenvalue = 7.503), the second factor contributes 15% (eigenvalue = 2.393), the third factor contributes 9.7% (eigenvalue = 1.553), and the fourth factor contributes 8.3% (eigenvalue = 1.173). As shown in Figure 1, the line graph continues horizontally after the fourth factor, and there is no significant drop between factors. These factors' contribution to the total variance has decreased. A four-

factor model structured the scale, accounting for 51.4% of the total variance. Akbulut (2010) states that the explained variance should be higher than the unexplained variance. Therefore, the variance of the scale explained is suitable.

EFA was conducted on 38 items using the oblimin rotation method. Items to be removed from the scale were determined based on a minimum factor loading of .40. Items were not allowed to load significantly on more than one factor. Additionally, attention was paid to factors consisting of three or more items. Consequently, items 15, 16, 18, 29, 33, and 38 were removed due to overlap; items 8, 9, 10, 12, 19, 24, and 35 were removed due to factor loadings below .40; and items 7, 11, 28, and 31 were removed because their contents did not match the relevant factor. In this direction, the scale was reduced to 21 items.

In the second step, goodness-of-fit indices and scale item fit indices were used for CFA. Robust Maximum Likelihood (MLR) estimation was preferred for CFA. MLR estimation is recommended when normality assumptions are unmet (Muthén & Muthén, 2007). The MLR estimator is resistant to distributional deviations, such as skewness and kurtosis, adjusting parameter estimates and standard errors accordingly to provide more reliable outcomes despite violations of normality (Brown, 2015). For reliability analysis, McDonald's omega analysis was conducted to measure the internal consistency among items at the scale level. Additionally, split-half reliability analysis and item discriminability analysis were performed. Voluntary consent forms were used for this study, and all participants who gave their consent completed the scale form in full. To prevent incomplete data, items were set as mandatory fields in the Google Form system. The data collected for the study were analyzed using the open-source JASP statistical software (R Core Team, 2019).

#### 2.2.9. *Optimize scale length*

The scale length was optimized based on EFA and CFA results presented in the results section.

## 3. RESULTS

The findings section presents the EFA, CFA, and reliability analysis results in order.

### 3.1. Exploratory Factor Analysis

EFA included 21 items, and four factors were obtained. The oblimin rotation method was used to better explain the factor structure in exploratory factor analysis. This method is preferred when factors are correlated (Tabachnick & Fidell, 2007). Table 3 shows the item loadings under these factors due to the oblimin rotation.

The pattern matrix reflects the strength of each item within a specific factor. Field (2009) states that item loadings in the pattern matrix should be greater than .40. These findings deem the factor loadings in the pattern matrix suitable. The total explained variance should be 50% or higher (Thompson, 2004). According to the findings in Table 3, the factor loadings and explained variance in the pattern matrix are deemed suitable. The developed curriculum leadership scale clustered under four factors with eigenvalues greater than 1, explaining 51.4% of the total variance.

The first factor consists of eight items (17, 20, 21, 22, 23, 25, 26, and 27), with loading values ranging from .505 to .775. This factor is labeled "curriculum organization." The second factor consists of six items (1, 2, 3, 4, 5, and 6), with loading values ranging from .559 to .844, indicating "curriculum mastery." The third factor consists of three items (13, 14, and 30), with loading values ranging from .480 to .843, related to "collaboration with colleagues." The fourth factor consists of four items (32, 34, 36, and 37), with loading values ranging from .490 to .695, labeled as "openness to being a curriculum leader" (Table 4). Factors were named considering the literature on curriculum leadership and items (Akbaş & Keskin, 2021; Çelik, 2012; Harris *et al*., 2020; Henderson *et al*., 2018; Noddings, 2020; Wiles, 2016).

**Table 3.** *The items under the factors and the variance they explain as a result of oblimin rotation.*

| | | Factors | | | |
|---|---|---|---|---|---|
| Factors | Items | 1 | 2 | 3 | 4 |
| Curriculum organization | I27 | .775 | | | |
| | I22 | .766 | | | |
| | I26 | .743 | | | |
| | I23 | .637 | | | |
| | I17 | .628 | | | |
| | I20 | .578 | | | |
| | I21 | .550 | | | |
| | I25 | .505 | | | |
| Curriculum mastery | I2 | | .844 | | |
| | I3 | | .717 | | |
| | I1 | | .716 | | |
| | I4 | | .648 | | |
| | I5 | | .645 | | |
| | I6 | | .559 | | |
| Collaboration with colleagues | I14 | | | .843 | |
| | I13 | | | .789 | |
| | I30 | | | .480 | |
| Openness to being a curriculum leader | I36 | | | | .695 |
| | I37 | | | | .641 |
| | I32 | | | | .521 |
| | I34 | | | | .490 |
| Eigenvalues | | 7.503 | 2.393 | 1.553 | 1.173 |
| Explained variance (%) | | 18.4 | 15 | 9.7 | 8.3 |
| Total variance (%) | | 51.4 | | | |

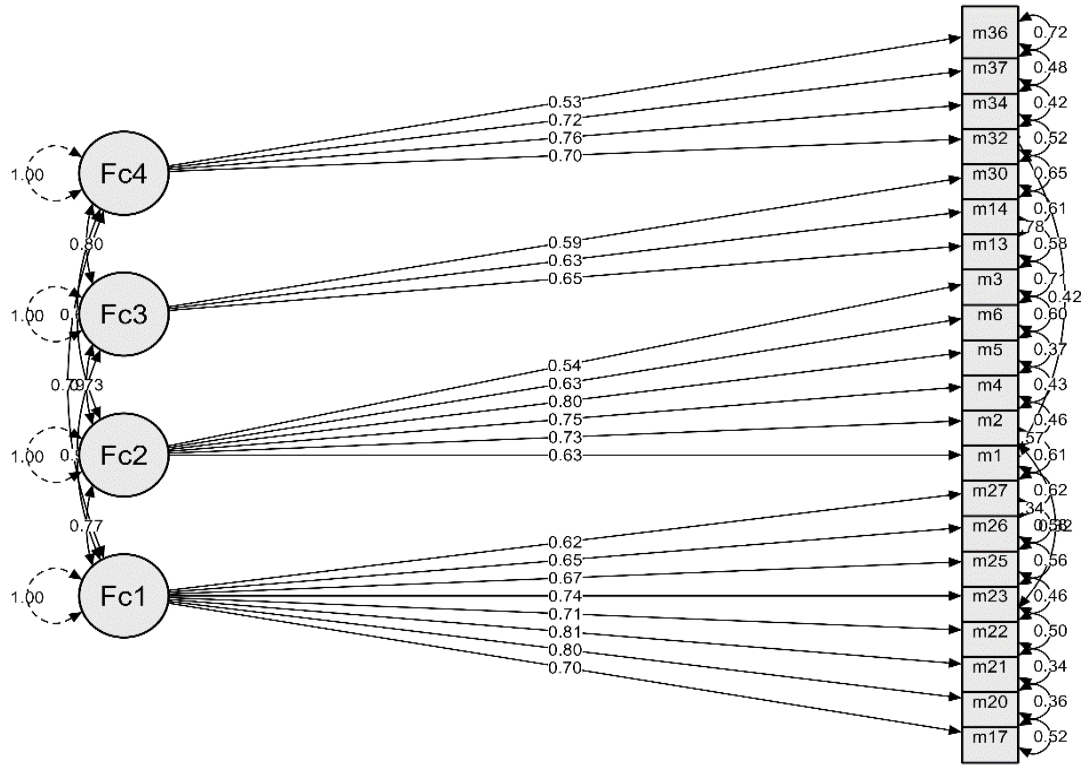
**Table 4.** *Sample items on the scale.*

| Curriculum leadership perception scale statements | Rotated factor loadings |
|---|---|
| 26 I make arrangements in the curriculum in line with the individual development needs of students. | .743 |
| 5 I have a good command of current curriculum design and teaching best practices. | .648 |
| 30 I guide my colleagues in the implementation of the curriculum. | .480 |
| 37 I think I should participate in the curriculum's leading, directing, and decision-making processes. | .641 |

### 3.2. Confirmatory Factor Analysis

The curriculum leadership scale, previously determined to have four dimensions in EFA, was tested using CFA with a different sample of 130 participants. The analysis results showed the goodness-of-fit indices and the covariance between items “1 and 2,” “13 and 14,” “2 and 22,” “26 and 27,” and “27 and 34” in the model presented in Figure 2.

**Figure 2.** *The model for Confirmatory Factor Analysis of the scale items.*




The chi-square test was statistically significant ($\chi^2$ = 267.442, *df* = 178, *p* <.001). However, it is well documented that the chi-square test is sensitive to large sample sizes and tends to produce significant results even when the model is acceptable (Kline, 2014). Consequently, additional fit indices were examined. The Comparative Fit Index (CFI = .940) and Tucker-Lewis Index (TLI = .929) exceeded the recommended threshold of ≥ .90, suggesting a good model fit (Hu & Bentler, 1999; Kline, 2014). Similarly, the Incremental Fit Index (IFI = .941) indicated a strong model fit. Furthermore, the Root Mean Square Error of Approximation (RMSEA) was calculated as .062 (90% CI [.046, .070]), signifying an acceptable level of fit. RMSEA values below .08 are considered satisfactory, whereas values ≤ .06 indicate a good fit (Browne & Cudeck, 1993). Thus, the obtained RMSEA (.062) falls within acceptable boundaries. The Standardized Root Mean Square Residual (SRMR = .056) also indicated an acceptable model fit below the .08 threshold (Hu & Bentler, 1999). Although the Goodness of Fit Index (GFI = .842) slightly fell below the conventional threshold of ≥ .90, it remained acceptable. Meanwhile, McDonald's Fit Index (MFI=.709) indicated moderate fit. Additionally, Hoelter's critical values at both .05 (103.134) and .01 (110.277) significance levels suggested an adequate sample size.

Based on the fit values in Table 5, the confirmatory factor analysis results indicate that the proposed model demonstrated acceptable to a good fit based on recommended fit index criteria in the literature (Browne & Cudeck, 1993; Hu & Bentler, 1999; Kline, 2014).

**Table 5.** *The fit values obtained in CFA.*

| N | $\chi^2$ | *df* | $\chi^2/df$ | RMSEA | SRMR | TLI | GFI | MFI | CFI | IFI |
|---|---|---|---|---|---|---|---|---|---|---|
| 130 | 267.442 | 178 | 1.502 | .062 | .056 | .929 | .842 | .709 | .940 | .941 |

**Table 6.** *Summary of confirmatory factor analysis results.*

| Factors | Item No | Factor Loading | Standard Error | *z* | *p* |
|---|---|---|---|---|---|
| F1. Curriculum organization | m17 | .346 | .039 | 8.828 | < .001 |
| | m20 | .456 | .043 | 10.702 | < .001 |
| | m21 | .485 | 044 | 11.004 | < .001 |
| | m22 | .412 | .045 | 9.172 | < .001 |
| | m23 | .394 | .041 | 9.525 | < .001 |
| | m25 | .427 | .051 | 8.330 | < .001 |
| | m26 | .333 | .042 | 8.017 | < .001 |
| | m27 | .358 | .047 | 7.594 | < .001 |
| F2. Curriculum mastery | m1 | .356 | .048 | 7.405 | < .001 |
| | m2 | .398 | .043 | 9.250 | < .001 |
| | m4 | .436 | .046 | 9.506 | < .001 |
| | m5 | .449 | .044 | 10.300 | < .001 |
| | m6 | .398 | .053 | 7.540 | < .001 |
| | m3 | .282 | .045 | 6.243 | <.001 |
| F3. Collaboration with colleagues | m13 | .358 | .051 | 7.014 | < .001 |
| | m14 | .357 | .052 | 6.815 | < .001 |
| | m30 | .388 | .060 | 6.472 | < .001 |
| F4. Openness to being a curriculum leader | m32 | .362 | .043 | 8.401 | < .001 |
| | m34 | .416 | .044 | 9.527 | < .001 |
| | m37 | .425 | .049 | 8.746 | < .001 |
| | m36 | .307 | .051 | 6.025 | < .001 |
| Correlation among factors | F1 ↔ F2 | .773 | .049 | 15.689 | < .001 |
| | F1 ↔ F3 | .976 | .075 | 13.075 | < .001 |
| | F1 ↔ F4 | .789 | .051 | 15.615 | < .001 |
| | F2 ↔ F3 | .732 | .094 | 7.813 | < .001 |
| | F2 ↔ F4 | .614 | .075 | 8.204 | < .001 |
| | F3 ↔ F4 | .805 | .092 | 8.758 | < .001 |

The CFA indicated strong and significant factor loadings for all items ($p$ < .001, Table 6). Additionally, also the correlations among the factors were moderate to high: Factor 1 showed strong correlations with Factor 2 ( $r$= .773) and Factor 4 ($r$ = .789), Factor 2 had moderate correlations with Factor 3 ($r$ = .732) and Factor 4 ($r$ = .614), and Factor 3 was strongly correlated with Factor 4 ($r$ = .805). However, the correlation between Factor 1 and Factor 3 was very high ($r$=.976, $p$<.001). Although this correlation exceeds the typical limit (.85) recommended by Kline (2014), it does not necessarily imply that these two factors measure the same concept. As explained in the discussion and conclusion section, theoretical evidence and practical considerations demonstrate that Factors 1 and 3 measure distinct aspects of the broader construct. In social science research, closely related factors often have high correlations, yet they still represent separate, meaningful dimensions (Brown, 2015; Hair *et al.*, 2019). Combining these two factors could lead to a loss of valuable detail and reduce the clarity of the findings. Therefore, it is considered that the factors should remain separate.

These results indicate a coherent structure among the factors, indicating related yet distinguishable dimensions, which is common in social sciences research. Accordingly, the

correlations between the factors support the construct validity of the model (Brown, 2015; Hair *et al*., 2019).

### 3.3. Findings Regarding the Reliability of the Scale

In light of this study's second research question, the reliability of the scale's items and factors was analyzed using McDonald's omega coefficient, the item discriminability method by comparing the 27% sub-upper groups, and split-half reliability analyses. Regarding scale reliability, McDonald's omega coefficient value should be equal to or greater than .70 (Frankel & Wallen, 2008).

**Table 7.** *Factors' reliability coefficients.*

| Factors | McDonald's omega ($\alpha$) | Number of Items |
|---|---|---|
| 1 | .893 | 8 |
| 2 | .850 | 6 |
| 3 | .799 | 3 |
| 4 | .774 | 4 |
| Total | .933 | 21 |

The reliability coefficients for the factors were calculated as follows: Factor 1 (curriculum organization), .893; Factor 2 (curriculum mastery), .850; Factor 3 (collaboration with colleagues), .799; Factor 4 (openness to being a curriculum leader), .774. The overall McDonald's omega coefficient for the scale was .933 (Table 7). The scale's reliability coefficients and its factors above .70 indicate that it is a reliable measurement tool.

### 3.4. Item Discrimination Analysis

An independent sample t-test examined the scale's item discrimination feature. The perception scores of the 27% lower and upper groups were compared (Table 8).

**Table 8.** *Item discrimination analysis.*

| Measure | N | Mean | *SD* | *df* | *t* | *p* |
|---|---|---|---|---|---|---|
| Lower %27 | 35 | 2.72 | 0.159 | 68 | 19.23 | .000 |
| Upper %27 | 35 | 3.63 | 0.229 | | | |

The analysis showed a statistically significant difference between the 27% lower and upper groups ($t$ = 19.23, $p$ = .00, Table 8). The developed curriculum leadership perception scale can significantly differentiate between the lower and upper groups, demonstrating strong discriminative power.

### 3.5. Split-Half Reliability

The Split-Half method, another internal consistency coefficient calculation method, was used to determine the scale's reliability. This method was conducted separately for the EFA-applied group and the CFA-applied group.

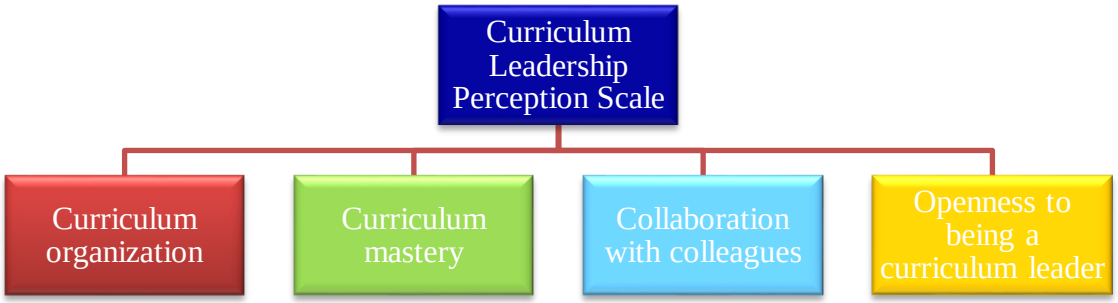
**Table 9.** *Split-half analysis for both the EFA and CFA groups.*

| Analyses | Half | Cronbach's Alpha | Split-Half Reliability | |
|---|---|---|---|---|
| | | | Guttman Coefficient | Spearman-Brown Coefficient |
| EFA | First | .848 | .840 | .841 |
| | Second | .837 | | |
| CFA | First | .890 | .880 | .881 |
| | Second | .875 | | |

Table 9 shows that Cronbach's Alpha internal consistency coefficients ranged from .837 to .890, the Guttman coefficients from .840 to .880, and the Spearman-Brown coefficients from .841 to .881. In scale development processes, reliability coefficients of .70 and higher are considered to have sufficient reliability (Carpenter, 2017; Hooper *et al*., 1992; Frankel & Wallen, 2008). The values determined for the EFA and CFA groups show that the EPL scale has good internal consistency.

The analyses concluded that the CLPS for branch teachers is a valid and reliable measurement tool. The scale's sub-dimensions, as shown in Figure 3, align with the study's third research question.

**Figure 3.** *Sub-dimensions of the Curriculum Leadership Perception Scale.*




## 4. DISCUSSION and CONCLUSION

This study aims to create a valid and reliable measurement tool to measure middle school branch teachers' perceptions of curriculum leadership. 358 middle school branch teachers participated in the validity and reliability study of the developed scale. Six experts evaluated the 45-item scale draft, leading to the removal of seven items. A pilot study was conducted with ten branch teachers on the 38-item trial form of the scale, and items that were not understood were revised based on feedback. The pilot application did not result in any items being removed from the scale. Following expert evaluation and pilot testing, the scale underwent refinement to enhance clarity and relevance.

### 4.1. Psychometric Properties

The scale development process involved exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) to ensure construct validity. The KMO value of .883 indicated that the sample size was sufficient for EFA, and Bartlett's test of sphericity confirmed the sample's suitability for factor analysis. Care was taken to ensure that each factor consisted of at least two items. Four factors emerged from the oblimin rotation analysis: curriculum organization, mastery, collaboration with colleagues, and openness to being a curriculum leader. The EFA results showed that the four factors explained 51.40% of the scale variance. Consequently, items 15, 16, 18, 29, 33, and 38 were removed due to overlap; items 8, 9, 10, 12, 19, 24, and 35 were removed due to factor loadings below .40; and items 7, 11, 28, and 31 were removed because their contents did not match the relevant factor.

CFA was conducted to confirm the four-factor model. The fit indices of the four-factor model are as follows: $\chi^2/df$ = 1.502; RMSEA = .062; SRMR = .056; GFI = .842; TLI = .929; CFI = .940; MFI = .709; IFI = .941. These values indicate a good fit for the model. We calculated the overall reliability coefficient of the scale as .933, which indicates satisfactory internal consistency. Internal consistency analysis results for each subscale are as follows: The values for Factor 1 are .893, Factor 2 is .850, Factor 3 is .799, and Factor 4 is .774. These values reflect acceptable levels of internal consistency for each factor. These findings suggest that the scale is theoretically and statistically sound for evaluating curriculum leadership perceptions among middle school teachers.

### 4.2. Comparison with Existing Instruments and Factor Structure

Several existing instruments measure curriculum leadership but primarily focus on administrative or principal-led perspectives. For instance, Wang and Chou (2016) proposed curriculum leadership indicators using the Delphi methodology, emphasizing administrative roles. Similarly, Bayırlı and Balcı (2021) developed a scale targeting principals' curriculum leadership roles within a hierarchical school structure. Additionally, Chen *et al*. (2021) investigated teacher curriculum leadership through field dynamic theory, highlighting the influence of individual and environmental factors. Bolat and Baş (2023) also emphasized the significance of curriculum leadership for educators, considering it a sub-dimension within their scale. Unlike these instruments, the CLPS developed in this study explicitly evaluates teachers' perspectives on curriculum leadership. This distinction is critical, as teachers play a pivotal role in curriculum implementation and adaptation. The scale fills a crucial gap in research by focusing on teacher agency. As a result of the analyses, the final version of the scale consists of 21 items and four dimensions (Appendix A1).

Curriculum leadership is critical to educational effectiveness and influences how the curriculum is designed, implemented, and improved. Curriculum leadership includes developing a vision, exceeding standards, aligning curriculum to stakeholders, conducting authentic assessments, establishing a functioning curriculum team, engaging administrators, teachers, parents, and other school stakeholders, and planning and managing change (Wiles, 2016, p. 10). The CLPS, developed to assess teachers' perceptions of curriculum leadership, identifies four factors that closely overlap with these tasks at the core of curriculum leadership. These dimensions align with the literature on curriculum leadership (Akbaş & Keskin, 2021; Harris *et al*., 2020; Henderson *et al*., 2018; Noddings, 2020; Wiles, 2016).

One of the most fundamental aspects of curriculum leadership is organizing and structuring the curriculum effectively (Wiles, 2016). The curriculum organization factor in the scale aligns with this principle, emphasizing teachers' roles in adapting and structuring curricula. Wiles (2016, p. 37) argues that curriculum leaders must define the school's curriculum by identifying goals and ensuring alignment with educational standards. Similarly, Jacobs (2004, p. 45) highlights the importance of stakeholder involvement in shaping curriculum structures, reinforcing the necessity of organized curriculum leadership.

Beyond organization, effective curriculum leadership requires mastery of the curriculum itself. Teachers must have deep knowledge of curriculum design, instructional strategies, and assessment methods to effectively implement and refine curricula (Harris, Jones, & Crick, 2020). Effective curriculum leaders should have expertise in curriculum content and pedagogy to guide their teaching practices (Wiles, 2016). This aligns with Posner's (1995, p. 56) assertion that teachers play a critical role in interpreting and realizing the curriculum. Moreover, Henderson *et al*. (2018, p. 82) emphasize the importance of continuous professional learning in enhancing teachers' curricular expertise. As curriculum leaders, teachers must continuously update their knowledge and skills to ensure their instructional practices remain practical and relevant.

Equally crucial in curriculum leadership is fostering collaboration among educators. Collaboration among educators is a cornerstone of curriculum leadership, fostering shared decision-making and collective improvement (Ho, 2010). Collaboration with colleagues underscores the necessity of teamwork in curriculum development and implementation. Wiles (2016) suggests that curriculum leaders should actively promote collaboration to create cohesive instructional practices. Furthermore, research indicates that collaboration enhances curriculum coherence, facilitates professional learning, and contributes to overall educational improvement (Henderson *et al*., 2018). When teachers work together, they can share insights, refine their approaches, and create a stronger learning environment for students.

The willingness to assume curriculum leadership responsibilities is essential for fostering educational change (Ho, 2010). Leading curriculum development involves cooperation, critical thinking, and creativity (Noddings, 2013). Wiles (2016) further emphasizes the role of leadership in driving curriculum reforms and fostering instructional innovation. Teachers must recognize their potential as curriculum leaders and actively engage in curriculum development, implementation, and evaluation for meaningful change.

### 4.3. Contribution to International and National Literature

This research's primary contribution to the national and international literature is the development of a teacher-centered curriculum leadership scale. While existing studies have explored curriculum leadership from an administrative or institutional perspective, the current study highlights teachers' roles as active curriculum leaders. The findings underscore the necessity of empowering teachers as curriculum decision-makers, fostering a more decentralized and dynamic approach to curriculum leadership. Furthermore, by providing a reliable and valid instrument to assess perceptions of curriculum leadership, the scale may facilitate longitudinal studies examining the changing role of teachers in curriculum leadership. Given the increasing emphasis on teacher agency in curriculum development, this study offers a new perspective that bridges the gap between theoretical leadership models and practical classroom applications.

Since the proclamation of the Republic, many curriculum development studies have been carried out for different levels and disciplines in Türkiye, and numerous curricula have been implemented. However, each curriculum has encountered a variety of critiques. Consequently, minor adjustments or fundamental changes have been made in a single discipline, at one level of education, or across all levels from primary to high school (Çobanoğlu & Yıldırım, 2021). This indicates that the curriculum implemented in Türkiye has not achieved stability or has had short-lived stability. The stability of an implemented curriculum depends on its ability to meet the needs of individuals, society, and the subject area; reflect national culture; fulfill contemporary requirements; and, most importantly, have competent teachers to implement it.

Teachers are informed and supported about the curriculum implemented through pre-service education at the undergraduate level and in-service training during their careers. However, the constant changes in curriculum and the large number of teachers make this process challenging. In this context, teachers must master the curriculum to implement it correctly, make decisions and adjustments when necessary, inform and guide their colleagues about it, and ultimately lead it. Determining the status of all teachers in terms of leading the curriculum and identifying their needs may enable the preparation of in-service training programs to develop them.

### 4.4. Interpretation of Scale Scores

The scale is scored on a 4-point Likert scale as “1: It does not reflect me at all”, “2: It does not reflect me”, “3: It reflects me”, and “4: It reflects me completely”. Scores for the factors of the scale should be calculated by averaging the scores of the items belonging to each factor. The following cut-off points can be used in the interpretation of average scores: 1.00-1.74 should be interpreted as “Very Low Perception,” 1.75-2.49 as “Low Perception,” 2.50-3.24 as “High Perception,” and 3.25-4.00 as “Very High Perception.”

Accordingly, the high scores obtained from the scale indicate that the participants' perceptions of curriculum leadership are high. In particular, each of the dimensions, such as “Curriculum Organization,” “Curriculum Dominance,” “Collaboration with Colleagues,” and “Openness to Being a Curriculum Leader,” should be interpreted separately. The average score of each dimension should be used to determine the strengths and weaknesses of individuals in the relevant dimension. Interpreting the scale in this way will provide important data for improving curricula and strengthening educational leadership practices.

### 4.5. Future Research Directions

As a consequence of the CLPS application, there are some recommendations for future research. Future research can extend the application of the CLPS by exploring its use in diverse educational contexts, including elementary and high school settings. Additionally, studies examining the impact of professional development initiatives on teachers' curriculum leadership perceptions can further validate the scale's applicability. Individual differences in curriculum literacy, gender, seniority, education level, and age can also be examined in future studies. This study contributes to ongoing discussions on educational reform and instructional innovation by positioning teachers as central figures in curriculum leadership.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Kırıkkale University, Social Sciences and Humanities Research Ethics Committee, 12-02.

## Contribution of Authors

**Cemile Keski**: Investigation, Literature review, Data collection, Data interpretation, and Writing-original draft. **İlkay Doğan Taş**: Literature review, Methodology, Data analysis, Data interpretation, Writing-original draft, Supervision, and Validation.

## Orcid

Cemile Keski https://orcid.org/0000-0002-6660-5016
İlkay Doğan Taş https://orcid.org/0000-0002-1418-1688

## REFERENCES

Akbaş, O., & Keskin, A. (2021). Eğitim programı liderliği [Curriculum leadership]. In G. Ekici, & S.M. Özdemir (Eds.), *Eğitimde program geliştirme ve değerlendirme* [*Curriculum development and evaluation in education*]. Nobel Academic Publishing.

Akbaş, O., Keskin, A., & Duman, S.N. (2021). Eğitim programı liderliği yeterlilikleri üzerine bir araştırma [A study on curriculum leadership competencies]. *Turkish Journal of Primary Education*, *6*(2), 157-179. https://doi.org/10.52797/tujped.1014824

Akbulut, Y. (2010). *Sosyal bilimlerde SPSS uygulamaları [SPSS applications in social sciences]*. İdeal Kültür Publishing.

Akpınar, B. (2017). *Eğitimde program geliştirme* [*Curriculum development in education*]. Data Publishing.

Bayirli, A., & Balcı, A. (2021). Okul müdürlerinin eğitim programı liderliğini belirlemeye yönelik bir ölçek geliştirme çalışması [Developing a scale to determine curriculum leadership of school principals]. *The Journal of Turkish Educational Sciences*, *19*(2), 1252-1276. https://doi.org/10.37217/tebd.938818

Bolat, Y., & Baş, M. (2023). İki önemli öğretmenlik mesleği yeterliği: eğitim programı okuryazarlığı ve eğitim programı liderliği [Two important teaching professional competencies: Curriculum literacy and curriculum leadership]. *Journal of Education for Life*, *37*(2), 294-330. https://doi.org/10.33308/26674874.2023372541

Bentler, P.M., & Chou, C.P. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Berinsky, A. (2004). *Silent voices: Public opinion and political participation in America*. Princeton University Press.

Brown, T.A. (2015). *Confirmatory factor analysis for applied research*. Guilford Press.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage Publications.

Carpenter, S. (2017). Ten steps in scale development and reporting: A guide for researchers. *Communication Methods and Measures, 12*(1), 25-44. https://doi.org/10.1080/19312458.2017.1396583

Chen, J., Yang, Y., Xu, F., Xu, W., Zhang, X., Wang, Y., & Zhang, Y. (2021). Factors Influencing Curriculum Leadership of Primary and Secondary School Teachers from the Perspective of Field Dynamic Theory: An Empirical Investigation in China. *Sustainability, 13*, 12007. https://doi.org/10.3390/su132112007

Cherryholmes, C.H. (1988). *Power and criticism: Poststructural investigations in education*. Teachers College Press.

Chuan-Hsing, W., & Mei-Ju, C. (2016). A study on curriculum leadership indicators regarding elementary school dean of academic affairs. *European Journal of Research and Reflection in Educational Sciences, 4*(8), 20–48.

Creswell, J., & Guetterman, T.C. (2018). *Educational research: planning, conducting, and evaluating quantitative and qualitative research*. (6th Ed.). Pearson Education Limited.

Çelik, V. (2012). *Eğitimsel liderlik* [*Educational leadership*]. Pegem Publishing.

Çobanoğlu, R., & Yıldırım, A. (2021). Türkiye'de Program Geliştirme Çalışmaları: Cumhuriyet'in İlanından Günümüze Tarihsel Bir Analiz [Curriculum Development Studies in Türkiye: A Historical Analysis from the Declaration of Republic to Present] *The Journal of Turkish Educational Science*, *19*(2), 810-830. https://doi.org/10.37217/tebd.912329

de Winter, J.C.F., & Dodou, D. (2012). Response to commentary on "The driver behaviour questionnaire as a predictor of accidents: A meta-analysis". *Journal of Safety Research, 43*(1), 85–90. https://doi.org/10.1016/j.jsr.2011.06.013

Demirel, Ö. (2021). *Eğitimde program geliştirme* [*Curriculum development in education*]. Pegem Academic Publishing.

DeVellis, R.F., & Thorpe, C.T. (2021). *Scale development: Theory and applications*. Sage Publications, Thousand Oaks.

Dewey, J. (1897). My pedagogic creed. *The School Journal, 54*, 77–80.

Ertürk, S. (2013). *Eğitimde program geliştirme* [*Curriculum development in education*]. Edge Academic Publishing.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C., & Strahan, E.J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272-299.

Field, A. (2009) Discovering Statistics Using SPSS. 3rd Edition, Sage Publications Ltd., London.

Frankel, J.R., & Wallen, N.E. (2008). *How to design and evaluate research in education*. McGraw Hill Higher Education.

Giroux, H.A. (1991). Border pedagogy and the politics of postmodernism. *Social Text, 28*, 51–67. https://doi.org/10.2307/466376

Glatthorn, A.A., Boschee, F., & Whitehead, B.M. (2005). *Curriculum leadership: Development and implementation*. Sage Publications.

Harris, A., Jones, M., & Crick, T. (2020). Curriculum leadership: a critical contributor to school and system improvement. *School Leadership & Management, 40*(1), 1-4. https://doi.org/10.1080/13632434.2020.1704470

Henderson, J.G., Castner, D.J., & Schneider, J.L. (2018). *Democratic curriculum leadership: critical awareness to pragmatic artistry*. Rowan & Little.

Ho, D.C.W. (2010). Teacher participation in curriculum and pedagogical decisions: Insights into curriculum leadership. *Educational Management Administration & Leadership, 38*(5), 613–624. https://journals.sagepub.com/doi/10.1177/1741143210373739

Ho, R. (2006). *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. Chapman and Hall/CRC.

Hooper, D., Coughlan, J., & Mullen, M.R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, *6*(1): 53–60. https://doi.org/10.21427/D7CF7R

Horn, J.L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hutcheson, G., & Sofroniou, N. (2006). *The multivariate social scientist: introductory statistics using generalized linear models* (2nd ed.). Sage

Jacobs, H.H. (2004). *Getting results with curriculum mapping*. Association for Supervision.

Johns, R. (2005). One Size doesn't fit all: Selecting response scales for attitude items. *Journal of Elections, Public Opinion and Parties*, *15*(2), 237-264. https://doi.org/10.1080/13689880500178849

Kline, P. (2014). *An easy guide to factor analysis* (1st ed.). Routledge. https://doi.org/10.4324/9781315788135

Kline, R.B. (2023). *Principles and practice of structural equation modeling* (4th ed.). The Guilford Press.

Krosnick, J. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, *5*, 213-36. https://doi.org/10.1002/acp.2350050305

Krosnick, J.A., & Presser, S. (2010). Question and questionnaire design. In P.V. Marsden & J.D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263–313). Emerald Group Publishing.

Mardia, K.V. (1970). Measures of Multivariate Skewness and Kurtosis with Applications. *Biometrika*, *57*, 519-530. https://doi.org/10.1093/biomet/57.3.519

Millî Eğitim Bakanlığı [MEB]. (2017). *Öğretmenlik mesleği genel yeterlikleri* [*General competencies for teaching profession*]. https://oygm.meb.gov.tr/dosyalar/StPrg/Ogretmenlik_Meslegi_Genel_Yeterlikleri.pdf

Muthén, L.K., & Muthén, B.O. (2007). *Mplus user's guide*. Muthén & Muthén.

Noddings, N. (2013). *Education and democracy in the 21st. century*. Teachers College, Columbia University.

Park, H.S., Dailey, R., & Lemus, D. (2002). The use of exploratory factor analysis and principal component analysis in communication research. *Human Communication Research*, *28*(4), 562–577. https://doi.org/10.1111/j.1468-2958.2002.tb00824.x

Pinar, W., Reynolds, W., Slattery, P., & Taubman, P. (1995). *Understanding curriculum: An introduction to the study of historical and contemporary curriculum discourses*. Peter Lang.

Posner, G.J. (1995). *Analyzing the curriculum*. McGraw-Hill, Inc.

R Core Team. (2019). *R: A language and environment for statistical computing* [*Computer software manual*]. *http://www.R-project.org/*

Schwab, J.J. (1971). The practical: Arts of eclectic. *The School Review*, *79*, 493-542. https://doi.org/10.1086/442998

Taba, H. (1962). *Curriculum development: Theory and practice*. Brace ve World, Inc.

Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics*. Allyn & Bacon.

Thompson, B. (2004). Exploratory and confirmatory factor analysis: Understanding concepts and applications. *Applied Psychological Measurement*, *31*(3), 245-248. https://doi.org/10.1177/0146621606290168

Wiles, J. (2016). *Eğitim program liderliği* [*Leading curriculum development*]. (B. Acat, Trans. Ed.). Nobel Academic Publishing.

## APPENDIX

### A1. The Perception Scale for Curriculum Leadership (Eğitim Program Liderliği Algı Ölçeği)

| Maddeler (Orijinal Ölçek Madde Numaraları) | Beni hiç yansıtmıyor | Beni yansıtmıyor | Beni yansıtıyor | Beni tamamen yansıtıyor |
|---|---|---|---|---|
| 1 (1). Öğretim programlarına etki eden güncel eğitim politikalarına hâkimim. | | | | |
| 2 (2). Öğretim programının özel amaçlarına hâkimim. | | | | |
| 3 (3). Öğretim programında yer alan kazanımlara hâkimim. | | | | |
| 4 (4). Program tasarımı ve öğretimdeki güncel iyi uygulamalara hâkimim. | | | | |
| 5 (5). Öğretim programlarına etki edebilecek eğitimdeki eğilimleri takip ederim. | | | | |
| 6 (6). Öğretim programı politikalarını, öğrencilere ve velilere uygun bir şekilde açıklarım. | | | | |
| 7 (13). Meslektaşlarımla, öğretim programları hakkında anahtar yetkinliklere göre fikir alışverişinde bulunurum. | | | | |
| 8 (14). Meslektaşlarımla, öğretim programındaki kök değerlerin kazandırılmasına yönelik fikir alışverişinde bulunurum. | | | | |
| 9 (30). Öğretim programının uygulanmasında meslektaşlarıma rehberlik yaparım. | | | | |
| 10 (17). Öğrencilerin bireysel gelişimlerine rehberlik etmek için öğretim programının uygulanmasında öğretimle ilgili yeniliklerden yararlanırım. | | | | |
| 11 (20). Öğretim programının uygulanma sürecinde öğrencide meydana gelen gelişimi programın özel amaçlarına göre değerlendiririm. | | | | |
| 12 (21). Öğretim programının etkililiğini öğretimle ilgili verilere (akademik başarı, öğrenci gelişimi vb.) dayalı olarak değerlendiririm. | | | | |
| 13 (22). Öğrenme ve öğretme süreçlerini öğretim programının vizyon ve misyonuna göre düzenlerim. | | | | |
| 14 (23). Öğretim programıyla ilgili planlamalar yaparken, programın uygulamadaki yansımalarını dikkate alırım. | | | | |
| 15 (25). Öğretim programının iyileştirilmesine yönelik uzun vadeli planlamalar yaparım. | | | | |
| 16 (26). Öğrencilerin bireysel gelişim ihtiyaçları doğrultusunda öğretim programında düzenleme yaparım. | | | | |
| 17 (27). Dersin kazanımlarına ve alanımdaki diğer ilgili gelişmelere göre öğretim programı üzerinde düzenlemeler yaparım. | | | | |
| 18 (32). Öğretim programlarını uygulama sürecinde yeni fikirlere açığım. | | | | |
| 19 (34). Öğretmenlerin eğitim programına liderlik etme, yönlendirme ve karar alma süreçlerine katılmasının öğrenci yetkinliklerinin geliştirilmesine katkı sağlayacağını düşünüyorum. | | | | |
| 20 (36). Öğretmenlerin, idari pozisyonlara sahip olmasalar bile eğitim programına liderlik etme, yönlendirme ve karar alma süreçlerine katılabileceğini düşünüyorum. | | | | |
| 21 (37). Eğitim programına liderlik etme, yönlendirme ve karar alma süreçlerine katılmam gerektiğini düşünüyorum. | | | | |

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 387–413
https://doi.org/10.21449/ijate.1478068
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# Examining the level of culturally responsive school culture

**Gürkan Sarıdaş**[1*], **Funda Nayır**[2]

[1]Ministry of National Education, Denizli, Türkiye
[2]Ağrı İbrahim Çeçen University, Faculty of Education, Department of Educational Sciences, Ağrı, Türkiye

**ARTICLE HISTORY**
*Received: May 3, 2024*
*Accepted: Nov. 6, 2024*

**Keywords:**
Culturally responsive school culture,
Changing school culture,
Mixed method.

**Abstract:** Culturally responsive school culture is a type of school culture in which culturally responsive education is implemented. The purpose of the study was to define culturally responsive school culture, to identify its components, and to determine a roadmap for school administrators to transform school culture into a culturally responsive school culture. The research was designed with mixed methods. In accordance with the sequential exploratory design, qualitative data were first collected through a semi-structured interview form and in-depth research was conducted on an unknown situation. Then, quantitative data were collected through the Culturally Responsive School Culture Scale. Through the mixed method, the generalizability of qualitative data with quantitative data was examined. The research was conducted on teachers working in randomly selected provinces from Statistical Regional Units. In the qualitative part, there were 91 participants in total, including school administrators, teachers, students, and parents, while there were 541 teachers and 397 administrators in the quantitative part. As a result of the research, it was found that culturally responsive school culture consists of three dimensions: technical, institutional, and managerial, and has three outputs: success, development, and universalization. A definition of culturally responsive school culture was developed, and its values, assumptions and elements were identified. Various suggestions were made to researchers and practitioners within the scope of this model.

## 1. INTRODUCTION

There are unspoken rules governing interpersonal relations in a community. Since these rules are handled from various angles, the concept of culture has a very wide scope. The concept of culture is a definition that is very difficult to define due to its wide scope, but there are also many definitions. Robbins and Judge (2017) define culture as a system of meaning that distinguishes individuals in a particular region from other individuals and is shared by these individuals. Although culture affects people's behaviors, it is also effective on people's learning. Concepts such as beliefs, attitudes and values that are necessary for a person to evaluate, organize and assimilate what they perceive through their senses as a mental process are referred to as culture (Shade *et al.*, 1997). This is also referred to as a mental model in the sense that it gives meaning to the world (Senge, 2014). Therefore, everyone is different in terms of worldview. This difference is also found in schools. Culturally responsive school culture is an

*CONTACT: Gürkan SARIDAŞ ✉ theapeiron@gmail.com Ministry of National Education, Denizli, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

important concept in terms of ensuring the continuity of teachers' professional development processes, ensuring that students receive education in environments where cultural differences are used as a tool, ensuring that parents develop and maintain positive attitudes towards society, school and the group they are in, and creating a development-oriented school culture with cultural and transformational leadership behaviors of administrators Ladson-Billings (1995).

The educational philosophy recommended to address this diversity in schools is culturally responsive education. Its simplest definition is to create channels for more effective learning by utilizing the cultural characteristics of culturally diverse students. Culturally responsive education is a design that contributes to the academic and social development of the student and the professional development of the teacher while using the cultural differences of the student as a tool in the education and training process (Sarıdaş & Nayir, 2020).

When considered across schools, culturally responsive education practices should be supported by a culturally responsive management approach. One of the primary needs related to culturally responsive management is to provide equal opportunities. School administrators should adopt instructional and transformational leadership styles to ensure equal opportunities (Khalifa *et al.*, 2016). Proper use of resources for the academic and social development of culturally diverse students, contributing to the professional development processes of teachers in this direction, and directing the institutional culture to use this diversity as a richness and an important tool are the behaviors expected from school administrators. These behaviors are grouped by Khalifa *et al.* (2016) under the headings of creating personal awareness, preparing the curriculum and teachers, making the school environment sensitive and inclusive of cultural values, and including students and parents in this culture.

Culturally responsive education is grounded in critical theory, racial identity theory, and social cognitive theory. Critical theory directly adopts the views of equality, justice, and non-alienation as it focuses on constant questioning, criticism, and examination (Blake & Masschelein, 2002). This theory focuses on students choosing knowledge as a guide for themselves. Racial identity theory is defined by the complex interplay between emotions, attitudes, beliefs, behaviors, and cognition to solve identity problems (Helms, 1986). This theory focuses on the unification of students around universal values and the creation of harmonious communities. Social cognitive theory includes students' perceptions of organizing and specializing in academic learning activities, perceptions of developing and managing interpersonal communication, and perceptions of resisting social pressure (Bandura, 2002). With this theory, students organize their learning through their own culture and focus on the impact of the social environment at school on learning. The intersection point of these three theories is the culturally responsive school culture.

School culture is an important tool for setting clear and higher education goals within the school, creating environments conducive to learning, creating unity around the school mission and vision, training teacher leaders, and maintaining positive attitudes towards students and parents (Jones, 2009). Putting this concept into practice in schools and ensuring its continuity points to the development of students, teachers, parents, and administrators based on schools. This study is important because it will reveal the creation of a culturally responsive school culture in schools, how cultural change will be achieved, the perspectives of teachers, students, parents and administrators on culturally responsive school culture, and the effects of culturally responsive school culture for the future.

When culturally responsive school culture is examined within the framework of various theories, various features come to the fore. In a culturally responsive school culture, the school interacts with its environment and rehearses the cultural differences in the environment in the classroom, uses its resources for development and change, aims to reflect the diversity of the society in the classroom, aims to be a mirror of the society, has the goals of being integrated with the society, analyzes the requests that may come from the society and produces appropriate

ways of development, works for the professional development of teachers, supports the personal development of teachers, enables them to meet universal and human values, creates high goals that are achievable by students (Gay, 2000; Khalifa *et al.*, 2016; Ladson-Billings, 1995; Shade *et al.*, 1997; Villegas & Lucas, 2007; Weinstein *et al.*, 2003). These characteristics also reveal the differences from other school culture theories.

In the research, culturally responsive school culture is examined. When the studies are examined, many studies have been conducted on school culture, culturally responsive leadership, culturally responsive pedagogy and culturally responsive education, but no research has been found on culturally responsive school culture. The concept of culturally responsive education is a new concept for Türkiye (Kotluk & Kocakaya, 2018). For this reason, while the research contributes to the relevant literature in terms of culturally responsive school culture, it is also important for teachers, students, school administration and parents with the examinations made based on sub-objectives. In this context, culturally responsive school culture emerges as a critical concept for fostering a development-focused school environment. It facilitates active participation in educational activities among individuals from diverse cultural backgrounds, promotes the continuous expansion of comprehensive developmental programs, and utilizes cultural diversity as a resource for advancing widespread educational practices. Additionally, it supports increased parental engagement, fosters positive attitudes toward the school and its community, and ensures the sustainability of this approach. Through the integration of cultural diversity and transformative leadership, culturally responsive school culture plays a pivotal role in creating an inclusive and growth-oriented educational atmosphere. The implementation of this concept in schools and ensuring its continuity indicates the development of students, teachers, parents, and administrators based on schools (Gay, 2000). This study is important because it will reveal the creation of a culturally responsive school culture in schools, how cultural change will be achieved, the perspectives of teachers, students, parents and administrators about a culturally responsive school culture, and the future effects of a culturally responsive school culture.

Through the findings obtained from the research conducted, a conceptual literature integrity regarding culturally responsive school culture, evaluations obtained in terms of school administrators, teachers, students, and parents, and within the scope of the findings, schools are encouraged to create a culturally responsive school culture, to create a culturally responsive school culture. Contributions have been made to the literature in terms of the process of cultural change and the contributions of a culturally responsive school culture. With the study carried out, practitioners will feel the difference between the culture that exists in schools and the culturally responsive school culture, identify the duties of the school administrator, teacher, student and student's parent, describe how to transform the school culture into a culturally responsive school culture, and at the end of this process, the school culture will be determined. They will observe the contributions of the administrator, teacher, student, and student's parent. Thus, the research carried out in this context is important.

The problem of the study is to define the concept of a culturally responsive school culture, how to create it, its effects on administrators, teachers, students, and parents, and to determine the variables affecting culturally responsive school culture. In line with the identified problem, answers to the following questions were sought:

Sub-problems of the qualitative method:

1. With which concepts do school administrators explain culturally responsive school culture?
2. With which concepts do teachers explain culturally responsive school culture?
3. With which concepts do students explain culturally responsive school culture?
4. With which concepts do parents explain culturally responsive school culture?
5. What are the components of a culturally responsive school culture?

Sub-problems of the quantitative method:

6. What are school administrators' views on culturally responsive school culture?
7. What are teachers' views on culturally responsive school culture?
8. What are the variables affecting culturally responsive school culture?

Sub-problems of the mixed method:

9. To what extent can the findings obtained from qualitative data be generalized with the findings obtained from quantitative data?

## 2. METHOD

### 2.1. Research Design

A mixed methods design was employed in the study. Mixed methods design integrates both qualitative and quantitative data collection and analysis (Creswell, 2009). Specifically, a sequential exploratory design was adopted within the mixed methods approach. This design is particularly suitable when the relationships between variables are unknown (Baki & Gökçek, 2012). Within the scope of the research, mixed methods in sequential exploratory design was used because the variables related to the culturally responsive school culture were not known and the relationship between the variables could not be fully determined.

For the first phase of the study, a grounded theory design, one of the qualitative research methodologies, was utilized. Grounded theory involves systematically analyzing data collected from participants' perspectives to generate explanations and ideas (Creswell, 2009). In the second phase, a culturally responsive school culture scale was developed, and the factors influencing culturally responsive school culture were examined.

### 2.2. Population and Sample

Criterion sampling was employed for the interviews conducted in the qualitative part of the study. The purpose of criterion sampling is to examine situations that meet one or more criteria established by the researcher (Yıldırım & Şimşek, 2008). The criteria determined within the scope of the research were as follows: for school administrators, having foreign national students who are consistently present and enrolled in their schools; for teachers, having foreign national students who are consistently present and able to speak Turkish in their classes; for students, being foreign nationals who are consistently present and able to speak Turkish in their classes; and for parents, having children who are foreign national students, consistently present, and able to speak Turkish in their classes In line with these criteria,, data were collected from secondary schools in a randomly selected province within the Statistical Regional Units, ensuring maximum diversity and to be compatible with the quantitative part of the study. The randomly selected provinces are as follows: Istanbul (TR1), Tekirdağ (TR2), İzmir (TR3), Bursa (TR4), Ankara (TR5), Antalya (TR6), Kayseri (TR7), Samsun (TR8), Trabzon (TR9), Erzurum (TRA), Van (TRB) and Mardin (TRC). The participants of the qualitative part included school administrators and teachers working in these provinces, students studying in these provinces, and the students' families.

Each interview was assigned a unique code to distinguish the participants' information. The coding started with "A" for administrators, "T" for teachers, "P" for parents and "S" for students, and the letter "M" was used for men and "F" for women to indicate their gender. The number of interviews conducted was advanced starting from 1. Information on the regional distribution of participants in the qualitative part of the study is presented in Table 1. According to Table 1, there were 26 administrators, 26 teachers, 15 students and 24 parents in the study group.

The proportional stratified sampling method was used for the quantitative component of the study. Stratified sampling is a method that identifies subgroups within the population and ensures their representation based on their proportions (Özen & Gül, 2007). The formula suggested by Cochran (1962; cited in Balcı, 2013) was used to determine the sample. In line

with the number of teachers working in Statistical Regional Units (SRU) (MEB, 2021), the sample size obtained for the quantitative part of the research was determined as 383.

**Table 1.** *Information on the distribution of participants in the qualitative section by region.*

| Region | School Administrator | Teacher | Parent | Student |
|---|---|---|---|---|
| TR1 | 3 | 4 | 2 | 2 |
| TR2 | 2 | 2 | 2 | 2 |
| TR3 | 3 | 2 | 2 | 2 |
| TR4 | 2 | 2 | 1 | 2 |
| TR5 | 2 | 2 | 1 | 2 |
| TR6 | 2 | 2 | 1 | 2 |
| TR7 | 2 | 2 | 1 | 2 |
| TR8 | 2 | 2 | 1 | 2 |
| TR9 | 2 | 2 | 1 | 2 |
| TRA | 2 | 2 | 1 | 2 |
| TRB | 2 | 2 | 1 | 2 |
| TRC | 2 | 2 | 1 | 2 |
| Total | 26 | 26 | 15 | 24 |

Considering that there is an average of 1 principal and 2 deputy principals working in each institution (MEB, 2021) as an approximate value throughout Turkey, 51,627 people were estimated as the population. Using this estimate, the sample size was recalculated and determined to be 381 participants. The distribution of the required sample size and the actual number of teachers and administrators included in the sample, categorized by region, is presented in Table 2. According to the information given in Table 2, 541 teachers and 397 administrators were included in the sample group due to the inability to evaluate the scale results, the possibility that the scales were not filled in correctly and the possibility of discarding extreme values.

**Table 2.** *Number of secondary school teachers, administrators and samples by SRU.*

| Region | Teacher | Teacher to be Sampled | Sampled Teacher | Administrator | Manager to be Sampled | Sampled Manager |
|---|---|---|---|---|---|---|
| TR1 | 160644 | 64 | 79 | 7119 | 52 | 58 |
| TR2 | 42356 | 17 | 23 | 3675 | 28 | 30 |
| TR3 | 138200 | 56 | 61 | 3960 | 29 | 29 |
| TR4 | 104701 | 42 | 51 | 4758 | 35 | 36 |
| TR5 | 113018 | 46 | 55 | 4521 | 33 | 36 |
| TR6 | 155222 | 63 | 69 | 4749 | 35 | 36 |
| TR7 | 59139 | 24 | 31 | 3681 | 27 | 25 |
| TR8 | 60849 | 25 | 29 | 3504 | 26 | 26 |
| TR9 | 36843 | 15 | 22 | 3297 | 25 | 29 |
| TRA | 35004 | 14 | 21 | 3465 | 25 | 28 |
| TRB | 65015 | 26 | 31 | 3924 | 29 | 23 |
| TRC | 144851 | 58 | 69 | 4974 | 37 | 41 |
| Total | 1115842 | 450 | 541 | 51627 | 381 | 397 |

### 2.3. Data Collection Tools

In the qualitative portion of the study, data were collected through a semi-structured interview form. The interview form was designed to enhance the depth of responses during the interview process (Türnüklü, 2000). The interview form was developed based on insights gained from the literature review and submitted to three language experts and fourteen field experts for evaluation. The interview forms were finalized by making the necessary arrangements according to the feedback received from eight field experts and three language experts. A pilot study was conducted by applying the interview forms to two school administrators, teachers, students, and parents.

In the quantitative phase of the study, a culturally responsive school culture scale was developed. The culturally responsive school culture scale was first analyzed by exploratory factor analysis and then by first and second level confirmatory factor analysis. The validity and reliability of the scale, with its construct validity established, were thoroughly examined. The variables affecting the culturally responsive school culture were also examined through the developed scale.

### 2.4. Validity and Reliability

In qualitative research, the validity and reliability of the research are met by the criteria of credibility (internal validity), transferability (external validity), consistency (internal reliability) and confirmability (external reliability) (Guba & Lincon, 2005; Yıldırım & Şimşek, 2008). For the credibility of the research, coding was conducted separately by both the researcher and the advisor. To ensure the transferability of the findings, participants' perspectives on the themes were presented using direct quotations. For the consistency of the study, all research processes were described in detail, including the participants, the methods employed, and the coding of the data. Lastly, to establish confirmability, an external researcher reviewed and verified the collected data.

First and second level Confirmatory Factor Analysis (CFA) was conducted for the scale used in the study. Accordingly, the first level fit indices $\chi^2/df$ value was calculated as 1.81, RMSEA = .048, CFI = .914 and NFI = .876. The second level fit indices for the scale were calculated as RMSEA = .049, CFI = .922 and NFI = .892 with $\chi^2/df$ value of 1.935. Cronbach's Alpha coefficient of the scale was found to be .807.

Validity in mixed methods research is categorized under four main headings: content validity, internal validity, utility criterion, and external validity (Zohrabi, 2013). In this context, to ensure content validity in the research, data collection tools were developed in consultation with field experts and language experts. To ensure internal validity, the data collection tools used during the research were diversified. To ensure the criterion of utility, attention was paid to the applicability of the research findings. To ensure external validity, data were collected according to Statistical Regional Units and a highly generalizable study was ensured.

Creswell (2021) suggested various strategies for validity in mixed methods. These strategies are having the person check, triangulation, negative situation analysis, peer supervision, external supervisor, and being in the field. To ensure validity within the scope of the study, qualitative findings were checked by a third researcher (self-checking). At the same time, the whole process was monitored by various researchers and necessary corrections were made (external supervision). Since the researcher was a teacher and was in contact with the research subject, the responses were checked for relevance in the field (field presence). Qualitative and quantitative data were combined throughout the research, and qualitative and quantitative data were collected from various people with various data collection tools (triangulation).

### 2.5. Data Collection Process

The research was conducted with the permission of the Ministry of National Education dated 28/03/2023 and numbered E-70297673-605.01-73265586, and the ethics committee

permissions of Pamukkale University dated 09/03/2022 and numbered 68282350/2022/G05, dated 27/02/2023 and numbered 68282350/2023/04. Since the researchers and the participants were in different provinces, contact information of the participants was obtained and interviews were conducted online or by phone. The scale form including the items of the culturally responsive school culture scale and demographic information was created and made ready through Google Forms. The form was sent to the participants.

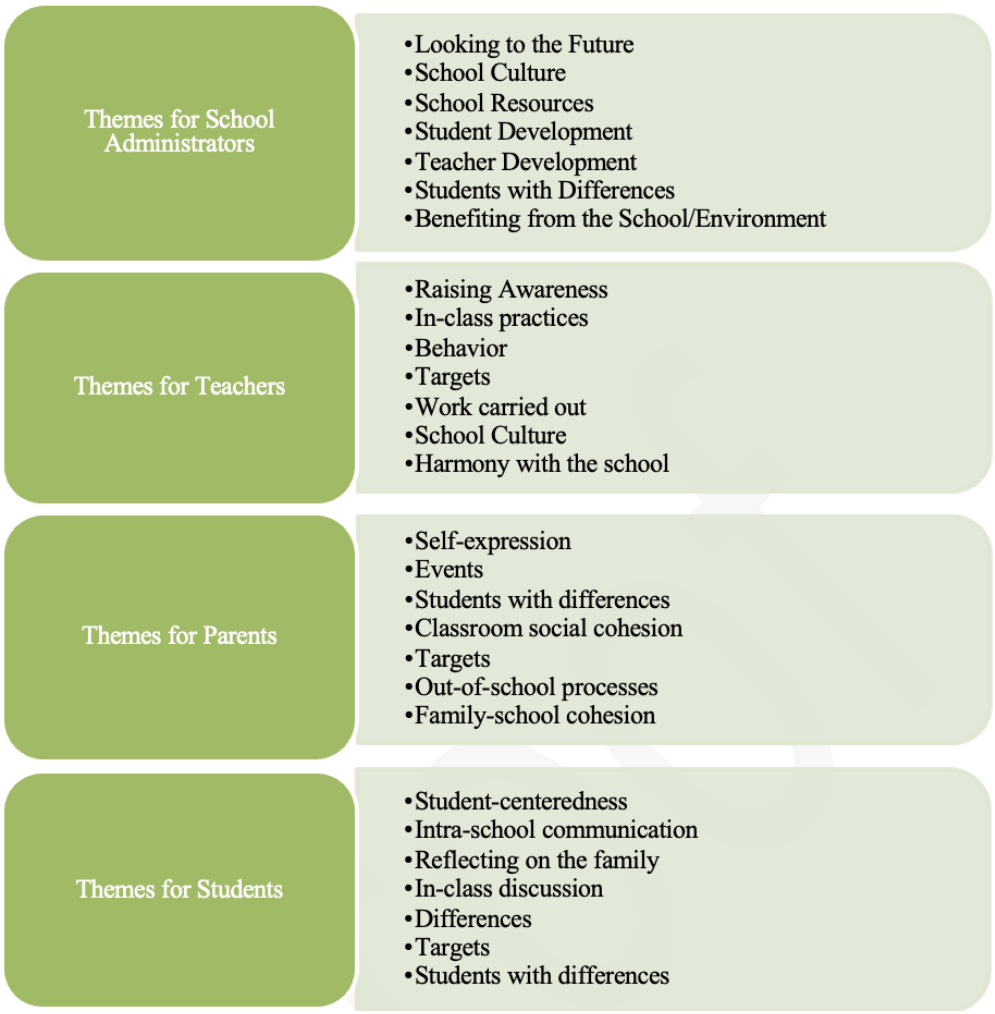
### 2.6. Data Analysis

In the qualitative part of the research, three stages were carried out as initial coding, focused coding, and theoretical coding in accordance with the grounded theory (Charmaz, 2006; Charmaz, 2008). For the quantitative part, various statistical analysis were conducted for the purpose of the study. CHAID (Chi-squared Automatic Interaction Detection) Analysis, which is one of the decision tree methods, was applied to determine the independent variables predicting culturally responsive school culture. In this analysis, one dependent variable is analyzed with more than one independent variable (Kayri, & Boysan, 2007). Since the research was based on an exploratory design, the association analysis recommended for exploratory designs was used. The views on culturally responsive school culture were used to explain the data obtained from the culturally responsive school culture scale. Therefore, the generalizability of the views on culturally responsive school culture was also examined.

## 3. FINDINGS

The generated themes are shown in Figure 1 below to make sense within the integrity.

**Figure 1.** *Themes of culturally responsive school culture school administrators, teachers, students, and parents.*



As indicated in Figure 1, the themes that emerged from the analysis conducted in the grounded theory design because of the interviews were divided into school administrators, teachers, parents, and students. Some of the findings are included for the readability of the research report.

## 3.1. Findings Related to School Administrators

### 3.1.1. *Findings related to the theme of students with diversity*

The results of the analysis of school administrators' answers to the questions and sub-questions are given in Table 3.

**Table 3.** *School administrators' codes for the theme of students with diversity.*

| Theme | Sub Theme | Code | Frequency |
|---|---|---|---|
| Students with diversity | Participation in education | Can participate | 18 |
| | | Different schools | 7 |
| | | Controlled participation | 2 |
| | Curriculum | Direct participation | 24 |
| | | Different curricula | 1 |
| | Behavior | Equal behavior | 25 |
| | Don't pose a problem | No problem | 19 |
| | | Discipline problem | 11 |
| | | Behavior problem | 7 |
| | Exemplary behavior | Equal behavior | 22 |
| | | Optimistic | 13 |
| | | Solution-oriented | 5 |

According to Table 3, when the opinions of school administrators are analyzed, most school administrators stated that they could participate in education ($f$ = 18). AM25 stated, "*The question of whether these students can participate in education should not even be asked. When these students do not participate, they will become harmful to society. We need to include them in education and training so that they can adapt to society*." AM25 stated that they can participate in education and training. Similarly, AM19 said, "*We do not know what will happen if these students do not participate in education. But we know that they will be registered when they enroll with the school. For this reason, they should definitely participate*.", he reveals the necessity of their participation in education. AM16, on the other hand, states that "I advocate that especially foreign students should study in special schools for themselves. Let's not think of it as isolation, this situation is only to keep them under control.". AM11, on the other hand, citing the region he lives in as an example, said, "*It is difficult for foreign nationals to be included in the society here. I also believe that they should be educated in a region like a concentration camp on the outskirts of the country due to their status*.", emphasizing the code of different schools. With a different perspective to these views, AF5 stated that "*Foreign students should enter the country in a controlled manner. For example, if there is no language, they should learn a language. If there is no basic level education, they should attend pre-school education. Or they should start at a lower level of the class appropriate to their level. The education of these students should be kept under constant control. Thus, cultural degeneration would be quite low*." emphasizes the code of controlled participation ($f$ = 2).

When the curriculum of students with diversity is analyzed, it is observed that a great majority of them express the opinion of direct participation in the curriculum ($f$ = 24). AM9 stated "*With the inclusion of the student in the education system, the student starts to be developed in line with the goals of the country. For this reason, the student should have the same curriculum*." emphasizes direct participation in the curriculum. AF3 emphasizes direct participation in the current curriculum with the statement "*We have a very intensive curriculum that can develop*

*all aspects of students, so they can develop themselves with this curriculum*." *On the other hand, AF5 contradicts these views by saying, "You say it has differences, and you ask about the curriculum. If they have differences, I think they should be subject to a different curriculum*." emphasizing the code of different curriculum ($f$ = 1) and drawing attention to the individual curriculum.

When the behaviors towards students with diversity were examined, all participants emphasized the code of equal treatment ($f$ = 25). AM24 emphasized the code of equal treatment with the statement "*As a teacher, I cannot think of anything other than treating all students equally*." Likewise, AF20 emphasized the code of equal treatment with the statement "*It is necessary to correctly apply the principle of equality among the general principles of the education system*."

When it was analyzed whether students with diversity constitute a problem in the education system, the majority emphasized the code "not a problem" ($f$ = 19). AM6 stated that "*When these students receive enough attention, they develop smoothly. As long as we take care of them*." AM6 stated that it does not pose a problem. AM10 said, "*These are actually children, provided that the exceptions do not break the rule. Which child is problematic. In fact, let's think about whether the child who is not average is a problematic child*." Despite these views, AF4 stated, "*Actually, these children do not cause problems. But some behavioral problems, although not often, are very difficult for us. It is necessary to maintain discipline*." With this statement, he emphasizes the code that they do not cause problems and cause discipline problems ($f$ = 11). AM18 also emphasized that "*These students usually challenge us in terms of discipline. Other than that, they do not cause any problems. We can't even call them a problem anyway*." emphasizes the code of discipline problem. There are also school administrators who see them as behavior problems ($f$ = 7) as opposed to discipline problems. AF20 said, "*The most common problem we encounter with these students is behavioral problems. These students affect the behavior of other students as well*." and draws attention to the behavior problem.

When exemplary behaviors for students with diversity were examined, the majority again emphasized the code of equal behavior ($f$ = 22). It is noteworthy that school administrators both perform a behavior at the point of equal treatment and show equal treatment as an exemplary behavior. AF7 "*The only thing we need to do is to know that the students are the same in the classroom and that the students behave the way we treat them because they take our behavior as an example. After all, if we treat students equally, they treat their friends equally*." emphasizes the code of equal treatment. In addition to this majority, the code of being optimistic ($f$ = 13) also draws attention. AM9 said, "*Let's be optimistic towards all students. They do not think anything bad. Maybe they just can't think. It is enough to make them feel good and show them that everyone has equal rights*." emphasizes the codes of being optimistic and acting equally. In addition to these views, AM12 said, "*What is our purpose with these students? To beat the winegrower or to eat grapes? Then let's identify ways to find a solution. They are all world citizens and our children. They are all equal in our eyes. We should treat them all well*." emphasizes the codes of treating them equally, being optimistic and being solution-oriented ($f$ = 5).

## 3.2. Findings Related to Teachers

### 3.2.1. *Findings related to the theme of school culture*

The results of the analysis of the teachers' answers to the questions and sub-questions are given in Table 4. According to Table 4, teachers state that they adapt to the school culture ($f$ = 25). TF19 stated "*Teachers in the school act collectively. Since our school is a successful school, outsiders have to get used to it immediately*." TM22 "*As in every school, there are incompatible teachers here too. But in general, teachers are harmonious and help each other*." TF27 "*In our school, there is a voice from every head. Teachers always want what they want. The one who*

*is close to the administration gets his job done. There is no justice or anything like that*." with the statement "There is no harmony against the school culture (*f* = 2).

**Table 4.** *Teachers' codes for the theme of school culture.*

| Theme | Sub Theme | Code | Frequency |
|---|---|---|---|
| School culture | Adaptation to school culture | Harmonizing | 25 |
| | | Incompatible | 2 |
| | Trust environment | Safe | 14 |
| | | Neutral | 10 |
| | | Not safe | 3 |
| | Professional development | Does not support | 19 |
| | | Supporting | 5 |
| | | Summit support | 3 |
| | Competitive environment | No | 17 |
| | | Supportive competition | 7 |
| | | Overwhelming competition | 3 |
| | Relationship between teachers | Strong | 15 |
| | | Supportive | 11 |
| | | Business friendship | 7 |
| | | Forced relationship | 3 |

When asked about the trust environment among teachers, most of the teachers stated that it was safe (*f* = 14). In this regard, TM18 said, "*There is very little gossip at school. Everyone minds their work. There is not much talk outside of work. Private life is usually between very sincere people. So, there is a trustworthy environment*." emphasizes the trustworthy environment. TF13 said, "*I don't know what to say about this. You can trust some people and some people you can't. It would be wrong to expect infinite trust just because they are teachers*." and emphasized the neutral (*f* = 10) code. Similarly, TF11 said, "*I usually do not share anything except with my close friends. It may come back to me as gossip. Therefore, I can say that it is safe with those I like and insecure with those I don't like*." and emphasized the neutral code. On the other hand, TF12 said, "*Compared to other professions, teachers usually undermine each other. Therefore, I cannot say it is safe. I think there is no safe environment*." emphasized the code of insecure (*f* = 3).

When asked about information sharing and professional development, most teachers did not support it (*f* = 19). TM5 stated that it is not supported with the statement "*In general, there are new teachers in our school, so I will say that it is not supported*.". On the other hand, TF1 said, "*We share information and activities as a group. We support each other in the subjects we are good at. We also share with other teachers in a professional sense*." (*f* = 5). TF7 "*In our school, the departments share within themselves. I do not witness two different departments sharing information*." emphasizes the code of department support (*f* = 3).

When asked about the competitive environment, most of the teachers stated that there was no competition (*f* = 17). TF9 stated "*There is no competition in our school. We are quite harmonious." and stated that there is no competition. Similarly, TM2 stated that there is no competition with the statement "Our administrators do not want a competitive environment*." Among the teachers who elaborated, TM6 said, "*There is a competitive environment in our school that raises each other up. It is like who will share more. This is a good thing. Everyone supports each other in what they are good at*." emphasizes the code of supportive competition (*f* = 7). Similarly, in detail, TF8 stated "*The school administration takes money for student*

*registration. Therefore, it supports competition among teachers. Teachers are in constant competition to show that they are better*." emphasized the code of overwhelming competition (*f* = 3).

When asked about the relationship between teachers, the teachers agreed that the relationship was strong (*f* = 15). "*Our friendship at school continues after school. Friendships and friendships are quite strong*." and expresses the opinion that relationships are strong. TF11 stated, "*Since the friendship between teachers is good, everyone wants the best for each other. Everyone rushes to help each other*." and stated that there is a supportive relationship (*f* = 11). Far from these views, TM3 stated that "*Actually, work friendship or social friendship is different. If you meet at school and in social life, you talk about school again. I think they should be separated*." and stated that the relationships should be in the form of work friendship (*f* = 7). In the same way, TF19 also said, "*If you are at school, you are friends at school. You don't need to carry the friendship*." emphasizes work friendship. On the other hand, TM24 said, "*At school, everyone is like strangers to each other. They come, go to class and leave. Therefore, they do not make friends*." emphasizes the code of obligatory relationship (*f* = 3).

## 3.3. Findings Related to Parents

### 3.3.1. *Findings related to the theme of students with diversity*

The results of the analysis of parents' answers to the question and sub-questions are given in Table 5.

**Table 5.** *Student parents' codes for the theme of students with diversity.*

| Theme | Sub Theme | Code | Frequency |
|---|---|---|---|
| Students with diversity | Education | Must be together | 11 |
| | | Can be together | 3 |
| | | Must be separate | 1 |

According to Table 5, while most of the parents think that students with diversity should be educated in the same environment (*f* = 11), some of them think that there is no problem with them being educated together (*f* = 3) and very few of them think that they should be educated separately (*f* = 1). PM2 argues that students should be educated together with the statement "*I do not think it is appropriate for students to receive education in different environments. I think there is a fundamental separation in such situations*." Likewise, PM12 argues that it should be together with the statement "*They are all children of this country, why should they be educated in different classes?*". In addition to these views, PM6 stated "*If it will be beneficial to have separate education, it can be. But if not, if the same education is to be received, it should be the same. Teachers know this better*." and defends that they can receive education. PM13 stated, "*Each student is different and should receive different education. For this reason, appropriate education should be given in separate environments*." and argues that they should be separate.

## 3.4. Findings Related to Students

### 3.4.1. *Findings related to the theme of students with diversity*

The results of the analysis of the students' answers to the questions and sub-questions are given in Table 6. According to Table 6, students (*f* = 24) responded similarly to each other in terms of the idea of educating with students with diversity. While some students found it appropriate to be educated together (*f* = 14), others thought that it was not appropriate (*f* = 14). SM4 emphasized the code "*All children can be educated in the same class. I think there is no problem.*", emphasizing the appropriate code. Expatriate teacher SF12 also emphasized the appropriate code with the statement "*It is not a problem for me to teach in the same environment. I do not think that I am an obstacle for them."* and emphasized the appropriate code. On the contrary, foreign national SF1 said, "*They do not want us because we are behind*

*compared to them. If they were in my country, they would also be behind us*." On the other hand, SF6 stated that it was not appropriate with the statement "*I think it is more correct for them to have the same lessons in different classes with fewer people than us*."

**Table 6.** *Students' codes for the theme of students with diversity.*

| Theme | Sub Theme | Code | Frequency |
|---|---|---|---|
| Students with diversity | Training together | Suitable | 14 |
| | | Not suitable | 10 |
| | Impact | Positive | 11 |
| | | Negative | 10 |
| | | Neutral | 3 |
| | Process | Yes | 14 |
| | | No. | 10 |
| | Curriculum | No. | 19 |
| | | Yes | 6 |
| | Measurement and evaluation | No. | 19 |
| | | Yes | 5 |

When asked about the effects of studying with students with diversity, some students mentioned positive effects ($f$ = 11) while others mentioned negative effects ($f$ = 10). In this regard, foreign national SF9 said, "*If there was no war in our country, we would not have come here anyway. I am happy that I came here, I made new friends. They taught me a lot of things*." whereas SM24 said, "*It is very nice to have children from another country as guests at our school. I am happy to have the opportunity to meet them*." SM20 emphasizes the positive effect with the statement "*I am very happy to meet different people. I think it is a great chance. My parents did not make such a friendship*." On the contrary, foreign national SF1 said, "*I don't like the exclusion of my friends. They always exclude me. They push me around. I wish I could go back to my country*." Likewise, SM5 emphasizes the negative effect with the statement "*I don't want them to be here. This is Turkey, they should go back home*." SF7 emphasizes the negative effect with the statement "*Foreigners being here does not affect anyone. How can it affect us when guests come to our house? If we stand upright, they cannot harm us*." ($f$ = 3).

When asked about creating different processes for students with diversity, some students stated that a different process should be managed ($f$ = 14), while some students stated that the same process should be continued ($f$ = 10). In this regard, SF10 said, "*It is ridiculous that they take the exam with us. Their purpose for coming to school is not to take the exam." and stated that a different process should be managed. When they fall behind, our teacher does not wait for them*." with the expression "*Different practices are needed." SM16 "Before the pandemic, a special class was opened for them. They were also taking lessons there. It could have continued that way*." emphasizes different practices. On the contrary, SF19 said, "*They can learn the same things in the same class as us. It would be unfair for them to get higher grades by asking them simpler questions*." and suggests the same educational practices. SM22 "*There is no need to do anything different for them. Everyone is taking the same courses. If they want to participate in these lessons, they can do so*." and supports the same practices.

When asked about the idea of changing the curriculum for students with diversity, the idea that a different curriculum should not be created ($f$ = 19) came to the forefront. SM22 does not support change with the statement "*Subjects can be changed for reading and writing, but I think there is no need to change them in other subjects*.". Likewise, SF19 stated, "*If we try to change the subjects according to them, we will have to change many things along with it. I think it is unnecessary*." and emphasizes the same curriculum. As a counter view, SM8 emphasizes the issue of different curricula ($f$ = 6) with the statement "*When I go to another country, I would want special subjects to be explained to me, so I think special subjects can be created for them*.". Similarly, SM11 also emphasized the issue of different subjects with the statement "*If they are*

*not going to have a profession by passing the class, different subjects should be made. But if they are going to get ahead of us with simpler subjects, they shouldn't*.", he draws attention to both the positive and negative aspects of the different curriculum by making it conditional.

When asked about measurement and evaluation, most of the students stated that there was no different measurement and evaluation practices ($f$ = 19). "*Exams are very important for us. For example, I will take the exam next year. My whole life will change accordingly, so everyone should take the same exam. There should not be a different exam for them*." emphasizes the same assessment and evaluation. Similarly, SM16 also emphasized the same assessment with the statement "*A different exam means that they take the exam more easily. It means that they get higher scores than us*" and expressed their opinions in the direction of the same assessment and evaluation. As a contrary opinion, foreign national SM15 stated "*We should not take the same exam with them. They are luckier than us*" and point to different assessment and evaluation practices. SM13 said, "*Not only them, but everyone should be tested according to their own level. In this way, everyone can know how much they have learned*" and draws attention to different assessment and evaluation systems.

In the quantitative part of the study, the developed scale was applied to 479 teachers and 378 administrators. After EFA and CFA, 120 participants were collected again and a total of 977 participants were obtained. Of the total 977 participants, 556 were teachers and 421 were administrators.

### 3.5. Teachers' and School Administrators' Views on CRSC

The means, standard deviations, and standard deviation error values of the total scores of the school administrators' responses for the scale and its sub-dimensions are given in Table 7 below. When Table 7 is analyzed, it is seen that school administrators "Strongly Agree" with the Culturally Responsive School Culture (CRSC) scale (Table 7). $\bar{X}$ = 112.55). Teachers, on the other hand, responded to the CRSC Scale at the "Strongly Agree" level ($\bar{X}$ = 110.69)

**Table 7.** *Mean and standard deviation values of CRSC school administrators and teachers.*

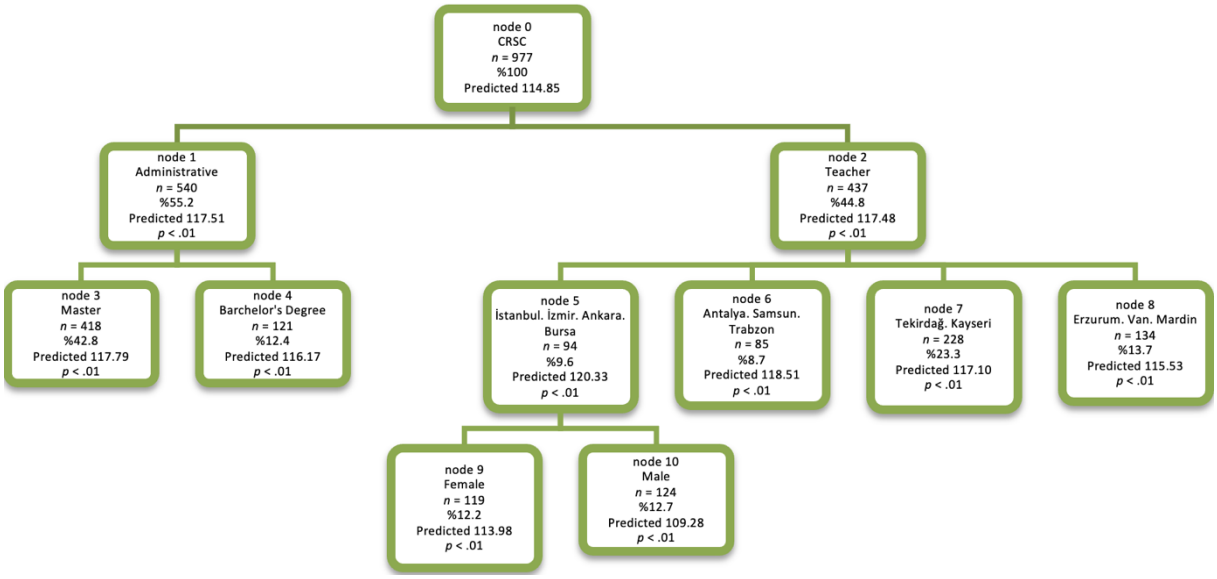
| | School Administrators | | | | Teacher | | | |
|---|---|---|---|---|---|---|---|---|
| | CRSC | Corporate | Administrative | Technical | CRSC | Corporate | Administrative | Technical |
| *n* | 421 | 421 | 421 | 421 | 556 | 556 | 556 | 556 |
| $\bar{X}$ | 112.52 | 47.28 | 37.66 | 27.61 | 110.69 | 44.49 | 41.03 | 25.17 |
| *SD* | 14.17 | 11.04 | 3.59 | 3.63 | 9.84 | 8.91 | 3.49 | 3.99 |
| *SE* | .69 | .54 | .18 | .18 | .42 | .38 | .15 | .17 |

### 3.6. Variables Affecting Culturally Responsive School Culture

CHAID analysis serves to reveal the independent variables affecting the dependent variable by analyzing other independent variables that are thought to be related to this dependent variable based on the answers given to a dependent variable (Horner *et al.*, 2010). The graph resulting from the analysis is given in Figure 2.

When Figure 2 is examined, it is observed that firstly, the position, secondly, the province where the person is located and finally, gender influence the culturally responsive school culture. Since the 4th level variable in the analysis was not significant ($p$ > .05), it was not included in the findings section.

**Figure 2.** *Variables affecting culturally responsive school culture.*




### 3.6. Findings Related to Mixed Method

In accordance with the mixed method, the generalizability of the views of teachers, students, administrators, and parents obtained through interviews about culturally responsive school culture and the verifiability of the views of teachers and administrators obtained through the developed culturally responsive school culture scale were analyzed. The relationship between the scale items in the administrative dimension and the qualitative interviews is given in Table 8. When Table 8 is examined, it is seen from the relationship between the interviews and the scale items that the qualitative interview results are compatible with the quantitative scale items, thus the generalizability of the qualitative interviews is revealed.

The relationship between the scale items in the institutional dimension and qualitative interviews is given in Table 9. When Table 9 is examined, it can be said that the quantitative items can be verified through the association in the overall dimension, although the items of having students with a scientific perspective, the school being in harmony with different cultural values in the society, and all parents being in positive communication in the institutional dimension are not clearly matched.

The relationship between the scale items in the technical dimension and the qualitative interviews is given in Table 10. When Table 10 is examined and the findings obtained from the interviews in the technical dimension are correlated with the averages obtained from the scale items, it can be said that the items for the quantitative part are verified.

**Table 8.** *Mixed analysis findings for managerial dimension.*

| Qualitative Analysis Themes and Subthemes | Quantitative Analysis Scale Items | Qualitative and Quantitative Part Association |
|---|---|---|
| The sub-theme of "Studies on diversity" under the theme of "Looking to the future" and the sub-theme of "Status of diversity" under the theme of "School culture" belonging to school administrators | School administration is sensitive to individual differences. ($\bar{X} = 4.66$) | In the interviews, it was stated that there was no work for students with diversity in the school, but that the school culture was inclusive. In the scale item, there was an opinion that the school administration was sensitive to diversity. |
| The "Professional Development" sub-theme under the "Teacher Development" theme for school administrators and the "Continuous Learning" sub-theme under the "School Culture" theme for teachers. | School administration conducts studies for the professional development of teachers sensitive to cultural values. ($\bar{X} = 4.61$) | In the interviews, it was stated that teachers contribute to their professional development through seminars (in-service training) and that the school culture supports continuous learning. In the scale item, the view that professional development activities are carried out for diversity emerged. |
| The sub-theme of "Recommended practices" under the theme of "Benefiting from the school/environment" for school administrators. | School administration makes efforts to ensure that the school environment is culturally responsive. ($\bar{X} = 4.56$) | To increase the culturally responsiveness of students, it is suggested that special education and orientation trainings should be provided, and the scale item reveals an opinion that school administrations make efforts to increase the sensitivity of the environment. |
| The sub-themes of "Participation in Education", "Behavior" and "Causing Problems" under the theme of "Students with Differences" belonging to school administrators. | The school administration is kind and friendly to students with diversity ($\bar{X} = 4.61$) | The interviews revealed that they do not see any problem in the participation of students with diversity in education and that they treat students equally. In the scale item, it was revealed that the administration was kind and friendly to the students. |
| "Practices" and "Benefit Effect" sub-themes under the theme of "Benefiting from the School/Environment" of school administrators. | School administration supports activities that recognize different cultures as a richness within the school. ($\bar{X} = 4.66$) | The view that extra training is provided for students with diversity to make better use of the school/environment and that this practice is useful was revealed. In the scale item, it was revealed that the school administration sees diversity as a richness. |
| "Continuous learning" sub-theme under the theme of "School culture" for school administrators and "Competitive environment" sub-theme under the theme of "School culture" for teachers. | School administration encourages innovative studies on cultural values. ($\bar{X} = 4.75$) | The interviews revealed that the school administration encourages continuous learning and that there is no bad competition between teachers within the school, but rather supportive competition. The scale item also revealed that innovative work is encouraged. |

| | | |
|---|---|---|
| "Academic development", "Social development", "Goals" sub-themes under the theme of "Student development" for school administrators, "School management" sub-theme under the theme of "Goals" for teachers, "Academic goals", "Social goals" sub-themes under the theme of "Goals" for parents, and "Teacher/administration support" sub-theme under the theme of "Goals" for students. | School administration sets high but achievable goals for the school in academic and social areas ($\bar{X} = 4.71$) | In the interviews, it was stated that courses are offered for students' academic development, activities such as excursions, observations and sports competitions are organized for social development, various goals are set for students and the school administration/teachers support these goals and the parents of the students find these goals achievable. In the scale item, it was stated that the school administration set high but achievable goals. |
| The sub-theme "Description of culture" under the theme "School culture", the sub-theme "Getting to know the teacher" under the theme "Teacher development". "Exemplary behavior" sub-theme under the theme of "Students with diversity" | School administration influences teachers with positive behaviors towards cultural differences. ($\bar{X} = 4.69$) | In the interviews, it was found that the school culture is focused on problem-free, and work, various interviews and observations are made to get to know the teacher, and students are treated equally and optimistically to create exemplary behaviors. In the scale item, it was revealed that the school administration influences teachers with positive behaviors. |
| The sub-themes of "Communication" under the theme of "Teacher development" for school administrators, and the sub-themes of "Trust environment" and "Relationship between teachers" under the theme of "School culture" for teachers. | There is an environment in the school where teachers with diversity feel safe ($\bar{X} = 4.67$) | The interviews revealed that various activities were organized to strengthen communication between teachers, that teachers felt safe, and that the relationship between teachers was strong and supportive. The scale item also revealed that there is an environment where teachers with diversity feel safe. |

**Table 9.** *Institutional dimension mixed analysis findings.*

| Qualitative Analysis Themes and Subthemes | Quantitative Analysis Scale Items | Qualitative and Quantitative Part Association |
|---|---|---|
| The sub-theme of "Adaptation to school culture" under the theme of "School culture" for teachers, and the sub-theme of "Valuing students" under the theme of "Student orientation" for students. | Teachers embrace the cultural diversity of the school ($\bar{X} = 4.13$) | In the interviews, it was revealed that teachers adapted to the school culture, that teachers valued students and were interested in their problems and provided material and moral support. In the scale item, it was revealed that teachers embraced the diversity of the school at the "Agree" level. |
| "Academic goals" sub-theme under the "Goals" theme for students. | Teachers strive for students to have a scientific outlook ($\bar{X} = 4.09$) | In the interviews, it was stated that teachers set high but achievable goals for students, but they did not set a clear goal in terms of having a scientific perspective. In the scale item, an opinion at the level of "I agree" was stated in terms of teachers having a scientific perspective. |
| "Exemplary behavior" sub-theme under the "Behavior" theme for teachers, "Giving importance" sub-theme under the "Reflecting the family" theme for students, and "Freedom of expression" sub-theme under the "In-class discussion" theme. | Teachers are respectful of students' differences ($\bar{X} = 4.54$) | In the interviews, it was stated that teachers behaved lovingly towards students with diversity, that students could express themselves freely and that the values in their families were reflected in the classroom environment. In the scale item, it was also stated that teachers were respectful towards students' differences. |
| "Out-of-class" sub-theme under the theme "Activities carried out" for teachers, and "Compulsory activities" and "Optional activities" sub-themes under the theme "Activities" for parents. | Teachers organize extracurricular activities in a way that is sensitive to cultural values ($\bar{X} = 4.29$) | In the interviews, it was found that teachers organize out-of-class activities such as celebrations and visits, and that mandatory activities are based on a specific plan, but optional activities are appropriate for students with diversity. In the scale item, the view that out-of-class activities are organized in a diversity-sensitive manner was also obtained. |
| The sub-theme of "Classroom practices" under the theme of "School culture" belonging to school administrators, the sub-themes of "Education and training practices" and "Providing high benefit" under the theme of "Classroom practices" belonging to teachers, and the sub-theme of "Classroom" under the theme of "Activities carried out". | Teachers organize in-class activities for diversity ($\bar{X} = 4.37$) | In the interviews, it was stated that the school culture supports sensitivity towards diversity in the classroom, but teachers do not implement any practices other than strengthening communication and providing support. In the scale item, it was stated that in-class activities were organized for diversity. |
| "Against teachers" and "Against the administration" sub-themes under the theme of "Expressing oneself". | Parents with diversity can express themselves freely within the school ($\bar{X} = 4.00$) | In the interviews, the parents of the students stated that they were able to express themselves to the teachers and the administration. In the scale item, it was determined at the level of "Agree" that the parents of the students could express themselves comfortably. |

| | | |
|---|---|---|
| "Values" sub-theme under the theme of "Class community cohesion", "Consistency" sub-theme under the theme of "Family school cohesion", and "Consistency" sub-theme under the theme of "Reflecting the family". | The school is in harmony with different cultural values in society ($\bar{X} = 4.07$) | In the interviews, the parents of the students stated that the school was consistent with the values of society, but not with the concepts, views and values of the family. Students also stated that it was not consistent. In the scale item, it is observed that compatibility with cultural values is stated at the "Agree" level. |
| "Parent support" sub-theme under the "Goals" theme for teachers, and "Trust in the process" sub-theme under the "Goals" theme for parents. | Culturally diverse students believe that the education they receive at school will improve them ($\bar{X} = 4.23$) | In the interviews, teachers stated that the education provided for students with diversity was supported by the parents of the students and that the parents of the students trusted the process. In the scale item, the view that students believe that the education they receive will improve themselves was obtained. |
| The sub-theme of "Student communication" under the theme of "Out-of-school processes" for parents, and the sub-theme of "Communication between students" under the theme of "In-school communication" for students. | All students are in positive communication ($\bar{X} = 4.04$) | While the parents of the students stated that the students come together outside the school for reasons such as games and homework, the students also stated that communication between the students within the school is smooth. In the scale item, it was revealed that they were in positive communication at the "Agree" level. |
| "Parent communication" and "Impact on students" sub-themes under the theme of "Out-of-school processes". | All parents are in positive communication ($\bar{X} = 3.91$) | In the interviews, it was stated that the parents of the students did not meet except for social activities, but that the meetings had a positive effect on the students. In the scale item, it is at the level of "I agree" that parents are in positive communication. |
| "Collaboration" and "Parent role" sub-themes under the "Goals" theme. | Parents trust the school's efforts to achieve diversity goals ($\bar{X} = 4.38$) | In the interviews, it was stated that the parents of the students cooperated to achieve the goals and provided financial/spiritual support and support at the point of control. In the scale item, it was also stated that they trusted these studies. |

**Table 10.** *Technical dimension mixed analysis findings.*

| Qualitative Analysis Themes and Subthemes | Quantitative Analysis Scale Items | Qualitative and Quantitative Part Association |
|---|---|---|
| "Freedom of expression" sub-theme under the theme of "Reflecting on family" and "Freedom of expression" sub-theme under the theme of "In-class discussion". | Teachers support students to express their opinions freely in the classroom $(\bar{X} = 4.60)$ | In the interviews, it was stated that the students could not express their opinions in the classroom, whereas it was stated that they could express their opinions freely in the classroom. In the scale item, the view that students can express their opinions freely was obtained. |
| "Values management" sub-theme under the theme of "Classroom practices" for teachers and "Special days" sub-theme under the theme of "Family reflection" for students. | Teachers ensure that students meet universal values in the classroom $(\bar{X} = 4.49)$ | In the interviews, the opinion was obtained that teachers provided values education for students to meet universal values, but students could not celebrate special days in the family in the classroom. In the scale item, the view that students are enabled to meet universal values was obtained. |
| "Exemplary behavior", "Behavioral competence", "Suggestions" sub-themes under the "Behavior" theme for teachers, and "Social goals" sub-theme under the "Goals" theme for students. | Teachers try to develop respect for differences in their students $(\bar{X} = 4.59)$ | In the interviews, it was found that teachers used various exemplary behaviors to develop respect for differences, thought that these behaviors were sufficient and suggested behaviors such as special attention and equal treatment. In the scale item, the view that teachers try to develop respect for differences in students was also obtained. |
| The "Continuous learning" sub-theme under the "School culture" theme for school administrators and the "Student reaction" sub-theme under the "Goals" theme for teachers. | Teachers provide the necessary conditions for students to be continuous learners $(\bar{X} = 4.27)$ | In the interviews, it was found that the school culture supports the behavior of being a continuous learner and teachers set the goal of being a continuous learner for students. In the scale item, the view that the necessary conditions are provided for students to be continuous learners was obtained. |
| The sub-themes of "Having awareness" and "The way of creating awareness" under the theme of "Creating awareness". | For teachers, individual differences are a tool to strengthen teaching and learning $(\bar{X} = 4.40)$ | In the interviews, it was stated that teachers were aware of the diversity in the classroom and that they gained awareness through in-service training etc. In the scale item, the view that teachers see individual differences as a tool to strengthen education was obtained. |
| "In-class" and "Out-of-class" sub-themes under the theme of "Activities carried out". | Teachers prepare an educational environment that allows cultural differences to emerge in the classroom $(\bar{X} = 4.53)$ | In the interviews, it was seen that teachers carried out activities such as strengthening communication and providing support for students with diversity and carried these activities outside the classroom with celebrations and visits. In the scale item, the view that they prepare an environment that enables cultural differences to emerge was also obtained. |

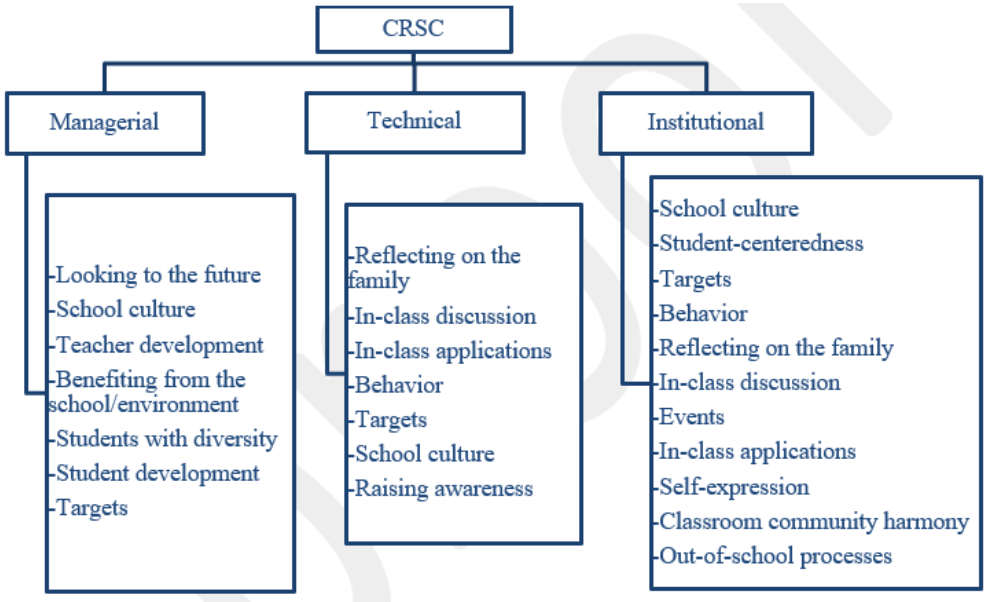
Some of the themes and sub-themes created within the scope of the findings obtained from the interviews could not be matched with the scale items. To examine these results from the interviews, the themes and sub-themes that could not be matched are given in Table 11.

**Table 11.** *Themes and sub-themes that could not be analyzed in the mixed method.*

| Working Group | Theme | Sub Theme |
|---|---|---|
| School Administrators | Looking to the Future | Mission, vision, and strategic plan |
| | School Resources | Cash Resources |
| | | In-kind Resources |
| | Students with Diversity | Curriculum |
| | Benefiting from the School/Environment | Not Recommended Practices |
| | | School Barriers |
| | | School Opportunities |
| Teachers | Harmony with School | Mission, vision, and strategic plan |
| | | Reference to studies |
| Parents | Students with Diversity | Education, training |
| Students | Differences | Students with Diversity |
| | | Empathy |
| | Students with Diversity | Education Together |
| | | Impact |
| | | Process |
| | | Curriculum |
| | | Measurement and Evaluation |

When Table 11 is examined, it can be said that practices that differ from school to school and practices that need to be implemented within certain patterns are not evaluated using the scale. As a result of the associations obtained, the distribution of qualitative themes into quantitative dimensions is shown in Figure 3. In Figure 3, culturally responsive school culture has three dimensions: administrative, technical, and institutional, and under these dimensions, there are various themes developed by the researcher using the qualitative method. These themes also provide information about the titles of the dimensions.

**Figure 3.** *Distribution of qualitative themes into quantitative dimensions.*



## 4. DISCUSSION and CONCLUSION

The findings obtained through the qualitative method are generally compatible with each other. Due to the application of a mixed method approach and the sequential exploratory design, the findings from the qualitative method facilitate the identification of variables related to culturally responsive school culture and the relationships between these variables. The components and dimensions of culturally responsive school culture were tried to be revealed. At this point, culturally responsive school culture can be can be conceptualized through components of development, universalization, and achievement. The component called development can be explained as all stakeholders renewing themselves and continuing their learning through school. The development component can be understood as all stakeholders engaging in continuous self-renewal and lifelong learning within the school context. The component called success is the achievement of all stakeholders' goals through the school.

During the interviews with school administrators, it was prepared at a basic level regarding the requirements of CRSC and were willing to work on diversity. Although there is no study directly related to the CRSC, it has been revealed that school administrators do not have negative opinions and attitudes towards immigrant students in studies (Altunay, & Dede, 2019; Koşar, & Aslan, 2020) when analyzed through immigrant students, where differences are most prominent. When considered together with the opinions obtained from the qualitative method, it can be said that school administrators have a narrow field of action because they are under centralized management and cannot create the necessary space to change school cultures. In the interviews with school administrators and in the results of the CRSC scale, it is observed that they are sensitive to diversity, recognize the needs of individuals with differences, and make efforts to display inclusive behaviors that support the adaptation of such individuals to society.

When the opinions of the teachers are examined, it is evident that teachers are aware of students with diversity, exhibit exemplary behaviors and consider these behaviors sufficient. Among the studies examining teachers' cultural responsiveness, various studies have been conducted on Social Studies teachers (Karaseyfioğlu & Çalışkan, 2021), classroom teachers (Batmaz, 2023) and Turkish teachers (Şengül & Yaprak, 2022), and in these studies, it was revealed that teachers' culturally responsiveness was high. Likewise, it is observed that teachers do not have negative opinions and attitudes in studies conducted on immigrant students, where differences are most evident (Özbek *et al.*, 2021; Yurdakul & Tok, 2018). It can be said that teachers exhibit various behaviors contributing to a culturally responsive school culture overall, but they often do so without systematic or conscious effort.

When the variables affecting the culturally responsive school culture were analyzed, it was found that the most important variable was the task variable. School administrators are more effective than teachers in making culturally responsive school culture. Çelikten (2003) also stated in his study that school culture is shaped in line with the visions of school principals and that school principals should follow various ways to shape school culture. School administrators with a postgraduate level of education are more effective in culturally responsive school culture compared to school administrators with undergraduate level of education. In a study conducted by Turhan and Yaraş (2013), it was found that postgraduate education provides support in specialization and leadership, contributes to problem solving, and creates awareness in professional development processes. Türkmenoğlu and Bülbül (2015), on the other hand, found that school administrators' postgraduate education has a significant effect on school culture. It is thought that the province where the teacher works influences the teacher. In a study conducted by Gül *et al.* (2018), it was stated that the level of development of the place of duty is effective in teachers having a human-centered attitude. Similarly, in the master's thesis conducted by Mutlu (2019), it was found that cultural and artistic activities, transportation, and climate type in the place of duty affect teachers' professional development and motivation. Female teachers are more compatible with a culturally responsive school culture than male teachers. In general, when the perceptions of school culture are examined, it is observed that gender does not create

a significant difference (Işık, 2017; Öztürk, 2015). Although the finding obtained within the scope of the research is not a finding of perception of school culture, teachers' professional development processes can be examined since it is related to culturally responsive school culture. In this context, when examined in terms of professional development, it is observed that there is a difference in favor of women regarding the gender variable (Ada & Ayık, 2008; Karahan *et al.*, 2018).

When the development, universalization and success, which are the components of the culturally responsive school culture obtained as a result of the interviews in the qualitative method, are interpreted with the culturally responsive school culture scale in the quantitative method, they can be interpreted as the outputs of the culturally responsive school culture. When interpreted with the findings obtained from the quantitative method, it is observed that these outcomes are supported by various concepts. Development is supported by the concepts of innovation and continuous learning when analyzed with quantitative method findings. Universalization is supported by the concepts of adaptation and increasing sensitivity. Success is supported by the concepts of trust and using diversity as a tool.

While it was stated in the interviews that no work was done for students with diversity in the school, it was stated in the scale that the school administration was sensitive to diversity. At this point, it can be interpreted that the school administration does not do any extra work but acts inclusively and respects the values, norms, beliefs and behaviors of students, teachers, and parents. Likewise, although there is an opinion that teachers' continuous learning and professional development processes are supported through in-service trainings, it is stated in the scale that studies are carried out for culturally responsive professional development. This can be explained by the fact that the school administration cannot organize a training within the school because it does not have a budget for in-school professional development, whereas the central administration conducts various studies for students with diversity and provides resources to teachers.

In the institutional dimension, it was found that the data obtained with the qualitative method could be generalized through the data obtained with the quantitative method. In the interviews, it was stated that there was no practice other than strengthening communication and providing support in classroom activities, while in the quantitative method, it was stated that in-class activities were organized for differences. This can be interpreted to mean that teachers have a limited curriculum in classroom activities and that the curriculum does not include flexibility for students with diversity. Similarly, it was determined that teachers and school administration did not set a clear goal for students to have a scientific perspective. Similarly, this may reflect the idea that the curriculum should set this goal. The fact that teachers do not perform any behavior for this purpose in in-class activities can be seen as a deficiency for teachers. During the interviews, the view that the school is in harmony with society but not in full harmony with the family emerged. In the quantitative method, the school's harmony with different cultural values was expressed at a low level. At this point, it can be suggested that the school should be flexible enough to provide a family-based harmony. A different result was obtained in communication between the parents of the students. During the interviews, it was stated that the parents of the students communicated within the scope of certain people during social activities, while the quantitative method revealed a low level of opinion that the parents were in positive communication. At this point, it is expected that while the school strengthens its communication with the parents of the students, it should also help the parents of the students to strengthen their communication within one other.

In the technical dimension, the generalizability of the findings obtained with the qualitative method through the findings of the quantitative method was observed. While students can express themselves freely in the classroom, they state that they cannot express their views freely in the family. This situation can be explained by the view that the school in the institutional dimension is compatible with the society but incompatible with the family. It can be thought

that the fact that both the curriculum and in-school communication are not flexible towards differences in the family causes this problem. The same situation can also be explained by the inability of students to bring special days in the family into the classroom.

Some findings obtained from the qualitative method could not be explained by the quantitative method and their generalizability could not be checked. These findings may be due to the items that were discarded both during scale construction and during exploratory and confirmatory factor analysis. When these non-generalizable issues are examined, it is observed that there are issues that differ by school. It can be said that the generalizability of these issues is also limited.

### 4.1. Conclusion

In the context of the findings obtained within the scope of the research and the related discussion, culturally responsive school culture is defined as a success-oriented school culture characterized by a high level of cultural competence within the school and its environment. It is a culture where the understanding of continuous learning is emphasized, individual differences are viewed as resources that enrich the educational environment, and universal values are embraced as fundamental principles. More briefly, a culturally responsive school culture can be defined as a school culture in which individuals achieve individual success, the school and its environment meet universal values, and continuous development is adopted within the school. At this point, the keywords universalization, development, and success can be considered as the outcomes of a culturally responsive school culture.

In general, a culturally responsive school culture is a school culture created under the leadership of school administrators and maintained by teachers. However, when it is compared with various organizational theories (learning school, effective school, productive school, etc.), it is observed that it has more inclusive features. At the same time, CRSC does not fit into the culture classification of support, bureaucratic, achievement and task (Terzi, 2005), which is generally accepted among culture classifications. Therefore, CRSC is not included in any type of culture, but it is a school culture built on various elements, levels, components, and values with its own unique characteristics.

For a school, complying with the CRSC standards will be a road map to improve both school culture and success. The general opinion about a successful school is that it is related to school culture or school climate (Balcı, 2002; Çobanoğlu, & Badan, 2016; Gürbüz *et al.*, 2013). In this context, it is recommended to carry out studies on CRSC to bring the school to a better point and to change the school culture into CRSC. CRSC takes place at four levels: individual, group, school, and environment. At the individual level, the individual adopts a continuous learning approach and is ready to do innovative work. At the group level, when people come together, they support each other's learning, create an environment of trust and act in harmony in supportive competition. At the school level, the focus is on continuous learning, using diversity as a tool, and doing innovative work. At the environmental level, the school engages in harmonious, continuous learning and development-oriented communication and interaction with the community and family. At this level, it is aimed to acquire the ability to be culturally responsive on a global scale.

### 4.2. Recommendations

#### 4.2.1. *Recommendations for Practitioners*

To create, support and develop a culturally responsive school culture, various suggestions are made to all stakeholders, especially teachers and school administrators.

Teachers and school administrators are recommended to promote different cultures in their schools, to organize cultural activities, to ensure diversity of resources in terms of courses, to take measures to prevent prejudices, to celebrate traditional days, to respect cultural labels, to use materials culturally responsive, to set goals to support students' social development, to

organize activities to ensure school-family harmony, and to continue postgraduate education for professional development.

#### 4.2.2. *Recommendations for Researchers*

Due to the limited research on CRSC, various suggestions were made to researchers. Researchers can apply the CRSC scale at various educational levels, create a guide for school administrators to develop culturally responsive school culture, and test the applicability of culturally responsive school culture standards.

## Acknowledgments

This study was produced from the doctoral thesis of the first author, supervised by the second author.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Pamukkale University, 09/03/2022 and 68282350/2022/G05 - 27/02/2023 and 68282350/2023/04.

## Orcid

Gürkan Sarıdaş https://orcid.org/0000-0002-7989-2130
Funda Nayır https://orcid.org/0000-0002-9313-4942

## REFERENCES

Altunay, E., & Dede, M. (2019). Göçmen öğrencilerin eğitim süreci açısından okullar arası işbirliğine ilişkin okul yöneticilerinin görüşleri [School administrators' views on inter-school cooperation in terms of the educational process of immigrant students]. *Ege Eğitim Dergisi*, *20*(1), 127-144. https://doi.org/10.12984/egeefd.474018

Baki, A., & Gökçek, T. (2012). Karma yöntem araştırmalarına genel bir bakış [An overview of mixed methods research]. *Elektronik Sosyal Bilimler Dergisi*, *11*(42), 1-21.

Balcı, A. (2002). *Etkili okul geliştirme [Development of effective school]*. Pegem Akademi.

Balcı, A. (2013). *Sosyal bilimlerde araştırma [Research in social sciences]*. Pegem Akademi.

Bandura, A. (2002). Social cognitive theory in cultural context. *Applied Psychology: An International Review, 51*(2), 269-290. https://doi.org/10.1111/1464-0597.00092

Batmaz, O. (2023). Sınıf öğretmenlerinin kültürel değerlere yönelik duyarlılıklarının incelenmesi [Investigation of classroom teachers' culturally responsiveness]. *Manisa Celal Bayar Üniversitesi Sosyal Bilimler Dergisi*, *21*(1), 127-141.

Blake, N., & Masschelein, J. (2002). Critical theory and critical pedagogy. In N. Blake, P. Smeyers, R. Smith, & P. Standish, *The blackwell guide to the philosophy of education* (pp. 38-56). Blackwell Publishers Ltd.

Charmaz, K. (2006). *Constructing grounded theory: a practical guide through*. Sage Pub.

Charmaz, K. (2008). Grounded theory as an emergent method. In S. N. Biber, & P. Leavy, *Handbook of emergent methods* (pp. 155-172). The Guilford Press.

Creswell, J.W. (2009). *Research design* (3rd ed.). SAGE Pub. Inc.

Creswell, J.W. (2021). *Karma yöntem araştırmalarına giriş [A concise introduction to mixed methods research]*. (M. Sözbilir, Trans.) Pegem Akademi.

Çelikten, M. (2003). Okul kültürünün şekillendirilmesinde müdürün rolleri [The roles of the principal in shaping school culture]. *Eğitim ve Bilim*, *31*(140), 56-61.

Gay, G. (2000). *Culturally responsive teaching*. Teachers College Press.

Guba, E.G., & Lincoln, Y.S. (2005). Paradigmatic controversies, contradictions and emerging confluences. In. N. K. Denzin, & Y. S. Lincoln (Eds.), *The sage handbook of qualitative research* (pp. 191-216). SAGE Publications, Inc.

Gül, S., Çobanoğlu, İ.H., Aydoğmuş, M., & Türk, H. (2018). Sınıf öğretmenlerinin çevreye yönelik tutumlarının incelenmesi: samsun ili örneği [Investigation of classroom teachers' attitudes towards the environment: the case of samsun province]. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi, 37*(2), 139-157, https://doi.org/10.7822/omuefd.427161

Gürbüz, R., Erdem, E., & Yıldırım, K. (2013). Başarılı okul müdürlerinin özellikleri [Characteristics of successful school principals]. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, (20), 167-179. https://dergipark.org.tr/tr/pub/zgefd/issue/47944/606579

Helms, J.E. (1986). Expanding racial identity theory to cover counseling process. *Journal of Counseling Psychology, 33*(1), 62-64. https://doi.org/10.1037/0022-0167.33.1.62

Horner, S.B., Fireman, G.D., & Wang, E.W. (2010). The relation of student behavior, peer status, race, and gender to decisions about school discipline using CHAID decision trees and regression modeling. *Journal of School Psychology*, *48*(2), 135-161. https://doi.org/10.1016/j.jsp.2009.12.001

Işık, H. (2017). Ortaokullarda okul kültürünün incelenmesi [Investigation of school culture in secondary schools]. *İstanbul Aydın Üniversitesi Dergisi*, *35*, 61-71.

Jones, L. (2009). The importance of school culture for instructional leadership. *International Journal of Educational Leadership Preparation*, *4*(4), 1-9. https://files.eric.ed.gov/fulltext/EJ1071384.pdf

Karahan, F., Gür-Erdoğan, D., Canan-Güngören, Ö., Kaya-Uyanık, G., Güler, N., Koç, B., Bayrakçı, M., Çalışkan, H., Demir-Uzunkol, E., & Horzum, M.B. (2018). Öğretmenlerin cinsiyet, kıdem, internet kullanım becerisi ve web tabanlı mesleki gelişime yönelik motivasyonlarının hizmet içi eğitimdeki başarılarına etkisi [The effect of teachers' gender, seniority, ınternet usage skills and motivation for web-based professional development on their success in ın-service training]. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, (36), 65-81.

Karaseyfioğlu, F., & Çalışkan, H. (2021). Sosyal bilgiler öğretmenlerinin kültürel değerlere yönelik duyarlılıklarının incelenmesi [Investigation of social studies teachers' culturally responsiveness]. *Journal of Multidisciplinary Studies in Education, 5*(3), 88-101.

Kayri, M., & Boysan, M. (2007). Araştırmalarda CHAID analizinin kullanımı ve baş etme stratejileri ile ilgili bir uygulama [The use of CHAID analysis in research and an application of coping strategies]. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 40*(2), 133-149.

Khalifa, M.A., Gooden, M.A., & Davis, J.E. (2016). Culturally responsive school leadership: a synthesis of the literature. *Review of Educational Research, 86*(4), 1272-1311.

Koşar, S., & Aslan, F. (2020). İlkokul ve ortaokula devam eden göçmen çocukların eğitim sorunlarına yönelik okul yöneticilerinin görüşleri [School administrators' views on the educational problems of immigrant children attending primary and secondary schools]. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi, 9*(4), 1799-1831.

Kotluk, N., & Kocakaya, S. (2018). Türkiye için alternatif bir anlayış: kültürel değerlere duyarlı eğitim [An alternative understanding for Turkey: culturally responsive education]. *YYÜ Eğitim Fakültesi Dergisi, 15*(1), 749-789.

Ladson-Billings, G. (1995). But that's just good teaching!: The case for culturally relevant pedagogy. *Theory Into Practice, 34*(3), 159-165.

MEB. (2021). *Milli eğitim istatistikleri [National education statistics]*. MEB. http://sgb.meb.gov.tr/meb_iys_dosyalar/2020_09/04144812_meb_istatistikleri_orgun_egitim_2019_2020.pdf

Mutlu, F.B. (2019). *Müzik öğretmenlerinin görev yaptıkları eğitim kurumunun çevresel koşullarının öğretmen motivasyonu üzerindeki etkisi (Osmaniye ili örneği) [The effect of the environmental conditions of the educational institution where music teachers work on teacher motivation (the case of Osmaniye province)]* [Yayımlanmamış Yüksek Lisans Tezi]. İnönü Üniversitesi.

Özbek, D., Özdaş, F., & Kavan, N. (2021). Beden eğitimi öğretmenlerinin göçmen öğrencilere ilişkin görüşlerinin incelenmesi: bir durum çalışması [Investigation of physical education teachers' views on immigrant students: a case study]. *Milli Eğitim Dergisi, 50*(Özel Sayı/2021), 785-806.

Özen, Y., & Gül, A. (2007). Sosyal ve eğitim bilimleri araştırmalarında evren örneklem sorunu [The universe sampling problem in social and educational sciences research]. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi,* (15), 394-422.

Öztürk, U.C. (2015). Örgüt kültürü algısında cinsiyet faktörünün etkisi ve bir uygulama [The effect of gender factor on the perception of organizational culture and an application]. *Süleyman Demirel Üniversitesi Vizyoner Dergisi*, *6*(12), 62-86.

Robbins, S.P., & Judge, T.A. (2017). *Örgütsel davranış [Organizational Behaviour]*. (İ. Erdem, Trans.) Nobel Akademik Yayıncılık.

Sarıdaş, G., & Nayir, F. (2020). Kültürel değerlere duyarlı eğitim ve uygulama örnekleri [Culturally responsive education and practice examples]. In F. Nayir (Eds.), *Kültürel Değerlere Duyarlı Eğitim [Culturally responsive education]* (pp. 256-257). Anı Yayıncılık.

Senge, P. (2014). *Öğrenen Okullar [Schools that learn]*. (N. Ü. Münevver Çetin, Trans.) Nobel Akademik Yayıncılık.

Shade, B.J., Kelly, C., & Oberg, M. (1997). *Creating culturally responsive classrooms*. American Psychological Association.

Şengül, K., & Yaprak., N. (2022). Türkçe öğretmenlerinin kültürel değerlere duyarlılıklarının bazı değişkenler açısından incelenmesi [Investigation of Turkish teachers' culturally responsiveness in terms of some variables]. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi, 11*(4), 1665- 1679.

Terzi, A.R. (2005), İlköğretim okullarında örgüt kültürü [Organizational culture in primary schools]. *Kuram ve Uygulamada Eğitim Yönetimi*, *43*, 423-442.

Turhan, M., & Yaraş, Z. (2013). Lisansüstü programların öğretmen, yönetici ve denetmenlerin mesleki gelişimine katkısı [The contribution of graduate programs to the professional development of teachers, administrators and supervisors]. *Elektronik Sosyal Bilimler Dergisi, 12*(43), 200-218.

Türkmenoğlu, G., & Bülbül T. (2015). Okul yöneticilerinin göreve geliş biçimlerinin okul kültürüne yansımaları [Reflections of school administrators' way of taking office on school culture]. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, *11*(2), 526-549.

Türnüklü, A. (2000). Eğitimbilim araştırmalarında etkin olarak kullanılabilecek nitel bir araştırma tekniği: görüşme [A qualitative research technique that can be used effectively in educational research studies: interview]. *Kuram ve Uygulamada Eğitim Yönetimi*, *24*, 543-559.

Villegas, A.M., & Lucas, T. (2007). The culturally responsive teacher. *Educational Leadership, 64(*6), 28-33.

Weinstein, C., Curran, M., & Tomlinson-Clarke, S. (2003). Culturally responsive classroom management: awareness into action. *Theory Into Practice, 42*(4), 269-276.

Yıldırım, A., & Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri [Qualitative research methods in social sciences]*. Seçkin Yayıncılık.

Yurdakul, A., & Tok, T.N. (2018). Öğretmen gözüyle mülteci / göçmen öğrenci [Refugee / immigrant student from the perspective of a teacher]. *Adnan Menderes Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Dergisi*, *9*(2), 46-58.

Zohrabi, M. (2013). Mixed method research: instruments, validity, reliability and reporting findings. *Theory and Practice in Language Studies*, *3*(2), 254-262.

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 414–429
https://doi.org/10.21449/ijate.1474855
journal homepage: https://dergipark.org.tr/en/pub/ijate *Research Article*

# The effect of polytomous item ratio on ability estimation in multistage tests

**Hasibe Yahsi Sari**[1*], **Hulya Kelecioglu**[2]

[1]Necmettin Erbakan University, Eregli Faculty of Education, Department of Educational Sciences, Konya, Türkiye
[2]Hacettepe University, Faculty of Education, Department of Educational Sciences, Ankara, Türkiye

**ARTICLE HISTORY**
*Received: Apr. 29, 2024*
*Accepted: Mar. 6, 2025*

**Keywords:**
Multistage testing,
Polytomous item,
Mixed test,
Ability estimation,
Pisa 2018.

**Abstract:** The aim of the study is to examine the effect of polytomous item ratio on ability estimation in different conditions in multistage tests (MST) using mixed tests. The study is simulation-based research. In the PISA 2018 application, the ability parameters of the individuals and the item pool were created by using the item parameters estimated from the dichotomous and polytomous items obtained in the field of reading skills. MST conditions; panel design, test lengths, routing methods, and polytomous item ratio. Simulation data, MST pattern and analysis were obtained with the help of WinGen, CPLEX, and the "mstR" package in the R Studio program. A total of 108 conditions and 100 replications were examined in the study. As a result of the simulations, RMSE, mean absolute bias and correlation values were calculated. As a result of the research, it is seen that when the ratio of polytomous items in the tests increases from 10% to 50%, the mean absolute bias and RMSE values decrease while the correlation values increase. As the test length increases, RMSE and mean absolute bias values decrease while correlation values increase. In terms of routing methods, MFI performed better than the NC routing method. In general, three-stage panel designs gave significantly better results than two-stage panel designs. In 1-2 and 1-4 panel designs, it does not matter which routing method is used.

## 1. INTRODUCTION

Computer-based measurement and assessment applications in education have been utilized since the 1980s (Weiss, 1982; Weiss & Kingsbury, 1984). Over time, advancements in technology and the integration of artificial intelligence have significantly increased the importance of technology in educational measurement and assessment. Among these innovations, computerized adaptive testing (CAT) and multistage testing (MST) have gained global prominence. The first large-scale application of CAT was conducted in the United States with the Graduate Record Examination (GRE) in 1993. Some of the exams that have been computerized adaptive include The Law School Admission Test (LSAT), the Test of English as a Foreign Language (TOEFL), the National Council of Architectural Registration Boards (NCARB), the National Assessment of Educational Progress (NAEP), and the U.S. Medical Licensure Examination (USMLE) (Hendrickson, 2007). In recent years, CAT has been replaced by MST in national and international large-scale exams. GRE was implemented as MST in 2011, the Programme for International Student Assessment (PISA) in 2018, and the Programme for the International

*CONTACT: Hasibe YAHSİ SARİ ✉ hsbyahsi@gmail.com Necmettin Erbakan University, Eregli Faculty of Education, Department of Educational Sciences, Konya, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

Assessment of Adult Competencies (PIAAC) in 2012 (Yamamoto *et al*., 2019). Notably, in 2023, the Scholastic Aptitude Test (SAT), a standardized large-scale international test taken by students who want to attend university in the United States, was administered in digital format as MST. Finally, in 2024, the SAT administered throughout the U.S. had fully transitioned to MST.

The most significant advantage provided by CAT is the establishment of a relationship between the individual's ability and item difficulty. Unlike fixed test lengths in traditional paper-and-pencil test applications, personalized test lengths are available in CAT applications (Wang, 2017; Weiss, 1983). However, criticisms of CAT include varying test lengths, the inability to revisit previous questions, and differences in questions presented to test-takers. MST emerged as a solution to address some limitations of CAT while maintaining its adaptive nature. MST organizes tests into modules, stages, and panels. Test-takers are directed to subsequent modules based on their performance, creating a dynamic yet structured pathway for assessment (Zheng *et al*., 2012). MST applications increase in popularity due to their ability to tailor to individuals, allow test developers to preview test forms in advance, and enable individuals taking the test to review their answers. In MST, control over content and other features provides increased test security and content management (Hendrickson, 2007). While CAT adapts to individuals at the item level, MST adapts at the module level. MST encourage individuals with lower ability levels while preventing boredom in individuals with higher ability levels. These features have made MST increasingly popular for large-scale assessments, particularly those requiring precise and efficient measurement across broad ability ranges.

Tests composed of both dichotomously and polytomously scored items are called mixed-format tests. Mixed-format tests are crucial both due to their increasing prevalence in international large-scale tests and because they offer higher levels of test information compared to dichotomous items (Rosa *et al*., 2001). In recent years, many large-scale assessments, such as PIAAC and PISA, now employ mixed-format tests combining dichotomous (e.g., multiple-choice) and polytomous (e.g., open-ended) items (Yamamoto *et al*., 2019). In addition to mixed-format tests, the proportions of polytomous items used also play a significant role in the accuracy of ability estimation (Kim & Dodd, 2014). Kim *et al*. (2013) examined various routing methods and panel designs within the context of classification testing in MST applications based on the partial credit model. Their findings indicated that longer tests yielded higher accuracy, whereas a 50% pass rate resulted in the lowest accuracy levels. Park *et al*. (2014) explored a new item pool usage method for mixed-format tests in MST. Three designs were implemented using the linear programming (LP) test assembly method, with item change rates of 0.22, 0.44, and 0.66 for each test combination. The findings demonstrated that the applied MST recombination conditions enhanced item pool utilization while maintaining the desired MST structure.

### 1.1. Purpose and Importance of Research

The aim of this study is to examine the impact of varying polytomous item ratios in individually adapted multistage tests on the accuracy of ability estimation under different testing conditions. A review of the literature reveals numerous studies on tests comprising both dichotomous and polytomous items, where MST is analyzed under various conditions, including test assembly methods, routing strategies, and test lengths (Dogruoz, 2018; Kim *et al*., 2010; Luecht & Nungester, 1998; Luo & Kim, 2018; Yahsi Sari & Kelecioglu, 2023; Wang, 2017). While the importance of mixed-format tests is well-recognized, research on their application within MST frameworks remains limited (Kim *et al*., 2012; Kim *et al*., 2013). This study seeks to bridge this gap by investigating the interaction between polytomous item ratios and other critical testing variables, thereby making a substantial contribution to the literature.

The item ratio in MST with mixed-format tests is anticipated to exhibit significant variability under different conditions, highlighting the theoretical and practical importance of this study. In the international context, simulation data based on the item parameters of dichotomous and

polytomous items from real-world MST applications in the reading domain of PISA 2018 have been utilized, ensuring ecological validity. This approach offers valuable insights for future applications of PISA and other large-scale assessments. The statistical properties of the routing module are especially crucial, as they significantly affect the overall measurement accuracy of the test (Kim & Plake, 1993). In the national context, there is a clear gap in research addressing mixed-format tests within MST applications. While real-world CAT studies have been conducted at the research level (Cikrikci *et al*., 2020; Senel & Kutlu, 2018), studies focusing on MST applications are still emerging. This research will not only contribute to the international literature but also advance MST applications within the national context. Furthermore, there is limited research on the performance of test assembly methods under varying conditions with mixed-format tests, particularly those based on the linear programming approach-one of the automatic test assembly methods (Park *et al*., 2014; Park, 2015). To address these gaps, future studies should explore the application of linear programming test assembly methods to mixed-format tests, examining the effects of different panel designs, routing methods, and proportions of polytomous items. Such research would contribute significantly to advancing both the theory and practice of MST, supporting its widespread adoption in educational assessment. This study seeks to address the following research questions:

1- Primary Research Question:

- How does the ratio of polytomous items in a test (10%, 30%, and 50%) affect individuals' ability estimations?

2- Secondary Research Questions:

- How do test length (20, 40, and 60 items) and panel design (“1-2”, “1-3”, “1-4”, “1-2-2”, “1-2-3”, and “1-3-3”) influence the relationship between the ratio of polytomous items and ability estimations?
- How do routing methods (Maximum Fisher Information vs. Number-Correct) interact with the ratio of polytomous items in affecting ability estimations?

By exploring these questions, the study aims to provide practical recommendations for optimizing MST design and improving the precision of ability estimates across diverse testing conditions.

## 2. METHOD

### 2.1. Research Method

The aim of the research is to examine the effect of polytomous item ratios in individually adapted multistage mixed tests on ability estimation under different conditions. Systematic research aimed at generating new products or new processes, or significantly improving existing ones, by utilizing existing knowledge gained from research or experience is experimental development research (OECD, 2002). In the research, new conditions were determined that will improve the systems produced by simulation. In this aspect, the research is experimental development research based on simulation.

### 2.2. Sample of Research

In PISA 2018, a multistage computer-based test was administered in the domain of reading skills. For this research, data were generated through simulation based on the parameters of dichotomous and polytomous items from individuals who participated in the PISA 2018 assessment in the reading skills domain. According to the *PISA 2018 Technical Report Final Annex A*, 244 items-72 trend items and 172 new items-were utilized in the computerized multistage reading skills test. The data were obtained from the website of the Organization for Economic Co-operation and Development (OECD).

### 2.3. Research Design

In this section, the manipulated and fixed conditions in the simulation study are explained. The fixed conditions in the study include the sample size (10,000 individuals), the distribution of

individuals' ability levels (normal distribution, *N* (0,1)), and the method used for ability estimation for the ability levels calculated at the end of MST (Expected a Priori-EAP). EAP (Bock & Mislevy, 1982) is one of the popular methods that belong to the Bayesian ability estimations. There are several studies in the literature that use normal ability distribution and the EAP ability estimation method (Park, 2015; Sahin & Ozturk, 2019). Therefore, the EAP ability estimation method was used in this study. The manipulated conditions in the study are panel design, total test length, routing method, and the ratio of polytomous items in the total test length. These conditions were determined considering the most frequently used conditions in both simulation studies and real MST applications according to the existing literature.

In the panel design, six different panel designs were used: "1-2", "1-3", "1-4", "1-2-2", "1-2-3", and "1-3-3", based on Patsula (1999) "1-3" and "1-3-3", Zenisky (2004) "1–3–3", "1–2–3", "1–3–2", and Sarı and Raborn (2018) "1-3", "1-2-2", "1-2-3", and "1-3-3". For test lengths, Wang (2017) 45 and 60; Sarı and Raborn (2018) used test lengths of 30 and 60. Based on the existing literature, three different test lengths were determined as 20, 40 and 60, representing short, medium and long test lengths.

When we examine the literature in terms of routing method: Approximate Maximum Information (AMI), Defined Population Intervals (DPI), Modified Approximate Maximum Information (M-AMI), Stage-Level DPI (SL-DPI), and Module-Level DPI (ML-DPI) method are frequently used routing methods in the literature (Kim *et al*., 2010; Kim *et al*., 2013; Wang, 2017; Zenisky, 2004). Also, there are some studies that provide the maximum amount of information based on Maximum Likelihood Estimation (MLE) and M-AMI of ability (Kim *et al*., 2012). Number of Correct Responses (NC) is another common routing method in MST (Weissman *et al*., 2007; Zenisky *et al*., 2010). In recent years, using the mstR (Magis *et al*., 2018) package, Maximum Fisher Information (MFI), Maximum Likelihood Weighted Information (MLWI), Maximum Posterior Weighted Information (MPWI), Kullback–Leibler (KL), and Posterior Kullback–Leibler (KLP) routing methods have also begun to be used (Boztunc-Ozturk, 2019; Sarı & Raborn, 2018). In this study, as routing methods, two directing methods were determined: MFI and NC.

Park (2015) determined the polytomous item rates as (10%, 30%, 50% and 70%) in his study. In this study, the ratio of polytomous items to the total test length was determined as 10%, 30% and 50%. All conditions were crossed with each other, resulting in 108 (6x3x2x3) different conditions examined in the study. The simulation was repeated 100 times for all conditions. The manipulated conditions of the research are summarized in Table 1.

**Table 1.** *Research condition.*

| Condition | Number of Levels | Lower Levels |
|---|---|---|
| Panel Design | 6 | 1-2<br>1-3<br>1-4<br>1-2-2<br>1-2-3<br>1-3-3 |
| Test Length | 3 | 20 items<br>40 items<br>60 items |
| Routing Methods | 2 | MFI<br>NC |
| Ratio of Polytomous Items | 3 | %10<br>%30<br>%50 |
| Total | 6x3x2x3=108 | |

### 2.4. Data Collection Process

In this stage, 200 polytomous and 400 dichotomous item parameters and ability levels (theta-$\theta$) of 10,000 individuals were generated using the WinGen3 program (Han, 2007). The three stages are explained in detail.

#### 2.4.1. *1. Stage: Generation of ability levels*

In the first stage, ability levels for a sample of 10,000 individuals were generated using the WinGen3 program according to a normal distribution ($N$ (0,1)) with a mean of 0 and a standard deviation of 1. 100 iterations were performed to generate ability levels. The probability of responding to each item for individuals generated in the simulation was calculated according to the Generalized Partial Credit Model (GPCM).

#### 2.4.2. *2. Stage: Obtaining data and conducting item analyses*

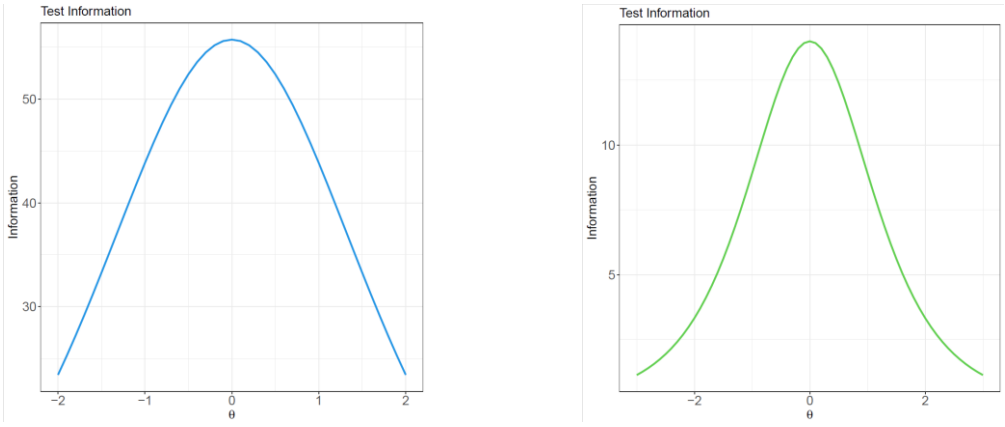
A simulation study based on the characteristics of item parameters used in the MST application in reading skills in PISA 2018 was conducted in the research. Among these items, 223 were dichotomous and 21 were polytomous. In this context, MST was also applied in the research, and parameters of both dichotomous and polytomous items were used. Descriptive statistics of item parameters were calculated using the R program (R Development Core Team, 2018). Descriptive statistics of the original items are provided in Table 2.

**Table 2.** *Descriptive statistics of item parameters.*

| | 2 categories | | 3 categories | | |
|---|---|---|---|---|---|
| | a | b | a | b1 | b2 |
| Min. | 0.15 | -1.91 | 0.40 | -0.40 | -2.02 |
| Max. | 1.83 | 2.66 | 1.25 | 2.07 | 2.02 |
| Mean | 1.01 | 0.03 | 0.72 | 0.65 | 0.20 |

The test information functions of the item pools consisting of items used in the MST application in the reading skills domain of PISA 2018 are shown in Figure 1. When comparing the test information functions in Figure 1, we observe that polytomous items provide more information, despite being fewer in number than dichotomous items.

**Figure 1.** *Test information functions for PISA data.*




Dichotomous items Polytomous items

#### 2.4.3. *3. Stage: Generation of Item Pool*

In this stage, item pools were generated using the estimated item parameters. Item parameters for dichotomous items were generated using the Two-Parameter Logistic (2PL) IRT model, while item parameters for polytomous items were generated using the GPCM. When generating

item parameters for the items in the item pool, the distribution of item parameters used in the reading skills domain of PISA 2018 was considered.

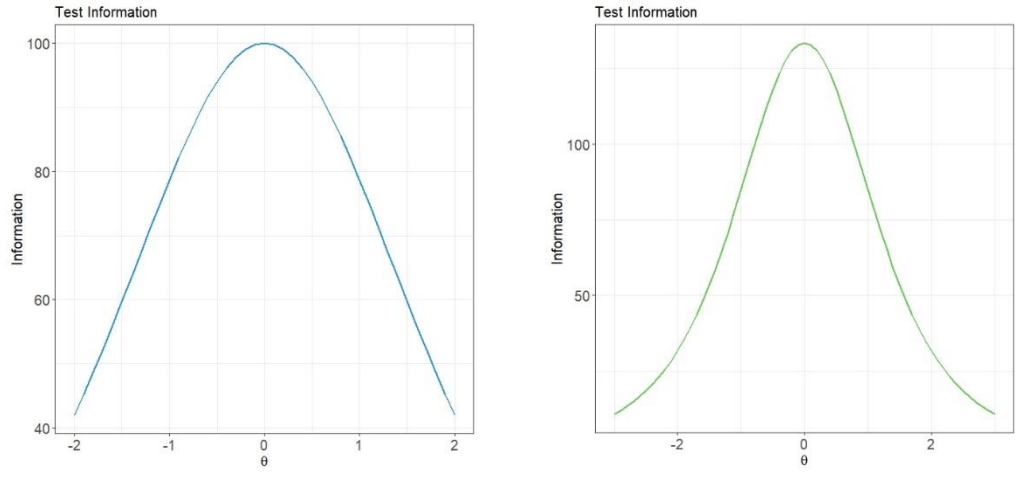
In MST studies conducted in the literature, the item pool size generally varies between 200 and 600 items (Lim, 2019; Xing & Hambleton, 2004; Zheng *et al*., 2012). In MST studies using mixed-item formats, 424 items have been used (Kim *et al*., 2012; Park *et al*., 2014). In the study by Kim *et al*. (2012), out of the 424-item pool, 244 items (57.55%) were dichotomous, 113 items (26.65%) were three-category, and 67 items (15.80%) were four-category items. In this study, a total of 600 items were generated for the item pool, with 400 being two-category and 200 being three-category items. During the panel construction stage, polytomous items were distributed into modules representing 10%, 30%, and 50% of the entire test. Descriptive statistics of the generated 400 and 200-item pools are provided in Table 3.

**Table 3.** *Descriptive statistics of item parameters in the 400 and 200 items pool.*

| | 400 items | | 200 items | | |
|---|---|---|---|---|---|
| | a | b | a | b1 | b2 |
| Min. | 0.15 | -1.90 | 0.40 | -0.40 | -0.34 |
| Max. | 1.82 | 2.65 | 1.25 | 1.98 | 2.07 |
| Mean | 0.97 | 0.44 | 0.85 | 0.42 | 1.21 |

When examining the descriptive statistics of the PISA items and the simulated item parameters, slight differences were observed, which are likely due to the size of the item pool. The test information functions of the generated item pools consisting of 400 two-category and 200 three-category items are shown in Figure 2. When we examine the distribution of test information functions of PISA and simulation data, it is observed that the amount of test information obtained from the simulation is higher due to the larger number of items.

**Figure 2.** *Test information functions for item pools.*




Dichotomous items Polytomous items

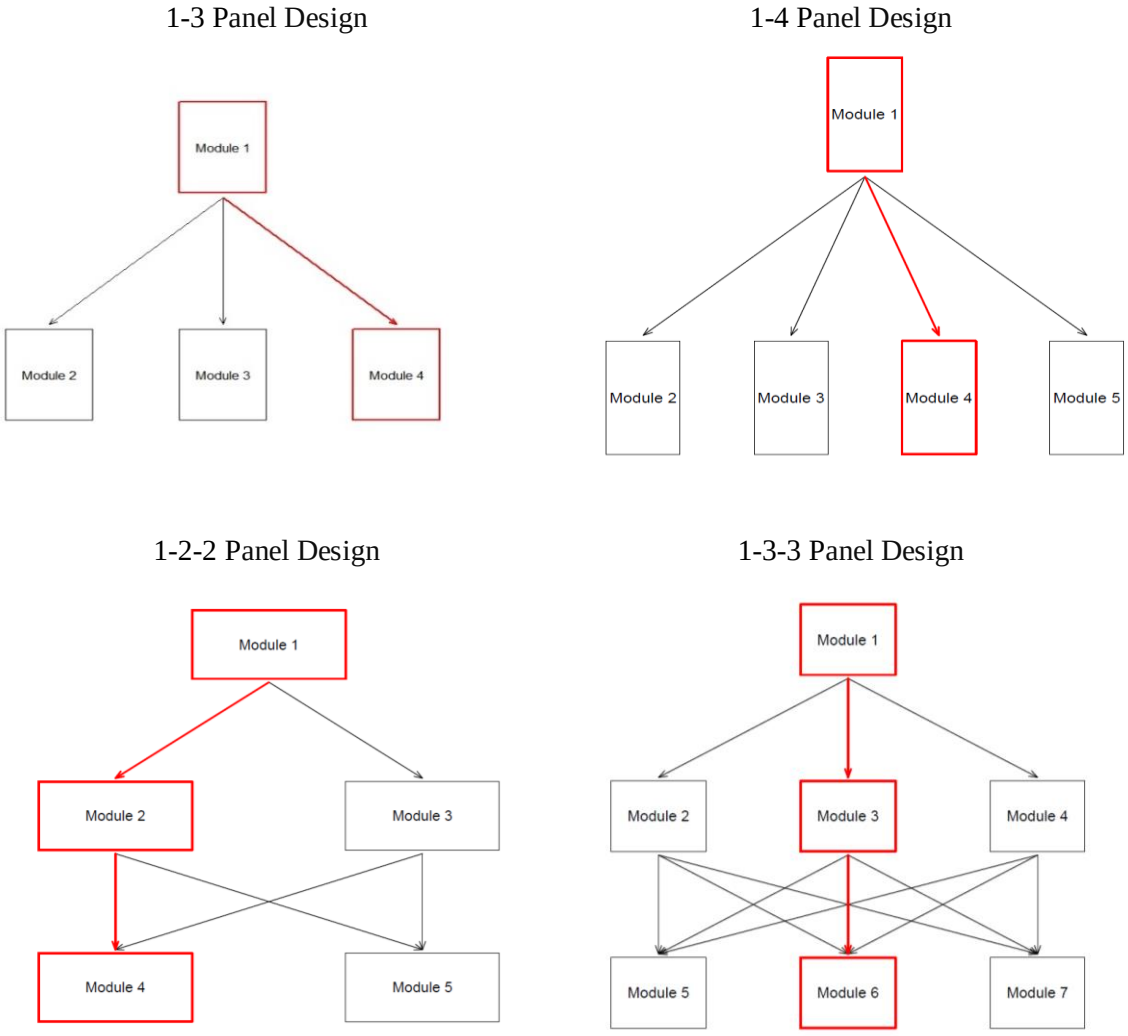
## 2.5. Data Analysis

An MST simulation was created as a data analysis tool. The MST simulation consists of modules and panels. There is a total of six-panel designs: 1-2, 1-3, 1-4, 1-2-2, 1-2-3, and 1-3-3. There are three test lengths: 20, 40, and 60. The number of items in the modules varies according to the total test length and panel designs.

### 2.5.1. *MST Simulation*

In the process of selecting items for panels in different designs in the MST, the following steps were taken. In the first stage, the information functions of items were calculated to place the items into panels as easy, medium, and difficult. The information functions for polytomous items were calculated using the GPCM, and for dichotomous items, they were calculated using the 2PLM. In the MST simulation, there are a total of six-panel designs: 1-2, 1-3, 1-4, 1-2-2, 1-2-3, and 1-3-3, and specific paths. Some sample panel designs and paths used in the study are shown in Figure 3.

**Figure 3.** *Examples of panel designs used in MST simulation.*




The creation of panels and test assembly process were carried out using the IBM CPLEX optimization program. Individuals were randomly assigned to panels in the study. The bottom-up test assembly method was used in creating panels. First of all, in order to maximize the module information, codes were written in the CPLEX program such that the item information levels were easy ($\theta$ = -1), low medium ($\theta$ =-0.3), medium ($\theta$ = 0), high medium ($\theta$ = 0.3) and difficult ($\theta$ = 1). Items that provided the highest information for each module were assigned to the modules based on the characteristics and numbers of each module. The selection of items designated as easy, medium, and difficult was carried out considering the total test length and the ratio of polytomous items. For example, the number of items in modules for the 1-2 panel design is detailed in Table 4.

**Table 4.** *Number of items in modules.*

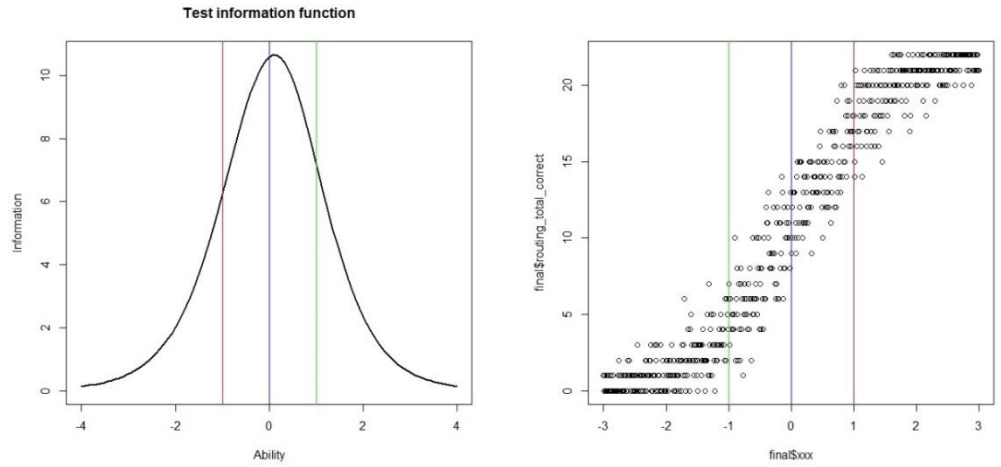
| | | Ratio of polytomous items | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | %10 | | | %30 | | | %50 | | |
| | Test length | 20 | 40 | 60 | 20 | 40 | 60 | 20 | 40 | 60 |
| Panel Design 1-2 | 1.Stage-M | 1*/9 | 2*/18 | 3*/27 | 3*/7 | 6*/14 | 9*/21 | 5*/5 | 10*/10 | 15*/15 |
| | 2.Stage -E | 1*/9 | 2*/18 | 3*/27 | 3*/7 | 6*/14 | 9*/21 | 5*/5 | 10*/10 | 15*/15 |
| | 2. Stage -H | 1*/9 | 2*/18 | 3*/27 | 3*/7 | 6*/14 | 9*/21 | 5*/5 | 10*/10 | 15*/15 |

*Indicates the number of polytomous items in the modules. M: Medium, E: Easy, H: Hard

When examining Table 4, for the special condition of 1-2 design, 60 items and 10% polytomous item ratio, 30 items were selected and 10% (3 of them) were polytomous, giving high information at the point $\theta$= 0. Similarly, under the condition $\theta$=-1, 30 items were selected to provide high information and placed in the second easy module. Finally, under the condition $\theta$=1, 30 items were selected to provide high information and placed in the second difficult module.

After creating modules and panels, MST simulations were continued with the R program. The simulation study for the MST was conducted using the 'mstR' package in R (Magis *et al*., 2018). The responses of individuals to all items were calculated using the GPCM. The EAP method was used for estimating abilities. In MST applications, commonly used routing methods include the NC (Weissman *et al*., 2007; Zenisky *et al*., 2010), ability estimation, and MFI (Weissman *et al*., 2007). NC routing method is an alternative method commonly used in IRT ability estimation in CAT (Armstrong, 2002). In this study, MFI and NC routing methods were used to transition from one stage to another. In the MFI routing method, the maximum amount of information is provided in one module and directed to the appropriate module in the next stage. In the NC-based routing method, an IRT model is used in the background to obtain a score distribution. Research indicates that the effectiveness of the NC directing method applied based on the raw number of correct responses is consistent with MST results designed for $\theta$ estimation (Armstrong, 2002). In the NC method, individuals are directed between stages based on the cumulative cut score point for each module. When determining the cut scores, the module information functions obtained by individuals according to MFI and their total number of correct responses were calculated for each stage. The correct numbers that provided the highest information were chosen as cut scores between modules. For example, for the 1-3 panel design with a 10% polytomous item ratio and a 40-item test, the graphs of the directing module information and total number of correct responses are shown in Figure 4.

**Figure 4.** *1-3 Panel design, 10% polytomous item ratio, routing module information and total numbers correct of the 40-item test.*



### 2.5.2. *Evaluation Criteria*

After obtaining simulation data under appropriate conditions for sub-problems, Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and correlations between actual and estimated ability levels were calculated to examine the accuracy of the obtained ability estimation. Bias is the square root of the average of the differences between the actual and estimated values. MAE is the absolute value of each difference between the actual and estimated values. RMSE is the square root of the average of the squares of the differences between the actual and estimated values. Low values of MAE and RMSE, and high correlation values indicate high accuracy of ability estimations. These values were calculated separately for each iteration for evaluation of the findings. Total values for each condition were divided by the number of iterations to report average values.

After obtaining MAE, RMSE, and correlation values, factorial ANOVA was conducted in SPSS using 100 iterations for each condition to determine if there was a significant difference among the condition variables. Effect sizes obtained from factorial ANOVA were examined. Three separate factorial ANOVA tests were conducted with MAE, RMSE, and correlation values as dependent variables, while all other conditions (ratio of polytomous items, test length, panel design, and routing method) were independent variables.

## 3. FINDINGS

### 3.1. Findings Related to Mean Absolute Error

The MAE findings for all conditions are shown in Table 5. The factorial ANOVA results on MAE are given in Table 6. Based on the overall findings, test length emerged as the most influential factor, accounting for 98% of the variance ($\eta^2_p$ =.98), followed by the ratio of polytomous items in the test ($\eta^2_p$ = .20). The routing method (MFI vs. NC) also showed a moderate effect ($\eta^2_p$ = .13), while the panel design exhibited a smaller yet significant influence ($\eta^2_p$ = .05). Differences across panel designs and routing methods were smaller but indicated a slight advantage for more complex panel designs and the MFI routing method. However, the ratio of polytomous items had minimal effect on MAE.

**Table 5.** *Findings of mean absolute error across all conditions.*

| Routing Method | Panel Design | Ratio of Polytomous Items in the Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | %10 | | | %30 | | | %50 | | |
| | | 20 | 40 | 60 | 20 | 40 | 60 | 20 | 40 | 60 |
| MFI | 1-2 | 0.24 | 0.18 | 0.15 | 0.23 | 0.17 | 0.15 | 0.23 | 0.17 | 0.15 |
| | 1-3 | 0.24 | 0.18 | 0.15 | 0.23 | 0.17 | 0.15 | 0.23 | 0.17 | 0.15 |
| | 1-4 | 0.24 | 0.18 | 0.16 | 0.23 | 0.18 | 0.15 | 0.23 | 0.17 | 0.15 |
| | 1-2-2 | 0.23 | 0.18 | 0.15 | 0.23 | 0.17 | 0.14 | 0.23 | 0.17 | 0.14 |
| | 1-2-3 | 0.23 | 0.18 | 0.15 | 0.23 | 0.17 | 0.15 | 0.23 | 0.17 | 0.14 |
| | 1-3-3 | 0.23 | 0.18 | 0.15 | 0.23 | 0.17 | 0.15 | 0.23 | 0.17 | 0.14 |
| NC | 1-2 | 0.24 | 0.18 | 0.15 | 0.23 | 0.17 | 0.15 | 0.23 | 0.17 | 0.14 |
| | 1-3 | 0.24 | 0.18 | 0.16 | 0.24 | 0.18 | 0.15 | 0.24 | 0.18 | 0.15 |
| | 1-4 | 0.24 | 0.18 | 0.16 | 0.23 | 0.17 | 0.15 | 0.23 | 0.17 | 0.15 |
| | 1-2-2 | 0.25 | 0.18 | 0.16 | 0.24 | 0.17 | 0.15 | 0.24 | 0.17 | 0.16 |
| | 1-2-3 | 0.24 | 0.18 | 0.15 | 0.24 | 0.17 | 0.15 | 0.24 | 0.18 | 0.14 |
| | 1-3-3 | 0.24 | 0.18 | 0.16 | 0.24 | 0.18 | 0.15 | 0.25 | 0.18 | 0.15 |

**Table 6.** *ANOVA Results for grand mean absolute bias.*

| Factor | Sum of Squares | *df* | Mean Square | *F* | *p* | $\eta^2_p$ |
|---|---|---|---|---|---|---|
| Routing | 0.05 | 1 | 0.05 | 1594.11 | 0.00 | 0.13 |
| TL | 14.00 | 2 | 7.00 | 228391.13 | 0.00 | 0.98 |
| PIR | 0.08 | 2 | 0.04 | 1297.90 | 0.00 | 0.20 |
| PD | 0.02 | 5 | 0.00 | 104.28 | 0.00 | 0.05 |
| Routing*TL | 0.01 | 2 | 0.00 | 112.19 | 0.00 | 0.02 |
| Routing*PIR | 0.00 | 2 | 0.00 | 18.52 | 0.00 | 0.00 |
| Routing*PD | 0.04 | 5 | 0.09 | 273.18 | 0.00 | 0.11 |
| TL*PIR | 0.00 | 4 | 0.00 | 10.24 | 0.00 | 0.00 |
| TL*PD | 0.01 | 10 | 0.00 | 24.61 | 0.00 | 0.02 |
| PIR*PD | 0.00 | 10 | 0.00 | 12.92 | 0.00 | 0.01 |
| Routing*TL*PIR | 0.00 | 4 | 0.00 | 12.23 | 0.00 | 0.01 |
| Routing*TL*PD | 0.01 | 10 | 0.00 | 32.32 | 0.00 | 0.03 |
| Routing*PIR*PD | 0.00 | 10 | 0.00 | 8.09 | 0.00 | 0.01 |
| TL*PIR*PD | 0.01 | 20 | 0.00 | 11.23 | 0.00 | 0.02 |
| Routing*TL*PIR*PD | 0.01 | 20 | 0.00 | 11.41 | 0.00 | 0.02 |
| Error | 0.33 | 10692 | 0.00 | | | |
| Total | 391.02 | 10800 | | | | |

*Note.* PIR represents the Polytomous Item Ratio, TL refers to the Test Length, and PD denotes the Panel Design

### 3.2. Findings Related to Root Mean Square Error (RMSE)

The RMSE findings for all conditions are shown in Table 7. The factorial ANOVA results on this outcome are given in Table 8. The findings reveal that test length is the most influential factor ($\eta^2_p$=.98), with longer tests consistently reducing RMSE values across all conditions. For instance, RMSE values decreased notably as the test length increased from 20 to 60 items, irrespective of the routing method or panel design. The ratio of polytomous items, while statistically significant, exhibited a minimal impact on RMSE ($\eta^2_p$=.19). Routing methods demonstrated a moderate effect ($\eta^2_p$=.19), with MFI consistently outperforming NC routing in reducing RMSE, particularly in tests with longer lengths or more complex panel designs. Similarly, panel design contributed to slight improvements in estimation accuracy.

**Table 7.** *Findings of RMSE across all conditions.*

| | | Ratio of Polytomous Items in the Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | %10 | | | %30 | | | %50 | | |
| Routing Method | Panel Design | 20 | 40 | 60 | 20 | 40 | 60 | 20 | 40 | 60 |
| MFI | 1-2 | 0.30 | 0.23 | 0.19 | 0.30 | 0.22 | 0.19 | 0.29 | 0.22 | 0.19 |
| | 1-3 | 0.30 | 0.23 | 0.19 | 0.30 | 0.22 | 0.19 | 0.29 | 0.22 | 0.19 |
| | 1-4 | 0.30 | 0.23 | 0.20 | 0.30 | 0.22 | 0.19 | 0.29 | 0.22 | 0.19 |
| | 1-2-2 | 0.30 | 0.22 | 0.19 | 0.29 | 0.22 | 0.18 | 0.29 | 0.22 | 0.18 |
| | 1-2-3 | 0.30 | 0.22 | 0.19 | 0.29 | 0.22 | 0.18 | 0.29 | 0.22 | 0.18 |
| | 1-3-3 | 0.30 | 0.22 | 0.19 | 0.29 | 0.22 | 0.18 | 0.29 | 0.22 | 0.18 |
| NC | 1-2 | 0.30 | 0.23 | 0.19 | 0.30 | 0.22 | 0.19 | 0.29 | 0.22 | 0.18 |
| | 1-3 | 0.30 | 0.23 | 0.20 | 0.30 | 0.23 | 0.19 | 0.30 | 0.22 | 0.19 |
| | 1-4 | 0.30 | 0.23 | 0.20 | 0.30 | 0.22 | 0.19 | 0.29 | 0.22 | 0.19 |
| | 1-2-2 | 0.32 | 0.24 | 0.20 | 0.31 | 0.22 | 0.20 | 0.31 | 0.22 | 0.20 |
| | 1-2-3 | 0.31 | 0.23 | 0.19 | 0.30 | 0.22 | 0.19 | 0.30 | 0.22 | 0.18 |
| | 1-3-3 | 0.31 | 0.23 | 0.21 | 0.31 | 0.22 | 0.19 | 0.32 | 0.23 | 0.19 |

**Table 8.** *ANOVA results for grand mean RMSE.*

| Factor | Sum of Squares | *df* | Mean Square | *F* | *p* | $\eta^2_p$ |
|---|---|---|---|---|---|---|
| Routing | 0.11 | 1 | 0.11 | 2437.64 | 0.00 | 0.19 |
| TL | 22.72 | 2 | 11.36 | 251592.07 | 0.00 | 0.98 |
| PIR | 0.11 | 2 | 0.06 | 1219.26 | 0.00 | 0.19 |
| PD | 0.03 | 5 | 0.01 | 107.25 | 0.00 | 0.05 |
| Routing*TL | 0.02 | 2 | 0.01 | 191.05 | 0.00 | 0.04 |
| Routing*PIR | 0.00 | 2 | 0.00 | 17.60 | 0.00 | 0.00 |
| Routing*PD | 0.09 | 5 | 0.02 | 409.14 | 0.00 | 0.16 |
| TL*PIR | 0.00 | 4 | 0.00 | 6.33 | 0.00 | 0.00 |
| TL*PD | 0.02 | 10 | 0.00 | 38.50 | 0.00 | 0.04 |
| PIR*PD | 0.01 | 10 | 0.00 | 15.59 | 0.00 | 0.01 |
| Routing*TL*PIR | 0.00 | 4 | 0.00 | 13.39 | 0.00 | 0.01 |
| Routing*TL*PD | 0.02 | 10 | 0.00 | 50.33 | 0.00 | 0.05 |
| Routing*PIR*PD | 0.01 | 10 | 0.00 | 11.43 | 0.00 | 0.01 |
| TL*PIR*PD | 0.02 | 20 | 0.00 | 21.16 | 0.00 | 0.04 |
| Routing*TL*PIR*PD | 0.02 | 20 | 0.00 | 18.73 | 0.00 | 0.03 |
| Error | 0.48 | 10692 | 0.00 | | | |
| Total | 633.56 | 10800 | | | | |

*Note.* See notes in Table 6.

### 3.3. Findings Related to Correlation

The correlation findings for all conditions are shown in Table 9. The factorial ANOVA results on this outcome are given in Table 10. As shown in Table 9, correlations increase with test length, consistently improving from 20 to 60 items, regardless of the routing method, panel design, or ratio of polytomous items. The ratio of polytomous items demonstrated minimal impact on correlation values, as observed in other metrics. Across all panel designs and routing methods, correlations remained virtually identical for tests with 10%, 30%, and 50% polytomous items. Routing methods showed slight differences, with MFI generally producing marginally higher correlations compared to NC routing, especially in longer tests. ANOVA reinforced that test length ($\eta^2_p$=.91) is the most influential factor in maximizing the alignment between estimated and true abilities, followed by the ratio of polytomous items ($\eta^2_p$=.07) and routing method ($\eta^2_p$=.07).

**Table 9.** *Findings of correlations across all conditions.*

| | | Ratio of Polytomous Items in the Test | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | %10 | | | %30 | | | %50 | | |
| Routing Method | Panel Design | 20 | 40 | 60 | 20 | 40 | 60 | 20 | 40 | 60 |
| MFI | 1-2 | 0.95 | 0.97 | 0.98 | 0.96 | 0.98 | 0.98 | 0.96 | 0.98 | 0.98 |
| | 1-3 | 0.95 | 0.97 | 0.98 | 0.95 | 0.98 | 0.98 | 0.96 | 0.98 | 0.98 |
| | 1-4 | 0.95 | 0.97 | 0.98 | 0.95 | 0.97 | 0.98 | 0.96 | 0.98 | 0.98 |
| | 1-2-2 | 0.96 | 0.97 | 0.98 | 0.96 | 0.98 | 0.98 | 0.96 | 0.98 | 0.98 |
| | 1-2-3 | 0.96 | 0.98 | 0.98 | 0.96 | 0.98 | 0.98 | 0.96 | 0.98 | 0.98 |
| | 1-3-3 | 0.96 | 0.97 | 0.98 | 0.96 | 0.98 | 0.98 | 0.96 | 0.98 | 0.98 |
| NC | 1-2 | 0.95 | 0.97 | 0.98 | 0.96 | 0.98 | 0.98 | 0.96 | 0.98 | 0.98 |
| | 1-3 | 0.95 | 0.97 | 0.98 | 0.95 | 0.97 | 0.98 | 0.95 | 0.97 | 0.98 |
| | 1-4 | 0.95 | 0.97 | 0.98 | 0.96 | 0.98 | 0.98 | 0.96 | 0.98 | 0.98 |
| | 1-2-2 | 0.95 | 0.97 | 0.98 | 0.95 | 0.98 | 0.98 | 0.95 | 0.98 | 0.98 |
| | 1-2-3 | 0.95 | 0.97 | 0.98 | 0.95 | 0.98 | 0.98 | 0.95 | 0.97 | 0.98 |
| | 1-3-3 | 0.95 | 0.97 | 0.98 | 0.95 | 0.97 | 0.98 | 0.95 | 0.97 | 0.98 |

**Table 10.** *ANOVA results for correlation.*

| Factor | Sum of Squares | *df* | Mean Square | *F* | *p* | $\eta^2_p$ |
|---|---|---|---|---|---|---|
| Routing | 0.01 | 1 | 0.01 | 810.48 | 0.00 | 0.07 |
| TL | 1.34 | 2 | 0.67 | 51715.76 | 0.00 | 0.91 |
| PIR | 0.01 | 2 | 0.01 | 389.50 | 0.00 | 0.07 |
| PD | 0.00 | 5 | 0.00 | 24.99 | 0.00 | 0.01 |
| Routing*TL | 0.01 | 2 | 0.00 | 244.91 | 0.00 | 0.04 |
| Routing*PIR | 0.00 | 2 | 0.00 | 14.16 | 0.00 | 0.00 |
| Routing*PD | 0.01 | 5 | 0.00 | 106.40 | 0.00 | 0.05 |
| TL*PIR | 0.01 | 4 | 0.00 | 99.69 | 0.00 | 0.04 |
| TL*PD | 0.00 | 10 | 0.00 | 24.70 | 0.00 | 0.02 |
| PIR*PD | 0.00 | 10 | 0.00 | 7.57 | 0.00 | 0.01 |
| Routing*TL*PIR | 0.00 | 4 | 0.00 | 3.54 | 0.01 | 0.00 |
| Routing*TL*PD | 0.00 | 10 | 0.00 | 34.33 | 0.00 | 0.03 |
| Routing*PIR*PD | 0.00 | 10 | 0.00 | 7.16 | 0.00 | 0.01 |
| TL*PIR*PD | 0.00 | 20 | 0.00 | 4.63 | 0.00 | 0.01 |
| Routing*TL*PIR*PD | 0.00 | 20 | 0.00 | 3.21 | 0.00 | 0.01 |
| Error | 0.14 | 10692 | 0.00 | | | |
| Total | 10158.54 | 10800 | | | | |

*Note.* See notes in Table 6.

## 4. DISCUSSION and CONCLUSION

This study investigates the impact of varying proportions of polytomous items, test length, panel design, and routing methods on the accuracy of ability estimation in MST. The findings highlight the critical role of test length and the ratio of polytomous items in improving estimation accuracy, with secondary influences from panel design and routing methods. These results contribute to a nuanced understanding of the conditions under which MST achieves optimal precision.

The results demonstrate that as the ratio of polytomous items in the test increases from 10% to 50%, the estimation of abilities improves, reflected by lower MAE and RMSE values and higher correlations between estimated and true abilities. This aligns with the theoretical premise that polytomous items provide more information per item than dichotomous ones (Embretson & Reise, 2000). The enhanced test information associated with a higher ratio of polytomous items enables more precise ability estimation, a finding corroborated by previous research on mixed-format tests (Kim *et al*., 2012). However, the differences between 10% and 30% proportions were relatively minor, suggesting that the benefits of polytomous items become more pronounced at higher ratios. In contrast to Park (2015), who found that an increased ratio of polytomous items reduced measurement precision due to challenges in test construction, this study observed consistent improvements in accuracy, likely attributable to well-constructed test panels that maintained sufficient test information.

Test length emerged as the most significant factor influencing ability estimation, as indicated by the ANOVA results, which showed that it explained the highest variance across MAE, RMSE, and correlations. Longer tests consistently yielded lower error rates and higher correlations, with tests comprising 40 or 60 items achieving the most accurate ability estimates. This finding is consistent with prior MST studies, which emphasize the role of test length in enhancing measurement precision (Erdem-Kara, 2019; Kim *et al*., 2013; Patsula, 1999). The results underscore the importance of prioritizing test length during test assembly to achieve optimal accuracy.

Panel design also influenced estimation accuracy, with three-stage panel designs generally outperforming two-stage designs. This result aligns with the broader MST literature, which indicates that additional stages in the panel design allow for more refined routing decisions, thereby reducing estimation error (Dogruöz, 2018; Patsula, 1999). However, the effect size was smaller compared to the test length and the ratio of polytomous items, suggesting that panel design plays a supportive but less critical role in overall accuracy.

Routing methods showed a modest but significant effect on estimation accuracy, with the MFI method consistently outperforming the NC method. This difference was particularly evident in conditions involving longer tests and higher proportions of polytomous items. Unlike Weissman *et al*. (2007), who found comparable performance between MFI and NC routing methods in terms of classification accuracy, this study suggests that MFI provides superior measurement accuracy in contexts requiring precise ability estimation. The superior performance of MFI in this study may stem from its ability to optimize information across items, whereas NC routing is more constrained by cutoff score determination.

Overall, this study highlights the critical interplay among test length, item composition, panel design, and routing methods in achieving accurate ability estimates in MST. The results emphasize the importance of prioritizing test length and leveraging the informational advantages of polytomous items while optimizing panel design and routing methods to further enhance precision. These findings provide practical guidance for test developers in designing MST frameworks that balance psychometric precision with operational constraints. Future research should explore these dynamics further in operational settings and extend the analysis to other adaptive testing contexts.

### 4.1. Recommendations

The findings of this study provide several practical and theoretical insights for optimizing MST frameworks. Key recommendations are presented in the following paragraphs:

First, test length emerged as the most influential factor in improving ability estimation accuracy. Therefore, it is recommended to prioritize longer tests in MST applications, particularly when measurement precision is critical. Tests with 40 or more items consistently yielded lower error rates and higher correlations, regardless of other factors. However, operational constraints such as time limitations and test-taker fatigue should also be considered when extending test lengths.

Second, the ratio of polytomous items in mixed-format tests significantly enhances the accuracy of ability estimates. While the effects of increasing polytomous item ratios from 10% to 30% were modest, the benefits became more pronounced at a 50% ratio. These findings suggest that incorporating a balanced ratio of polytomous items can maximize test information and improve measurement precision. However, test developers should ensure that the item pool remains sufficiently diverse and aligned with test specifications to maintain content validity.

Third, the panel design should be tailored to the specific conditions of the test. Under conditions with long tests (60 items) and high proportions of polytomous items (50%), differences in measurement accuracy across panel designs were minimal. This suggests that simpler designs (e.g., "1-2") could be employed in such cases to reduce operational complexity without compromising accuracy. For tests with shorter lengths or lower proportions of polytomous items, more complex designs (e.g., "1-2-3") may be beneficial to enhance measurement precision.

Fourth, the choice of routing method is critical for ensuring accuracy in ability estimation. The MFI routing method consistently outperformed the NC method across most conditions, particularly for longer tests and higher proportions of polytomous items. However, interactions between panel designs and routing methods were observed, such as with the "1-2" and "1-4" designs, where both methods performed similarly. In such cases, operational considerations,

such as ease of implementation or alignment with institutional goals, may guide the choice of routing method.

Future research should address the limitations of this study by expanding the scope of item parameters and test conditions. For instance, the data in this study were based on item parameters from the PISA 2018 Reading Skills domain test. Investigating other domains or disciplines with different item characteristics (e.g., difficulty, discrimination) could provide broader insights. Additionally, exploring how MST frameworks perform across diverse cultural and linguistic contexts would offer valuable perspectives on the generalizability of these findings.

Finally, this study assumed that individuals' ability levels followed a normal distribution. Future research could investigate how alternative distributions (e.g., uniform, skewed) influence ability estimation accuracy. Additionally, examining the effects of systematically arranging polytomous items across modules, rather than random distribution, could yield insights into how item pool characteristics interact with panel designs and routing methods. Such investigations could lead to more tailored and flexible MST designs that accommodate diverse testing needs. These recommendations provide actionable guidance for test developers while identifying key areas for future research to enhance the flexibility, accuracy, and fairness of MST applications.

## Acknowledgments

This study was produced from the doctoral thesis of the first author.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

## Contribution of Authors

**Hasibe Yahsi Sari**: Literature Review, Investigation, Data Collection, Visualization, Statistical Analysis, and Writing. **Hulya Kelecioglu**: Supervision, Methodology, and Writing.

## Orcid

Hasibe Yahsi Sari https://orcid.org/0000-0002-0451-6034
Hulya Kelecioglu https://orcid.org/0000-0002-0741-9934

## REFERENCES

Armstrong, R.D. (2002). *Routing rules for multi-form structures* (LSAC Computerized Testing Report No. 02-08). Law School Admission Council. https://searchworks.stanford.edu/view/6794803

Bock, R.D., & Mislevy, R.J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, *6*(4), 431-444. https://doi.org/10.1177/014662168200600405

Boztunç-Öztürk, N. (2019). How the length and characteristics of routing module affect ability estimation in CA-MST? *Universal Journal of Educational Research*, *7*(1), 164-170. https://doi.org/10.13189/ujer.2019.070121

Cıkrıkcı, N., Yalçın, S., Kalender, İ., Gül, E., *et al*. (2020). Development of a computerized adaptive version of the Turkish Driving Licence Exam. *International Journal of Assessment Tools in Education*, *7*(4), 570-587. https://doi.org/10.21449/ijate.716177

Dallas, A. (2014). *The effects of routing and scoring within a computer adaptive multi-stage framework* [Doctoral dissertation, University of North Carolina]. University of North Carolina Libraries. https://libres.uncg.edu/ir/uncg/f/Dallas_uncg_0154D_11394.pdf

Doğruöz, E. (2018). *Bireyselleştirilmiş çok aşamalı testlerin test birleştirme yöntemlerine göre incelenmesi* [Doktora tezi, Hacettepe Üniversitesi]. Hacettepe Üniversitesi Açık Erişim Sistemi. http://hdl.handle.net/11655/5298

Erdem-Kara, B. (2019). *Değişen madde fonksiyonu gösteren madde oranının bireyselleştirilmiş bilgisayarlı ve çok aşamalı testler üzerindeki etkisi* [Doktora tezi, Hacettepe Üniversitesi]. Hacettepe Üniversitesi Açık Erişim Sistemi. http://hdl.handle.net/11655/11968

Han, K.T. (2007). WinGen: Windows Software That Generates Item Response Theory Parameters and Item Responses. *Applied Psychological Measurement*, *31*(5), 457-459. https://doi.org/10.1177/0146621607299271

Hendrickson, A. (2007). An NCME instructional module on multistage testing. *Educational Measurement: Issues and Practice*, *26*(2), 44-52. https://doi.org/10.1111/j.1745-3992.2007.00093.x

Kim, J., Chung, H., & Dodd, B.G. (2010, April). *Comparing routing methods in the multistage test based on the partial credit model* [Conference presentation] Annual meeting of the American Educational Research Association, Denver, CO.

Kim, J., Chung, H., Dodd, B.G., & Park, R. (2012). Panel design variations in the multistage test using the mixed-format tests. *Educational and Psychological Measurement, 72*(4), 574-588. https://doi.org/10.1177/001316441142897

Kim, J., Chung, H., Park, R., & Dodd, B.G. (2013). A comparison of panel designs with routing methods in the multistage test with the partial credit model. *Behavior Research Methods, 45*(4), 1087-1098. https://doi.org/10.3758/s13428-013-0316-3

Kim, J., & Dodd, B.G. (2014). Mixed-format multistage tests: Issues and methods. In D. Yan, A.A. von Davier, & C. Lewis (Eds.), *Computerized multistage testing: Theory and applications* (pp. 55–67). CRC Press.

Kim, H., & Plake, B.S. (1993). Monte carlo simulation comparison of two-stage testing and computerized adaptive testing. *Annual Meeting of the National Council on Measurement in Education*, Atlanta, GA. https://eric.ed.gov/?id=ED357041

Luo, X., & Kim, D. (2018). A top-down approach to designing the computerized adaptive multistage test. *Journal of Educational Measurement*, *55*(2), 243-263. https://doi.org/10.1111/jedm.12174

Luecht, R.M., & Nungester, R.J. (1998). Some practical examples of computer-adaptive sequential testing. *Journal of Educational Measurement*, *35*(3), 229-249. http://www.jstor.org/stable/1435202

Magis, D., Yan, D., & von Davier A.A. (2018). *Package 'mstR': Procedures to generate patterns under multistage testing*. https://cran.r-project.org/web/packages/mstR/mstR.pdf

OECD (2002). *Proposed standard practice for surveys on research and experimental development: Frascati Manual 2002*. OECD Publishing. https://doi.org/10.1787/9789264199040-en

Park, R. (2015). *Investigating the impact of a mixed-format item pool on optimal test designs for multistage testing* [Doctoral dissertation, The University of Texas]. University of Texas Libraries. http://hdl.handle.net/2152/31011

Park, R., Kim, J., Chung, H., & Dodd, B.G. (2014). Enhancing pool utilization in constructing the multistage test using mixed-format tests. *Applied Psychological Measurement*, *38*(4), 268-280. https://doi.org/10.1177/0146621613515

Park, R., Kim, J., Chung, H., & Dodd, B. G. (2017). The development of mst test information for the prediction of test performances. *Educational and Psychological Measurement*, *77*(4), 570–586. https://doi.org/10.1177/0013164416662960

Patsula, L.N. (1999). *A comparison of computerized adaptive testing and multistage testing.* [Doctoral dissertation, University of Massachusetts]. University of Massachusetts Libraries. https://doi.org/10.7275/10994910

R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Rosa, K., Swygert, K.A., Nelson, L., & Thissen, D. (2001). Item response theory applied to combinations of multiple-choice and constructed-response items—scale scores for patterns of summed scores. In D. Thissen & H. Wainer (Eds.), *Test scoring* (1st ed., pp. 253–292). Routledge. https://doi.org/10.4324/9781410604729

Sahin, M.G., & Ozturk, N.B. (2019). Analyzing the maximum likelihood score estimation method with fences in ca-MST. *International Journal of Assessment Tools in Education*, *6*(4), 555-567. https://dx.doi.org/10.21449/ijate.634091

Sari, H.I., & Raborn, A. (2018). What information works best?: A Comparison of Routing Methods. *Applied Psychological Measurement*, *42*(6), 499-515. https://doi.org/10.1177/0146621617752990

Senel, S., & Kutlu, O. (2018). Computerized adaptive testing design for students with visual impairment. *Education and Science Journal*,*43*(194), 261-284. https://doi.org/10.15390/EB.2018.7515

Wang, K. (2017). *A fair comparison of the performance of computerized adaptive testing and multistage adaptive testing* [Doctoral dissertation, University of Michigan State]. University of Michigan State Libraries. https://doi.org/doi:10.25335/ypy5-6g68

Weiss, D.J. (1982). Improving measurement quality and efficiency with adaptive testing. *Applied Psychological Measurement*, *6*(4), 473-492. https://doi.org/10.1177/014662168200600408

Weiss, D.J. (1983). Latent trait theory and adaptive testing. In Weiss D.J. (Ed.), *New horizons in testing* (pp. 5-7). Academic Press.

Weiss, D.J., & Kingsbury, G.G. (1984). Application of computer adaptive testing to educational problems. *Journal of Educational Measurement*, *21*(4), 361-375. https://doi.org/10.1111/j.1745-3984.1984.tb01040.x

Weissman, A., Belov, D.I., & Armstrong, R.D. (2007). Information-based versus number-correct routing in multistage classification tests, *LSAC Computerized Testing Report, 7(*5). Law School Admission Council.

Yahsi Sari, H., & Kelecioglu, H. (2023). Ability estimation with polytomous items in computerized multistage tests. *Journal of Measurement and Evaluation in Education and Psychology*, *14*(3), 171-184. https://doi.org/10.21031/epod.1056079

Yamamoto, K., Shin, H., & Khorramdel, L. (2019). Introduction of multistage adaptive testing design in PISA 2018. *OECD Education Working Papers, No. 209*, OECD Publishing. https://doi.org/10.1787/b9435d4b-en

Zenisky, A.L. (2004). *Evaluating the effects of several multi-stage testing design variables on selected psychometric outcomes for certification and licensure assessment* [Doctoral dissertation, University of Massachusetts]. University of Massachusetts Libraries. https://doi.org/10.7275/18739572

Zenisky, A.L., Hambleton, R.K., & Luecht, R.M. (2010). Multistage testing: Issues, designs, and research. In W.J. van der Linden & C.A.W. Glas (Eds.), *Elements of adaptive testing* (pp. 355-372). Springer.

Zheng, Y., Nozawa, Y., Gao, X., & Chang, H.H. (2012). Multistage adaptive testing for a large-scale classification test: The designs, automated heuristic assembly, and comparison with other testing modes. *ACT Research Reports 2012-6.* https://www.act.org/content/dam/act/unsecured/documents/ACT_RR2012-6.pdf

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 430–456
https://doi.org/10.21449/ijate.1464011
journal homepage: https://dergipark.org.tr/en/pub/ijate *Research Article*

# Investigating the relationship among gifted students' spatial reasoning skills, their conceptual understanding of astronomy and academic achievement

**Pelin Ertekin**[1*], **Mustafa Serdar Köksal**[2]

[1]İnönü University, Faculty of Education, Department of Special Education, Malatya, Türkiye
[2]Hacettepe University, Faculty of Education, Department of Special Education, Ankara, Türkiye

**ARTICLE HISTORY**
*Received: Apr. 7, 2024*
*Accepted: Jan. 9, 2025*

**Keywords:**
Gifted students,
Spatial reasoning skill,
Science achievement,
Conceptual understanding,
Astronomy subjects.

**Abstract:** The purpose of this research is to examine the relationship among sixth-, seventh-, and eighth-grade gifted students' spatial reasoning skills, conceptual understanding of basic astronomy topics, and science achievement. To achieve this, the study was designed as a multi-factor predictive correlational study. The research sample consists of 642 gifted middle school students enrolled in Science and Art Centers across 12 cities in Turkey. Data were collected using the Conceptual Understanding of Basic Astronomy Subjects Test, the Mental Image-Focused Spatial Reasoning Skill Test, and the Science Achievement Test. Path analysis was employed as the statistical method, and a path diagram was used to illustrate the explanatory relationships among the observed variables. The analysis revealed that grade level positively and significantly predicts static and dynamic spatial reasoning skills, conceptual understanding of basic astronomy topics, and science achievement among gifted students. Additionally, both static and dynamic spatial reasoning skills positively and significantly predict students' conceptual understanding of basic astronomy topics. Similarly, science achievement is positively and significantly predicted by students' static and dynamic spatial reasoning skills as well as their conceptual understanding of basic astronomy topics. Overall, the findings indicate positive and significant direct and indirect relationships between gifted students' grade level, static and dynamic spatial reasoning skills, conceptual understanding of basic astronomy topics, and science achievement.

## 1. INTRODUCTION

The pursuit of answers to complex questions about the universe, Earth and nature plays a crucial role in fostering a strong connection between astronomy and the natural sciences (Gündoğdu, 2014). Moreover, the interdisciplinary nature of astronomy is evident in its involvement with time measurement, calendars, daily and seasonal weather changes, the formation of day and night, the effects of sunlight, and tidal patterns (Kurnaz, 2012; Percy, 2006; Percy, 2009). Consequently, advancements in astronomy have contributed to progress in fundamental natural sciences such as physics, chemistry, and biology, while also enhancing the potential of gifted students who aspire to pursue academic careers in these fields by fostering creativity and

*CONTACT: Pelin ERTEKIN pelin.ertekin@inonu.edu.tr İnönü University, Faculty of Education, Department of Special Education, Malatya, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

innovation. Gifted students possess advanced skills such as solving scientific problems, generating innovative scientific ideas, making connections between scientific concepts and observed phenomena, thinking creatively and critically, engaging in scientific discussions, and sharing scientific ideas with their peers (Bailey *et al*., 2016; Gardner & Sternberg, 1994; Gilbert & Newberry, 2007). Such children have a significantly greater potential to become scientists, driven by their curiosity about natural phenomena and their willingness to apply scientific skills (e.g., observing, measuring, classifying, experimenting) to satisfy their curiosity (Pyryt, 2000). Accordingly, astronomy provides a valuable context by incorporating challenges that demand high-level thinking (Wallace *et al*., 2013). Moreover, astronomy can foster the development of advanced skills such as critical analysis, data-driven decision-making, questioning, and both critical and creative thinking, particularly in relation to gifted individuals' experiences and observations of daily life (Goodman *et al*., 2011; Percy, 2006; Prather *et al*., 2009; Wallace *et al*., 2013).

Astronomy plays a significant role in fostering academic success and promoting a positive attitude toward science across all age groups (Percy, 2009). Similarly, Baker (1987) describes astronomy as an effective tool for science education, providing gifted students with the motivation and willingness necessary for their learning processes. However, due to the abstract nature of astronomical concepts and the difficulty in structuring scientific explanations for them, numerous studies have identified astronomy as one of the most challenging subjects for students at all grade levels (Gazit *et al*., 2005; Lelliot & Rollnick, 2010; Plummer *et al*., 2014; Yair *et al*., 2003). Even gifted individuals with advanced cognitive skills exhibit misconceptions about astronomical topics, similar to other learner groups. For instance, in a study conducted by Kim *et al*. (2011), the conceptual understanding of lunar observations among seventh-grade gifted students was examined.

The results showed that students primarily focused on a single reference frame, particularly when depicting the Moon's position in their drawings of its phases. Additionally, some students exhibited conceptual misconceptions, such as believing that the Moon rotates clockwise and that its phases are caused by Earth's shadow. Similarly, Subaşı *et al*. (2015) conducted semi-structured interviews to explore seventh-grade gifted students' conceptual perceptions of basic astronomy topics. The interview responses were categorized into five levels: understanding, limited understanding, not understanding, misunderstanding, and no response. Findings indicated that students' understanding of the concept of orbit, in particular, contained scientific misconceptions, which pose a significant challenge in the science education of gifted students. Such misconceptions can hinder further learning, especially in astronomy, a multidisciplinary field that encompasses key scientific concepts and phenomena (e.g., gravity, energy, the electromagnetic spectrum, entropy, life, light, and reflection), particularly in relation to fundamental sciences such as physics, chemistry, biology, and geology. Therefore, fostering a strong conceptual understanding of astronomy is essential, as it serves as a critical context for developing high-level cognitive skills, including hypothetical, inductive, and deductive reasoning, as well as problem-solving-key objectives in the science education of gifted students (Hannust & Kikas, 2007; Taber, 2007).

Numerous studies in the literature focus on understanding the formation processes of concepts and phenomena in astronomy, aiming to identify the conditions that influence these observed challenges (Atwood & Atwood, 1997; Barnett & Morran, 2002; Brunsell & Marcks, 2005; Frede, 2006; Kanlı, 2014; Kanlı, 2015; Küçüközer, 2007; Mulholland & Ginns, 2008; Ogan-Bekiroğlu, 2007; Sadler, 1992; Stover & Saunders, 2000; Trumper, 2000; Trumper, 2001a; Trumper, 2001b; Trumper, 2003; Vosniadou *et al*., 2004; Young & Shavl, 2013; Zeilik *et al*., 1998). These challenges in understanding astronomy arise from the difficulty of visualizing vast structures and immense distances in three dimensions. (Bailey & Slater, 2003; Bretones & Neto, 2011; Coble *et al*., 2013; Eriksson *et al*., 2014; Padalkar & Ramadas, 2011). This challenge is recognized as a spatial reasoning problem. The mental processes involved in

solving problems related to three-dimensional objects are referred to as spatial reasoning (NRC, 2006). Although spatial reasoning is a cognitive skill that can be learned at any age level, this situation may vary significantly among individuals and groups (NRC, 2006). Nowadays, it is possible that these differences can be identified with the help of frequently utilized psychometric tests such as "The Stanford-Binet Intelligence Scale-V (SB-5)" and "Wechsler Intelligence Scale for Children (WISC-V)". Spatial reasoning is recognized as a measure of giftedness in these tests. (Andersen, 2014). Accordingly, gifted students are expected to demonstrate high performance in spatial thinking processes such as visualizing, manipulating, and rotating objects presented in test content. However, the measurement of spatial reasoning skills remains problematic, as they are often assessed in isolation from context and are difficult to distinguish from general cognitive ability in both these tests and intelligence assessments. This highlights the need to revise the evaluation of spatial reasoning, a skill frequently utilized in daily life. Moreover, Maker (2005) and her colleagues' research studies on "Discovering Intellectual Strengths and Capabilities while Observing Varied Ethnic Responses" (DISCOVER) indicated that the current assessments of spatial ability have some problems in terms of figures' cultural bias. Therefore, there is a need to develop measurement tools that focus on measuring spatial performance in a specific context, both by focusing on mental images and using objects that are determined on the basis of proximodistal principle. According to Lee and Berdnarz (2012), this context can be provided with the help of scientific disciplines that include understanding of the nature, structure and functions of scientific phenomena on a microscopic and astronomical scale, apart from the objects interacted with in daily lives and living spaces. In this regard, spatial reasoning needs to be addressed by everyday objects and then by focusing on more distant objects and ultimately astronomical objects. The concept involving this situation is spatial reasoning based on mental imagination. Spatial reasoning based on mental imagination is the ability to decide on the static and dynamic states of three-dimensional objects in the minds of the individuals in a context, which is closely linked to the proximodistal principle (Al-Balushi & Coll 2013). So, in this study, spatial reasoning skill was handled as "spatial reasoning skill based on mental imagination" due to its more inclusive nature.

The spatial reasoning skill plays a crucial role in explaining the three-dimensional positions and movements of celestial bodies within the learning process of gifted students in basic astronomy subjects, such as the phases of the Moon, seasons, the formation of day and night, and solar and lunar eclipses (Bretones & Neto, 2011; Hegarty, 2010; Heyer, 2012; Plummer *et al.*, 2014). This is because gifted individuals' conceptual understanding and academic achievement in astronomy are closely linked to their spatial reasoning skills (DeRoche, 1967; Türk, 2016; Wellner, 1995). A study conducted by Türk (2016) supports the finding that there is a strong correlation between achievement in astronomy and spatial thinking skills. Similarly, Rudmann (2002) stated that students' ability to explain astronomical phenomena, such as the cause of the seasons, is constrained by their spatial abilities. Wellner (1995) also emphasized that students with strong spatial perception were more likely to accurately explain the cause of the Moon's phases. According to Kozhevnikov *et al.* (2007), individuals with high spatial reasoning skills can more easily develop their conceptual understanding of scientific principles. Myron Atkin (1961) stated that gifted students in the fourth, fifth, and sixth grades possess the ability to comprehend many important astronomy concepts. In a study conducted by DeRoche (1967) on sixth-grade gifted students, a significant relationship was found between individuals' intelligence levels and their achievement in astronomy subjects. In this context, cognitive skills such as spatial reasoning are thought to have substantial potential in enhancing gifted students' conceptual understanding of basic astronomy topics and their overall academic achievement in science. Additionally, the development of spatial reasoning skills, which are central to the formal operational stage, can be more precisely observed among middle school students at various grade levels.

Consequently, the relationship between gifted students' spatial reasoning skills, their conceptual understanding of basic astronomy topics, and their academic achievement remains insufficiently explored in previous studies. In particular, the investigation of this relationship across different grade levels has been largely overlooked. Considering these factors, understanding the explanatory links between spatial reasoning, conceptual understanding of basic astronomy topics, and academic achievement in science is a crucial step toward addressing gifted students' misconceptions about astronomy and their challenges in science learning.

### 1.1 Spatial Reasoning and Astronomy for Gifted Learners

Gifted students generally demonstrate higher academic achievement and spatial thinking skills than their peers (DeRoche, 1967; Myron Atkin, 1961; Shea *et al*., 2001; Taber, 2010). However, in science education, they often learn concepts in isolation rather than making interdisciplinary connections. Interdisciplinary knowledge transfer—the ability to apply skills and knowledge across domains—can significantly enhance learning outcomes (Bailey *et al*., 2016; Sasson & Dori, 2012; Schönborn & Bögeholz, 2009). Astronomy serves as an ideal interdisciplinary science, integrating concepts from physics, chemistry, biology, and more, such as heat, temperature, light, DNA, gravity, and atmosphere. Despite this, gifted students often struggle with spatially complex astronomical topics, such as the phases of the Moon and the changing seasons (Hollow, 2005). Without proper attention, they may develop misconceptions similar to those of their peers (Kim *et al*., 2011; Kolar & Ho-Wisniewski, 2009)

Recent research on astronomy education has focused on conceptual understanding across different age groups (Eriksson *et al*., 2014; Kalkan *et al*., 2014). Common misconceptions about basic astronomy topics include misunderstandings of the Earth-Sun-Moon system, the day-night cycle, the seasons, and the solar system. For example, many individuals mistakenly believe that the Sun revolves around the Earth or that the phases of the Moon result from Earth's shadow (Lelliott & Rolnick, 2010). Table 1 presents examples of notable topics and misconceptions identified in the literature.

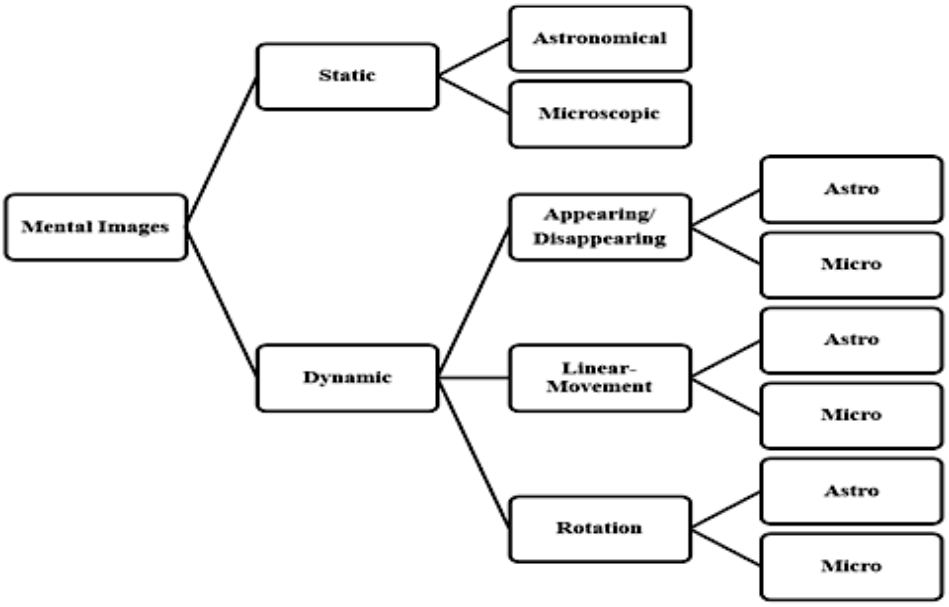
**Table 1.** *Students' common misconceptions of astronomy.*

| Subject | Misconceptions regarding Conceptual Understanding |
|---|---|
| The Earth-Sun-Moon System | - The Sun orbits the Earth.<br>- The phases of the Moon result from the Moon's revolution around the Sun.<br>- The phases of the Moon occur because the Moon enters Earth's shadow. |
| Day and Night Cycle | - Day and night occur due to Earth's revolution around the Sun.<br>- The day-night cycle is caused by Earth's movement around the Sun.<br>- The Moon influences the formation of the day-night cycle. |
| Seasons | - Earth's axial tilt varies from season to season.<br>- Summer is warmer than winter because Earth is closer to the Sun during summer months.<br>- The change of seasons is caused by variations in the distance between the Sun and Earth or by the relative positions of the Sun, Earth, and Moon. |
| The Solar System and Stars | - The sun is the center of the universe.<br>- The Milky Way Galaxy is the center of the universe.<br>- Stars reflect sunlight like planets. |

Table 1 outlines common misconceptions about key astronomy topics among students. These misunderstandings reflect difficulties in comprehending the spatial and conceptual relationships

of celestial phenomena, emphasizing the need for targeted educational interventions. Research indicates that many misconceptions arise from challenges in visualizing celestial phenomena in three dimensions (Al-Balushi & Coll, 2013; Plummer & Maynard, 2014). Figure 1 illustrates the concept of mental images as described by Al-Balushi and Coll (2013). Spatial reasoning plays a crucial role in overcoming these challenges, as it involves the mental manipulation of both static and dynamic states of objects—an essential skill for understanding astronomy (Barnett & Morran, 2002; Gazit *et al*., 2005)

**Figure 1.** *Mental images (Al-Balushi & Coll, 2013).*



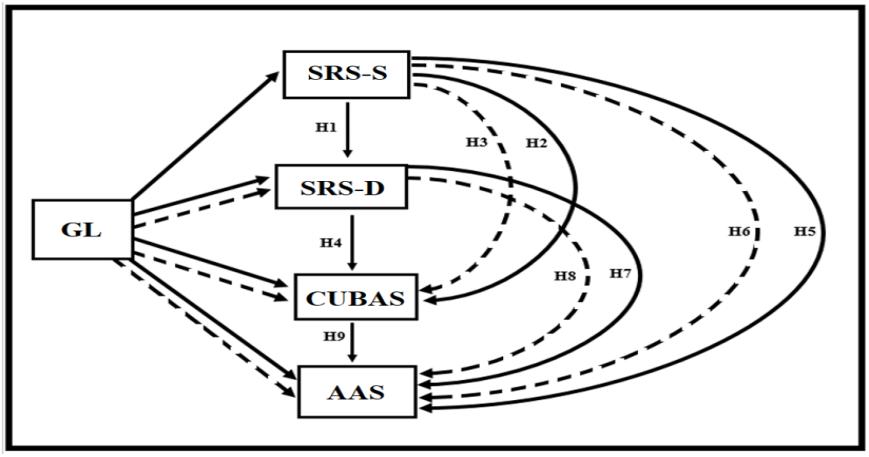

The model emphasizes the importance of assessing spatial reasoning within a specific context, progressing from familiar, everyday objects to distant, abstract astronomical entities. This approach aligns with the proximodistal principle, which supports students' ability to apply spatial reasoning progressively in increasingly complex scenarios. This model was chosen for the study due to its relevance in explaining spatial reasoning within scientific contexts, particularly its focus on the mental manipulation of objects that cannot be directly observed, such as celestial bodies. Gifted students exhibit varying levels of conceptual understanding. For instance, while some accurately associate Earth's rotation with the day-night cycle, others hold misconceptions about celestial movements (Plummer *et al*., 2011). Effective teaching methods incorporating kinesthetic and three-dimensional activities can enhance comprehension (Plummer *et al*., 2014). Furthermore, strengthening spatial reasoning skills at an early age has been linked to improved academic achievement in science (Shea *et al*., 2001).

Despite the importance of spatial reasoning, research on gifted students in this area remains limited. Existing literature indicates that spatial reasoning skills are often assessed in a decontextualized manner, detached from real-world experiences. Moreover, misconceptions about basic astronomy topics are not only prevalent among general students but also occur among the gifted. This study aims to examine the extent to which spatial reasoning contributes to gifted students' conceptual understanding and academic achievement in science, addressing a significant gap in the literature. Specifically, it explores the relationship between spatial reasoning, conceptual understanding, and academic achievement in science among gifted students. In conclusion, this research presents a unique model that investigates both direct and indirect relationships among gifted students' spatial reasoning skills, their conceptual understanding of basic astronomy topics, and their science achievement. The model developed in this study not only advances the literature but also fills a critical gap in understanding these interconnections. By doing so, it provides a framework for context-based assessments and instructional practices that foster deeper learning in astronomy and science education.

## 2. METHOD

This study employed a multifactorial predictive correlational design to test direct and indirect relationships between two or more predictive variables. Path analysis, a structural equation modelling technique, was used to examine explanatory relationships between variables within a confirmatory framework (Bayram, 2013; Büyüköztürk *et al*., 2016). The model, in which each hypothesis was tested according to the study's objectives, is presented in Figure 2.

**Figure 2.** *Hypothesis model.*




GL= Grade level; SRS-S= Mental imagery-focused spatial reasoning skill-static; SRS-D= Mental imagery-focused spatial reasoning skill-dynamic; CUBAS= Conceptual understanding related to basic astronomy subjects; AAS= Academic achievement in science.

In the model presented in Figure 2, hypotheses regarding indirect relationships between variables are represented by dashed arrows, while direct relationships are indicated by solid arrows. The hypotheses tested in the model are as follows:

H1: The static dimension of spatial reasoning skills in gifted sixth-, seventh-, and eighth-grade students directly and significantly predicts the dynamic dimension.

H2: The static spatial reasoning skills of gifted sixth-, seventh-, and eighth-grade students directly and significantly predict their conceptual understanding of basic astronomy topics.

H3: The static spatial reasoning skills of gifted sixth-, seventh-, and eighth-grade students indirectly and significantly predict their conceptual understanding of basic astronomy topics.

H4: The dynamic spatial reasoning skills of gifted sixth-, seventh-, and eighth-grade students directly and significantly predict their conceptual understanding of basic astronomy topics.

H5: The static spatial reasoning skills of gifted sixth-, seventh-, and eighth-grade students directly and significantly predict their academic achievement in science.

H6: The static spatial reasoning skills of gifted sixth-, seventh-, and eighth-grade students indirectly and significantly predict their academic achievement in science.

H7: The dynamic spatial reasoning skills of gifted sixth-, seventh-, and eighth-grade students directly and significantly predict their academic achievement in science.

H8: The dynamic spatial reasoning skills of gifted sixth-, seventh-, and eighth-grade students indirectly and significantly predict their academic achievement in science.

H9: The conceptual understanding of basic astronomy topics among gifted sixth-, seventh-, and eighth-grade students directly and significantly predicts their academic achievement in science.

According to the hypothesis model, grade level was identified as a predictor of all variables. Additionally, the SRS-S variable predicts the SRS-D variable, while both SRS-S and SRS-D

predict the CUBAS variable. Similarly, SRS-S, SRS-D, and CUBAS predict the AAS variable, both directly and indirectly.

Due to the complex interplay of direct and indirect explanatory relationships among the examined variables, classifying them strictly as dependent or independent variables is not feasible. Therefore, within the framework of path analysis modelling, the variables were categorized as internal and external variables, as presented in Table 2.

**Table 2.** *Internal and external variables related to the model.*

| Group | External Variables | Internal Variables |
|---|---|---|
| Gifted Students | Grade Level | SRS-S<br>SRS-D<br>CUBAS<br>AAS |

In Table 2, the external variable is modelled as dependent on the internal variables (SRS-S, SRS-D, CUBAS, and AAS). Internal variables are those that are explained directly or indirectly by other internal and external variables (Bayram, 2013).

## 2.1. Study Group

A total of 642 gifted students were selected for this study using the convenience sampling method. This method was chosen as it provides a practical and efficient way to access a specialized population, particularly when the target group is geographically dispersed, as recommended in educational and psychological research (Cohen *et al.*, 2018). Accordingly, students enrolled in Science and Art Centers (SACs) across various geographical regions of Turkey, including both large and small cities, were included in the study. These students were in the sixth, seventh, and eighth grades, attending SACs in 12 different cities (Elazığ, Erzincan, Malatya, Gaziantep, Şanlıurfa, Ankara, Adana, Antalya, Denizli, İzmir, Ordu, and Rize). They were officially identified as gifted through formal procedures and were eligible to enrol in these centers. Science and Art Centers are defined as "special education institutions that provide support education during extracurricular times to help students, who continue their formal education and are identified as gifted in the fields of general mental talent, visual arts, or music talent, to develop their capacities to the highest level" (General Directorate of Special Education and Guidance Services, 2019). Descriptive statistics on the grade level and gender distribution of the participants are presented in Table 3. According to Table 3, the majority of gifted students in the study are in the sixth and seventh grades, and the number of boys exceeds that of girls.

**Table 3.** *Descriptive statistics results related to the sample.*

| **Grade Level** | ***N*** | **%** |
|---|---|---|
| 6th | 287 | 44.7 |
| 7th | 264 | 41.1 |
| 8th | 91 | 14.2 |
| **Gender** | | |
| Girl | 249 | 38.8 |
| Boy | 393 | 61.2 |
| **Total** | 642 | 100 |

## 2.2. Data Collection Tools

The data collection tools in this study included the Personal Information Form, the Mental Image-Focused Spatial Reasoning Skill Test (SRST), the Conceptual Understanding of Basic Astronomy Subjects Test (CUBAST), and the Science Achievement Test (SAT). The Personal Information Form was completed by students before the administration of the three tests to

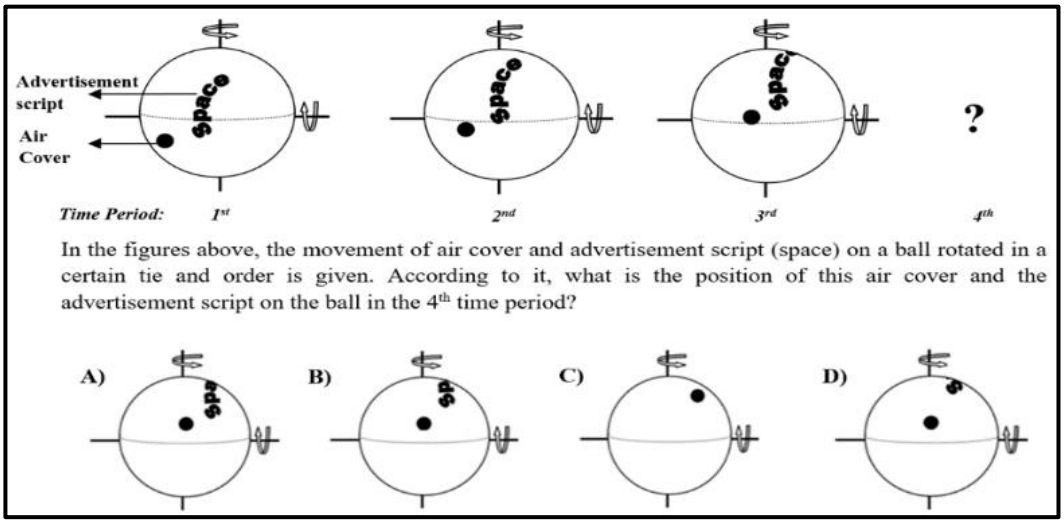
gather demographic information, such as gender, age, grade level, and the SAC they attended. To ensure the validity and reliability of the tests, a pilot study was conducted prior to the main implementation. A pilot study serves as a preliminary investigation to identify potential issues in the testing process and assess the suitability of the measurement tools for the research objectives (van Teijlingen & Hundley, 2002). This pilot study was carried out with 143 gifted students in the sixth, seventh, and eighth grades ($N$ = 143) enrolled at Malatya SAC.

In order to test the validity and reliability of the tests to be used in the research, a pilot study was performed before the actual implementation. Pilot study is the preliminary study that is significant in terms of identifying the problems that may arise in the implementation process before the main study and providing the opportunity to test the appropriateness of the measurement tools for the purpose of the research in a practical way (van Teijlingen & Hundley, 2002). The pilot study was conducted with gifted students in 6th, 7th, and 8th grade ($N$ = 143) who were studying in Malatya SAC.

### 2.2.1. *Mental image-focused spatial reasoning skills test*

To determine the subject scope of the test, studies in the literature on spatial thinking, mental imagery, spatial reasoning, and mental rotation were first examined (Al-Balushi & Coll, 2013; Bektaşlı, 2006; Heyer, 2012; Shipley & Gentner, 2013; Weakly, 2010). Based on this review, the theoretical framework of the measurement tool was developed according to the model proposed by Al-Balushi and Coll (2013), which focuses on the mental visualization of events that cannot be perceived by the naked eye at microscopic and astronomical scales. Unlike the original model, this test incorporates the synthesis of static and dynamic mental imagery in both the macro-world and astronomy, with a particular emphasis on high-level spatial reasoning skills, aligning with the study's objectives. The test items were designed using a three-stage, context-based approach grounded in the proximodistal principle. In the first step of these stages, the objects (such as apples, insects, pieces of faience, etc.) commonly encountered in daily life were used. While in the second step, larger-scale objects and environmental structures in the individual's surroundings (e.g., lakes, islands, seashores, fields) were used, in the third step, celestial and space objects located at the greatest distance from the individual (e.g., planets, constellations, satellites, meteorites). For each acquisition type, three question formats were developed, resulting in a total item pool of 36 questions. A sample question is presented in Figure 3.

**Figure 3.** *Sample question item (Dynamic-Rotation-Prediction).*



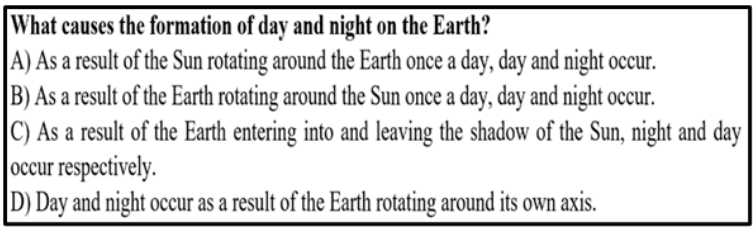

As a result of the item analysis, six questions (items 14, 31, 12, 35, 2, and 14) with discrimination index values below .20 were excluded from the test. In the final version of the test, item difficulty levels ranged from .25 to .97, while discrimination indices were between .20 and .49. Item analysis was re-conducted for the static (14 questions) and dynamic (16

questions) sub-dimensions separately. The results indicated that the discrimination index for the static sub-dimension ranged from .31 to .55, while for the dynamic sub-dimension, it varied between .23 and .53. The difficulty levels of the items remained consistent with the overall analysis, regardless of the number of items. Additionally, the sub-dimensions of the test demonstrated a significant, strong, and positive correlation with both the total test score and each other ($r_{static\text{-}dynamic}$ = .42, $p$ < .00; $r_{static\text{-}total}$ = .80, $p$ < .00; $r_{dynamic\text{-}total}$ = .88, $p$ < .00). Based on this information, the developed test demonstrates construct validity and effectively differentiates gifted students with varying levels of spatial reasoning ability in both static and dynamic dimensions. To assess the reliability of the total test score and sub-dimension scores, the KR-20 internal consistency coefficient, formulated by Kuder and Richardson (1937), was calculated. The KR-20 reliability coefficient for the overall test (SRST) was .72. Additionally, the reliability coefficients for the static (SRS-S) and dynamic (SRS-D) sub-dimensions were .55 and .60, respectively. Following the pilot study, the final version of the test includes 30 items, has a maximum completion time of 40 minutes, and demonstrates sufficient reliability (Wells & Wollach, 2003).

### 2.2.2. *Conceptual understanding of basic astronomy subjects test*

To determine the scope of the measurement tool, fundamental astronomy textbooks (Arny, 1994; Aslan *et al*., 1996; Fucili, 2009; Moche, 2009) and review studies on astronomy education across different time periods were examined (Bailey & Slater, 2003; Bretones & Neto, 2011; Lelliott & Rollnick, 2010). Based on these sources, the test content was structured around four core astronomy topics: The Earth-Sun-Moon System, Day and Night Cycle, Seasons, and The Solar System and Stars. Furthermore, studies on diagnostic tests designed to identify misconceptions and individuals' existing knowledge on these topics were reviewed, leading to the formation of a 20-item question pool (Agan, 2004; Atwood & Atwood, 1997; Baxter, 1989; Dove, 2002; Gündoğdu, 2014; Jeffrey, 2001; Küçüközer, 2007; Ogan-Bekiroğlu, 2007; Sharp, 1996; Trumper, 2000; Trumper, 2001a; Trumper, 2006; Trundle *et al*., 2002; Trundle *et al*., 2007; Young & Shawl, 2013; Zeilik *et al*., 1998). Content validity was ensured through expert opinions, a table of specifications, and a comprehensive literature review. The expert panel included a science education teacher and two doctoral researchers specializing in science education with a focus on astronomy. The table of specifications was prepared as an additional method to strengthen content validity. The distractor options in the test items were derived from common misconceptions about fundamental astronomical concepts, phenomena, and principles that individuals acquire from various sources (Agan, 2004; Baxter, 1989; Dove, 2002; Dunlop, 2000; Jeffery, 2001; Küçüközer, 2007; Sadler, 1992; Ogan-Bekiroğlu, 2007; Trundle *et al*., 2007; Trumper, 2001a; Trumper, 2006; Vosniadou & Brewer, 1990; Zeilik *et al*., 1998). Figure 4 presents a sample question from the CUBAST.

**Figure 4.** *Sample question item.*



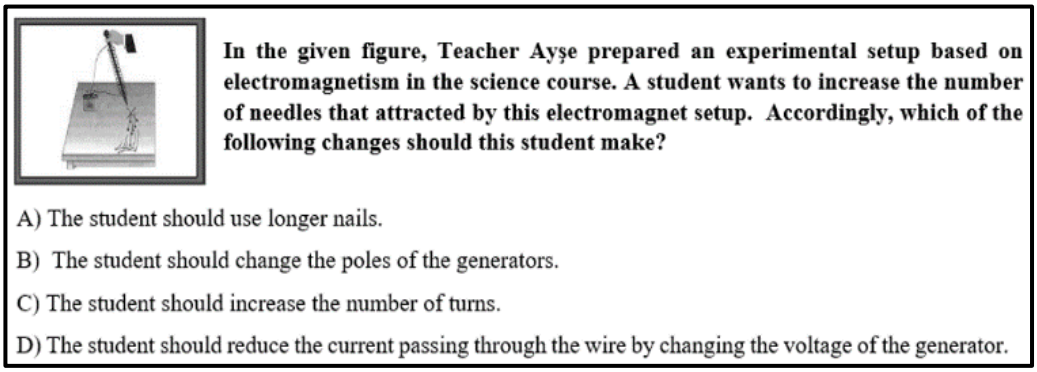

As a result of the item analysis, three questions (items 8, 14, and 20) with discrimination index values below .20 were excluded from the test. In the final version, discrimination indices ranged from .20 to .58, while item difficulty levels varied between .12 and .92. Considering these values collectively, the test demonstrates sufficient construct validity (Downing & Haladyna, 2006; Gronlund & Linn, 1990; Osadebe, 2015). The KR-20 reliability coefficient for the test was calculated as .61. Reliability and validity values for multiple-choice tests designed to assess mental processes should be interpreted together (Downing & Haladyna, 2006; Kubiszyn &

Borich, 2013). The Conceptual Understanding of Basic Astronomy Subjects Test (CUBAST) consists of 17 questions and has a maximum administration time of 20 minutes. Overall, these findings indicate that the test has a moderate difficulty level, exhibits internal consistency, and effectively differentiates gifted sixth-, seventh-, and eighth-grade students based on their conceptual understanding of basic astronomy topics.

### 2.2.3. *Science achievement test*

To assess the science achievement levels of gifted sixth-, seventh-, and eighth-grade students, the Science Achievement Test (SAT) developed by Aşut (2013) was used. The test consists of 45 multiple-choice questions, each with four answer options. Its content includes questions that assess high-level cognitive skills such as applying, analyzing, and evaluating, as well as fundamental skills like understanding. Thus, the test effectively measures the academic achievement of gifted students by requiring high-level cognitive processing. In the pilot study involving 143 participants, item analysis was conducted to determine the difficulty and discrimination indices of the test questions. Items with a discrimination index below .20 were re-evaluated, and five questions (items 1, 9, 11, 13, and 18) were removed from the original test. Figure 5 illustrates a sample question from the SAT.

**Figure 5.** *Sample question item (physical events- electricity in daily life- evaluating).*




The discrimination indices of the test items ranged from .20 to .51, while the item difficulty levels varied between .16 and .95. When these values are analyzed collectively, the test appears to be sufficient in terms of construct validity for assessing achievement (Downing & Haladyna, 2006; Osadebe, 2015). The KR-20 reliability coefficient for the test was calculated as .84, which is close to the .92 value reported by Aşut (2013). Additionally, the mean item difficulty index (.54) and mean discrimination index (.37) suggest that the measurement procedures for assessing academic achievement, a cognitive trait, are valid (Murphy & Davidshofer, 2005; Tuckman & Harper, 2012). As a result, the Science Achievement Test (SAT) consists of 40 items, with a maximum administration time of 40 minutes.

## 2.3. Implementation Process

This study aimed to examine the relationship among Spatial Reasoning Skills, Conceptual Understanding of Basic Astronomy Topics, and Academic Achievement in Science among gifted sixth-, seventh-, and eighth-grade students. To achieve this, three multiple-choice tests (SRST, CUBAST, and SAT), each with four answer options, were administered. Based on data from the pilot study, an item analysis was conducted, and the tests were finalized. Considering the challenges encountered during the pilot phase, the tests were revised and refined for the main implementation. In the main study, while the sequence of test administration remained consistent with the pilot phase, the duration of each test was adjusted based on the number of questions.

The test administration sequence during data collection was structured to align with students' cognitive load. The SRST was administered first, followed by the CUBAST, and finally, the SAT. Due to variations in the number and content of questions in each test, administration

durations were adjusted accordingly. Each student completed the tests independently, with completion times varying based on individual cognitive capacity. To minimize fatigue, a five-minute break was provided between tests.

The implementation was conducted in Science and Art Centers (SACs), specialized education institutions that provide supplementary instruction for gifted students. To ensure consistency and high-quality administration of the measurement instruments across different implementers, the researcher developed an Administration Guide and an Administration Process Checklist. The Administration Guide provided detailed instructions on test procedures and time allocation, ensuring clarity and uniformity in administration. The Administration Process Checklist served as a practical tool for implementers to verify the completion of each procedural step and provided the researcher with systematic feedback on the administration process. By following this structured approach, the study aimed to enhance the reliability and validity of the collected data, minimizing potential biases and procedural inconsistencies.

### 2.4. Data Analysis

Each test in the study was administered by SAC teachers in two sessions. The skewness-kurtosis coefficients and histogram curves of the average test scores indicated a normal distribution. Additionally, trimmed mean values for the lower and upper 5% of participants were compared with actual mean values, showing minimal differences. This confirmed that outliers had no strong effect on the mean, further supporting the normality of the data (Pallant, 2016; Yerdelen-Damar & Aydın, 2015). Consequently, parametric tests were used in inferential statistical analyses. Path analysis was conducted to develop a theoretical model by examining the direct and indirect relationships between internal and external variables. Before performing the path analysis, the Pearson Product-Moment Correlation Coefficient was calculated to determine the direction and strength of the relationships between variables. Additionally, Mardia's (1970) multivariate kurtosis coefficient was computed to test the assumption of multivariate normality. Bentler (2005) suggested that a value below 5 indicates a normal distribution of multivariate data. In this study, the calculated multivariate kurtosis coefficient (1.337) confirmed that the data met this assumption, allowing for path analysis. Further statistical analyses were conducted using AMOS 21 and LISREL 8.7. Path coefficients, representing standardized regression coefficients, were calculated to explain the interdependent variation of variables (Çokluk *et al*., 2012; Loehlin, 2004). Direct, indirect, and total effects between internal and external variables were analyzed using these coefficients. The alpha value for inferential statistical analyses was set at .05. To assess model fit, different indices ($\chi^2/df$, RMSEA, SRMR, GFI, AGFI, CFI, and NFI) were calculated. Model fit was evaluated based on these indices to determine its compatibility with the observed data.

## 3. FINDINGS

In this quantitative study, findings were presented under four main categories: (1) identifying the mental imagery-based spatial reasoning skills of gifted sixth-, seventh-, and eighth-grade students, along with their conceptual understanding of basic astronomy topics and academic achievement in science; (2) descriptive analysis findings examining the direct and indirect exploratory relationships among these variables; (3) path analysis findings; and (4) findings related to predictive relationships and the direct and indirect relationships between variables.

### 3.1. Findings of Descriptive Analysis

Descriptive analysis and reliability coefficient values for the average scores of gifted sixth-, seventh-, and eighth-grade students ($N$ = 642) on SRS, SRS-S, SRS-D, CUBAS, and AAS are presented in Table 4. An examination of Table 4 shows that the mean skewness and kurtosis values of participant scores for each measurement tool range from 0.01 to 0.18 and 0.05 to 0.21, respectively. Based on these values, the mean scores for each test exhibit a normal distribution (Tabachnick & Fidell, 2013). Additionally, the mean scores for the static and dynamic sub-

dimensions of the SRS were 10.42 out of 14 and 9.90 out of 16, respectively. Similarly, the mean CUBAS score was 10.50 out of 17, while the mean AAS score was 23.06 out of 40.

**Table 4.** *Descriptive analysis results.*

| Value | Test | | | | |
|---|---|---|---|---|---|
| | SRS | SRS-S | SRS-D | CUBAS | AAS |
| *N* | 642 | 642 | 642 | 642 | 642 |
| *M* | 20.33 | 10.42 | 9.90 | 9.90 | 23.06 |
| *SD* | 3.96 | 2.17 | 2.43 | 2.76 | 6.95 |
| Min | 5 | 1 | 1 | 1 | 4 |
| Max | 29 | 14 | 16 | 17 | 39 |
| Skewness | 0.17 | 0.18 | 0.13 | 0.14 | 0.01 |
| $SE_{skewness}$ | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 |
| Kurtosis | 0.20 | 0.21 | 0.17 | 0.05 | 0.17 |
| $SE_{kurtosis}$ | 0.19 | 0.19 | 0.19 | 0.19 | 0.19 |
| *KR-21* | .72 | .55 | .60 | .61 | .84 |

The descriptive analysis of the mean scores for SRS, SRS-S, SRS-D, CUBAS, and AAS by grade level is presented in Table 5. It shows that the mean scores for SRS, SRS-S, and SRS-D increase consistently with grade level. Additionally, the maximum scores of gifted students across different grade levels for CUBAS and AAS are similar, while their mean scores also show an upward trend with increasing grade level, similar to other measurement tools.

**Table 5.** *Descriptive analysis results according to grade levels.*

| Grade Level | Value | Test | | | | |
|---|---|---|---|---|---|---|
| | | SRS | SRS-S | SRS-D | CUBAS | AAS |
| 6 | *N* | 287 | 287 | 287 | 287 | 287 |
| | Mean | 19.58 | 10.17 | 9.41 | 9.69 | 18.78 |
| | *SD* | 3.94 | 2.26 | 2.37 | 2.61 | 5.18 |
| | Min | 6 | 1 | 2 | 2 | 4 |
| | Max | 28 | 14 | 16 | 16 | 36 |
| 7 | *N* | 264 | 264 | 264 | 264 | 264 |
| | Mean | 20.59 | 10.51 | 10.09 | 10.67 | 24.87 |
| | *SD* | 3.78 | 2.07 | 2.32 | 2.70 | 5.81 |
| | Min | 6 | 4 | 2 | 1 | 9 |
| | Max | 29 | 14 | 15 | 16 | 37 |
| 8 | *N* | 91 | 91 | 91 | 91 | 91 |
| | Mean | 21.90 | 10.98 | 10.92 | 12.56 | 31.29 |
| | *SD* | 4.00 | 2.06 | 2.57 | 2.20 | 4.88 |
| | Min | 5 | 3 | 1 | 5 | 12 |
| | Max | 28 | 14 | 15 | 17 | 39 |

### 3.2. Findings of Path Analysis

Before testing the hypothesis model, the relationships between variables were examined for multicollinearity, singularity, and linearity. Accordingly, the correlation matrix illustrating these relationships is presented in Table 6. An examination of Table 6 reveals that correlation values range from .22 to .88. Accordingly, a significant and positive relationship exists among SRS-S, SRS-D, CUBAS, and AAS. However, while SRS is significantly and positively

correlated with its sub-dimensions, SRS-S and SRS-D, the high correlation coefficients indicate a singularity issue (Pallant, 2016; Tabachnick & Fidell, 2013). To address this, SRS-S and SRS-D were analyzed separately instead of the overall SRS variable, as they demonstrated a strong relationship while ensuring the path analysis assumptions in the hypothesized model.

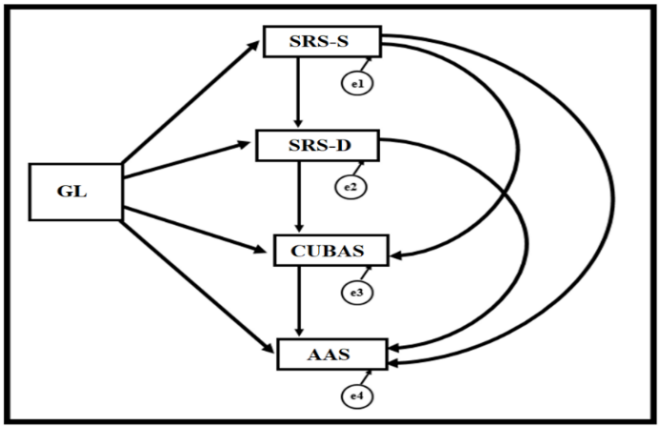
**Table 6.** *Correlation matrix.*

| Variable | SRS | SRS -S | SRS -D | CUBAS |
|---|---|---|---|---|
| SRS | - | | | |
| SRS -S | .85* | - | | |
| SRS -D | .88* | .48* | - | |
| CUBAS | .34* | .26* | .32* | - |
| AAS | .28* | .22* | .27* | .47* |

*$p<.05$

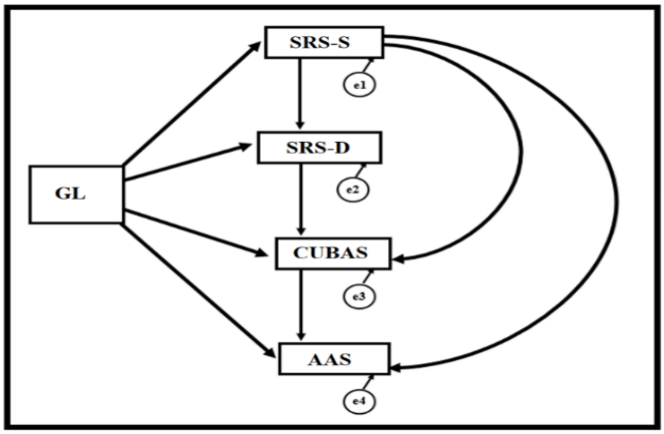
The path diagram illustrating the direct and indirect relationships between the examined variables is presented in Figure 6. An analysis of the fit values for the hypothesis model indicates that the data did not support the model ($\chi^2/df$ = 74.95; $p$ < .05; RMSEA = .34). Furthermore, an examination of the path coefficients reveals that SRS-S and SRS-D do not significantly predict AAS (βSRS-S =.06, $t$ = 1.81, $p$ > .05; βSRS-D = .04, $t$ = 1.24, $p$ > .05). Based on this finding, the model was revised by removing the path representing the direct relationship between SRS-D and AAS.

**Figure 6.** *Path diagram related to the hypothesis model.*




The modified model is presented in Figure 7. An examination of Figure 7 shows that, in the modified model, the explanatory relationship between SRS-D and AAS is established only indirectly, unlike in the hypothesis model.

**Figure 7.** *The modified model.*



To determine whether to accept or reject the modified model and to evaluate the fit between the covariance matrices of the model and the observed data, model fit indices were analyzed. Since fit indices are sensitive to sample size and each represents different criteria, they should be evaluated collectively (Byrne, 2010; Kline, 2011). Accordingly, the fit index values for the modified model are presented in Table 7.

**Table 7.** *Fit index values regarding the modified model.*

| Fit Indexes | *p* | $\chi^2$ /*df* | RMSEA | SRMR | GFI | AGFI | CFI | NFI |
|---|---|---|---|---|---|---|---|---|
| Modified Model | .22* | 1.53 | .03 (.000-.114)** | .01 | .99 | .98 | .99 | .99 |

*$p$ >.05, **90% probability confidence interval values

An examination of Table 7 indicates that, based on the chi-square goodness-of-fit index and significance level values, the modified model exhibits a perfect fit with the observed data ($\chi^2/df$ = 1.53; $p$ >.05) (Byrne, 2010).The Goodness-of-Fit Index (GFI) represents the ratio of the sum of squared differences between the theoretical model and the covariance matrices of the data, similar to the $R^2$ value in multiple regression (Schumaker & Lomax, 2010; Tabachnick & Fidell, 2013). The Adjusted Goodness-of-Fit Index (AGFI) adjusts the GFI for degrees of freedom and reduces the impact of sample size. GFI and AGFI values range from 0 to 1, with values of .95 and above indicating a perfect fit (Çokluk *et al.*, 2012). In this study, the GFI (.99) and AGFI (.98) confirm a strong model-data fit. The Root Mean Square Error of Approximation (RMSEA) index estimates the confidence interval for the mean square error of approximation, ranging between 0 and 1. Unlike GFI and AGFI, lower RMSEA values are preferred, with values close to 0 indicating better fit. RMSEA was calculated as .03, which, being below .05, signifies a perfect model-data fit (Bayram, 2013; Schumaker & Lomax, 2010). The Standardized Root Mean Square Residual (SRMR) measures the standardized difference between the observed data and the covariance matrices of the theoretical model. An SRMR value below .08 is acceptable, while values below .05 indicate a perfect fit (Byrne, 2010). The SRMR value in this study (.007) confirms that the model exhibits a perfect fit. Comparative Fit Index (CFI) and Normed Fit Index (NFI) evaluate model fit by comparing the theoretical model with an independent model assuming no relationships between variables (Çokluk *et al.*, 2012). A CFI value above .97 and an NFI value above .95 indicate a perfect fit (Byrne, 2010). The CFI (.99) and NFI (.99) values in Table 7 further confirm the strong fit between the modified model and the observed data.

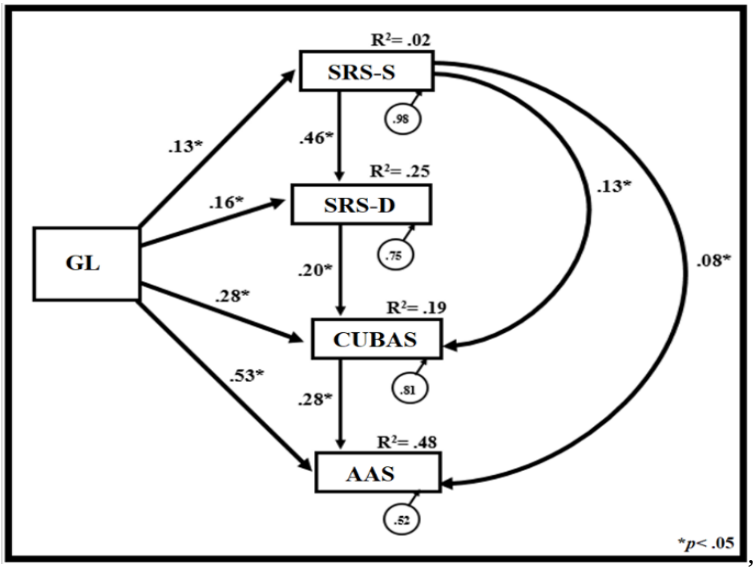
In conclusion, when evaluated collectively, the fit indices confirm that the modified model exhibits a perfect fit with the observed data. However, model fit alone is not sufficient to fully establish the validity of the modified model. Therefore, it is essential to further examine the degree, direction, and direct and indirect effects of the predictive relationships between the variables.

### 3.3. Findings for Predictive Relationships among Variables

To further examine the modified model, path coefficients among the re-analyzed variables and the explained variance ratios ($R^2$) for the predicted variables were calculated. The results are presented in Figure 8. An examination of Figure 8 indicates that the explanatory relationships among internal and external variables in the modified model are statistically significant, with path coefficients (β) ranging from .08 to .53 ($p$ < .05). In the model, the path coefficients among the external variable, grade level, and the internal variables (SRS-S, SRS-D, CUBAS, and AAS) are βGL→SRS-S = .13 ($t$ = 3.22, $p$ < .05), βGL→SRS-D = .16 ($t$ = 4.49, $p$ <.05), βGL→CUBAS = .28 ($t$ = 7.62, $p$ < .05), and βGL→AAS = .53 ($t$ = 17.37, $p$ <.05). Additionally, 2% of the variance in SRS-S is explained solely by grade level ($R^2$ = .02), while 48% of the variance in AAS is accounted for by the remaining variables ($R^2$ = .48). The path coefficient between SRS-S and SRS-D was found to be β = .46 ($t$ = 13.29, $p$ < .05), indicating a significant and

explanatory relationship between these two variables. Based on this finding, Hypothesis 1, formulated by the researcher, was confirmed. Furthermore, 25% of the variance in SRS-D is explained by both grade level and SRS-S ($R^2$ = .25)

**Figure 8.** *The path coefficients among the variables in the modified model and $R^2$ values (* p <.05).*




According to the path coefficients calculated in the modified model, independent changes in SRS-S and SRS-D were associated with significant changes in CUBAS, with βSRS-S→CUBAS = .13 ($t$ = 3.11, $p$ < .05) and βSRS-D→CUBAS =.20 ($t$ = 4.87, $p$ < .05). Based on these findings, Hypothesis 2 and Hypothesis 4, formulated by the researcher, were confirmed. Additionally, 19% of the variance in CUBAS is explained by both external (Grade Level [GL]) and internal (SRS-S and SRS-D) variables ($R^2$ = .19).

The path coefficients between AAS and its explanatory variables, SRS-S and CUBAS, were calculated as βSRS-S→AAS = .08 ($t$ = 2.59, $p$ < .05) and βCUBAS→AAS = .28 ($t$ = 8.83, $p$ < .05), respectively. These results indicate that AAS is directly predicted by both CUBAS and SRS-S, independently and in a statistically significant manner. Accordingly, Hypothesis 5 and Hypothesis 9, formulated by the researcher, were confirmed. However, in the modified model presented in Figure 8, SRS-D did not significantly and directly predict AAS. Therefore, Hypothesis 7 was not supported. Additionally, approximately half of the variance in AAS ($R^2$ = .48) was explained by external and internal variables in the model.

### 3.4. Findings regarding Analysis of Direct and Indirect Relations

In the model, the predictive relationship between two variables can occur directly or indirectly through a third variable. Direct effects are represented by path coefficients in the path diagrams, while indirect effects occur when an internal or external variable influences the explained variable via another variable. Indirect effect values are calculated by multiplying the path coefficients between the explanatory variable, the mediating variable, and the explained variable (Çokluk *et al*., 2012). The total effect of an explanatory variable on the explained variable is the sum of its direct and indirect effects. Accordingly, the direct, indirect, and total effects among the variables in the model are presented in Table 8. An examination of Table 8 reveals that the external variable, Grade Level (GL), has a direct impact on all internal variables and an indirect effect on SRS-D, CUBAS, and AAS. In the indirect effect of GL on SRS-D (βSD→SRS-D = .06), SRS-S serves as a mediating variable. Similarly, in the indirect effect of GL on CUBAS (βSD→CUBAS = .06), SRS-D acts as a mediator. Lastly, in the indirect effect of GL on AAS (βSD→AAS = .10), CUBAS functions as a mediating variable.

**Table 8.** *The direct, indirect and total effect values among variables.*

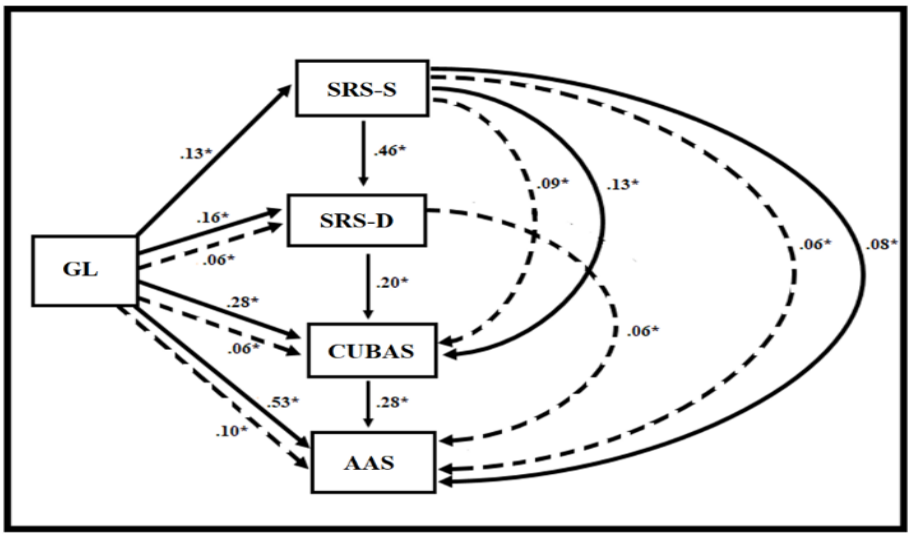
| The Explained Variable | | SRS-S | | | SRS-D | | | CUBAS | | | AAS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Impact Type | | Direct | Indirect | Total | Direct | Indirect | Total | Direct | Indirect | Total | Direct | Indirect | Total |
| Explanatory Variable | GL | .13* | - | .13* | .16* | .06* | .22* | .28* | .06* | .34* | .53* | .10* | .63* |
| | SRS-S | - | - | - | .46* | - | .46* | .13* | .09* | .22* | .08* | .06* | .14* |
| | SRS-D | - | - | - | - | - | - | .20* | - | .20* | - | .06* | .06* |
| | CUBAS | - | - | - | - | - | - | - | - | - | .28* | - | .28* |

*$p$< .05

The variables SRS-S and SRS-D have a total effect of .22 and .20, respectively, on CUBAS. Additionally, .09 of the total effect of SRS-S on CUBAS is indirect, mediated through SRS-D. Similarly, SRS-S indirectly influences AAS (.06) through SRS-D and CUBAS. Based on these findings, Hypotheses 3, 6, and 9, formulated by the researcher, were confirmed.

Finally, it was observed that all external and internal variables in the model had either direct or indirect effects on AAS. The SRS-D variable exerted an indirect effect on AAS (.06) through CUBAS, confirming Hypothesis 8, formulated by the researcher. As a result, CUBAS was identified as a mediating variable in the indirect predictive relationship between AAS and both SRS-S and SRS-D. The final model, developed based on the values in Table 8, was confirmed by its fit with the obtained data and the direct and indirect explanatory relationships between the variables. This model, determined by the researcher, is presented in Figure 9.

**Figure 9.** *The final model (*p <0 .05).*




In the final model presented in Figure 9, indirect relationships between variables are represented by dashed arrows. The model developed by the researcher demonstrates significant direct and indirect explanatory relationships between gifted students' grade levels, their static and dynamic spatial reasoning skills, their conceptual understanding of basic astronomy topics, and their academic achievement in science.

## 4. DISCUSSION and CONCLUSION

This study conducted an exploratory investigation to determine the relationship between the spatial reasoning skills of gifted sixth-, seventh-, and eighth-grade students, their conceptual understanding of basic astronomy topics, and their academic achievement in science. It makes

a significant contribution to the existing literature by examining the interplay between spatial reasoning, conceptual understanding of astronomy, and science achievement in gifted students. Unlike previous research, which often relies on decontextualized assessments, this study employs context-based tools specifically designed for astronomy, offering a more nuanced perspective on how spatial reasoning supports learning (Al-Balushi & Coll, 2013). The path analysis performed for this purpose yielded several findings regarding the explanatory relationships between the examined variables. One key finding revealed that grade level significantly explained changes in both static and dynamic spatial reasoning skills, positively predicting gifted students' spatial abilities as they advanced in grade level. Additionally, 25% of the variance in dynamic spatial reasoning was explained by both grade level and static spatial reasoning. Similarly, Kikas (2006) and Türk (2016), who conducted studies with pre-service science teachers, found that spatial reasoning skills improved as grade levels increased. Moreover, their findings indicated that conceptual understanding of basic astronomy topics also improved with grade progression. This result aligns with the findings of Göncü (2013), who examined fifth- and seventh-grade students, and Padalkar (2010), who studied fourth- and seventh-grade students. Both studies concluded that students in higher grades demonstrated a more sophisticated conceptual understanding of astronomy.

In this study, static spatial reasoning skills were identified as a predictor of dynamic spatial reasoning, both of which are defined as sub-dimensions of spatial reasoning. Accordingly, changes in static spatial reasoning- which involves thinking processes related to features such as the shape and size of three-dimensional objects-explain variations in dynamic spatial reasoning, which includes thinking processes related to movement, such as distance, appearance/disappearance, linear motion, and rotation. D'Oliveira (2004), in a study conducted with 104 university students, administered nine different spatial tests and applied explanatory and confirmatory factor analysis techniques to the collected data. Consistent with the findings of the present study, D'Oliveira (2004) proposed that spatial skills consist of two fundamental structures: static and dynamic.

Another key finding of this study was that gifted students' static and dynamic spatial reasoning skills positively and significantly predicted their conceptual understanding of basic astronomy topics. Additionally, static spatial reasoning skills indirectly influenced conceptual understanding through dynamic spatial reasoning. Similarly, Heyer (2012) found a strong relationship between spatial reasoning skills and conceptual understanding of astronomy among university students, with spatial skills accounting for approximately 25% of the variance in conceptual understanding. From this perspective, one of the key contributions of this study is its dual-dimensional approach, which examines both static and dynamic spatial reasoning and their distinct impacts on conceptual understanding and academic achievement. This expands upon previous research (Heyer, 2012) by focusing on gifted middle school students, highlighting developmental differences across grade levels. Kikas (2006) suggested that secondary school students' spatial skills play a crucial role in structuring astronomy-related concepts and phenomena scientifically. Similarly, Plummer *et al*. (2016) observed that students aged 7–9 with high levels of spatial reasoning could easily explain complex astronomical topics, such as the visible motion of the Sun and the seasons, by shifting between reference frames. Wilhelm (2009) also found that secondary school students' understanding of the phases of the Moon and the relative positions of the Sun, Earth, and Moon was linked to their spatial skills. Furthermore, Rudmann (2002) demonstrated that students' academic performance in astronomy was positively and significantly related to their spatial abilities. This finding aligns with Türk (2016), who reported a strong correlation between science teacher candidates' spatial skills and their academic achievement in astronomy. It also supports Wilhelm *et al*. (2013), who found that sixth-grade students needed advanced spatial skills to develop a scientific understanding of astronomical phenomena. Additionally, Yen *et al*. (2013) highlighted that university-level spatially focused applications effectively facilitated learning about the phases of the Moon.

Similarly, Wilhelm *et al*. (2017) found that instructional approaches emphasizing spatial skills in sixth-grade students significantly enhanced their scientific understanding of basic astronomy topics. Overall, these findings reinforce the importance of spatial reasoning as a crucial variable in developing an understanding of fundamental astronomy concepts. Moreover, this study addresses the limited research on gifted populations in the literature, providing insights into how their advanced cognitive abilities influence learning in complex scientific domains. The results align with and extend the work of Rudmann (2002) and Wilhelm *et al*. (2013), underscoring the essential role of spatial reasoning in science education.

Another variable that explains static and dynamic spatial reasoning skills is the academic achievement of gifted students in science. Static spatial reasoning skills have both a direct effect on academic achievement in science and an indirect effect through conceptual understanding of astronomy. In contrast, dynamic spatial reasoning skills influence academic achievement in science only indirectly through conceptual understanding of astronomy. These findings suggest that effectively structuring astronomy topics involving dynamic processes can contribute to increased academic achievement in science. These results support the conclusion of Black (2005), who found that university students' spatial skills were a crucial factor in developing a scientific understanding of astronomy and improving science achievement. Similarly, Wai *et al*. (2009), in a long-term longitudinal study conducted with students aged 9–12, highlighted that spatial skills play a critical and significant role in science achievement, particularly among gifted students. Hegarty (2010) also emphasized that university students' spatial thinking skills were linked to their academic performance in fundamental science fields such as chemistry, physics, biology, and geology. Additionally, studies by Taylor and Huttn (2013) with fourth-grade students and Miller and Halpern (2013) with university students suggested that spatially focused instructional practices enhance academic achievement in science. Maker (2020) further indicated that spatial ability assessments could be a practical tool for identifying gifted students in science. Moreover, numerous studies in the literature support these findings, demonstrating a positive and significant relationship between spatial skills and academic achievement in science (Hegarty *et al*., 2010; Jones & Burnett, 2008; Kozhevnikov *et al*., 2007; Sorby, 2001; Webb *et al*., 2007; Wu & Shah, 2004).

Therefore, the positive changes in static and dynamic spatial reasoning skills of gifted students—linked to their grade levels—contribute to an enhanced conceptual understanding of basic astronomy topics, ultimately leading to higher academic achievement in science. As a result, the hypothetical model derived from the analysis of this study provides a crucial framework for understanding the nature of these variables and their explanatory relationships. This model is instrumental in addressing challenges gifted students face in spatial reasoning within the fields of astronomy and science. Moreover, the validated path model offers a theoretical foundation for understanding the direct and indirect relationships among these variables, enriching both theoretical discourse and practical applications in cognitive development and gifted education. It also supports the integration of spatially rich, interdisciplinary learning environments to optimize the educational experiences and STEM potential of gifted students (Wai *et al*., 2009).

This study makes a significant contribution to the literature by providing a context-specific exploration of spatial reasoning and its role in enhancing gifted students' understanding of astronomy and their academic achievement in science. Unlike previous studies that focus on general populations or rely on decontextualized assessments, this research highlights the importance of context-based spatial reasoning (Al-Balushi & Coll, 2013). By examining static and dynamic subdimensions within gifted learners, the study uncovers developmental nuances and their implications for tailored instructional design. The validated model further deepens the theoretical understanding of cognitive variable interactions, offering a robust framework for future research and practice in gifted education.

### 4.1. Discussion of Study Limitations

The study's limitations and potential avenues for future research are closely intertwined, highlighting areas for both improvement and further exploration. This section comprehensively addresses these limitations to provide a cohesive understanding of the challenges encountered and the opportunities they present for advancing research in this field.

The hypothetical model developed through the path analysis technique primarily focused on spatial reasoning and conceptual understanding. However, it could be expanded by incorporating additional cognitive factors, such as memory and problem-solving skills, as well as affective factors like motivation, attitudes, and interest in astronomy. Integrating these elements could offer a more comprehensive understanding of the relationships between observed and latent variables, further enriching the model's explanatory power.

While this study primarily focused on gifted sixth-, seventh-, and eighth-grade students, it did not explicitly examine the potential influence of gender differences on spatial reasoning and astronomy understanding. Future research could explore this aspect by developing and testing models that integrate gender as a variable, allowing for comparisons across different grade levels to uncover nuanced differences in spatial reasoning and conceptual understanding.

This study included gifted students attending Science and Art Centers (SACs) in 12 provinces (Elazığ, Erzincan, Malatya, Gaziantep, Şanlıurfa, Ankara, Adana, Antalya, Denizli, İzmir, Ordu, and Rize). To enhance generalizability, future research could expand the sample by including SACs from additional provinces or incorporating students from other educational institutions. Additionally, comparative studies involving non-gifted students could highlight significant contrasts and provide broader insights into the role of spatial reasoning in astronomy education. Furthermore, the Conceptual Understanding of Basic Astronomy Subjects Test (CUBAST) used in this study was limited to specific astronomy topics. Expanding its content to cover additional topics could provide a more comprehensive measure of students' conceptual understanding. Moreover, incorporating qualitative methodologies, such as interviews or observations, could complement the quantitative findings, offering deeper insights into gifted students' thought processes and learning strategies.

This study was conducted within a specific educational and cultural context, focusing on Turkish gifted students. Future research could replicate the study in different cultural or educational settings to examine how contextual factors influence the relationships explored in the model. Such an approach would enhance the cross-cultural applicability and relevance of the findings, providing deeper insights into the role of spatial reasoning in astronomy education across diverse learning environments. By addressing these limitations, future research could refine the proposed model and contribute more comprehensively to the literature on gifted education, spatial reasoning, and astronomy learning. Expanding the scope, increasing sample diversity, and incorporating varied methodological approaches would offer a richer and more holistic understanding of the factors that shape gifted students' learning experiences in science education.

In conclusion, this study makes a significant contribution to the literature by offering a context-specific exploration of spatial reasoning and its essential role in enhancing gifted students' understanding of astronomy and science achievement. Unlike previous research, which often focuses on general populations or employs decontextualized assessments, this study highlights the importance of context-based spatial reasoning (Al-Balushi & Coll, 2013). By examining the static and dynamic subdimensions within gifted learners, it uncovers developmental nuances and their implications for tailored instructional design. The validated model further enriches the theoretical understanding of cognitive variable interactions, providing a robust framework for future research and educational practice in gifted education. Despite its critical role in identifying gifted students, spatial reasoning often receives less emphasis than verbal or quantitative skills in curricula. This study underscores the need for context-based enrichment

learning environments that prioritize spatial reasoning development, leveraging the explanatory relationships identified to better support the cognitive and academic growth of gifted students.

## Acknowledgments

This article is derived from the doctoral dissertation of the first author, who conducted the research under the supervision of Prof. Dr. Mustafa Serdar Köksal at İnönü University. The first author's doctoral studies were supported by TÜBİTAK's 2211 National PhD Scholarship Program.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Inönü University, 27250534-605-E.13323632, 25.12.2015.

## Contribution of Authors

**Pelin Ertekin**: Research Design, Literature Review, Methodology, Data Collection, Data Analysis, and Writing. **Mustafa Serdar Köksal**: Research Design, Methodology, Data Analysis, Supervision, Writing and Critical Review.

## Orcid

Pelin Ertekin https://orcid.org/ 0000-0001-8011-0178
Mustafa Serdar Köksal https://orcid.org/ 0000-0002-2185-5150

## REFERENCES

Agan, L. (2004). Stellar ideas: Exploring students' understanding of stars. *Astronomy Education Review, 3(*1), 77-97. http://dx.doi.org/10.3847/AER2004008

Al-Balushi, S.M., & Coll, R.K. (2013). Exploring verbal, visual and schematic learners' static and dynamic mental images of scientific species and processes in relation to their spatial ability. *International Journal of Science Education, 35*(3), 460-489. https://doi.org/10.1080/09500693.2012.760210

Andersen, L. (2014). Visual–spatial ability: Important in STEM, ignored in gifted education. *Roeper Review, 36*(2), 114-121. https://doi.org/10.1080/02783193.2014.884198

Arny, T.T. (1994). *Explorations an introduction to astronomy*. Mosby-Year Book.

Aslan, Z., Aydın C., Demircan, O., Derman E., & Kırbıyık, H. (1996). *Liseler için astronomi ve uzay bilimleri ders kitabı [Astronomy and space sciences textbook for high schools]*. Tekışık Yayıncılık.

Aşut, N. (2013). *Üstün yetenekli öğrencilerin epistemolojik inançlarının fen öğrenmeye yönelik motivasyon düzeyi ve fen başarısıyla ilişkisi [Relationship of gifted students' epistemological beliefs with science achievement and motivation towards science learning]* (Publication No. 350922) [Master's thesis, İnönü University]. National Thesis Center.

Atwood, R.K., & Atwood, V.A. (1997). Effects of instruction on preservice elementary teachers' conceptions of the causes of night and day and the seasons. *Journal of Science Teacher Education, 8* (1), 1–13. https://doi.org/10.1023/A:1009455201314

Bailey, J.M., & Slater, T.F. (2003). A review of astronomy education research. *Astronomy Education Review (AER), 2*(2), 20 – 45. http://dx.doi.org/10.3847/AER2003015

Bailey, L.M., Morris, L.G., Thompson, W.D., Feldman, S.B., & Demetrikopoulos M.K. (2016). Historical contribution of creativity to development of gifted science education in formal and informal learning environments. In M.K. Demetrikopoulos & J.L. Pecore (Eds.), *Interplay of creativity and giftedness in science* (pp. 3-15). Sense Publishers.

Baker, H.E. (1987). *Astronomy and space science activities for fifth and sixth grade gifted and talented students* [Unpublished master's thesis]. California State University.

Barnett, M., & Morran, J. (2002). Addressing children's alternative frameworks of the moon's phases and eclipses. *International Journal of Science Education, 24*(8), 859–879. https://doi.org/10.1080/09500690110095276

Baxter, J. (1989). Children's understanding of familiar astronomical events. *International Journal of Science Education*, *11*(5), 502-513. https://doi.org/10.1080/0950069890110503

Bayram, N. (2013). *Yapısal eşitlik modellemesine giriş: AMOS uygulamaları [Introduction to structural equation modeling: AMOS applications]* (2nd ed.). Ezgi Kitabevi.

Bektasli, B. (2006). *The relationships between spatial ability, logical thinking, mathematics performance and kinematics graph interpretation skills of 12th grade physics students* (Publication No. 3226336) [Doctoral dissertation, The Ohio State University]. ProQuest Dissertations Publishing.

Bentler, P.M. (2005). *EQS 6 Structural equations program manual*. Multivariate Software.

Black, A. (2005). Spatial ability and earth science conceptual understanding. *Journal of Geoscience Education*, *53*(4), 402-414. https://doi.org/10.5408/1089-9995-53.4.402

Bretones, P.S., & Neto, J.M. (2011). An analysis of papers on astronomy education in proceedings of IAU meetings from 1988 to 2006. *Astronomy Education Review (AER), 10*(1), 10-18. https://doi.org/10.3847/AER2010010

Brunsell, E., & Marcks, J. (2004). Identifying a baseline for teachers' astronomy content knowledge. *Astronomy Education Review, 3*(2), 38-46. https://doi.org/10.3847/AER2004015

Bryne, B.M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). Routledge.

Büyüköztürk, Ş., Çakmak, E, Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2016). *Bilimsel araştırma yöntemleri [Scientific research methods].* Pegem Akademi.

Coble, K., Camarillo, C.T., Trouille, L.E., Bailey, J.M., Cochran, G L., Nickerson, M.D., *et al*. (2013). Investigating student ideas about cosmology I: Distances and structure. *Astronomy Education Review (AER), 12*(1), 1-54. https://doi.org/10.3847/AER2012038

Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge. https://doi.org/10.4324/9781315456539

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları [Multivariate statistics for social sciences: SPSS and LISREL applications]* (2nd ed.). Pegem Akademi.

DeRoche, E.F. (1967). *A study of the effectiveness of selected creative exercises on creative thinking and the mastern of a unit in elementary Science* (Publication No. 6704542) [Doctoral dissertation, University of Connecticut]. ProQuest Dissertations Publishing.

D'Oliveira, T.C. (2004). Dynamic spatial ability: An exploratory analysis and a confirmatory study. *The International Journal of Aviation Psychology*, *14*(1), 19-38. https://doi.org/10.1207/s15327108ijap1401_2

Dove, J. (2002). Does the man in the moon ever sleep? An analysis of student answers about simple astronomical events: A case study. *International Journal of Science Education*, *24*(8), 823-834. https://doi.org/10.1080/09500690110066935

Downing, S.M., & Haladyna, T.M. (2006). *Handbook of test development*. Lawrence Erlbaum Associates Publishers.

Dunlop, J. (2000). How children observe the universe. *Publications of the Astronomical Society of Australia*, *17*(02), 194-206. https://doi.org/10.1071/AS00194

Eriksson, U., Linder, C., Airey, J., & Redfors, A. (2014). Who needs 3D when the universe is flat?. *Science Education. 98*(3), 412-442. https://doi.org/10.1002/sce.21109

Frede, V. (2006). Pre-service elementary teacher's conceptions about astronomy. *Advances in Space Research, 38*, 2237– 2246. https://doi.org/10.1016/j.asr.2006.02.017

Fucili, L. (2009). Implementing astronomy education research. In J. M. Pasachoff, & J. R. Percy (Eds.), *Teaching and learning astronomy* (pp. 66-79). Cambridge University Press.

Gardner, M.K., & Sternberg, R.J. (1994). Novelty and intelligence. In R. Sternberg & R. Wagner (Eds.), *Mind in context: Interactionist perspectives on human* intelligence (pp. 38-74). Cambridge University Press.

Gazit, E., Yair, Y., & Chen, D. (2005). Emerging conceptual understanding of complex astronomical phenomena by using a virtual solar system. *Journal of Science Education and Technology*, *14*(5/6), 459-470. https://doi.org/10.1007/s10956-005-0221-3

General Directorate of Special Education and Guidance Services (September 2019). *Bilim ve sanat merkezleri öğrenci tanılama ve yerleştirme kılavuzu 2019-2020 [Student identification and placement guide for science and art centers 2019-2020]*. https://orgm.meb.gov.tr/meb_iys_dosyalar/2019_11/15173608_TanYlama_KYlavuzu__YeYitek_Ekli_2.pdf

Gilbert, J.K., & Newberry, M. (2007). The characteristics of the gifted and exceptionally able in science. In K. S. Taber (Ed.), *Science education for gifted learners (pp*. 15-31). Routledge.

Goodman, A.A., Udomprasert, P.S., Kent, B., Sathiapal, H., & Smareglia, R. (2011). Astronomy visualization for education and outreach. In I.N. Evans, A. Accomazzi, D.J. Mink & A.H. Rots (Eds.), *Astronomical data analysis software and systems XX* (pp. 659-662). Astronomical Society of the Pacific.

Göncü, Ö. (2013). *İlköğretim beşinci ve yedinci sınıf öğrencilerinin astronomi konularındaki kavram yanılgılarının tespiti [Determining of astronomical misconception in fifth and seventh grade students]* [Master's thesis, Mehmet Akif Ersoy University]. National Thesis Center.

Gronlund, N.E., & Linn, R.L. (1990). *Measurement and evaluation in teaching* (6th ed.). Macmillan Publishing.

Gündoğdu, T. (2014). *8. sınıf öğrencilerinin astronomi konusundaki başarı ve kavramsal anlama düzeyleri ile fen dersine yönelik tutumları arasındaki ilişkinin incelenmesi [Investigation of relationship between 8th grade students' achievement and conceptual understanding levels in astronomy and attitudes towards science course]* (Publication No. 372278) [Master's thesis, Marmara University]. National Thesis Center.

Hannust, T., & Kikas, E. (2007). Children's knowledge of astronomy and its change in the course of learning. *Early Childhood Research Quarterly*, *22*(1), 89-104. https://doi.org/10.1016/j.ecresq.2006.11.001

Hegarty, M. (2010). Components of spatial intelligence. *Psychology of Learning and Motivation, 52*, 265–296. https://doi.org/10.1016/S0079-7421(10)52007-3

Hegarty, M., Crookes, R.D., Dara-Abrams, D., & Shipley, T.F. (2010, August). Do all science disciplines rely on spatial abilities? Preliminary evidence from self-report questionnaires. In C. Hölscher, T.F. Shipley, M.O. Belardinelli & J.A. Bateman (Eds.), *Spatial Cogniton VII: International Conference on Spatial Cognition 2010* (pp. 85-94). Springer.

Heyer, I. (2012). *Establishing the empirical relationship between non-science majoring undergraduate learners' spatial thinking skills and their conceptual astronomy knowledge* (Publication No. 3507367) [Doctoral dissertation, University of Wyoming]. ProQuest Dissertations Publishing.

Hollow, R. (2005). Engaging gifted science students through astronomy. *Highlights of Astronomy, 13*, 1041-1043. https://doi.org/10.1017/S1539299600018062

Jeffery, D.J. (2001). *Introductory astronomy problems*. Portpentagram Publishing.

Johnes, M.G., & Broadwell, B. (2008). Visualization without vision: Students with visual impairment. In J.K. Gilbert, M. Reiner & M. Nakhleh (Eds.), Visualization: Theory and practice in science education (pp. 283-294). Springer Science & Business Media.

Jones, S., & Burnett, G. (2008). *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments, 4*(1), 47-61. https://doi.org/10.17011/ht/urn.200804151352

Kalkan, H., Kiroğlu, K., Türk, C., Bolat, M., Kalkan, S., & Aslantürk, A. (2014). Basic astronomy concepts in the footsteps of Eratosthenes. *Procedia-Social and Behavioral Sciences, 116*, 3731-3739. https://doi.org/10.1016/j.sbspro.2014.01.832

Kanlı, U. (2014). A study on identifying the misconceptions of pre-service and in-service teachers about basic astronomy concepts. *Eurasia Journal of Mathematics, Science & Technology Education, 10*(5), 463-471. https://doi.org/10.12973/eurasia.2014.1120a

Kanlı, U. (2015). Using a two-tier test to analyse students' and teachers' alternative concepts in astronomy. *Science Education International, 26*(2), 148-165. https://files.eric.ed.gov/fulltext/EJ1064041.pdf

Kikas, E. (2006). The effect of verbal and visuo-spatial abilities on the development of knowledge of the Earth. *Research in Science Education, 36*(3), 269-283. https://doi.org/10.1007/s11165-005-9010-5

Kim, E.S., Yu, H.W., & Choe, S.U. (2011). Investigation of the 7th grade science-gifted students' understanding about the lunar phase through their own observation and interpretation. *Journal of The Korean Earth Science Society, 32*(5), 514-520. https://doi.org/10.5467/JKESS.2011.32.5.514

Kline, R.B. (2011). *Principles and practice of structural equation modeling* (3th ed.). The Guilford Press.

Kolar, C.G., & Ho-Wisniewski, E. (2009, 04 January). *Assessing high school gifted student progress in science through misconceptions and MOSART.* Paper presented at the 2009 American Educational Research Association Annual Meeting. https://digitalcommons.imsa.edu/cgi/viewcontent.cgi?article=1003&context=oir_pr

Kozhevnikov, M., Motes, M., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science, 31*, 549–579. https://doi.org/10.1080/15326900701399897

Kubiszyn, T., & Borich, G. (2013). *Educational testing and measurement: Classroom application and practice* (10th ed.). Wiley.

Kurnaz, M.A. (2012). Turkish students' understandings about some basic astronomy concepts: A cross-grade study. *World Applied Sciences Journal, 19*(7), 986-997. https://idosi.org/wasj/wasj19(7)12/11.pdf

Küçüközer, H. (2007). Prospective science teachers' conceptions about astronomical subjects. *Science Education International, 18*(2), 113-130. https://files.eric.ed.gov/fulltext/EJ1065856.pdf

Lee, J., & Bednarz, R. (2012). Components of spatial thinking: Evidence from a spatial thinking ability test. *Journal of Geography, 111*(1), 15-26. https://doi.org/10.1016/j.sbspro.2011.07.048

Lelliott, A., & Rolnick, M. (2010). Big ideas: a review of astronomy education research 1974-2008. *International Journal of Science Education, 32*(13), 1771-1799. https://doi.org/10.1080/09500690903214546

Loehlin, J.C. (2004). *Latent variable models.* Lawrence Erlbaum Associates, Publishers.

Maker, C.J. (2005). *The DISCOVER project: Improving assessment and curriculum for diverse gifted learners*. National Research Center on the Gifted and Talented.

Maker, C.J. (2020). Culturally responsive assessments of spatial analytical skills and abilities: development, field testing, and implementation. *Journal of Advanced Academics, 31*(3), 234-253. https://doi.org/10.1177/1932202X20910697

Mardia, K.V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*(3), 519-530. https://doi.org/10.1093/biomet/57.3.519

Miller, D.I., & Halpern, D.F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates?. *Learning and Individual Differences*, *26*, 141-152. https://doi.org/10.1016/j.lindif.2012.03.012

Moche, D.L. (2009). *Astronomy: A self-teaching guide* (7th ed.). John Wiley & Sons.

Mulholland, J., & Ginns, I. (2008). College MOON Project Australia: Preservice teachers learning about the Moon's phases. *Research in Science Education*, *38*(3), 385–399. https://doi.org/10.1007/s11165-007-9055-8

Murphy, K.R., & Davidshofer, C.O. (2005*). Psychological testing: Principles and applications (6th ed.).* Pearson Education International.

Myron Atkin, J. (1961). Teaching concepts of modern astronomy to elementary-school children. *Science Education*, *45*(1), 54-58. https://doi.org/10.1002/sce.3730450117

National Research Council (NRC). (2006). *Learning to Think Spatially*. National Academy Press.

Ogan-Bekiroglu, F. (2007). Effects of model-based teaching on preservice physics teachers' conceptions of the Moon, Moon phases, and other lunar phenomena. *International Journal of Science Education*, *29*(5), 555-593. https://doi.org/0.1080/09500690600718104

Osadebe, P.U. (2015). Construction of valid and reliable test for assessment of students. *Journal of Education and Practice*, *6*(1), 51-56. https://files.eric.ed.gov/fulltext/EJ10838 61.pdf

Padalkar, S. (2010). *Spatial cognition and visualization in elementary astronomy education* (Unpublished doctoral dissertation). Tata Institute of Fundamental Research, Mumbai.

Pallant, J. (2016). *SPSS survival manual: A step by step guide to data analysis using SPSS for Windows*. Allen-Unwin.

Percy, J.R. (2006). Teaching astronomy: Why and how?. *The Journal of the American Association of Variable Star Observers (JAAVSO)*, *35*, 248-254. https://adsabs.harvard.edu /full/2006JAVSO..35..248P

Percy, J.R. (2009). Why astronomy is useful and should be included in the school curriculum. In J.M. Pasachoff and J.R. Percy (Eds.), *Teaching and learning astronomy: Effective strategies for education worldwide* (pp. 10-13). Cambridge University Press.

Plummer, J., Wasko, K., & Slagle, C. (2011). Children learning to explain daily celestial motion. *International Journal of Science Education*, *33*(14), 1963-1992. https://doi.org/10. 1080/09500693.2010.537707

Plummer, J.D., Kocareli, A., & Slagle, C. (2014). Learning to explain astronomy across moving frames of reference: Exploring the role of classroom and planetarium-based instructional contexts. *International Journal of Science Education*, *36*(7), 1083-1106. https://doi.org/10. 1080/09500693.2013.843211

Plummer, J.D., & Maynard, L. (2014). Building a learning progression for celestial motion: An exploration of students' reasoning about the seasons. *Journal of Research in Science Teaching*, *51*(7), 902-929. https://doi.org/10.1002/tea.21151

Plummer, J.D., Bower, C.A., & Liben, L.S. (2016). The role of perspective taking in how children connect reference frames when explaining astronomical phenomena. *International Journal of Science Education*, *38*(3), 345-365. https://doi.org/10.1080/09500693.2016.114 0 921

Prather, E.E., Rudolph, A.L., & Brissenden, G. (2009). Teaching and learning astronomy in the 21st century. *Physics Today*, *62*, 41-47. https://home.miracosta.edu/rfrench/astro/rte2010/a ssets/PrathRudBriss_PT_Oct09.pdf

Pyryt, M.C. (2000). Talent development in science and technology. In K. A. Heller, F. J. Mönks, R. J. Sternberg, & R. F. Subotnik, *International handbook of giftedness and talent* (2nd ed., pp. 427-438). Elsevier Science.

Rudmann, D. (2002, 02 April). *Solving astronomy problems can be limited by intuited knowledge, spatial ability, or both*. Paper presented at the annual American Educational Research Association. http://files.eric.ed.gov/fulltext/ED468815.pdf

Sadler, P.M. (1992). The initial knowledge state of high school astronomy students (Publication No. 9228416) [Doctoral dissertation, Harvard University]. ProQuest Dissertations Publishing.

Sasson, I., & Dori, Y.J. (2012). Transfer Skills and Their Case-Based Assessment. In B., Fraser, K. Tobin, & C.J. McRobbie (Eds.), *Second international handbook of science education* (pp. 691-710). Springer Science & Business Media.

Schönborn, K.J., & Bögeholz, S. (2009). Knowledge transfer in biology and translation across external representations: experts' views and challenges for learning. *International Journal of Science and Mathematics Education*, *7*(5), 931-955. https://doi.org/10.1007/s10763-009-9153-3

Schumacker, R.E., & Lomax, R.G. (2010). *A beginners guide to structural equation modeling*. Routledge.

Sharp, J.G. (1996). Children's astronomical beliefs: a preliminary study of year 6 children in south-west England. *International Journal of Science Education*, *18*(6), 685-712. https://doi.org/10.1080/0950069960180604

Shea, D.L., Lubinski, D., & Benbow, C.P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93*(3), 604–614. https://doi.org/10.1037/0022-0663.93.3.604

Shipley, T.F., & Gentner, D. (2013). Introduction to the special issue on spatial learning and reasoning processes. *Cognitive Processing, 14*(2), 103-104. https://doi.org/10.1007/s10339-013-0565-1

Sorby, S.A. (2001). A course in spatial visualization and its impact on the retention of female engineering students. *Journal of Women and Minorities in Science and Engineering*, *7*, 153–172. https://doi.org/10.1615/JWomenMinorScienEng.v7.i2.50

Stover, S., & Saunders, G. (2000). Astronomical misconceptions and the effectiveness of science museums in promoting conceptual change. *Journal of Elementary Science Education*, *12*(1), 41-52. https://doi.org/10.1615/10.1007/BF03176897

Subaşı, M., Aydın, S., & Koçak, G. (2015). Gifted students' perceptions on basic astronomy concepts. *Journal of Emerging Trends in Educational Research and Policy Studies (JETERAPS), 6*(6), 444-451.

Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Taber, K.S. (Ed.). (2007). *Science education for gifted learners*. Routledge.

Taber, K.S. (2010). Challenging gifted learners: General principles for science educators and exemplification in the context of teaching chemistry. *Science Education International, 21*(1), 5-30. https://files.eric.ed.gov/fulltext/EJ890658.pdf

Taylor, H.A., & Hutton, A. (2013). Think3d!: Training spatial thinking fundamental to STEM education. *Cognition and Instruction*, *31*(4), 434-455. https://doi.org/10.1080/07370008.2013.828727

Trumper, R. (2000). University students‟ conceptions of basic astronomy concepts. *Physics Education*, *35*(1), 9-15. https://doi.org/10.1088/0031-9120/35/1/301

Trumper, R. (2001a). A cross-age study of junior high school students‟ conceptions of basic astronomy concepts. *International Journal of Science Education*, *23*(11), 1111-1123. https://doi.org/10.1080/09500690010025085

Trumper, R. (2001b). Across-age study of senior high school students‟ conceptions of basic astronomy concepts. *Research in Science and Technological Education*, *19*(1), 97-109. https://doi.org/10.1080/02635140120046259

Trumper, R. (2003). The need for change ın elementary school teacher training-a cross-college age study of future teachers‟ conceptions of basic astronomy concepts. *Teaching and Teacher Education*, *19*, 309-323. https://doi.org/10.1016/S0742-051X(03)00017-9

Trumper, R. (2006). Teaching future teachers basic astronomy concepts-seasonal changes-at a time of reform in science education. *Journal of Research in Science Teaching*, *43*(9), 879-906. https://doi.org/10.1002/tea.20138

Trundle, K.C., Atwood, R.K., & Christopher, J.E. (2002). Preservice elementary teachers' conceptions of moon phases before and after instruction. *Journal of Research in Science Teaching*, *39*(7), 633–658. https://doi.org/10.1002/tea.10039

Trundle, K.C., Atwood, R.K., & Christopher, J.E. (2007). A longitudinal study of conceptual change: preservice elementary teachers' conceptions of moon phases. *Journal of Research in Science Teaching*, *44*(2), 303–326. https://doi.org/10.1002/tea.20121

Tuckman, B.W., & Harper, B.E. (2012). *Conducting educational research*. Rowman & Littlefield Publishers.

Türk, C. (2016). The correlation between pre-service science teachers' astronomy achievement, attitudes towards astronomy and spatial thinking skills. *Journal of Education and Learning*, *5*(2), 187-199. https://doi.org/10.5539/jel.v5n2p187

Van Teijlingen E., & Hundley, V. (2002). The impotance of pilot studies. *Nursing Standard*, *16*(40), 33-36. https://doi.org/10.7748/ns2002.06.16.40.33.c3214

Vosniadou, S., & Brewer, W.F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*(4), 535-585. https://doi.org/10.1016/0010-0285(92)90018

Vosniadou, S., Skopeliti, I., & Ikospentaki K. (2004). Modes of knowing and ways of reasoning in elementary astronomy. *Cognitive Development*, *19*, 203-222. https://doi.org/10.1016/j.cogdev.2003.12.002

Wai, J., Lubinski, D., & Benbow, C.P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817-835. https://doi.org/10.1037/a0016127

Wallace, C.S., Prather, E.E., Mendelsohn, B.M. (2013). Astro 101 students' perceptions of science: Results from the thinking about science survey instrument. *Astronomy Ecuation Review (AER)*, *12* (1), 1-28. https://doi.org/10.3847/AER2012042

Weakley, K.D. (2010). *The effects of an inquiry-based earth science course on the spatial thinking of pre-service elementary teacher education students* (Publication No. 3410420) [Doctoral dissertation, Western Michigan University]. ProQuest Dissertations Publishing.

Webb, R.M., Lubinski, D., & Benbow, C.P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*(2), 397–420. https://doi.org/10.1037/0022-0663.99.2.397

Wellner, K. (1995). *A correlational study of seven projective spatial structures with regard to the phases of the Moon* (Publication No. 9536267) [Doctoral dissertation, University of Iowa]. ProQuest Dissertations Publishing.

Wells, C.S., & Wollack, J.A. (2003). *An instructor's guide to understanding test reliability*. Testing & Evaluation Services, University of Wisconsin.

Wilhelm, J. (2009). Gender differences in lunar-related scientific and mathematical understandings. *International Journal of Science Education*, *31*(15), 2105-2122. https://doi.org/10.1080/09500690802483093

Wilhelm, J., Jackson, C., Sullivan, A., & Wilhelm, R. (2013). Examining differences between preteen groups "spatial-scientific understandings: A quasi-experimental study. *The Journal of Educational Research*, *106*(5), 337-351. https://doi.org/10.1080/00220671.2012.753858

Wu, H., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465–492. https://doi.org/10.1002/sce.10126

Yair, Y., Schur, Y., & Mintz, R. (2003). A "thinking journey" to the planets using scientific visualization technologies. *Journal of Science Education and Technology*, *12*, 43-49. https://doi.org/10.1023/A:1022107627775

Yen, J.C., Tsai, C.H., & Wu, M. (2013). Augmented reality in the higher education: Students' science concept learning and academic achievement in astronomy. *Procedia-Social and Behavioral Sciences*, *103*, 165-173. https://doi.org/10.1016/j.sbspro.2013.10.322

Yerdelen-Damar, S.Y., & Aydın, S. (2015). Relations of approaches to learning with perceptions of learning environment and goal orientations. *Education and Science*, *40*(179), 269-293. https://doi.org/10.15390/EB.2015.4332

Young, A., & Shavl, S.J. (2013). Multiple choice testing for introductory astronomy: Design theory using Bloom's taxonomy. *Astronomy Education Review, 12*(1). https://doi.org/10.3847/AER2012027

Zeilik, M., Schau, C., & Mattern, N. (1998). Misconceptions and their change in university-level astronomy courses. *The Physics Teacher, 36*(2), 104-107. https://doi.org/10.1119/1.880056

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 457–475
https://doi.org/10.21449/ijate.1548724
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# Validity and reliability study of the school technology leadership scale according to teachers' perceptions

**Tugba Ögücü**[1*], **Aytaç Göğüş**[2]

[1]Istanbul Okan University, Graduate School & Milli Eğitim Bakanlığı, İstanbul, Türkiye
[2]Istanbul Okan University, Faculty of Education, Educational Sciences, İstanbul, Türkiye

**ARTICLE HISTORY**
*Received: Oct. 1, 2024*
*Accepted: Mar. 4, 2025*

**Keywords:**
School technology leadership,
School principals' leadership,
Teacher perceptions,
Technology leadership skills,
Scale adaptation.

**Abstract:** School principals should have school technology leadership skills in order to manage technology integration into teaching and learning activities effectively and efficiently. The 'School Technology Leadership Scale' was developed by Grace in 2020 to assess principals' technology leadership skills from the perspective of the teachers. The scale was presented in five dimensions, but Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) were not conducted by Grace in 2020. In the present study, the scale was adapted and modified for a different national school system and language in Türkiye by following an eight-stage process. The factor structure and factor loadings were determined by conducting EFA without removing any items from the original scale, and a two-factor structure was obtained. The validity and reliability of the adapted scale were examined with Cronbach's alpha analysis, item analysis, composite reliability of the new scale structure, intraclass correlation coefficient (ICC) tests, EFA and CFA. The findings indicate that there is a concomitant increase in the perception of technological leadership as the score on the scale increases. The adapted scale is a suitable tool to assess the perceptions of teachers in relation to their principals' technology leadership skills. Determining the indicators of technology leadership skills are important since school principals should develop technological leadership skills to 1) facilitate access to novel technologies in an effective teaching and learning process, 2) support effective learning, 3) render the school management process more practical, transparent, rational and data-driven and 4) promote effective utilization of technology.

## 1. INTRODUCTION

Technology has entered every aspect of our lives and has influenced society through many changes from the past to the present. The Covid-19 pandemic has accelerated the process of change. Grace (2020) states that the 21st century generation includes students who started kindergarten in the fall of 2018 and all students thereafter, while all educators and principals were born in the 20th century. The digital natives, who were born during technological advancements, are digital immigrants who are resistant to change, and digital hybrids who are adapting to change. Education systems are that can adapt society to change. Schools have important impacts on the society. As open systems, schools constantly interact with their

*CONTACT: Tugba ÖGÜCÜ ✉ tugbaogucu@gmail.com Istanbul Okan University, Gradute School, İstanbul, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

environment and have a crucial responsibility in preparing qualified individuals with the skills that society needs. According to Ghavifekr and Wong (2022), schools must educate students with 21st-century skills to meet the demands of Industry 4.0 and the rapidly developing technologies. Nowadays, schools are institutions where digital natives, digital hybrids, and digital immigrants coexist. Therefore, the developing economy, technology, globalization and social changes have transformed schools into complex structures and processes, leading to the need for quick decision-making in school management since 2000 (Gümüşeli, 2001).

School leadership is important to integrate Information and Communication Technology (ICT) into school system effectively (Hamzah *et al*., 2014). School principals have a crucial role for effective and efficient use of the technology due to the impact of technology on society and education (Esplin, 2017). According to Torrato *et al*. (2021), the roles of school principals are critical in ensuring students' learning. Therefore, school principals should provide a sustainable and dynamic digital learning culture (Ghavifekr & Wong, 2022). The school principal has two important roles, management and leadership (Akın-Mart & Tulunay-Ateş, 2021). In 21st-century schools, leadership and technology play a crucial role (Grace, 2020). Therefore, theories and models that examine leadership and the leadership process in achieving organizational excellence have emphasized the link of technology with leadership, highlighting the emergence of technology leadership (Hamzah *et al*., 2014).

The increasing use of technology in all areas has made technology leadership an important issue in school management (Weng & Tang, 2014; Mendoza & Catiis, 2022). According to Sincar (2009), technology leadership is a determining factor in school leadership. Research on technology leadership, as noted by Chang *et al*. (2008), began in the United States in the 1990s and has not lost its importance since then. In recent years, research on the use of technology in schools has gained momentum (Gyeltshen, 2021). As Chang (2008) argues that technology leadership is not only about the use of technology, but also about developing and changing school culture. Therefore, technology leadership remains an important and current concept due to the constantly evolving technology (Kesim, 2020).

In the study context in Türkiye, the Ministry of Education (MEB) Strategic Plan focuses on the development of technological infrastructure, the integration of technology into the education system, the increase of teacher competencies for the use of digital content, and the increase of quality in education (MEB Strategic Plan, 2024). The Twelfth Development Plan (2023) emphasizes the development of the integration of information and communication technologies into education in a way that allows to increase quality, and the increase of digital competencies of students and educators. Among the long-term goals,

> "As digital technologies, personalized learning experiences, virtual and augmented reality, and artificial intelligence become prevalent in education, there will be a simultaneous recognition of the enduring significance of comprehensive teaching, guidance, and social interaction in shaping individual education. The education of the future will prioritize accessibility, diversification, and personalization, with curricula geared toward fostering the development of novel knowledge and skills" (The Twelfth Development Plan, 2023, p. 14).

In addition, 'providing digital competencies to educators' has been determined as the most important strategy of the report in the Türkiye Digital Roadmap (Saçak *et al*., 2020).

With the widespread use of Web 2.0 applications in the 2020s, there is a need for school technology leaders who are aware of the importance of educational technologies in effective time management for teachers and student learning (Grace, 2020). Dexter and Barton (2021, p. 368) define school technology leadership as "the ability of school leaders to select effective technologies to support student learning and create effective opportunities for teachers to learn how to integrate them." A'mar and Eleyan (2022) emphasize that technology leadership encompasses all activities related to technology in schools, including decision-making, policies

and implementation. Also, Grace (2020) points out that school technology leadership focuses on the integration of new technologies into the school environment.

Morgan (2014, p. 8) defines school technology leadership as "a series of instructional leadership practices related to technology integration that are distributed to multiple members of each school team and are consistent with three broad categories, direction setting, people development and organizational redesign." Therefore, technology changes the roles of school principals in terms of instructional leadership (Kwatubana, 2023). Instructional leadership, which includes technology leadership, is developing in a way that the success of computer and information technology applications in schools is a result of effective instructional leadership (Kwatubana, 2023). Technology leadership brings transparency, reliable data provision, practicality, and speed to instructional leadership practices. However, Akada and Şahin-Fırat (2022) argue that school technology leadership can speed up and facilitate school management processes. School technology leadership plays a key role in integrating technology into education because of directing the technology integration program of the school (Torrato *et al.*, 2021; Wu *et al*., 2015). Additionally, school technology leadership is a type of leadership that requires effective use of technology in school management practices (Topçu & Ersoy, 2020). Within this scope, rational implementation and maintenance of management processes at the organizational level can be achieved. Therefore, it can be claimed that school technology leadership can also increase administrative effectiveness in the school. Ultimately, school technology leadership is a functional type of leadership that encompasses all necessary leadership skills to prepare the school for the 21st century and ensure the beneficial application of technology (Rayray, 2023). It is one of the essential characteristics that school principals should possess (Anderson & Dexter, 2005). The continuous development of technology and its impact on the education system are rapidly creating change in the education process. In this context, it is believed that verifying the instrument that assesses technology leadership skills, adapting the scale by considering different cultures, emerging technologies, and changing practices, would be beneficial for applications in the field.

### 1.1. The Importance of School Technology Leadership Study

Eight components of technology leadership in schools are identified as follows (Apsorn *et al.*, 2019), 1) developing a vision and operational guidelines for emerging technologies in schools, 2) establishing strategies and providing support to direct teachers and students towards the use of technology in the learning process, 3) preparing plans for the development of teachers' and staff's technological skills, 4) creating a suitable environment by providing resources for the use of information technology, 5) investigating problems that may arise during the use of new technologies and developing solutions to these problems, 6) serving as a model for the use of information technology in education and daily life, 7) creating learning communities by ensuring the information sharing and the establishment of information technology culture among school personnel, 8) developing decision-making strategies in solving problems that may arise at school.

According to Anderson and Dexter (2005), effective school formation is influenced significantly by the school principal. In this regard, as technology advances and becomes digital, it is necessary for school principals to acquire technology leadership skills to increase students' potential (Mendoza & Catiis, 2022). According to Banoğlu (2012), school principals who are technology leaders can ensure the integration of education and technology by integrating them seamlessly. Banoğlu (2012) summarizes the effective educational leader roles as 1) provides technology for educational activities and management while efficiently managing resources, 2) monitors and supports the use of educational technology and professional development of teachers and staff, 3) evaluates educational activities in the school using technology and 4) uses technology appropriately in communication and relationships with the school's environment.

As technology leaders, school principals should be innovative, resourceful, crisis managers, guides, coordinators, knowledge managers, producers and managers (Durnalı, 2018). According to Hsieh *et al.* (2014, p. 113), "the school principal should be literate in information technology to enable the school staff to use technology effectively in teaching, create a learning environment that facilitates students' learning motivation, and achieve the goal of becoming a highly acclaimed school." School principals who are also technology leaders should possess a technology culture to be able to meet the needs of digital natives growing up in a technology culture (Öztaban, 2020).

Researchers have identified essential leadership skills in technology as 1) allocating time for technological education, 2) establishing goals for the administrative and educational use of technology, 3) planning the use of educational technologies, 4) providing equal access to technological resources (Demiraçan, 2019; Durnalı, 2018). Researchers have also determined the key objectives related to technology, tools and infrastructure as 1) sourcing and finding resources for technology, tools and infrastructure, 2) developing strategies for technology integration, 3) ensuring the effective use of technology in teaching and learning, 4) providing technical support to faculty and staff, 5) developing policies and procedures related to technology, tools and infrastructure, 6) establishing goals for the administrative and educational use of technology, 7) planning for the use of educational technologies, 8) recognizing successful technology implementations, 9) understanding the elements that are effective in evaluation, 10) providing visionary leadership, 11) sourcing and finding resources for technology, tools and infrastructure, 12) preparing policies for voluntary participation in technological innovation and usage that align with societal values and 13) advocating for the benefits of technology and determine priorities for its usage (Demiraçan, 2019; Durnalı, 2018). Grady (2011) outlines the roles of a school technology leader as 1) setting the school's technology vision and goals and representing technology, 2) modelling technology use and supporting its use, 3) participating in professional development for the technology integration, 4) providing professional development opportunities for school stakeholders to facilitate technological learning activities, 5) supporting the use and integration of technology by providing technological tools, 6) advocating technology that enhances students' learning activities, 7) being aware of national technology standards and supporting their adoption, encouraging school stakeholders to do the same, 8) informing school stakeholders about the importance of using technology to enhance students' learning skills.

Improving the technology leadership skills of school administrators can contribute to the quality of schools and support adaptation to rapid technological development. In fact, the Türkiye Digital Roadmap report by Saçak *et al*. (2020) emphasizes that "when the strategy of 'providing digital competencies to educators' is implemented alone, more benefits are obtained than when other strategies are implemented alone" (p. 344). This is because technology can create qualified outputs as an element of quality education. According to Gonzales (2020), the technology leadership skills demonstrated by school principals reflect the quality of school leadership, while Nababan *et al.* (2021) argue that quality students are the output of a good school system. School principals' technology leadership supports teachers and students in integrating technology into education in a way that allows effective and efficient learning.

Technology leadership has been analyzed from different perspectives in national and international literature. In this context, the first of these studies was conducted by Aten (1996) and by Anderson and Dexter (2005) in USA, by Tanzer (2004) in Türkiye (cited in Durnalı, 2022). In Türkiye, Sincar's (2009) research on the technological leadership roles of primary school principals is a turning point and the scale developed in this study has been used in eight studies (Durnalı, 2022). In the international literature, the literature on school technology leadership has grown in the first decade of the 21st century and continued to grow in the second decade (Grace, 2020).

According to Grace (2020), scale development in the international literature began with the 'Teaching, Learning and Computing Survey' in 1998, and numerous measurement tools have been developed since then (e.g. Chang *et al*, 2008). Researchers in the international literature have developed scales for their own studies, based on the International Society for Technology in Education's (ISTE, 2018) standard for administrators, which is called the Educational Technology Standards for Administrators (NETS-A) (Grace, 2020). NETS-A, known as technology leadership standards, defines the knowledge and skills that school principals should have in the role of technology leadership in primary and secondary education with five parameters including digital citizenship, equity, visioneering, team and systems building, continuous improvement and professional growth (ISTE, 2018).

According to Akada and Şahin-Fırat (2022), various scale development studies in Türkiye are based on the NETS-A standards prior to 2018 (e.g. Akbaba-Altun, 2002; Banoğlu, 2012; Hacıfazlıoğlu *et al*., 2011; Şişman-Eren, 2010). In addition to these scales, scales have been developed regarding the school principal's level of technology use, attitudes, behaviours, roles and self-efficacy (e.g. Afshari *et al.*, 2012; Bülbül & Çuhadar, 2012; Durnalı, 2018; Ermiş & Somuncuoğlu-Özerbaş, 2021; Sezer, 2011; Seferoğlu & Akbıyık, 2005; Sincar, 2009; Stuart *et al*., 2009). The scale developed by Akada and Şahin-Fırat (2022) is based on the updated 2018 NETS-A standards and the perceptions of middle school teachers. The scale developed by Durnalı (2022) focuses on the behaviors of technology leaders and targets the perceptions of school teachers.

Grace (2020) conducted a thematic literature review for about a year and sought the opinions of educational technology experts and graduate students to develop a new instrument called 'School Technology Leadership Scale.' Grace (2020) comprehensively analyzed all aspects of school technology leadership and scales, including NETS-A standards and quantitative measures of school technology leadership. Grace (2020) obtained reliable results with a small number of items in the scale by focusing on assessing school principals' educational and technology leadership skills.

This present study is important in order to adapt the 'School Technology Leadership Scale' developed by Grace (2020) to different languages and cultures to assess principals' technology leadership skills according to the perceptions of both principals and teachers in ongoing studies. Furthermore, technology leadership should be examined through new perspectives to contribute to the literature in terms of developing and adapting scales as opinions on technology leadership change.

## 2. METHOD

In this study, a scale adaptation and a modification method (Gökdemir & Yılmaz, 2023) were employed, comprising the execution of EFA and CFA, the identification of scale dimensions, and the establishment of a new nomenclature that is aligned with the theoretical framework

### 2.1. Research Model

The objective of this study is to adapt the 'School Technology Leadership Scale', originally developed by Grace (2020), into Turkish. To ascertain the psychometric properties of the scale, a series of item statistics and reliability tests, in addition to validity tests, EFA and CFA analyses were conducted during the eight-stage adaptation process.

### 2.2. Participants

The participants included 280 volunteer primary school teachers who worked in official schools in a metropolitan city in the 2023-2024 academic year. The participants were determined through the application of the simple random sampling technique. As stated by Tabachnick and Fidell (2001), a sample size of approximately 150 is deemed sufficient for exploratory factor analysis (Çokluk *et al.*, 2023). As Çokluk *et al.* (2023) note, the sample size should be at least twice the number of items. In this regard, the construct validity of the scale was evaluated

through EFA, utilizing data obtained from 180 primary school teachers. The existing structure of the scale was then assessed through CFA, employing data from 100 primary school teachers, distinct from those included in the EFA analysis. The demographic information of the teachers is presented in Table 1.

**Table 1.** *Participants' demographic information.*

| Demographics (n:280) | | *n* | *%* |
|---|---|---|---|
| Gender | Male | 91 | 32.5% |
| | Female | 189 | 67.5% |
| Education level | Bachelor's degree | 228 | 81.4% |
| | Master's degree | 45 | 16.1% |
| | Doctoral degree | 7 | 2.5% |
| Average Age | 42.18±9.62 | | |
| Average Tenure | 18.18±9.63 | | |

Table 1 indicates that the average age of the participants is 42 years, with the average tenure of 18 years. The data revealed that 67.5% of the teachers were female, 32.5% were male. Furthermore, 81.4% of the teachers held a bachelor's degree, 16.1% held a master's degree and 2.5% held a doctoral degree.

### 2.3. Data Collection Tool

Grace (2020) developed the "School Technology Leadership Scale According to Teachers' Perceptions" by examining published scales in the field and creating a 20-item scale with five factors based on the study by Chang (2008). As a result, the items in the dimensions served as indicators (Grace, 2020). The scale was designed in a four-point Likert scale, responses to the items ranged from '*Strongly Agree*' to '*Strongly Disagree*.'

Grace (2020) preferred to use the classical test theory instead of EFA when there was insufficient data to ensure accurate statistical analysis. Grace (2020) determined the analysis techniques based on the literature and data from scale development experts. Additionally, Grace (2020) conducted item analyses, Cronbach's alpha, inter-item correlation range, mean inter-item correlation and corrected item-total correlation. According to the findings of Grace (2020) , the inter-item correlations for each dimension were as follows, 1) "the vision, planning and management dimension" ranged from .511 to .810, (item 1, 2, 3, 4), 2) "the personnel development and training dimension" ranged from .519 to .741, (item 5, 6, 7, 8), 3) "the technological and infrastructure support dimension" ranged from .670 to .815, (item 9, 10,11,12), 4) "the evaluation and research dimension" ranged from .604 to .934, (item 13, 14, 15, 16) and 5) "the interpersonal and communication skills dimension" ranged from .226 to .787, (item 17, 18, 19, 20). However, the Cronbach's alpha value varied between .805 and .940 for the sub-dimensions. The findings indicated that the dimensions were largely reliable. Grace (2020) suggested conducting further studies to test the reliability and validity of all dimensions. Additionally, Grace (2020) recommended to identify a larger study group to apply other analysis types such as EFA. As proposed by Grace (2020), future studies might encompass dimension analysis, modification of substances identified within dimensions and the utilization of diverse statistical measurement techniques in data analysis.

#### 2.3.1. *A Turkish adaptation of the data collection tool and modification of the scale dimensions*

Researchers employ a variety of methods in the context of scale adaptation studies. In this context, the utilization of EFA, and then, the utilization of CFA in scale adaptation studies is subject to variation. As Orçan (2018) notes, EFA and CFA are employed concurrently in some studies, whereas in others, only CFA is utilized and the use of only CFA in adaptation studies may give rise to certain issues. To address this, multiple CFA analyses can be conducted to

ensure compliance with the model. It is essential to perform EFA in order to identify cultural differences and to detect any potential errors (Orçan, 2018). Consequently, EFA and then CFA were conducted in this study in accordance with the recommendation of the scale owner.

The objective of scale adaptation studies is to examine both psycholinguistic features (language adaptation) and psychometric properties (Karaçam, 2019). In the present study, firstly, the most appropriate scale for the study was determined through a comprehensive literature review. Subsequently, the scale owner was contacted to ascertain whether the scale had previously been adapted into Turkish. Permission was then sought for the adaptation and utilization of the scale. As the original study (Grace, 2020) did not employ EFA, the objective of this scale adaptation study was to ascertain the scale sub-dimensions resulting from EFA analyses, resulting in the determination of the scale dimensions and the introduction of new nomenclature that is aligned with the theory. The adaptation and modification of scale dimensions (Gökdemir & Yılmaz, 2023) were completed in an eight-stage process, the procedures of which are described in the following section.

*In the first stage (Translation into Turkish),* the translation into Turkish was conducted by four experts in English, who were both linguistically fluent and specialized in this field. Four experts translated the original language text into Turkish. The translations were compared and analyzed, and the most appropriate translation was determined.

*In the second stage (Expert Opinions),* the opinions of three experts in Turkish and foreign languages were sought regarding expressions that were unclear and contained errors. While the foreign language expert evaluated the translation, the Turkish language expert evaluated the suitability of the translation for Turkish. An information technology teacher analyzed the translation in terms of its terminological appropriateness and the necessary arrangements were made accordingly.

*In the third stage (Back Translation),* the Turkish expressions were translated into English, the original language. The back-translators were different experts from the first translators. Then, three linguists compared the original text with the back-translated text, and Turkish translations were revised by seeking expert opinion to investigate the reasons for the differences.

*In the fourth stage (Pilot Application with the Real Participants),* the items of the scale were discussed with 5 primary school teachers and 5 school principals who had completed postgraduate in the educational administration field and their opinions were taken.

*In the fifth stage (Expert Opinions),* the opinions of three teachers and one school principal with a doctorate degree in educational administration field, and four academicians from two different universities, were obtained, and the necessary arrangements were made. Since the scale adaptation process was followed without removing items, the question "What are your opinions regarding the sub-dimensions and scale items of the School Technology Leadership Scale?" was asked. The participants stated that the items reflected current skills. It was seen that there was no negative opinion regarding the scale items in general.

*In the sixth stage (Initial Application Process),* the scale was administered to a sample of 40 primary school teachers to assess the comprehensibility of the statements in the 'School Technology Leadership' scale. The comprehensibility of the statements was evaluated through statistical analyses, which indicated that no issues were identified. At this stage, linguistic equivalence was not determined. It was checked whether the scale items were perceived correctly by each participant. The scale was subsequently administered to additional teachers.

*In the seventh stage (Second Implementation Process),* the scale adapted by the researchers to measure school leaders' technology leadership skills was administered to 280 primary school teachers. The data were collected using Google Forms and face-to-face interviews. After the applications were completed, the incorrect ones from the face-to-face questionnaires were

excluded from the analysis. The collected data were used to determine the psychometric properties of the School Technology Leadership Scale.

*In the eighth stage (Analyses of the Second Application)*, Cronbach's alpha and item-total correlations were evaluated for reliability testing. EFA was employed to determine the structure of the items in the scale pool. CFA was conducted to verify the theoretical structure obtained by EFA. The CFA analysis was applied to a different data set than the one used in the EFA study.

### 2.4. Data Analysis

Item analysis and factor analysis were examined in the validity studies of the scale (Karasar, 2020). The construct validity of the scale was tested using EFA with data from 180 primary school teachers, and the existing structure of the scale was confirmed using CFA with data from 100 primary school teachers. According to Cohen and Cohen (1983), data from at least 10 participants for each item is sufficient in CFA analysis. Sample adequacy was analyzed using Kaiser-Meyer-Olkin (KMO) and the suitability of the data for factor analysis was analyzed using the Bartlett test. The number of factors was determined through the use of a scree plot and an examination of the eigenvalues. The reliability and internal consistency of the scale were analyzed with Cronbach's alpha coefficient, item-total correlation, composite reliability and intra-class correlation coefficient (ICC) tests. CFA was used to test the compatibility of the current structure of the scale with the applied study group. In this context, goodness-of-fit criteria were considered when explaining the model using CFA (Osborne, 2014). A path diagram was created based on the model results, and the standardized factor loadings of the included items were examined. The LISREL 8.5 software was used for CFA and IBM SPSS 28th version was used EFA and other tests.

## 3. FINDINGS

### 3.1. Reliability Analyses

Prior to the EFA analysis, tests were conducted to evaluate the reliability of the items, the results of which are presented in Table 2. As Karasar (2020) notes, reliability refers to the extent to which similar results are obtained when processes are repeated. It is therefore a crucial concept in any scientific investigation. Cronbach's alpha test is one of the most widely employed tests of scale reliability in the social sciences (Osborne, 2014). In accordance with the Cronbach's alpha test, a reliability alpha coefficient of 'α' between .80 and 1.00 indicates a highly reliable scale, while a coefficient between .60 and .79 denotes a quite reliable scale. A reliability coefficient of .40-.59 indicates low reliability, while a coefficient of .00-.39 indicates that the scale is not reliable (Alpar, 2020). A high reliability coefficient indicates that the scale is internally consistent (Karaca-Akarsu & Özdemir, 2021). The item-total correlation serves to determine the internal consistency of the scale (Çamur & Göğüş, 2023). In this context, internal consistency refers to the agreement between test items, whereas item-total correlation denotes the relationship between the total scores of the test and the scores obtained from the items of the test (Büyüköztürk, 2020; Karasar, 2020). The items exhibited high reliability according to Cronbach's alpha (.971) (Osborne, 2014). Upon examining the item-total correlations of the scale presented in Table 2, it can be observed that all items had item-total correlation values above .3. Therefore, it can be stated that the necessary conditions were met in terms of item-total correlations (Cristobal *et al.*, 2007). The internal consistency was high due to the high and positive item-total correlation, indicating the relationship between the total scores of the test and the scores obtained from the test items (Büyüköztürk, 2020; Karasar, 2020). Consequently, a high and positive item-total correlation indicated that the behaviors were similar, and the internal consistency was high (Büyüköztürk, 2020).

**Table 2.** *Item statistics and reliability test results for the scale*.

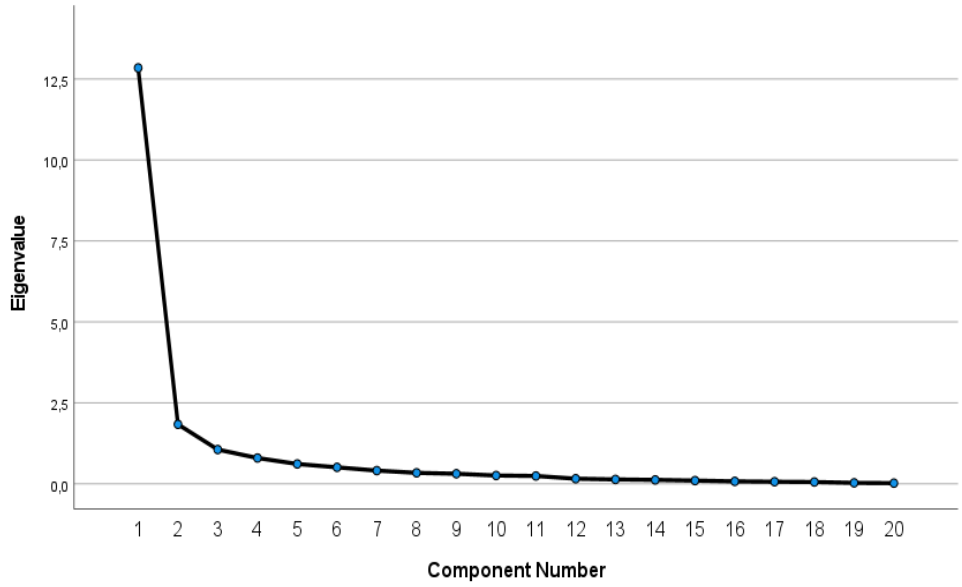
| Items No. | Item mean | Item standard deviation. | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
|---|---|---|---|---|
| Item 1 | 3.427 | 0.616 | .736 | .970 |
| Item 2 | 3.514 | 0.574 | .653 | .971 |
| Item 3 | 3.451 | 0.599 | .727 | .970 |
| Item 4 | 3.443 | 0.598 | .762 | .970 |
| Item 5 | 3.243 | 0.696 | .795 | .969 |
| Item 6 | 3.212 | 0.722 | .825 | .969 |
| Item 7 | 3.141 | 0.745 | .849 | .969 |
| Item 8 | 3.137 | 0.727 | .856 | .969 |
| Item 9 | 3.400 | 0.586 | .793 | .970 |
| Item 10 | 3.243 | 0.690 | .825 | .969 |
| Item 11 | 3.459 | 0.579 | .688 | .971 |
| Item 12 | 3.267 | 0.681 | .825 | .969 |
| Item 13 | 3.098 | 0.765 | .846 | .969 |
| Item 14 | 3.337 | 0.631 | .834 | .969 |
| Item 15 | 3.263 | 0.719 | .847 | .969 |
| Item 16 | 3.196 | 0.716 | .847 | .969 |
| Item 17 | 3.082 | 0.802 | .760 | .970 |
| Item 18 | 3.278 | 0.702 | .634 | .971 |
| Item 19 | 3.075 | 0.736 | .762 | .970 |
| Item 20 | 3.208 | 0.646 | .762 | .970 |
| Cronbach's Alpha | (.971) | | | |

### 3.2. Exploratory Factor Analysis

Exploratory Factor Analysis (EFA) is a sophisticated multivariate statistical technique that is widely employed in social sciences, education and psychology (Taherdoost, *et al.*, 2020). As defined by Osborne (2014, p. 4), "EFA is a group extraction and rotation technique designed to model completely unobserved or hidden constructs". Sub-dimensions formed by the scale items are determined by this method (Akcay et al., 2018). Conversely, the suitability of the data for exploratory factor analysis is determined through the utilization of the Kaiser-Meyer-Olkin KMO and Barlett Tests. While Kaiser-Meyer-Olkin (KMO) analysis assesses the suitability of the sampling for factor analysis, Bartlett's test determines the suitability of the data for factor analysis (Akcay *et al.*, 2018; Seçer, 2013; Taherdoost, *et al.*, 2020). The Bartlett's test of sphericity (suitability for factor analysis) is conducted to examine the verticality of the relationships between the items included in the analysis. The adequacy of the sample provides information that is useful for both grouping the questionnaire items as factors and for explaining the structures in a more effective manner (Taherdoost *et al.*, 2020). In the present study, in order to determine the EFA, the Varimax rotation method was employed, with consideration given to Principal Component Analysis. Principal component analysis represents one of the general dimension analyses and is a simplified version in terms of calculation (Osborne, 2014). As posited by Castello and Osborne (2005), the objective of the rotation process is to streamline and elucidate the data structure. Varimax rotation is the most commonly employed rotation method, with the objective of maximizing the variance within a factor, thereby increasing larger loadings and minimizing smaller loadings (Castello & Osborne, 2005; Osborne, 2014). In this regard in the study, EFA was conducted in accordance with the Varimax rotation method, with the results presented in Table 3. The number of factors to be included in the scale was determined on the basis of the scree plot and eigenvalue results, which are presented in Figure 1. The eigenvalues were employed in the calculation of the variance explained by the factors and in the determination of the number of factors. An eigenvalue exceeding one indicated that the factor was stable (Çokluk *et al.*, 2023). The scree plot facilitated factor reduction by

revealing the dominant factors (Çokluk *et al.*, 2023). The results of the EFA are presented in Figure 1.

**Figure 1.** *Scree plot of the Eigenvalue for EFA.*




Upon examination of the scree plot in Figure 1, it became evident that the eigenvalues exhibited a flattening trend after two factors. This observation led to the conclusion that a two-factor structure was present. However, the final factor structure can be determined through exploratory factor analysis (EFA). The factor loadings of the items for EFA are presented in Table 3.

**Table 3.** *Factor loads for EFA.*

| Item No. | Factor | |
|---|---|---|
| | 1. Factor | 2. Factor |
| Item 13 | .871 | *<.3* |
| Item 19 | .820 | *<.3* |
| Item 20 | .809 | *<.3* |
| Item 15 | .807 | *.402* |
| Item 16 | .802 | *.420* |
| Item 17 | .787 | *<.3* |
| Item 12 | .744 | *.439* |
| Item 8 | .728 | *.465* |
| Item 10 | .725 | *.518* |
| Item 7 | .706 | *.512* |
| Item 14 | .704 | *.465* |
| Item 11 | .619 | *.517* |
| Item 2 | *<.3* | .852 |
| Item 4 | *<.3* | .851 |
| Item 3 | *<.3* | .824 |
| Item 9 | *.441* | .768 |
| Item 1 | *.425* | .761 |
| Item 5 | *.473* | .726 |
| Item 6 | *.498* | .671 |
| Item 18 | *.425* | .583 |

KMO: .857. Bartlett Test $\chi^2$: 1068.7 ($p$<.05)
Variance Explaining Ratio: %73.42
Extraction Method: Principal Component Analysis

Upon reviewing Table 3, it was determined that the Turkish version of the scale, consisting of 20 items, exhibits two-dimensional structure based on the collected data. Upon examining the factor loads, it was found that all factor loads were above .3. The KMO test (KMO: .857) indicated that the sample size was sufficient and the Bartlett Test ($\chi^2$:1068.7 ($p$<.05)) was also conducted. The KMO value was .857, higher than .60, indicating an adequate level of sampling adequacy (Büyüköztürk, 2020). Furthermore, Bartlett's test yielded a statistically significant result ($p$<.05).

The results demonstrated that the assumption of sphericity was met, with a value less than .05. The Kaiser-Meyer-Olkin (KMO) result and Bartlett's test support the suitability of exploratory factor analysis (EFA). Ultimately, it was established that the two-factor structure accounted for 73.42% of the variance. The items pertaining to the first factor are associated with the theoretical field of "*managing applications*," whereas the items pertaining to the second factor are related to "*guiding and supporting*" school technology leadership.

### 3.3. Confirmatory Factor Analysis

Confirmatory Factor Analysis (CFA) was used to test the model validity for the compatibility of the current structure of the scale with the applied sample, and the results are presented in Table 4. CFA is an essential analytical tool for structure validation and other aspects of psychometric evaluation and is a type of structural equation modelling (Brown & Moore, 2012). According to Brown and Moore (2012), CFA confirms the number of factors and factor loadings of the scale.

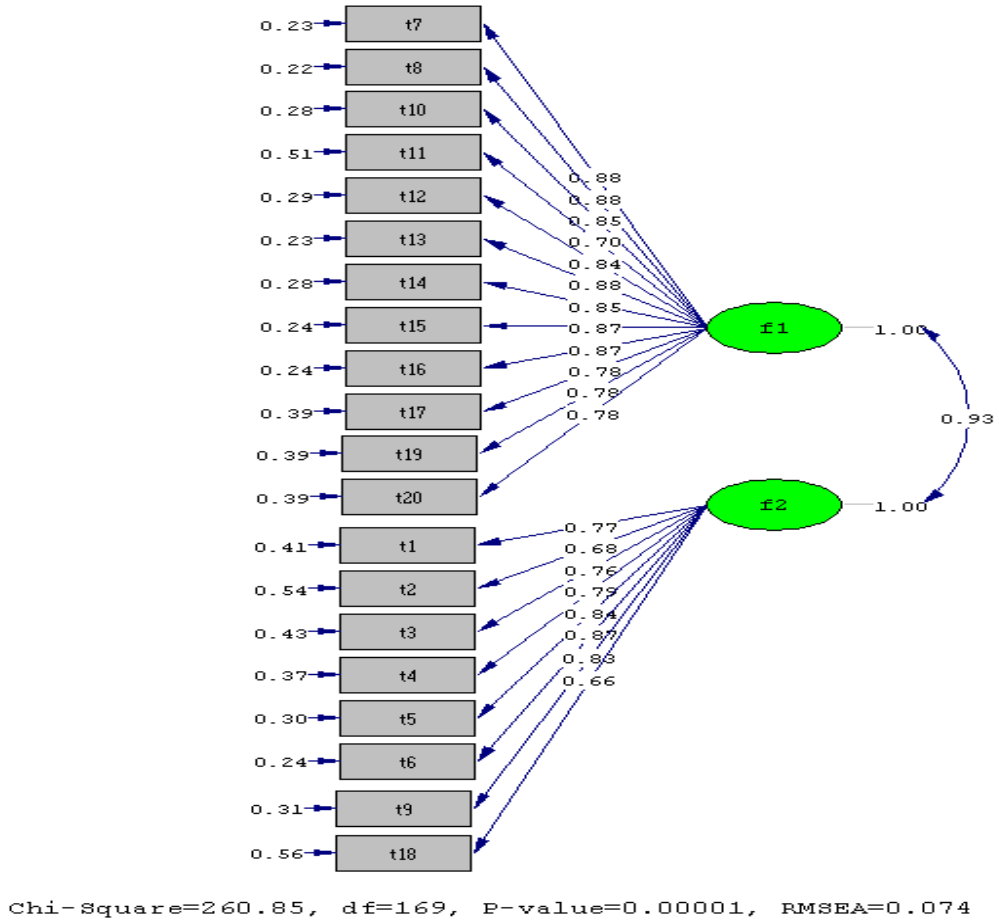
**Table 4.** *Fit Values for CFA.*

| Criteria | Goodness of Fit Values | Good | Acceptable |
|---|---|---|---|
| RMSEA | .074 | 0 ≤ RMSEA ≤ .05 | .05 ≤ RMSEA ≤.08 |
| $\chi^2/df$ | 1.543 | 0≤ $\chi^2/df$≤2 | 0≤ $\chi^2/df$ ≤3 |
| SRMR | .042 | 0 ≤ SRMR ≤ .05 | .05 ≤ SRMR ≤ .10 |
| NFI | .99 | .95 ≤ NFI ≤ 1.00 | .90 ≤ NFI ≤ .95 |
| GFI | .99 | .95 ≤ GFI ≤ 1.00 | .90 ≤ GFI ≤ .95 |
| AGFI | .99 | .90 ≤ AGFI ≤ 1.00 | .85 ≤ AGFI ≤ .90 |
| PGFI | .80 | .95 ≤ PGFI ≤ 1.00 | .50 ≤ PNFI ≤.95 |
| CFI | .98 | .95 ≤ CFI ≤ 1.00 | .90 ≤ NFI ≤ .95 |
| RFI | .99 | .95 ≤ RFI ≤ 1.00 | .90 ≤ RFI ≤.95 |

*df*:169 $\chi^2$:260.85 n: 100

Table 4 illustrates that the goodness of fit index (GFI: .99), the adjusted goodness of fit index (AGFI: .99), the consistent goodness of fit index (PGFI: .80) and the normed fit index (NFI: .99) were all highly satisfactory. Additionally, the comparative fit index (CFI: .98), relative fit index (RFI: .99), root mean square of standardized error squares (SRMR: .042) and root mean square of approximate errors (RMSEA: .074) were also considered (Schumacker & Lomax, 2004). The path diagram resulting from the model is presented in Figure 2. However, the path diagram is also a schematic representation of the results of the structural equation modelling analysis, showing summary outputs of the model such as factor loadings, *t*-values, unexplained variances and goodness-of-fit indices (Çapık, 2014).

**Figure 2**. *PATH diagram for factor loads.*




Upon examining Figure 2, the standardized factor loadings of the items included in the model through CFA are greater than .3 (Schumacker & Lomax, 2004). The results demonstrate that the generated model meets the necessary criteria in terms of standardized factor loadings. Based on the findings of the analysis, the adapted scale is a valid and reliable scale. The scale operates on a 4-grade Likert measurement level, with a minimum achievable score of 20 and a maximum score of 80. Given that all expressions within the scale embody positive constructs, one can infer that as the score derived from the scale rises, there is an increase in the perception of technological leadership. The results of the test conducted to ascertain the reliability and internal consistency levels of the novel factor structure are presented below. Table 5 presents the results of the reliability analyses conducted for the School Technology Leadership Scale. Item statistics and general reliability indicators were evaluated on two dimensions of the scale, Managing and Directing Applications and Guiding and Supporting.

As indicated in Table 5, the overall Cronbach's alpha value of the scale was determined to be .975. This demonstrates that all items of the scale exhibit high internal consistency. Upon analysis of the sub-dimensions, the Cronbach's Alpha value for the Managing Applications dimension was determined to be .966, while the Guiding and Supporting dimension exhibited a value of .941. As both values exceed .90, it can be concluded that the sub-dimensions are highly reliable. Upon the analysis of the Composite Reliability values, it was determined that the composite reliability value was .943 in the Managing Practices dimension and .930 in the Guiding and Supporting dimension. These values provide evidence to support the construct reliability of each dimension of the scale, with values exceeding the recommended threshold of .70. Furthermore, the item-total correlation values for all items fell within the range of .662 to .912. The findings demonstrate that each item is consistently associated with the overall scale and is capable of distinguishing between different levels of the construct. Upon the removal of

each item, there was no appreciable decline in the Cronbach's Alpha value of the scale. This demonstrates that the items contribute to the scale in a balanced manner, and that the exclusion of any item will not negatively impact the overall reliability. The results of the analyses indicate that the 'School Technology Leadership Scale' exhibits a notably high degree of internal consistency and reliability. Furthermore, the sub-dimensions of the scale demonstrate the capacity to provide a valid measurement. The findings provide substantial support for the usability and psychometric properties of the scale.

**Table 5.** *Item statistics and reliability test results of the School Technology Leadership Scale*

| Dimension | Item No. | Item mean | Item standard deviation. | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted | Reliability |
|---|---|---|---|---|---|---|
| Managing and Directing Applications | t7 | 3.160 | 0.748 | .872 | .961 | |
| | t8 | 3.150 | 0.770 | .895 | .961 | |
| | t10 | 3.260 | 0.691 | .813 | .963 | |
| | t11 | 3.460 | 0.610 | .717 | .965 | |
| | t12 | 3.270 | 0.723 | .820 | .963 | Cronbach's |
| | t13 | 3.080 | 0.787 | .879 | .961 | Alpha=.966 |
| | t14 | 3.360 | 0.659 | .809 | .963 | Composite |
| | t15 | 3.280 | 0.712 | .873 | .961 | Reliability=.943 |
| | t16 | 3.220 | 0.733 | .912 | .960 | |
| | t17 | 3.130 | 0.799 | .773 | .964 | |
| | t19 | 3.050 | 0.757 | .763 | .964 | |
| | t20 | 3.160 | 0.692 | .742 | .965 | |
| Guiding and Supporting | t1 | 3.440 | 0.671 | .765 | .934 | |
| | t2 | 3.510 | 0.628 | .759 | .935 | |
| | t3 | 3.370 | 0.661 | .851 | .929 | Cronbach's |
| | t4 | 3.390 | 0.650 | .874 | .927 | Alpha=.941 |
| | t5 | 3.220 | 0.719 | .823 | .930 | Composite |
| | t6 | 3.230 | 0.750 | .820 | .931 | Reliability=.930 |
| | t9 | 3.410 | 0.621 | .776 | .934 | |
| | t18 | 3.260 | 0.733 | .662 | .942 | |
| Cronbach's Alpha for the Scale: .975 | | | | | | |

Table 6 presents the findings of the Intraclass Correlation Coefficient (ICC) analysis, which was conducted to evaluate the reliability of the scale. In the aforementioned analyses, the measurement stability and reliability of the scale were examined at two distinct levels, single measures and average measures. The results are presented in the following section.

**Table 6.** *The results of the interclass correlation coefficient (ICC) analysis are presented below.*

| | | 95% Confidence Interval | | *F* Test (with True Value 0) | | | |
|---|---|---|---|---|---|---|---|
| | ICC | Lower Limit | Upper Limit | Statistics | *Fd1* | *Fd2* | *p* |
| Single measurements | .645 | .578 | .713 | 40.599 | 99 | 1881 | <.001 |
| Average Measurements | .973 | .965 | .980 | 40.599 | 99 | 1881 | <.001 |

The confidence interval, 95%, ranged from .578 to .713, providing support for the consistency of the result. Furthermore, the *F*-test result was statistically significant ($p$<.001), demonstrating that the observed differences in the measurements were not random and that the model was indeed significant.

Upon analysis of Table 6, the ICC value for the average measurements was calculated to be .973. This is a notably high value, indicating that the average measurements have an excellent reliability level. The confidence interval, 95%, was between .965 and .980, providing substantial evidence in support of the consistency of the measurements. The *F*-test was also statistically significant in this case ($p<.001$). Furthermore, a two-way mixed-effects model was employed in the analysis, wherein individuals were regarded as random effects and measurements as fixed effects. The definition of absolute agreement between measurements was utilized.

The ICC results demonstrated that the scale exhibited high reliability. While the reliability was moderate at the individual measurement level, it was found to be excellent at the average measurement level. The use of the scale at the group level will result in more consistent outcomes, thereby reducing measurement errors. It can be posited that the scale provides a sufficient degree of reliability in individual evaluations; however, it is imperative to exercise caution when utilizing it.

## 4. DISCUSSION and CONCLUSION

The "School Technology Leadership Scale", developed by Grace in 2020 with a thematic literature review, assesses teachers' perceptions of school principals' technology leadership skills. In this study, the 'School Technology Leadership Scale', originally developed by Grace (2020), was adapted into Turkish through an eight-stage process. The modification involved renaming the two sub-dimensions in accordance with the theory. The process was conducted in parallel with the recommendation of Grace (2020) as EFA should be employed in future studies. Grace (2020) was unable to apply EFA due to insufficient sampling and suggested that EFA be conducted for future studies. According to EFA and CFA results in this present study, the analysis without item extraction revealed the two-factor structure. Two factors of the *school technology leadership* include 1) *managing and directing applications* and 2) *guiding and supporting*. The results imply that the score derived from the scale increase, there is an increase in the perception of teachers regarding to school technological leadership of school principals. The adapted scale to the Turkish language and national school system was found to be valid and reliable for this study group.

The adapted scale is suitable to assess the technology leadership skills of school principals based on teacher perceptions. The scale reflects the identified essential leadership skills in technology by previous researchers (Demiraçan, 2019; Durnalı, 2018; Grady 2011; Grace 2020). In this context, it can be claimed that these roles are based on the roles of the school technology leader of this two-factor scale. The items in the adapted scale, the factors reflect the performance indicators in the dimensions of equity and digital citizenship advocacy, visionary planning, empowering leadership, system design and connected learner, which are the updated ISTE (2018) standards. Therefore, the school technology leader possesses characteristics such as being an advocate for equal opportunities, having a visionary approach to planning, being an empowering leader, a system designer and a continuous learner (ISTE, 2018). It can be stated that studies form the basis for the role of school principals, the components of technology leadership and technology leadership standards. Therefore, identifying and learning the levels of these skills is important for school leadership. Technology skills emerge as a determining factor in the quality of the education and management system in schools. The formation of intellectual capital based on 21st-century skills is effective. Additionally, it can be concluded that it is an important factor in meeting the needs of digital natives and digital hybrids.

In the last two years, three different scales have been developed in Türkiye (Akada & Şahin-Fırat, 2022; Durnalı, 2022; Ermiş & Somuncuoğlu-Özerbaş, 2021). Akada and Şahin-Fırat's (2022) scale focuses on ISTE 2018 standards, Durnalı's (2022) on technology leader behaviors, and Ermiş and Somuncuoğlu-Özerbaş's (2021) on technology leader self-efficacy. The study by Ermiş and Somuncuoğlu-Özerbaş (2021) is based on school principals' perceptions, while

other studies are based on middle school teachers' perceptions. All three scales have more than two factors. Unlike these scales, the adapted scale in Turkish language reflects the perceptions of primary school teachers. It is a previously validated and reliable scale used in the research. The scale can contribute to the literature on educational management, create perspectives on school technology leadership, and serve as a source for further studies. However, determining technology leadership levels can be an effective indicator for school principals to identify and develop areas for improvement, establish an innovative school culture, change society and improve the quality of education. The school principals should develop technological leadership skills, 1) to facilitate access to novel technologies in an effective teaching and learning process, 2) to support effective learning, 3) to render the school management process more practical, transparent, rational and data-driven and 4) to promote effective utilization of technology by teachers and students. Technology leadership may continue to be relevant and maintain its currency due to constantly evolving technology in the coming years and therefore, the development and adaptation process of such scales should continue. This is because technology has a dynamic structure and affects all elements of society.

The adapted scale is based on the perceptions of teachers working in official primary schools in Istanbul. Therefore, the scale can be conducted in different study groups. The limitation of this study is that the pilot application was not conducted with two language versions on a group of participants who speak both languages fluently within the scope of linguistic equivalence, the correlation between the scores was not determined, and it was not used for equivalence. Recommendations for further studies also include that the relationship between school technology leadership skills and different organizational concepts can be examined, Data and evidence-based analyses can be created on the technology leadership skills of school principals, and efforts can be made to develop the technological leadership skills of technology leaders based on the analysis results. Postgraduate education programs, leadership academies and certification programs can be created for school technology leadership. Recommendations for practitioners include that the technology leadership skill requirements of school principals can be identified. In this way, educational programs can be organized by shedding light on which skills need to be developed.

## Acknowledgments

We would like to thank Dr. Hakkı Polat for his sincere support during statistical analysis, thanks valuable faculty members for expert opinions, and thanks all teachers who voluntarily participated in the study.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Istanbul Okan University, Ethics Committee, 165, 556 April 12th, 2023. Additionally, permission was obtained from the Istanbul Provincial Directorate of National Education on June 1st, 2023, with approval number E-59090411. The research was conducted with volunteer teachers who signed informed consent forms obtained from the state schools. This study is associated with the first author's doctoral thesis titled "The Impact of School Principals' Technology Leadership Skills on Their Competency-Based Job Performance" and the second author is thesis advisor.

## Contribution of Authors

The authors contributed to the entire study. The study was carried out and reported in collaboration. This study is associated with the first author's **(Tugba Ögücü)** doctoral thesis (Ögücü, 2025) and the second author **(Aytaç Göğüş)** is thesis advisor.

## Orcid

Tugba Ögücü https://orcid.org/0000-0003-4026-6279

Aytaç Göğüş https://orcid.org/0000-0001-8215-3294

## REFERENCES


Afshari, M., Abu Bakar, K., Luan, W.S., & Siraj, S. (2012). Factors affecting the transformational leadership role of principals in implementing ICT in schools. *TOJET: The Turkish Online Journal of Educational Technology, 11*(4), 164-176. https://eric.ed.gov/?id=EJ989265

Akada, T., & Şahin Fırat, N. (2022). Development of school principles' technology leadership scale based on ISTE 2018 standards. *Western Anatolian Journal of Educational Sciences 13*(2), 1262-1289. https://doi.org/10.51460/baebd.1191300

Akbaba-Altun, S. (2002). Okul yöneticilerinin teknolojiye karşı tutumlarının incelenmesi. [Analyzing school administrators' attitudes towards technology]. *Journal of Contemporary Education, 286*, 8-14.

Akçay, B., Gelen, B., Tiryaki, A., & Benek, İ. (2018). An analysis of scale adaptation studies in science education: Meta-synthesis study. *Journal of Education in Science, Environment and Health, 4*(2), 227-245. https://doi.org/10.21891/jeseh.439150

Akın-Mart, Ö., & Tulunay-Ateş, Ö. (2021). Investigation of technological leadership of the school administrators in Turkey: A meta-analysis study. *Bartın University Journal of Faculty of Education, 10*(1), 169-186. https://doi.org/10.1016/buefad.740794

Alpar, R. (2020). *Uygulamalı istatistik ve geçerlik-güvenirlik [Applied statistics and validity and reliability]*. Detay.

A'mar, F., & Eleyan, D. (2022). Effect of principal's technology leadership on teacher's technology integration. *International Journal of Instruction, 15*(1), 781-798. https://doi.org/10.29333/iji.2022.15145a

Anderson, R., & Dexter, S. (2005). School technology leadership: An empirical investigation of prevalence and effect. *Educational Administration Quarterly, 41*, 49-82. http://dx.doi.org/10.1177/0013161X04269517

Apsorn, A., Sisan, B., & Tungkunanan, P. (2019). Information and communication technology leadership of school administrators in Thailand. *International Journal of Instruction, 12*(2), 639-650. https://doi.org/10.29333/iji.2019.12240a

Banoğlu, K. (2012). Technology leadership competencies scale for educational administrators: Development, validity and reliability study. *Inonu University Journal of the Faculty of Education 13*(3), 43-65. https://toad.halileksi.net/wp-content/uploads/2022/07/ptla-okul-mudurlerinin-teknoloji-liderligi-olcegi-toad.pdf

Brown, T.A., & Moore, M.T. (2012). Chapter: Confirmatory factor analysis. In R.H. Hoyle (Eds.) *Handbook of structural equation modelling*, (pp. 361-379). https://www.researchgate.net/publication/251573889_Hoyle_CFA_Chapter_-_Final

Bülbül, T., & Çuhadar, C. (2012). Analysis of the relationship between school administrators' perceptions of technology leadership self-efficacy and their acceptance of ICT. *Mehmet Akif Ersoy University Journal of Faculty of Education, 1*(23), 474-499.

Büyüköztürk, Ş. (2020). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum [Data analysis handbook for social sciences: Statistics, research design, SPSS applications and interpretation]*. PegemAkademi.

Chang, I.H., Chin, J.M., & Hsu, C.M. (2008). Teachers' perceptions of the dimensions and implementation of technology leadership of principals in Taiwanese elementary schools. *Educational Technology & Society, 11*(4), 229-245. https://eric.ed.gov/?id=EJ825113

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. L. NJ Eribaum.

Çokluk, Ö., Şekercioğlu, G. & Büyüköztürk, Ş. (2023). *Sosyal bilimler için çok değişkenli istatisitik Spss ve Lisrel uygulamaları [Multivariate statistical Spss and Lisrel applications for social sciences]*. PegemAkademi.

Costello, A.B., & Osborne, J.W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research & Evaluation, 10*(7), 1-9. https://doi.org/10.7275/jyj1-4868

Cristobal, E., Flavian, C., & Guinaliu, M. (2007). Perceived e-service quality (PeSQ): Measurement validation and effects on consumer satisfaction and web site loyalty. *Managing Service Quality, 17*(3), 317-340. https://doi.org/10.1108/09604520710744326

Çamur, H., & Göğüş, A. (2023) Validity and Reliability Study of School Principals' Competence-Based Job Performance Perceptions Scale. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 52*(1), 207-246. https://doi.org/10.14812/cuefd.1205795

Çapık, C. (2014). Geçerlik ve güvenirlik çalışmalarında doğrulayıcı faktör analizinin kullanımı [The use of confirmatory factor analysis in validity and reliability studies]. *Anatolian Journal of Nursing and Health Science, 17*(3), 196-205.

Demiraçan, A. (2019). *Okul yöneticilerinin teknoloji liderliği stratejileri ile yenilik yönetimi yeterlik inançları arasındaki ilişkinin incelenmesi [Investigation of the relationship between technology leadership strategies and innovation management beliefs of school administrators]* [Unpublished master's thesis]. Trakya University.

Dexter, S., & Barton, E. (2021). The development and impact of team-based school technology leadership. *Journal of Educational Administration, 59*(3), 367-384. https://doi.org/10.1108/JEA-12-2020-0260

Durnalı, M. (2018). *Öğretmenlere göre okul müdürlerinin teknolojik liderlik davranışları ve bilgi yönetimini gerçekleştirme düzeyleri [The views of teachers on technological leadership behaviours of school principals and school knowledge management]* [Unpublished doctoral dissertation]. Hacettepe University.

Durnalı, M. (2022). The development and validation of technological leadership behavior instrument for school principal. *Journal of Learning and Teaching in Digital Age, 7*(2), 210-221. https://doi.org/10.53850/joltida.1035591

Ermiş, U.F., & Somuncuoğlu-Özerbaş, D.H. (2021). The development of the self-efficacy form for school administrators' use of information and communication technologies in education. *Bartın University Journal of Faculty of Education, 10*(3), 663-680. https://doi.org/10.1016/buefad.983795

Esplin, N. (2017). *Utah elementary school principals' preparation as technology leaders.* [Unpublished doctoral dissertation]. Utah State University.

Ghavifekr, S., & Wong, S.Y. (2022). Technology leadership in Malaysian schools: The way forward to education 4.0 – ICT utilization and digital transformation. *International Journal of Asian Business and Information Management, 13*(2), 1-18. https://doi.org/10.4018/IJABIM.20220701.oa3

Gonzales, M. (2020). School technology leadership vision and challenges: Perspectives from American school administrators. *International Journal of Educational Management, 34*(4), 697-708. https://doi.org/10.1108/IJEM-02-2019-0075

Gökdemir, F., & Yılmaz, T. (2023). Processes of using, modifying, adapting and developing likert type scales. *J Nursology, 26*(2), 148-160. https://doi.org/10.5152/JANHS.2023.22260

Grace, G. (2020). *The development of a new instrument comparing teacher and principal perceptions of school technology leadership* [Doctoral dissertations]. University of Kentucky. https://uknowledge.uky.edu/edsc_etds/61

Grady, M.L. (2011). The principal as technology leader. https://www.sagepub.com/sites/default/files/upmbinaries/41065_Grady_Leading_the_Technology_Powered_School___cH1.pdf

Gümüşeli, A.İ. (2001). Çağdaş okul müdürünün liderlik alanları [Leadership areas of the contemporary school principal]. *Educational Management in Theory and Practice, 28*, 531-548.

Gyeltshen, L. (2021). Principals' technology leadership behavior and teachers' use of information and communication technology (ICT) in Bhutan. *Suranaree Journal of Social Science, 15*(2), 125-135. https://doi.org/10.55766/WIRG3818

Hacıfazlıoğlu, Ö., Karadeniz, Ş., & Dalgıç, G. (2011). Eğitim yöneticileri teknoloji liderliği öz-yeterlik ölçeğinin geçerlik ve güvenirlik çalışması [Validity and reliability study of educational administrators' technology leadership self-efficacy scale]. *Educational Administration: Theory and Practice, 17*(2), 145-166.

Hamzah, M.I.M., Juraime, F., Hamid, A.H.A, Nordin, N., & Attan, N. (2014). Technology leadership and its relationship with school-Malaysia Standard of Education Quality (School-MSEQ). *International Education Studies, 7*(13), 278-285. https://doi.org/0.5539/ies.v7n13p278

Hsieh, C.C., Yen, H.C., & Kuan, L.Y. (2014, December) *The relationship among principals' technology leadership, teaching innovation and students' academic optimism in elementary schools.* International Conferences on Educational Technologies 2014 and Sustainability, Technology and Education, New Tapei City, Taiwan. https://eric.ed.gov/?id=ED557326

ISTE (2018). *ISTE Standards for Education Leaders*. International Society for Technology in Education. https://iste.org/standards/iste-standards-for-education-leaders

Karaçam, Z. (2019). Ölçme araçlarının Türkçeye uyarlanması [Adaptation of measurement tools into Turkish]. *Journal of Midwifery and Health Sciences, 2*(1), 28-37.

Karasar, N. (2020). *Bilimsel araştırma yöntemleri Kavramlar-İlkeler-Teknikler [Scientific research methods Concepts-Principles-Techniques]*. Nobel Publ.

Kesim, E. (2020). Teknoloji liderliği [Technology Leadership]. K. Yılmaz (Eds.), *Liderlik kuram-araştırma-uygulama [Leadership theory-research-practice]* (pp. 741-758). PegemA Publ.

Kwatubana, S. (2023, June). *The role of principals as instructional leaders in the implementation of ICT curriculum in South African Public Schools.* BCES Conference Books, (21) Part 3: School Education: Policies, Innovations, Practices & Entrepreneurship. https://files.eric.ed.gov/fulltext/ED629371.pdf

MEB Strategic Plan (2024). MEB 2024-2028 Strategic Plan https://sgb.meb.gov.tr/yayinlarimiz/yayin/112

Mendoza, M.T.E., & Catiis, G.B.E. (2022). Administrators' technology leadership: Its influence on teachers' technology proficiency. *International Research Journal of Science, Technology, Education, and Management, 2*(3), 67-75. https://doi.org/10.5281/zenodo.7136432

Morgan, M.A. (2014). *Assessing school technology leadership practices: perceptions from 1-to-1 schools in the context of continuous improvement* [Unpublished doctoral dissertation]. University of Virginia.

Nababan, T.M., Zainuddin, Purba, S., Lumban Batu, J.S., & Sianipar, G. (2021). School leadership strategies in the digital era. *Advances in Social Science, Education and Humanities Research, 591*, 103-106. https://doi.org/10.2991/assehr.k.211110.068

Orçan, F., (2018). Exploratory and confirmatory factor analysis: which one to use first? *Journal of Measurement and Evaluation in Education and Psychology, 9*(4), 413-421. https://doi.org/10.21031/epod.394323

Osborne, J.W. (2014). *Best practices in Unpublished exploratory factor analysis.* https://www.researchgate.net/publication/265248967_Best_Practices_in_Exploratory_Factor_Analysis

Ögücü, T. (2025). *Okul müdürlerinin teknoloji liderlik becerilerinin yetkinliğe dayali iş performanslarina etkisi [The impact of school principals' technology leadership skills on their competency-based job performance]* [Unpublished doctoral thesis]. Istanbul Okan University.

Öztaban, A. (2020). *Okul yöneticilerinin teknoloji liderliği rollerini yerini getirme düzeyleri [Fulfilment levels of school administrators' technology leadership roles]* [Unpublished master's thesis]. Adnan Menderes University.

Rayray, M.R. (2023). Technology leadership competence of public-school heads. *Psychology and Education: A Multidisciplinary Journal, 9(9),* 1108-1123. https://ejournals.ph/article.php?id=21452

Saçak, R., Gür, Ş., & Eren, T. (2020). Türkiye'nin dijital dönüşüm yol haritasında yer alan stratejilerin topsis yöntemi ile sıralanması [Topsis method sorting of the strategies in the digital transformation road of Turkey]. *Journal of Dokuz Eylül University Faculty of Business Administration, 21*(2), 335-346. https://doi.org/10.24889/ifede.839791

Schumacker, R.E., & Lomax, R. (2004). *A beginner's guide to structural equation modelling.* Lawrence Erlbaum Associates.

Seferoğlu, S.S., & Akbıyık, C. (2005). İlköğretim öğretmenlerinin bilgisayara yönelik öz yeterlik algıları üzerine bir çalışma [A study on primary school teachers' perceived computer self-efficacy]. *Eurasian Journal of Educational Research, 19*, 89-101.

Sezer, B. (2011). *İlköğretim yöneticilerinin teknoloji liderliği rollerine ilişkin yeterlilikleri [Primary school administrators' competences on technology leadership roles]* [Unpublished master's thesis]. Ankara University.

Stuart, L.H., Mills, A.M., & Remus, U. (2009). School leaders, ICT competence and championing innovations. *Computers & Education*, *53*(3), 733-741. https://doi.org/10.1016/j.compedu.2009.04.013

Sincar, M. (2009). *İlköğretim okulu yöneticilerinin teknoloji liderliği rollerine ilişkin bir inceleme. Gaziantep ili örneği [Elementary teachers' views about school administrators' technology leadership roles]* [Unpublished doctoral dissertations]. Inonu University.

Şişman-Eren, E. (2010). *İlköğretim okul müdürlerinin eğitim teknolojilerini sağlama ve kullanmada gösterdikleri liderlik davranışları [Technological leadership behaviors of turkish elementary school principals in providing and using educational technologies]* [Unpublished doctoral dissertations]. Anadolu University.

Taherdoost, H., Sahibuddin, S. & Jalaliyoon, N. (2014) Exploratory factor analysis; Concepts and theory. *Advances in Applied and Pure Mathematics, 27*, 375-382. https://hal.science/hal-02557344

Twelfth Development Plan (2023). Twelfth Development Plan 2024-2028. https://www.sbb.gov.tr/wp-content/uploads/2024/06/Twelfth-Development-Plan_2024-2028.pdf

Topcu, İ., & Ersoy, M. (2020). Eğitim yönetiminde teknoloji kullanımına ilişkin okul yöneticilerinin görüşleri [School administrators' opinions on the use of technology in educational administration]. *International Journal of Society Researchs, 15*(10), 4930-4955. https://doi.org/10.26466/opus.652611

Torrato, J.B., Aguja, S.E., & Prudente, M.S. (2021) Examining school administrators' technology integration leadership. *International Journal of Learning and Teaching*, *7*(1), 38-42. https://doi.org/10.18178/ijlt.7.1.38-42

Weng, C.H., & Tang, Y. (2014). The relationship between technology leadership strategies and effectiveness of school administration: An empirical study. *Computers & Education, 76*, 91–10. https://doi.org/10.1016/j.compedu.2014.03.010

Wu, A., Yang, H.H., & Sha Zu, Y.S. (2015). *Development of school technology leadership: Cases in the United Kingdom and United States*. 2015 International Symposium on Educational Technology (ISET), Wuhan, China, 25-29. https://doi.org/10.1109/ISET.2015.14

International Journal of Assessment Tools in Education
2025, Vol. 12, No. 2, 476–488
https://doi.org/10.21449/ijate.1457753
journal homepage: https://dergipark.org.tr/en/pub/ijate
*Research Article*

# Assessing the reliability of the Beck Depression Inventory adapted for Turkish culture: A meta-analysis

**Yasemin Duygu Esen**[1*], **Filiz İsnaç**[2], **H. Deniz Gülleroğlu**[3]

[1]Ministry of National Education, Eskisehir, Türkiye
[2]Ministry of National Education, Mersin, Türkiye
[3]Ankara University, Faculty of Education, Department of Educational Sciences, Ankara, Türkiye

**ARTICLE HISTORY**
*Received: Mar. 28, 2024*
*Accepted: Nov. 25, 2024*

**Keywords:**
Meta-analysis,
Beck Depression Scale,
Validity,
Reliability.

**Abstract:** This study aims to determine the reliability of the Beck Depression Inventory (BDI), a widely used tool for diagnosing depression—a condition that significantly impacts individuals' lives—through meta-analysis, an advanced statistical technique. To achieve this objective, studies conducted in Turkey between 1961 and 2021 that utilized the Beck Depression Inventory were included. Relevant studies were identified through searches in the "Higher Education Council (HEC) National Thesis Center" and Google Scholar databases. The inclusion criteria focused on studies that reported validity and reliability analyses and contained the keywords “Beck Depression Inventory” or “Beck Depression Scale.” The effect size was calculated using the Cohen's d formula. The meta-analysis revealed an overall reliability coefficient of .88 for the Beck Depression Inventory, which, based on Cohen's d classification, represents a large effect size.

## 1. INTRODUCTION

Scientific progress is cumulative, with each researcher building upon the knowledge and findings of their predecessors. The systematic accumulation and replication of empirical research underpin the steady advancement of scientific knowledge. However, scientists often face limitations in storing, organizing, and synthesizing empirical findings, which can impede this progress. Consequently, repeated studies in a particular field or on a specific topic often yield varying results. To better understand the reasons for these discrepancies and to evaluate the overall findings, a comprehensive summary or review of the studies is essential. This is where meta-analysis plays a critical role. Meta-analysis, a quantitative research method, can be defined as the process of combining the results of multiple studies into a single, comprehensive outcome (Borenstein *et al*., 2009). Arthur *et al*. (2001) describe meta-analysis as a set of statistical procedures used to derive general conclusions from primary sources and synthesize the results of multiple primary studies. Similarly, Dinçer (2014) defines meta-analysis as grouping studies with similar subjects, themes, or fields under specific criteria and interpreting their quantitative findings collectively. Meta-analysis offers several advantages, including integrating findings from independent studies, enhancing the validity of results, generalizing findings from studies with small sample sizes to larger populations, assessing whether

*CONTACT: Yasemin Duygu ESEN yaseminduyguesen@gmail.com Ministry of National Education, Eskisehir, Türkiye

The copyright of the published article belongs to its author under CC BY 4.0 license. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/

**e-ISSN: 2148-7456**

variations between studies are due to chance, identifying the sources of variation, and resolving contentious issues (Glass, 1976; Hedges & Olkin, 1985).

Meta-analysis, widely utilized in the field of health, has increasingly gained significance in the field of education (Demiray, 2013; McDermott & Ebmeiern, 2009). It can be applied to various types of quantitative studies, including controlled clinical trials, quasi-experimental studies, and observational studies. While most meta-analyses focus on studies that aim to explain causal relationships, they can also be conducted on non-causal survey studies as well as on validity and reliability studies of diagnostic methods (Abramson, 1994).

Depression is a significant illness that profoundly impacts an individual's quality of life, engagement with life, and both professional and family dynamics (Belsher & Costello, 1988; Burcusa & Iacono, 2007; Monroe & Harkness, 2012; Solomon, 2000; Zis & Goodwin, 1979). If left undiagnosed and untreated, it can cause direct or indirect harm, including increasing the likelihood of alcohol and drug use and acting as a trigger for various medical conditions. The prevalence of suicide due to depression continues to rise, making it one of the leading causes of death. While depression is treatable, accurate diagnosis is a crucial first step. This relies on the validity and reliability of the measurement tools used for assessment. The Beck Depression Inventory (BDI) is a self-report scale developed by Beck in 1961, widely utilized for diagnosing depression and comprehensively evaluating the severity of depressive symptoms. According to the American Psychiatric Association (2000), the BDI is officially recognized for determining the initial diagnostic level of depression. It is one of the most frequently used tools in clinical studies as well as in health and education research due to several advantages. These include its applicability to both healthy individuals and psychiatric patients, its ease of use without requiring special training, its concise format, straightforward scoring and interpretation, and its well-established validity and reliability.

Due to its numerous advantages, the Beck Depression Inventory (BDI) has been adapted for use in various cultures and is widely utilized globally. The literature includes numerous studies examining the psychometric properties of the scale across diverse cultural contexts and populations, consistently demonstrating its high reliability (García-Batista *et al*., 2018; Hamidi *et al*., 2015). For instance, Wiebe *et al*. (2005) conducted a comprehensive analysis comparing the original English version of the BDI with its Spanish adaptation. The original version of the scale reported a Cronbach's alpha coefficient of .89, while the Spanish adaptation demonstrated a slightly higher coefficient of .91, reflecting high reliability in both versions. Additionally, the test-retest reliability method confirmed the scale's reliability at an acceptable level.

In their extensive review, Wang and Gorenstein (2013) analyzed approximately 118 articles that assessed the BDI's psychometric properties. Their findings revealed that the Cronbach's alpha coefficient reported in these studies ranged from .84 to .94, while test-retest reliability values varied from .73 to .96. This study concluded that the BDI is a reliable tool for diagnosing depression. A broader review of the international literature further supports these findings, indicating that the scale's reliability values are consistently high and comparable across different cultural adaptations. The Beck Depression Inventory (BDI) was adapted to Turkish culture by Şahin in 1988, with its reliability initially evaluated by Hisli using various methods. The split-half reliability coefficient was reported as .74, and the Cronbach's alpha coefficient was determined to be .80. However, reliability values for the BDI in studies conducted within Turkish samples vary widely, ranging from .53 to .98. For example, Orhan (2019) reported a reliability coefficient of .88, while Alver and Uğuryol (2016) noted a reliability value of .98. This considerable variation in reliability values, differing significantly from those reported in other cultures, forms the basis of this research.

This variability may be attributed to factors such as differences in sample characteristics and implementation conditions. It highlights the need to synthesize findings from various studies to derive a more generalizable reliability estimate for the Turkish adaptation of the BDI. This study aims to address this gap by providing a more precise and comprehensive understanding

of the scale's reliability. It is anticipated that the findings will not only contribute to a deeper understanding of the BDI's psychometric properties in Turkish contexts but also enable more accurate interpretation of studies utilizing this scale. A review of the literature reveals that while numerous meta-analyses focus on group compari-son types, such as transactional effectiveness and group differences (Gözübebek, 2012; Kablan *et al*., 2013; Şahin & Tekdal, 2005; Temel *et al*., 2020; Ulubey & Toraman, 2015), there is only one correlation meta-analysis study addressing the validity and reliability of diagnostic methods. This study, conducted by Eser and Aksu (2021), aimed to investigate the reliability generalization of the Beck Depression Inventory-II (BDI-II) adapted to Turkish culture.

Eser and Aksu included only English-language articles published between 2011 and 2019 in their meta-analysis. Using the Varying Coefficient Meta-Analytical Method (VC), they analyzed 40 studies and reported an average reliability level of .898 (Cronbach's alpha). However, their study had several limitations, including the exclusion of Turkish-language publications, a relatively small sample size, and a restricted time frame of 2011-2019. No meta-analyses have yet been conducted that address these limitations by examining a broader time range, incorporating a larger sample size, and including Turkish-language publications. This study aims to fill this gap in the literature, providing a more comprehensive and representative analysis of the reliability of the Beck Depression Inventory-II in Turkish contexts.

Based on this context, the primary aim of this study is to determine the reliability of the Beck Depression Scale, a widely used tool for diagnosing depression, a condition that significantly impacts an individual's life. The study employs meta-analysis, an advanced statistical technique, to achieve this objective. To address this overarching aim, the following research question was formulated: *"What is the reliability coefficient obtained by collectively analyzing the samples from studies that utilized the Beck Depression Scale?"*

## 2. METHOD

### 2.1. Research Model

This study is a quantitative research project designed using the meta-analysis method. Meta-analysis involves the synthesis and interpretation of the results from multiple studies on a specific subject by statistically combining their findings (Atalmış & Köse, 2018; Borenstein *et al*., 2009; Cooper *et al*., 2009). Field (2001) describes meta-analysis as a statistical technique based on the findings derived from the results of independent studies.

The following stages were followed in conducting the meta-analysis for this research (Borenstein *et al*., 2009; Hedges & Olkin, 1985; Higgins & Green, 2011; Higgins & Thompson, 2002; Sutton & Higgins, 2008):

1. Reviewing the literature
2. Determining the inclusion criteria
3. Coding
4. Calculating the effect size
5. Performing the heterogeneity test
6. Determining the model according to the heterogeneity test result
7. Calculating the Overall Effect Size
8. Interpreting the results

### 2.2. Data Collection Process

#### 2.2.1. *Research sample*

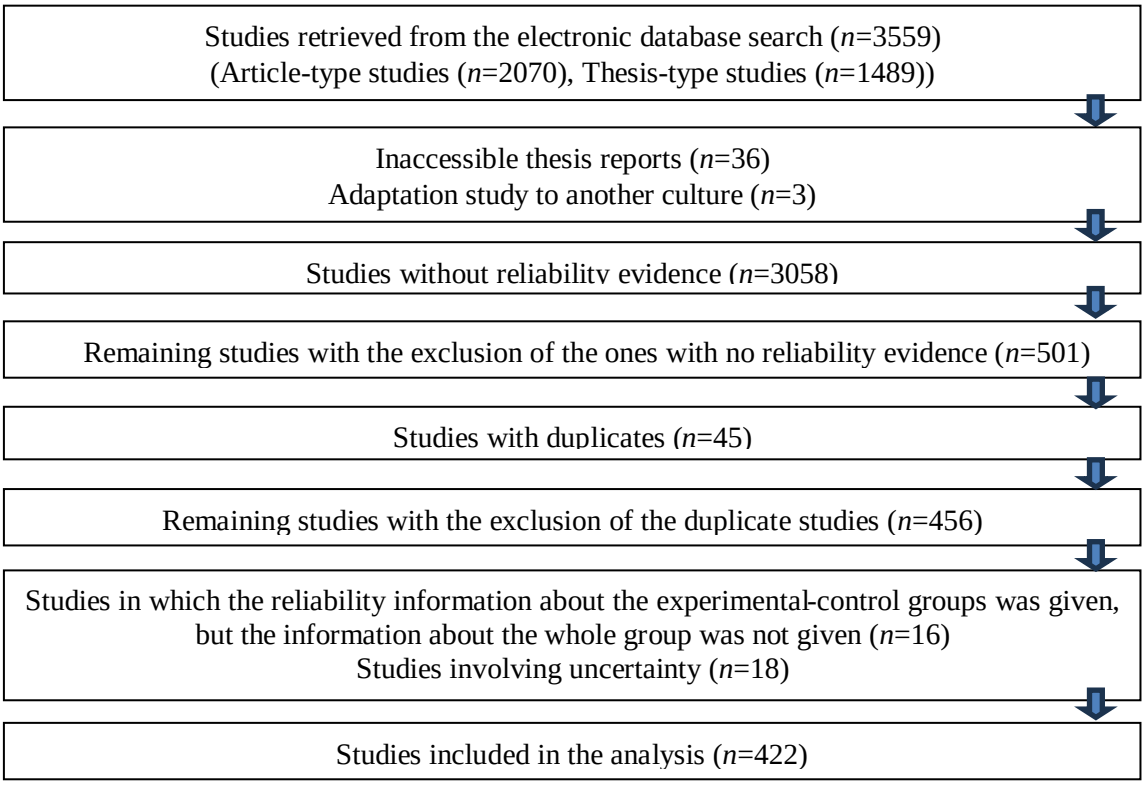
Studies utilizing the Beck Depression Scale were identified through a systematic search of the ProQuest (2021), HEC National Thesis Center (2021), and Google Scholar (2021) databases. The initial search was conducted on June 19, 2021, followed by a second search on July 31, 2021, to ensure completeness and accuracy. The keywords *"Beck Depression Inventory"* and *"Beck Depression Scale"* were used during the screening process.

In meta-analysis research, it is essential to establish clear inclusion criteria for the studies prior to conducting the analysis (Glass, 1976). For this study, the following criteria were applied to determine which studies would be included:

- Having been conducted in Turkey between 1961 and 2021
- Being a Master's / Doctoral thesis or an article published in peer-reviewed journals
- Having reported the reliability coefficient (Cronbach alpha) regarding the Beck Depression Scale
- Having open access permission of the full text of the studies

A total of 3,559 studies were retrieved from the electronic database search, with 41% sourced from the National Thesis Database and 59% from Google Scholar. Of these, 2,070 articles and 1,489 theses were assessed based on the predetermined criteria. Ultimately, 422 studies that reported the Cronbach's alpha coefficient as the internal consistency reliability estimation method were included in the analysis. The process of selecting these studies is depicted in the flowchart provided in Figure 1.

**Figure 1.** *Flowchart of data collection process.*




### 2.2.2. *Data coding*

To facilitate accurate data examination and minimize errors due to oversight, the researchers developed a coding form. This form included information such as the study's title, publication year, author details, and psychometric properties. To validate the coding form, consultations were held with two measurement experts. Based on their feedback, the form was refined, and its inter-coder reliability was evaluated using Kappa statistics.

The Kappa statistic assesses the degree to which the observed agreement between independent coders exceeds the agreement expected by chance (Sim & Wright, 2005). For this evaluation, 40 studies were randomly selected and independently coded by two researchers. The calculated Kappa value was .91, which, according to Landis and Koch (1977), indicates "perfect" agreement. These findings confirm that the coding form demonstrates high validity and reliability.

### 2.3. Data Analysis

Microsoft Excel was used for data preparation, while the Comprehensive Meta-Analysis (CMA) program was utilized for data analysis. The effect sizes for all studies included in the analysis were calculated, and the "average effect size" was derived from these calculations. Subsequently, a heterogeneity test (Q statistic) was conducted on the effect sizes to determine whether the Fixed Effects Model (FEM) or the Random Effects Model (REM) would be more appropriate for the analysis.

The Fixed Effects Model assumes that the effect size is equal and constant across all studies included in the analysis. In this model, variability arises solely from sampling error. Conversely, the Random Effects Model allows for the possibility that the effect sizes of individual studies may differ (Borenstein *et al.*, 2009). This model considers two sources of variability: sampling error at the individual level and random effects variance (Borenstein *et al.*, 2009; Cooper *et al.*, 2009; Şen & Yıldırım, 2020; Üstün & Eryılmaz, 2014). Card (2011) notes that results obtained using the Random Effects Model are generally more representative and widely applicable than those derived from the Fixed Effects Model.

In this study, bias analyses were performed following the selection of the appropriate model. The analyses included Fail-Safe N, Duval and Tweedie's Trim-and-Fill Method, and Egger's Regression Test. Funnel and forest plots were also examined to supplement and provide additional context to the test results. Effect sizes were interpreted using the criteria outlined by Thalheimer and Cook (2002), as shown in Table 1.

**Table 1.** *Interpretation of Cohen's d effect sizes.*

| Cohen's *d* Benchmarks | Interpretation of Effect Size |
|---|---|
| [ (-0.15) – 0.15) | Insignificant effect |
| [ 0.15 – 0.40 ) | Small effect |
| [ 0.40 – 0.75 ) | Moderate effect |
| [ 0.75 – 1.10) | Large effect |
| [1.10 – 1.45 ) | Extremely large effect |
| [1.45 - ) | Strong effect |

The statistic used to calculate effect size was Cohen's *d*, chosen for its computational simplicity, the availability of standardized reference ranges for interpretation, and the convenience these standards provide (Cohen, 1988). Additionally, Cohen's *d* was selected because the study involved a large sample size. Cohen's d is calculated by dividing the difference between group means by the pooled standard deviation.

## 3. FINDINGS

Meta-Analysis findings regarding the reliability level of Beck Depression Scale are presented below.

### 3.1. Heterogeneity Test

Before conducting the data analysis, the homogeneity or heterogeneity of the effect sizes from the included articles and theses was assessed using the Q-Test. The *Q* statistic is a widely used metric in meta-analysis to determine whether the variance between studies is significantly greater than what would be expected from random error alone. It evaluates the extent to which observed effect sizes deviate from the average effect size.

The *Q* statistic is interpreted alongside its degrees of freedom (*df*) and *p* value. If the *Q* value is high and the *p* value is low (typically $p < .05$), it indicates significant heterogeneity among the studies. However, it is important to note that a high *Q* value does not always confirm heterogeneity, as it can also result from small sample sizes. Therefore, it is recommended to

interpret the *Q* statistic in conjunction with other measures, such as the $I^2$ statistic, to provide a more comprehensive understanding of heterogeneity. The presence of heterogeneity influences the choice of the statistical model. If significant heterogeneity is detected, the Random Effects Model is preferred, as it accounts for variability both within and between studies. Conversely, if heterogeneity is low and all studies are assumed to represent the same effect size, the Fixed Effects Model may be appropriate (Borenstein *et al*., 2009; Higgins & Thompson, 2002).

In this study, the results of the Q-Test indicated significant heterogeneity among the effect sizes, leading to the selection of the Random Effects Model for the analysis. The detailed results of the heterogeneity test are presented in Table 2.

**Table 2.** *Findings regarding effect sizes (Pearson r) and heterogeneity test.*

| Model | *k* | *SE* | *Z* | *p* | %95 CI | | *Q* | *df*(*Q*) | *p* |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Limit | Upper Limit | | | |
| Fixed | 422 | 0.003 | 452.01 | .00 | 1.37 | 1.38 | 4511.51 | 421 | <.001 |
| Random | 422 | 0.010 | 134.12 | .00 | 1.37 | 1.41 | | | |
| Overall Effect Size | 0.88 | | | | | | | | |

Upon examining Table 2, the number of studies included in the analysis (*k*) is 422. The standard error (*SE*) of the effect size was calculated as 0.003 for the Fixed Effects Model and 0.010 for the Random Effects Model. The *Z* statistic was 452.01 for the Fixed Effects Model and 134.12 for the Random Effects Model. The heterogeneity test conducted to determine the appropriate model for calculating the overall effect yielded a *Q* value of 4511.51 ($p < .05$). Given this *Q* value, which was significant at 421 degrees of freedom (*df*) and a 95% confidence level, it was concluded that the studies exhibited significant heterogeneity. Consequently, the Random Effects Model was selected for the analysis. Using the Random Effects Model, the overall effect size was calculated as 0.88. According to Cohen's (1988) classification, this represents a large effect size. Therefore, the reliability coefficient for all studies included in the meta-analysis was determined to be 0.88, reflecting a high level of reliability for the Beck Depression Scale.

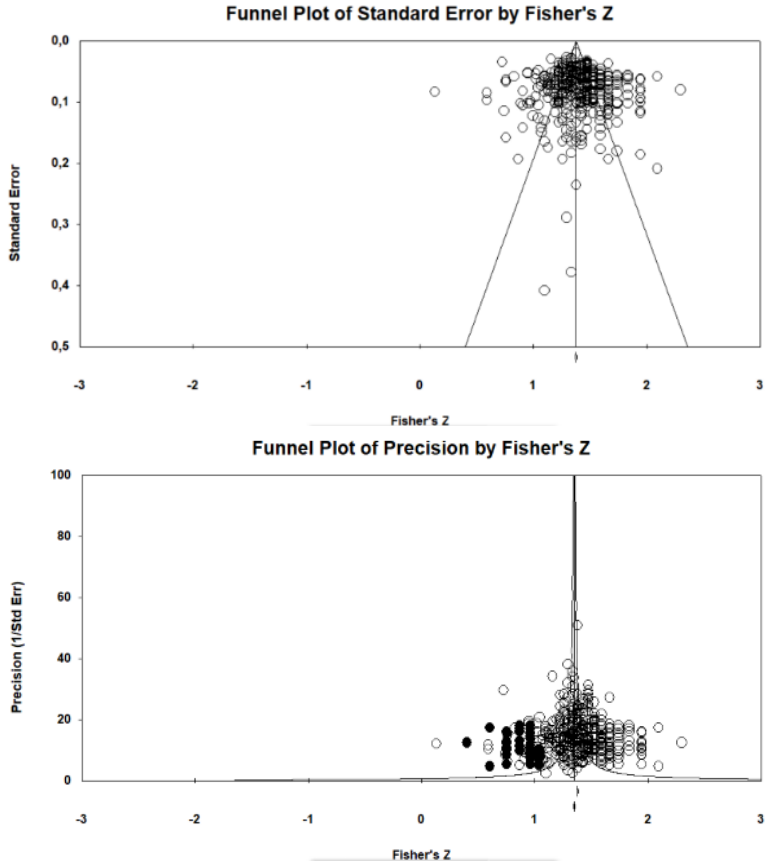
### 3.2. Publication Bias

Publication bias occurs when the studies included in a meta-analysis fail to represent the entirety of available studies in the literature (Rothstein *et al*., 2005). This bias is especially likely if the analysis exclusively includes studies with "statistically significant" findings while excluding those reporting nonsignificant results (Borenstein *et al*., 2009). To evaluate publication bias in this study, Funnel Scatterplots were analyzed alongside statistical tests such as Duval and Tweedie's Trim-and-Fill Test, Fail-Safe N, and Egger's Regression Test. Duval and Tweedie's Trim-and-Fill Test examines the symmetry of the study distribution in the meta-analysis. Asymmetric distributions may arise due to unpublished studies. This method corrects for bias by first "trimming" the studies that create asymmetry and then "filling" the gaps by imputing missing studies to achieve symmetry. The adjusted results provide a revised estimate of the effect size, accounting for the potential impact of unpublished studies (Duval & Tweedie, 2000).

Fail-Safe N assesses the robustness of meta-analysis findings by calculating the number of null-effect studies required to negate the significant results. Rosenthal's Fail-Safe N focuses on *p* values, determining the minimum number of additional studies needed to render the findings nonsignificant. Orwin's Fail-Safe N focuses on effect size, estimating the number of additional studies needed to reduce the observed effect size below a specified threshold. Higher Fail-Safe N values suggest that the meta-analysis results are more resistant to publication bias (Orwin, 1983; Rosenthal, 1979).

Egger's Regression Test uses a regression approach to evaluate the symmetry of effect sizes against standard errors. In a well-balanced meta-analysis, effect sizes are expected to form a symmetric funnel-shaped distribution. If the studies cluster asymmetrically in the Funnel Plot, it suggests the presence of publication bias (Egger *et al.*, 1997). The Funnel Scatterplots illustrating the distribution of effect sizes and indicating whether publication bias exists are presented in Figure 2.

**Figure 2.** *Funnel scatterplots for effect sizes.*




Upon examining Figure 2, the data in the Funnel Plot appear to exhibit a symmetrical distribution. According to Sterne *et al*. (2011), if the effect sizes, represented as dots for each study, are symmetrically distributed around the vertical line in the Funnel Plot, it suggests the absence of publication bias. In this case, the symmetrical distribution observed in Figure 2 supports the conclusion that publication bias is unlikely.

Furthermore, the Fail-Safe N value was calculated as 7,328, indicating that at least 7,328 null-effect studies would need to be added to invalidate the significant findings of this meta-analysis. This robust result provides additional evidence against the presence of publication bias. However, symmetrical distribution in the Funnel Plot and a high Fail-Safe N value alone may not provide conclusive evidence. To strengthen the findings, Duval and Tweedie's Trim-and-Fill Test results are presented in Table 3, offering further analysis to confirm the absence of publication bias.

**Table 3.** *Test results regarding publication bias.*

| Variable | Duval and Tweedie's Trim-and-Fill Method | | Egger's Regression Test (p) | | |
|---|---|---|---|---|---|
| | Trimmed Study | Observed / Corrected | Deg. Free. | *t* | *p* |
| Studies | 43 | $1.39_{(1.36 - 1.41)}$ / $1.34_{(1.32 - 1.36)}$ | 42 | 1.44 | .15 |

According to Duval and Tweedie's Trim-and-Fill Method, presented in Table 3, the number of studies that need to be trimmed to correct asymmetry in the Funnel Plot is determined, and the

effect size is recalculated after these studies are removed. If the difference between the observed and corrected effect sizes is not statistically significant, it provides evidence against the presence of publication bias (Pamuk *et al.*, 2015). In this study, the analysis revealed that 43 studies would need to be trimmed to achieve symmetry. However, the difference between the observed and corrected effect sizes was not statistically significant ($p = .15$; $p > 0.05$), indicating the absence of publication bias. Similarly, the results of Egger's Regression Test further support this conclusion, with a $p$ value of .15 ($p > .05$). These findings collectively confirm that this study is free from publication bias, ensuring the robustness and reliability of its meta-analytic results.

### 3.3. Distribution of Confidence Intervals for Effect Sizes

The distribution of confidence intervals and study weights for the effect sizes of the research reports included in this analysis is presented in Table 4. Given the impracticality of listing all 422 studies in a single table, the results for a selection of studies are provided as representative samples.

**Table 4.** *Confidence intervals and study weights of effect sizes of studies included in the analysis.*

| Study Name | Publication Date | Effect Size | Sd. Error | Var. | Confidence Intervals | | *Z* | *p* |
|---|---|---|---|---|---|---|---|---|
| | | | | | Lower Limit | Upper Limit | | |
| Ağca, Z. | 2017 | 1.528 | 0.098 | 0.010 | 1.335 | 1.720 | 15.578 | <.001 |
| Ağdemir, B. | 2017 | 1.293 | 0.071 | 0.005 | 1.155 | 1.432 | 18.336 | <.001 |
| Ahçıoğlu, A. | 2020 | 1.157 | 0.061 | 0.004 | 1.038 | 1.276 | 19.079 | <.001 |
| Ak, G. | 2001 | 1.333 | 0.089 | 0.008 | 1.159 | 1.507 | 15.023 | <.001 |
| Aka, B. T. | 2011 | 1.333 | 0.044 | 0.002 | 1.248 | 1.418 | 30.603 | <.001 |

The effect sizes and confidence intervals for each study are presented in Table 4, Figure 3, and Figure 4. The forest plots in Figure 3 display squares with horizontal lines extending on both sides, representing the confidence intervals. The horizontal position of each square corresponds to the effect size of the respective study. Larger squares indicate larger sample sizes, reflecting a greater contribution of that study to the overall average effect size. The horizontal lines represent the standard error, with longer lines indicating higher standard error values (Sen & Yıldırım, 2020). In Figure 3 and Figure 4, it is evident that most studies, except for those conducted by Yıldırım (2010), Yılmaz (2017) and Yılmaz (2019) exhibit relatively short lines, indicating low standard errors and high contributions to the overall analysis. Since study weight is closely related to sample size, it can also be inferred that the weights of these studies are similar.

**Figure 3.** *Forest plot regarding confidence intervals and overall effect.*

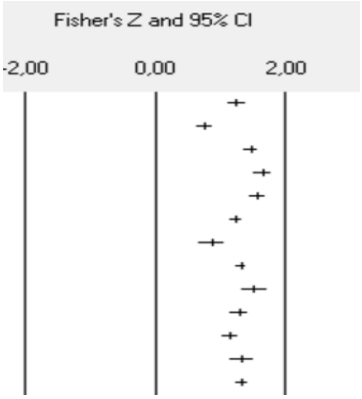
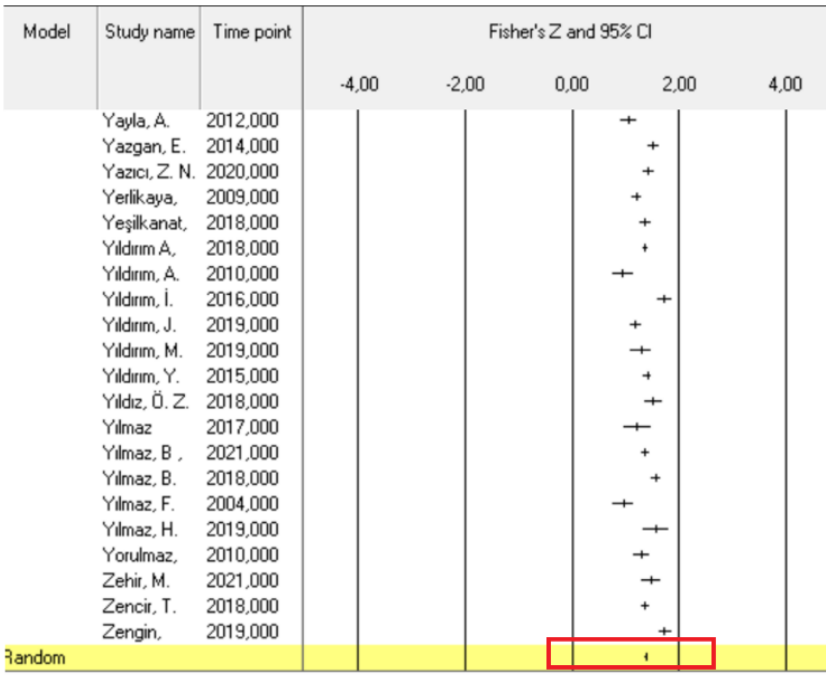


**Figure 4.** *Forest chart regarding studies.*




The "Random" line at the bottom of Figure 4 represents the overall meta-analysis result. Upon examining this line, it is evident that the confidence interval is quite narrow, suggesting high estimation precision of the effect sizes. Additionally, the average effect size is substantial, as indicated by the position of the line at the center.

## 4. DISCUSSION and CONCLUSION

This study aimed to derive a generalizable reliability coefficient for the Beck Depression Scale by synthesizing the reliability coefficients reported in primary studies conducted with the scale. Using meta-analysis, an advanced statistical method, the quantitative findings of similar studies were combined, enabling the generalization of the results to the population. Moreover, the meta-analysis approach mitigates the potential disadvantages arising from variations in sample sizes and implementation conditions across individual studies.

The analysis revealed that when the samples of the studies included in the meta-analysis were considered collectively, the reliability coefficient for the Beck Depression Scale was .88. This value demonstrates a very high level of reliability for the scale. These findings align with existing literature (Beck *et al.*, 1996; Steer & Clark, 1997). For instance, Dozois and Covin (2004) reported an average reliability coefficient of .91 in their analysis of 13 studies, while Eser and Aksu (2021), using the VC meta-analysis model with 40 articles, calculated a Cronbach's alpha value of .898. These results corroborate the findings of this study, further confirming the reliability of the Beck Depression Scale.

Regarding the time frame covered, Dozois and Covin (2004) analyzed studies published since 1996, while Eser and Aksu (2021) focused on studies published from 2011 onward. In contrast, this research spans a broader period, covering studies conducted between 1961 and 2021, thereby encompassing the time frames addressed in both of these prior studies. To assess potential bias in the study's results, Duval and Tweedie's Trim-and-Fill Test, Fail-Safe N, Egger's Regression Test, and funnel scatterplots were analyzed. These analyses concluded that no bias was present. With the exception of three studies - Yıldırım and Çevik (2010), Yılmaz (2017), and Yılmaz (2019) the weights of the studies included in the meta-analysis were found to be very similar. Additionally, the confidence interval observed in the forest plot for study weights was narrow, indicating high precision in the estimation of effect sizes.

The reliability coefficient calculated in this research was $r$ = .88. This coefficient matched the reliability reported in 34 studies and was comparable to the findings of 149 studies that estimated reliability coefficients between .86 and .90 (Bozyel, 2017; Dağ, 2012; Kahveci, 2020; Keleş, 2019; Sevilmiş, 2019; Turan, 2020). However, this result differs from 11 studies that reported reliability coefficients below .70, such as those by Orhan (2019), Alan (2015), Türkmen (2019), and Abdulkerim (2019).

This discrepancy may not be attributable to the measurement tool itself but rather to differences in the characteristics of the samples or the conditions under which the scale was administered. These factors could significantly influence the reliability outcomes observed in different studies. This study focused on the Beck Depression Scale, a widely used tool for diagnosing depression. However, other psychological assessment tools were not considered. It is recommended that future research employ the meta-analysis method to assess and verify the reliability of other commonly used psychological diagnostic tools. While this study determined the reliability of the Beck Depression Scale using the meta-analysis method, it did not examine its validity. Future research could address this gap by conducting a meta-analysis that includes all available validity studies of the Beck Depression Scale. Additionally, this study employed the internal consistency reliability estimation method, using the Cronbach's alpha coefficient as a criterion for including studies. Future studies might consider evaluating the reliability of the Beck Depression Scale using other reliability estimation methods through meta-analysis.

It should also be noted that some studies remain unpublished because they yield nonsignificant results, a phenomenon referred to as the “file drawer problem” (Rosenthal, 1979). This issue raises the possibility that the studies included in this research may not represent all relevant studies in the literature. To address this, future research on the reliability of the Beck Depression Scale should strive to include and publish nonsignificant results to ensure a more comprehensive and unbiased representation of the available data.

## Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Ankara University Social Sciences Ethics Committee, 27/06/2022-177.

## Contribution of Authors

All authors have equally contributed to all section of this study. The authors read and approved the final manuscript.

## Orcid

Yasemin Duygu Esen https://orcid.org/0000-0002-9300-0389
Filiz İsnaç https://orcid.org/0000-0002-2578-912X
H. Deniz Gülleroğlu https://orcid.org/0000-0001-6995-8223

## REFERENCES

Abdulkerim, Ş. (2019). *Determination of depression and anxiety symptom levels and related factors in individuals with chronic obstructive pulmonary disease* [Unpublished master’s thesis]. Gazi University.

Abramson, J.H. (1994). *Making sense of data* (2nd ed.). Oxford University Pres.

Alan, H. (2015). *Effect of progressive muscle relaxation exercises and reflexology applied during chemotherapy treatment on pain, fatigue, anxiety, depression and gynecologic cancer patients' quality of life in randomized control study* [Unpublished master’s thesis]. Hacettepe University.

American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (Revised 4th ed.). American Psychiatric Association.

Arthur Jr, W., Bennett, W., & Huffcutt, A.I. (2001). *Conducting meta-analysis using SAS*. Psychology Press.

Atalmış, E.H., & Köse, A. (2018). Turkish prospective teachers' attitudes towards the teaching profession: a meta-analysis study. *Journal of Measurement and Evaluation in Education and Psychology*, *9*(4), 393-413. https://doi.org/10.21031/epod.410287

Beck, A.T., Steer, R.A., & Brown, G.K. (1996). *Manual for the Beck Depression Inventory-II*. Psychological Corporation.

Belsher, G., & Costello, C.G. (1988). Relapse after recovery from unipolar depression: a critical review. *Psychological Bulletin 104*, 84–96.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H.R. (2009). *Introduction to meta-analysis*. Wiley.

Bozyel, G. (2017). *An examination of the mediating role of metacognition in the relationship between childhood trauma and psychological symptoms among undergraduate students* [Unpublished master's thesis]. Dokuz Eylül University.

Burcusa S.L., & Iacono W.G. (2007). Risk for recurrence in depression. *Clinical Psychology Review 27*, 959–985.

Card, N.A. (2011). *Applied meta-analysis for social science research*. The Guilford Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.

Cooper, H., Hedges, L.V., & Valentine, J.C. (2009). *The handbook of research synthesis and meta-analysis* (2nd ed.). Russell Sage.

Dağ, B. (2012). *Sleep behaviour of adolescents and the factors affecting it* [Unpublished master's thesis]. İstanbul University.

Demiray, P. (2013). *The effect of project based learning method: A meta analytic study* [Unpublished master's thesis]. Ankara University.

Dinçer, S. (2014). *Eğitim bilimlerinde uygulamalı meta analiz*. Pegem Akademi.

Dozois, D.J.A., & Covin, R. (2004). The Beck Depression Inventory-II (BDI-II), Beck Hopelessness Scale (BHS), and Beck Scale for Suicide Ideation (BSS). In M.J. Hilsenroth & D.L. Segal (Eds.), *Comprehensive handbook of psychological assessment* (Vol. 2, pp. 50–69). John Wiley & Sons Inc.

Duval, S., & Tweedie, R. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*(449), 89-98. https://doi.org/10.1080/01621459.2000.10473905

Egger, M., Smith, G.D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*(7109), 629-634. https://doi.org/10.1136/bmj.315.7109.629

Eser, M.T., & Aksu, G. (2021). Beck Depression Inventory-II: A study for meta-analytical reliability generalization. *Pegem Journal of Education and Instruction, 11*(3), 88-101. https://doi.org/10.14527/pegegog.2021.013

Field, A.P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed and random-effects methods. *Psychological Methods*, *6*(2), 161.

García-Batista, Z.E., Guerra-Peña, K., Cano-Vindel, A., Herrera-Martínez, S.X., & Medrano, L.A. (2018). Validity and reliability of the Beck Depression Inventory (BDI-II) in general and hospital population of Dominican Republic. *PloS one*, *13*(6). https://doi.org/10.1371/journal.pone.0199750

Glass, G.V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3-8.

Gözüyeşil, E. (2012). *The effect of brain based learning on academic achievement: A meta analytic study* [Unpublished master's thesis]. Niğde University.

Hamidi, R., Fekrizadeh, Z., Azadbakht, M., Garmaroudi, G., Taheri Tanjani, P., Fathizadeh, S., & Ghisvandi, E. (2015). Validity and reliability Beck Depression Inventory-II among the

Iranian elderly population. *Journal of Sabzevar University of Medical Sciences, 22*(1), 189-198. https://doi.org/10.21859/jsums-2201189

Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Academic Press.

Higgins, J.P.T., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions.* Wiley-Blackwell. https://doi.org/10.1002/9780470712184

Higgins, J.P.T., & Thompson, S.G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539-1558.

Kablan, Z., Topan, B., & Erkan, B. (2013). The effectiveness level of material use in classroom instruction: A meta-analysis study. *Educational Sciences: Theory & Practice, 13*(3), 1629-1644. https://doi.org/10.12738/estp.2013.3.1692

Kahveci, M. (2020). *Sexual function status, depression and marital adjustment in women in the climacteric period* [Unpublished master's thesis]. İzmir Katip Çelebi University]

Keleş Gözütok, Ç. (2019). *The effect of anxıety, depression and fear of childbirth on labor results in last trımester pregnancy* [Unpublished master's thesis]. İstanbul Medipol University.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. https://doi.org/10.2307/2529310

McDermott, L.M., & Ebmeier, K.P. (2009). A meta-analysis of depression severity and cognitive function. *Journal of Affective Disorders, 119*(1-3), 1-8. https://doi.org/10.1016/j.jad.2009.04.022

Monroe, S.M., & Harkness, K.L. (2012). Is depression a chronic mental illness? *Psychological Medicine, 42*(5), 899-902. https://doi.org/10.1017/S0033291711002066

Orhan, Ç. (2019). *Evaluation of depression and quality of life of women with breast cancer.* [Unpublished master's thesis]. Trakya University.

Orwin, R.G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*(2), 157-159.

Rosenthal, R. (1979). The 'file-drawer problem' and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641.

Rothstein, H.R., Sutton, A.J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments.* Wiley.

Steer, R.A., & Clark, D.A. (1997). Psychometric characteristics of the Beck Depression Inventory-II with college students. *Measurement and Evaluation in Counseling and Development, 30*(3), 128–136.

Şahin, M.C., & Tekdal, M. (2005). İnternet tabanlı uzaktan eğitimin etkililiği: Bir meta-analiz çalışması [The Effectiveness of Internet-Based Distance Education: A Meta-Analysis Study]. *Proceedings of the Academic Informatics 2005 Conference, 2*(4), 467–474.

Şen, S., & Yıldırım, İ. (2020). *CMA ile meta-analiz uygulamaları.* Anı Yayıncılık.

Sevilmiş, R. (2019). *The relationship between internet addiction and depression among university students* [Unpublished master's thesis]. Üsküdar University.

Sim, J., & Wright, C.C. (2005). The Kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*(3), 257-268. https://doi.org/10.1093/ptj/85.3.257

Solomon, D.A., Keller, M.B., Leon, A.C., Mueller, T.I., Lavori, P.W., Shea, M.T., … Endicott, J. (2000). Multiple recurrences of major depressive disorder. *American Journal of Psychiatry, 157*, 229–233. https://doi.org/10.1176/appi.ajp.157.2.229

Sutton, A.J., & Higgins, J.P.T. (2008). Recent developments in meta-analysis. *Statistics in Medicine, 27*(20), 2077-2090. https://doi.org/10.1002/sim.3108

Temel, S., Yildiz, T., & Aslan, F.E. (2020). The effect of marital status on burnout levels of nurses: A meta-analysis study. *Journal of Clinical Medicine of Kazakhstan, 4*(58), 51-56. https://doi.org/10.23950/jcmk/vol04iss58/a9

Thalheimer, W., & Cook, S. (2002). How to calculate effect sizes from published research articles: A simplified methodology. *Educational and Psychological Measurement, 62*(3), 503–514. https://doi.org/10.1177/0013164402042001

Turan, Y. (2020). *The effect of functional dependence of stroke patients on depression and quality of life* [Unpublished master's thesis]. Tekirdağ Namık Kemal University.

Türkmen, E. (2019). *Examination of the relationship between personality traits, anger expression styles and depressive symptoms in university* [Unpublished master's thesis]. Maltepe University.

Ulubey, Ö., & Toraman, Ç. (2015). The effect of creative drama on students' academic achievement: A meta-analysis study. *Mustafa Kemal University Journal of Social Sciences Institute, 12*(32), 195-220. https://doi.org/10.16992/mku.209027

Wang, Y.P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: a comprehensive review. *Brazilian Journal of Psychiatry, 35*, 416-431. https://doi.org/10.1590/1516-4446-2012-1048

Wiebe, J.S., & Penley, J.A. (2005). A psychometric comparison of the Beck Depression Inventory-II in English and Spanish. *Psychological assessment, 17*(4), 481. https://doi.org/10.1037/1040-3590.17.4.481

Yıldırım, A., & Çevik, B. (2010). Orem's self-care deficit theory-based assessment on the nursing interventions for a diabetic foot case. *Journal of Continuing Medical Education, 25*(5), 205-212. https://doi.org/10.23950/jcmk/vol25iss5/a9

Yılmaz, B.T. (2017). *Effects of aerobic exercises on the life quality, self-esteem and depression levels of the individuals with depression.* [Doctoral Dissertation, Cumhuriyet University].

Yılmaz, H. (2019). *Care burden, depression and related factors of caregivers with bipolar disorder* [Unpublished master's thesis]. Aydın Adnan Menderes University.

Zis, A.P., & Goodwin, F.K. (1979). Major affective disorder as a recurrent illness: A critical review. *Archives of General Psychiatry, 36*(3), 293-303.