International Journal of Assessment Tools in Education

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).

hosted by
Turkish **JournalPark**
ACADEMIC

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)

• Education Resources Information Center (ERIC)

• TR Index (ULAKBIM),

• EBSCOhost,

• SOBIAD,

• JournalTOCs,

• MIAR (Information Matrix for Analysis of the Journals),

• idealonline,

• CrossRef,

• ResearchBib,

• Index Copernicus International

CONTENTS

Research Article

Student engagement with a novel online assessment strategy

Sally E. Jordan [1*], **John P.R. Bolton** [1]

[1]School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, United Kingdom

Abstract: The study investigated the impact on student engagement and achievement of a "formative thresholded" continuous assessment strategy in which students had to meet a modest threshold, but their continuous assessment marks did not contribute to their final grade. Students were free to choose their own blend of tutor-marked and computer-marked assignments provided they met the threshold in a minimum number of assignments. For students reaching this threshold, the final grade was determined solely by an end-of-year examination. This assessment strategy was compared with one in which summative assignments contributed to the final grade alongside the final examination. The methodology took the form of data analysis of assessment records from more than 3500 anonymized students over a period of 7 years. Evidence was found for improved student engagement under the formative thresholded assessment strategy, with students choosing to complete many more computer-marked assignments, balanced by slightly fewer tutor-marked assignments. There was no change in overall retention. Many students made use of the capability to repeat different versions of the computer-marked questions. There was some evidence of students gaming the system (repeating questions until they obtained a version they had seen before), though most students did not do this at all and only a small number did so regularly. Students appeared to make better use of model answers, provided in response to three incorrect tries at a question, than of hints provided after each try. It was noted that the formative thresholded assessment strategy has potential for wider use.

1. INTRODUCTION

During the Covid-19 pandemic, much university teaching and assessment that previously took place in face-to-face settings was moved online. Despite some concerns, students were generally satisfied with the quality of the assessment they received (Şenel & Şenel, 2021). The adjustments made during the pandemic were remarkable, and it is now recognized that, beyond the immediate crisis response, there are longer-term lessons for sustainable online learning (Adedoyin & Soykan, 2023; Yang & Xin, 2022). While the early focus was on the immediate affordances brought by online assessment, it is now appropriate to consider longer term implications (St-Onge et al., 2022). In order to do this, it is necessary to move beyond a consideration of which assessment tools to use and instead to start from evidence-based assessment design and strategy (Morris et al., 2021).

---

*CONTACT: Sally Jordan ✉ sally.jordan@open.ac.uk 🖃 School of Physical Sciences, The Open University, Milton Keynes, MK7 6AA, United Kingdom

This paper draws lessons from earlier (pre-Covid) data obtained at a university which has taught at a distance for more than 50 years and has made increasing use of online teaching and assessment through recent decades. In this environment, which much of the sector found itself engaging with seriously for the first time in 2020, it is particularly important that students are supported and well engaged in their learning. It has been argued that regular assignments, in particular computer-marked assignments with instantaneous and targeted feedback, might provide distance-learning students with the equivalent of a tutor sitting right beside them (Ross et al., 2006). For this reason, modules at the university in the study are usually assessed by way of both continuous assessment through the year and an end-of-module assessment, frequently an examination. The continuous assessment element has learning and motivation as its primary purpose, and comprehensive feedback is provided, but historically it usually also contributed towards a student's grade for the module and hence to their overall result. However, as discussed further in Section 2, when assessment has both formative and summative functions, the relationships between the two are often complex and sometimes confusing, so it was decided to trial a different assessment strategy.

The new "formative thresholded" assessment strategy studied here required students to reach a modest threshold on continuous assessment tasks, but the marks for these did not contribute to the final grade. The strategy was designed to enable students and their tutors to focus on the formative function of the assignments, freed from anxiety over the minutiae of the score received. The formative focus meant that concerns about plagiarism were reduced, so assignments could be re-used from year to year and computer-marked assignments could be repeated as many times as a student wished. The detailed feedback and indicative scores enabled students to monitor their progress and self-regulate. Thus, in several senses of the word, the assessment could be considered to be sustainable (Boud & Soler, 2016).

As Holmes (2018) suggests, assessment can be used to increase student engagement and motivation and thus to contribute to the quality of the overall student experience, but concern has been expressed that student engagement with assessment might decrease without the perceived incentive of a contribution towards a student's final module grade (Kibble, 2007). Thus, the key research question to be addressed was: What is the impact on student engagement and achievement of a modest threshold on assignments, compared with a previous situation in which some assignments had been summative and some had been purely formative?

The assignments under consideration were electronically submitted tutor-marked assignments and interactive computer-marked assignments. A variety of feedback was provided on both and, for the computer-marked assessment, the feedback was instantaneous, and students had an opportunity to try the questions again, bringing alignment with Gibbs and Simpson's (2005) conditions under which assessment supports learning. The secondary research question was thus: What use do students make of different types of feedback on computer-marked assignments, varying from hints to model answers, with the opportunity to attempt questions again?

The remainder of the paper is structured as follows. Section 2 reviews relevant literature. Section 3 gives further details of the computer-marked assignments and the formative thresholded assessment architecture in which they were embedded. This section also describes the datasets and outlines the detailed research questions addressed in the paper. Section 4 contains the results of the data analysis and Section 5 discusses the results and draws conclusions.

2. LITERATURE REVIEW

The literature underpinning the work we describe is wide-ranging and extensive. For this reason, our review is necessarily selective, concentrating on significant recent contributions, meta-analyses, and independent literature reviews.

2.1. Student Engagement

In recent years, much has been written about student engagement, especially in an online environment (Bond et al., 2020). The meaning of "student engagement" has evolved over time, with widely accepted components including time on task, quality of effort and involvement in productive learning activities (Kuh, 2009). Student engagement has been recognized as one of the most important drivers of academic success (Kahu, 2013) and is associated with improved achievement and retention (Kuh et al., 2008), with disengagement having negative consequences on a student's outcomes and being a predictor of drop-out (Finn & Zimmer, 2012; Ma et al., 2015). It is widely accepted that digital educational technologies offer the potential to improve engagement (Chen et al., 2010), but the relationships are complex (Rashid & Asghar, 2016) and care must be taken in the design and integration of the technology to ensure that it is effective (Bond & Bedenlier, 2019; Englund et al., 2017; Kirkwood, 2009). Guo et al. (2023) identified eight specific factors which can encourage engagement or act as a barrier to it.

2.2. A Role for Computer-Marked Assessment?

The term "e-assessment" can refer to any use of a computer as part of the assessment process and terms such as "digital assessment", "online assessment", "computer-based assessment" and "technology-enhanced assessment" are similarly broad. More specifically, "computer-marked assessment" refers to situations in which students' responses are automatically marked. It is important to note that, while early computer-marked assessment relied on students indicating their responses to multiple-choice questions on a machine-readable form, for marking at a later stage, by the 21st Century, the focus had become online computer-marked assessment, enabling instantaneous interaction between a student and the system on which the assessment sits (Jordan, 2023). At the same time, the range of question types that are available has extended beyond selected-response question types such as multiple-choice, multiple-response and drag-and-drop to include automatically marked constructed-response question types with answers comprising numbers and algebraic expressions (Sangwin, 2013; Sangwin & Harjula, 2017), words and sentences (Burrows et al., 2015; del Gobbo et al., 2023), and even essays (Ramesh & Sanampudi, 2022; Süzen et al., 2020).

The interactivity of modern computer-marked assessment brings potential for it to motivate and engage students (Holmes, 2015). Riegel & Evans (2021) report on positive emotions experienced by students following a computer-marked assessment, contrasting with the negative emotions experienced following a conventional quiz. Feedback can be provided instantaneously, and, in principle, students can then repeat the question or a similar one, thus enabling them to learn from the feedback while it is still fresh in their minds (Gibbs & Simpson, 2005; Nicol & McFarlane-Dick, 2006). Students report that they prefer receiving feedback from a computer because it is perceived to be unbiased, non-judgmental and impersonal (Sim et al., 2004) and enables them to make mistakes in private (Miller, 2008). However, the mere act of taking online tests has been shown to improve subsequent performance more than additional study of the material, even when the tests are given without feedback. This is the so-called "testing effect"; Roediger and Karpicke's (2006) review of work in this area has been updated in Roediger and Karpicke (2018) and Yang et al. (2019).

2.3. Importance of Assessment Strategy

Although computer-marked assessment brings potential to improve student engagement, it is not uniformly well received by either students or educators (Ghabraie, 2020; Kumar & Sajja, 2021). Factors that have been identified as significant include the choice of an appropriate question type for each item, and the need to take care in writing and checking questions (Benson & Brack, 2010; Jordan, 2023). However, here we concentrate on the importance of assessment design and strategy, which we take to include the relationship with the rest of the assessment on the module and program, whether the focus is formative or summative, and the way in which

the computer-marked assessment operates (e.g. whether students can repeat questions and assignments). Nguyen et al. (2017) found that assessment design and strategy can have a significant impact on student engagement and pass rate, while highlighting the dangers inherent when assessment design alters between different but closely associated modules. It is also important that assessment is well aligned with learning activities (Barthakur et al., 2022).

Many authors have highlighted the complex interrelationships between the formative and summative functions of assessment (e.g. Lau, 2016). Where formative and summative functions co-exist, there is a danger that the purpose will be confused (Boud, 2000) and, in particular, that the summative function will dominate (Lipnevich & Smith, 2009). There may be poor alignment of student and staff perceptions, with staff seeing the purpose as primarily formative, while students are primarily concerned with the grade (Winstone & Boud, 2022). When assessment is delivered or marked by electronic means there can be further confusion and lack of trust (Rolim & Isias, 2019).

When assessment has a formative focus, concerns over plagiarism are reduced, which is particularly relevant given the growth in contract cheating (Ahsan et al., 2022) and in the use of generative artificial intelligence (Farrelly & Baker, 2023). In formative use, computer-marked assignments can be repeated multiple times, which means that students can act on the feedback provided by immediately revisiting the same question or a similar one (Lefevre & Cox, 2017). Formative assessment can help students to monitor their progress, encourage further study and increase their learning and understanding (McCallum & Milner, 2021).

3. METHOD

The investigation was based on a detailed analysis of student interactions with assessment collected over a period of seven years. The context in which the study took place is described in Section 3.1 while the data analysis methodology is described in more detail in Section 3.2.

3.1. Context

The investigation was conducted at the Open University, which is based in the UK, and focused on two of its upper-stage undergraduate modules: *The Quantum World* and *Electromagnetism*, key components of any physics degree. These modules piloted the formative thresholded assessment strategy prior to Faculty-wide adoption. Each of the modules counted for 30 points in the UK Credit Accumulation and Transfer Scheme (CATS), meaning that each was equivalent to a quarter of a year's study time for a full-time student, spread over a nine-month period. The modules could be studied alone or concurrently with each other or with other modules. Most Open University students study part-time alongside employment and/or caring responsibility and therefore study at an intensity of 30 or 60 CATS points per year. Students on the two modules investigated here had a wide range of ages from 18 to over 65, with the distribution peaking around age 30. The male: female ratio on both modules was about 4:1.

The Open University offers supported distance learning to students who may be studying anywhere in the world. Primary teaching resources are supplied in printed form and online, and students are supported by tutors who offer non-compulsory tutorials (usually online) as well as marking assignments and providing feedback comments electronically. Continuous assessment is completed in students' own homes and at a time of their choosing. Furthermore, although the students on the modules in question were strongly advised to study recommended lower-level mathematics and physics modules, the University's "open" mission extends to it having no entry qualifications and students could not be forced to study prerequisite modules.

The two modules shared the same overall assessment pattern. In each case, the continuous assessment consisted of four tutor-marked assignments and six interactive computer-marked assignments, and in order to pass the threshold, students were required to score more than 30% in seven of the ten assignments, two of which had to be tutor-marked assignments. Students

passing the threshold could then sit a final examination, and their overall module grade was determined solely by the result of this.

In guidance material, students were encouraged to use the continuous assessment as a resource to help them to develop skills and understanding, with the focus being on learning rather than marks. Each of the tutor-marked assignments included several questions assessing key physics learning outcomes, and required written answers from students, usually with extensive mathematical content. Tutors provided detailed feedback on each student's work as well as a summary of the main points requiring attention. Each of the six computer-marked assignments included eight questions, comprising a variety of question types, with constructed response questions (i.e. free-text entry) being favored over selected response question types such as multiple-choice. In line with their pedagogic function, the computer-marked assignments gave students many opportunities to re-try questions after receiving feedback. Each student received a particular version of each question and had three opportunities to get the correct answer. The opportunities were interspersed with increasingly detailed hints, wherever possible tailored to the errors made, and a fully worked solution was provided whenever a student completed a question, either by getting it right or by having had three tries. The functioning of a sample question is illustrated in Figure 1.

Figure 1. *A computer-marked question as attempted by a student, showing the tailored feedback generated after incorrect student responses at (a) the first try and (b) the second try; (c) the complete solution offered after a third incorrect try or a correct response.*

After completing a question, students could immediately make an attempt at a different version of the question. For brevity, we refer to attempting a fresh version as repeating a question. All versions of a question covered the same knowledge and skills, and were of similar difficulty, but they differed sufficiently to make students think afresh, even after seeing a model solution, as illustrated in Figure 2. Each question had between three and eight versions of similar difficulty, arranged in a cyclic order so that a student was presented with fresh versions for as long as is possible, until the cycle repeated itself. The initial version of each question was chosen at random so repeating an assignment generally led to a different set of question versions for a student to tackle. Students were encouraged to make fresh attempts until they were confident of having the relevant skills and knowledge. To promote this, the mark awarded for the question was that for the most recent attempt; the system therefore rewarded eventual success rather than any initial failure.

Figure 2. *A different version of the question illustrated in Figure 1.*

In a one-dimensional system at time $t = 0$, the wave function of a particle is given by the function

$$\Psi(x,0) = \begin{cases} -Ax^{1/2} & \text{for } 0 \leq x \leq L \\ 0 & \text{elsewhere} \end{cases}$$

as sketched in the diagram, where A is a positive constant.

If the position of the particle is measured at time $t = 0$, what is the probability of finding it somewhere in the interval $0 \leq x \leq L/2$?

Specify your answer as a fraction or as a decimal correct to 2 significant figures.

probability = [   ]

Check

Classification: pencil and paper required.

References: Book 1 Section 2.4.1 pp 49-51, Section 3.1.6 pp 70-72.

At any stage, students could click a "Finish" button, whether or not they had answered all the questions in the assignment. This generated an overall score and a summary report of their strengths and weaknesses. However, students could then tackle the whole assignment again, as many times as they wished and it was their highest overall score that was used to determine whether the threshold had been met. Although the threshold was modest, students were encouraged to aim well above this by submitting all the assignments and repeating individual questions and/or whole assignments until they achieved scores of around 75% or more.

3.2. Data Analysis

In this study, the extent of student engagement was measured by: the number of completed assignments, the number of times questions and entire assignments were repeated and the effects of these repetitions in boosting marks above the modest required threshold. The depth of engagement was evaluated by considering the extent to which hints and model answers were acted on. Student achievement was measured by the marks obtained in the continuous assessment and the examination, and by the overall retention rate.

Data were analyzed from a number of presentations of the two modules which, for *The Quantum World*, bridged the changes in the assessment strategy, as shown in Table 1. Unfortunately, the Open University's student population also altered, in particular as a result of significant changes to higher education funding in England in 2012. To minimize the impact of the changing student

population, only presentations of the modules up to those starting in October 2013 and February 2014 were considered in the study. The dates were chosen so as to maximize the dataset while not considering presentations likely to have attracted students who had commenced their studies in October 2012 or later, acknowledging that students usually studied the modules in question towards the end of their degrees. These two modules were selected for the study because they used both tutor-marked and computer-marked assignments and were the initiators and first users of the new assessment strategy. In addition, it was reasonable to assume that by this stage in their study, most students would be familiar with the University's approach and would have taken the recommended prerequisite modules, although the open entry policy meant that this could not be forced. The modules had the same recommended pre-requisites and belonged to the same degree structures. Because of student choice over study paths and study intensity, the modules did not share a single cohort, though many students studied both modules, either concurrently or in different (usually consecutive) presentations.

Most of the reported analysis relates to the module *The Quantum World*, before and after the changes to its assessment strategy, with later presentations included for verification purposes. We have only included comparison with the module *Electromagnetism* where interesting differences were found between the modules. Where outcomes on subsequent presentations were found to be similar, data were combined to increase the size of the dataset.

Table 1. *Key features of the presentations of the modules included in the investigation.*

Module	Presentation abbreviation	Start date	Number of students	Characteristics of continuous assessment
The Quantum World	QW Old 1	Feb 2007	383	Summative tutor-marked assignments, no computer-marked assignments
	QW Old 2	Feb 2009	380	Summative tutor-marked assignments, purely formative computer-marked assignments
	QW New 1	Feb 2010	207	Formative thresholded tutor-marked assignments and computer-marked assignments
	QW New 2	Feb 2011	290	
	QW New 3	Feb 2012	243	
	QW New 4	Feb 2013	323	
	QW New 5	Oct 2013	365	
Electro-magnetism	EM New 1	Feb 2010	286	Formative thresholded tutor-marked assignments and computer-marked assignments
	EM New 2	Feb 2011	182	
	EM New 3	Feb 2012	333	
	EM New 4	Feb 2013	278	
	EM New 5	Feb 2014	313	

Several specific investigations were conducted, as outlined in the following sub-sections.

3.2.1. *Module completion rates*

The overall completion rates for students on *The Quantum World* were compared for the presentations before and after the change in assessment strategy i.e. the presentations abbreviated to QW Old 1 and QW Old 2 (see Table 1) were compared with QW New 1-5.

3.2.2. *Assignment submission rates and grades*

Analyses were conducted, again for *The Quantum World* over a range of presentations, into:

1. The impact of the switch to formative thresholded assessment on overall submission rates.
2. The extent to which students did more than the bare minimum required, as measured both by the number of assignments submitted and the grades achieved relative to the threshold.

3. Whether there was a relationship between the number of assignments completed and the student's final module result, based on examination performance. Note that we would expect there to be a strong correlation between the number of assignments submitted and the final examination mark, but this is not necessarily evidence of a causal link between these two factors; conscientious students are likely to study harder, resulting in both completion of more assignments and a better mark in the final examination. However earlier work (Bolton, 2010a) had shown that the single most important predictor of success in *The Quantum World* was performance on the recommended mathematics prerequisites. This allowed us to explore how completing a given combination of assignments increases or decreases a student's examination mark, on average, relative to other comparable students. The details of this analysis are given in Section 4.2.3.

3.2.3. *Repetition of computer-marked questions and the use of feedback*

Analyses were conducted into the extent to which students repeated computer-marked questions and the use they made of the feedback provided. Because some differences in behavior were noticed between the students on the different modules, *The Quantum World* and *Electromagnetism*, we report and discuss both, in considering:

1. The use made of hints by students prior to a second and/or third try at a question.

2. The extent to which students repeated complete questions (in different versions).

3. The extent to which students were "gaming the system" by repeating a question without making a serious attempt to answer it until they had seen the full answers to all versions of the question, at which point their next attempt was correct at the first try. As a proxy for this, we looked for a pattern of behavior which generated a repeated score of zero within the system for exactly the same number of attempts at the question as there were versions of that question, immediately followed by an answer that obtained full marks.

4. RESULTS

4.1. Module Completion Rates

60.3% of the students who commenced study of QW Old 1 (summative tutor-marked assignments and no computer-marked assignments) completed it, while 70.9% of the students who started the next presentation, QW Old 2 (summative tutor-marked assignments, purely formative computer-marked assignments), completed the module. This is a significant result ($p = 0.0034$ that such an outcome might arise by chance) and there is no reason to suspect that the student population changed significantly between the two presentations. It therefore seems likely that the introduction of the computer-marked assignments had a positive effect.

In contrast, the completion rates following the change to formative thresholded assessment (70.5% for QW New 1 and 69.5% for the 1221 students on all remaining presentations in the study) were very close to those for QW Old 2, indicating that the move to formative thresholded assessment had no significant impact on student retention. Completion and pass rates showed almost no correlation with age or gender.

4.2. Assignment Submission Rates and Grades

4.2.1. *Overall submission rates*

Figure 3 shows the assignment submission rates for QW Old 2 (summative tutor-marked assignments, purely formative computer-marked assignments) compared with the remaining five presentations in the study (formative thresholded tutor-marked and computer-marked assignments). The new thresholded assessment strategy gave students more flexibility than the previous strategy of summative tutor-marked assignments and optional computer-marked assignments. Figure 3 shows that students used this flexibility to complete many more computer-marked assignments, balanced by slightly fewer tutor-marked assignments. The

number of students attending the examination was unaffected by the change in continuous assessment strategy.

Figure 3. *Submission rates for assignments on the module The Quantum World. Figure 3(a) shows the four tutor-marked assessments (TMA1 – TMA4) and the examination. Figure 3(b) shows the six computer-marked assignments (CMA1 – CMA6). The labeling of the different presentations of the module is as given in Table 1.*
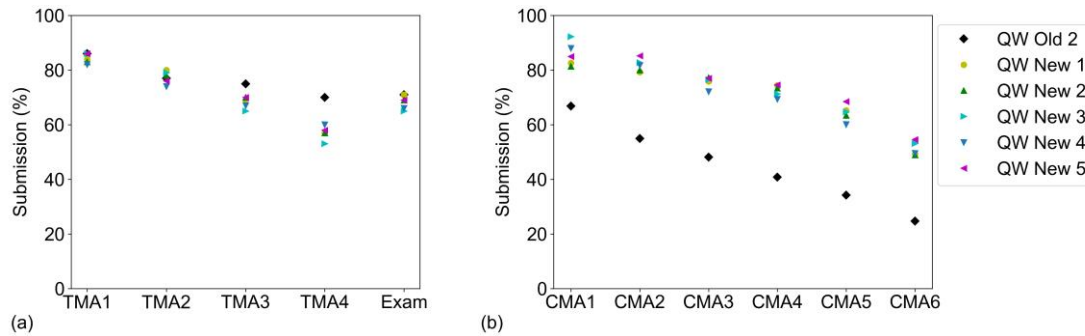
4.2.2. *Are students doing more than the bare minimum?*

Of the 1036 students who submitted and achieved the threshold of 30% in seven assignments, on the presentations of *The Quantum World* from QW New 1 to QW New 5, 865 (83.5%) did more than the minimum i.e. submitted at least eight assignments, and 499 (48.2%) of the students submitted all ten assignments available.

The scores recorded for assignments were also, on average, very much higher than the minimum of 30% required. Figure 4 illustrates that there was no consistent difference between the percentage scores for summative tutor-marked assignments (QW Old 2) and formative thresholded tutor-marked assignments (QW New 1), and the outcomes have been similar for all for subsequent presentations. Since QW New 1, mean scores of 80% or above have been observed on all computer-marked assignments, with a small but consistent increase in mean score in the years following the move to formative thresholded assessment.

Figure 4. *Mean percentage scores for tutor-marked assignments (TMA 1 – TMA 4) before and after the change in assessment strategy on the module The Quantum World.*
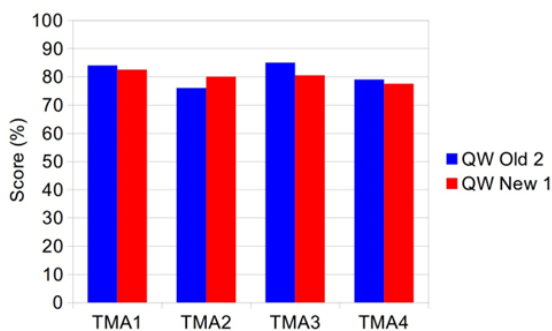
4.2.3. *Relationship between the number of assignments submitted and examination score*

Table 2 shows the mean examination marks for the students who were in each assignment submission category for QW New 1 and QW New 2 combined. Cells left blank in the table correspond to combinations that did not meet the continuous assessment threshold (so students could not obtain credit), though it is noteworthy that very few students who studied to the end of the module were in these categories. Non-zero entries along the uppermost diagonal correspond to the minimum allowed number of assignments with scores over 30%. Note, for example, that the 58 students who had submitted just three tutor-marked assignments (TMAs) and four computer-marked assignments (CMAs) had a mean examination mark of 49.4% while

the 478 students who had submitted all four tutor-marked assignments and all six computer-marked assignments had a mean examination mark of 74.1%. As expected, students who did more assignments during the year also did better on the examination, but it would be dangerous to attribute this outcome to a simple causal link between engagement with continuous assessment and final examination score.

Table 2. *Mean examination score for students who successfully completed given combinations of assignments in The Quantum World. In each cell n is the number of students. Blank cells indicate combinations that fall short of the threshold requirement for sitting the examination.*

	2 TMAs	3 TMAs	4 TMAs
3 CMAs			55.8% ($n = 20$)
4 CMAs		49.4% ($n = 58$)	62.7% ($n = 49$)
5 CMAs	52.0% ($n = 58$)	56.2% ($n = 54$)	66.8% ($n = 93$)
6 CMAs	44.0% ($n = 55$)	53.9% ($n = 88$)	74.1% ($n = 478$)

Table 3 presents the results of the more subtle investigation based on the difference between actual and reference examination scores, averaged over students in each cell of the table. The reference score for each student is the average examination score for *all* students with the same grade in the mathematics prerequisite module. A previous study (Bolton, 2010a) showed that this is a strong predictor of final outcome. Because the analysis is based on a redistribution of marks, cells can have positive or negative values, averaging to zero across the whole student cohort. Cells with positive (or negative) values indicate combinations of assignments that deliver better (or worse) results than predicted on the basis of prior performance. For example, the 478 students who completed all available assignments had an average uplift in their examination score of 6.6 percentage points compared to the average for all students who matched their achievement in the prerequisite module. By contrast, the 58 students who submitted just three tutor-marked assignments and four computer-marked assignments did worse than their peers by 12.0 percentage points.

Table 3. *Average differences between actual and reference examination scores for students with given combinations of completed assignments. In each cell n is the number of students.*

	2 TMAs	3 TMAs	4 TMAs
3 CMAs			–7.0 ($n = 20$)
4 CMAs		–12.0 ($n = 58$)	+0.1 ($n = 49$)
5 CMAs	–8.7 ($n = 58$)	–7.8 ($n = 54$)	+2.4 ($n = 93$)
6 CMAs	–15.4 ($n = 55$)	–8.3 ($n = 88$)	+6.6 ($n = 478$)

4.3. Repetition of Computer-Marked Questions and Use of Feedback

4.3.1. *Use of hints*

Table 4 shows the percentage of all computer-marked question attempts which were correct at the first, second and third try. Overall, 61.4% of question attempts on *The Quantum World* and 51.7% of question attempts on *Electromagnetism* were correct by the third try. It is noteworthy that most correct question attempts were correct at the first try, without any need for hints. For students who initially got the answer wrong, the first hint was more effective than the second, though many students failed to benefit from either. It is also noteworthy that, on average, students were less likely to get the correct answer for *Electromagnetism* questions, even after three tries.

Table 4. *Percentage of responses that were correct at first, second and third try, for all available questions on The Quantum World (QW) and Electromagnetism (EM).*

	QW (as % of all question attempts)	QW (as % of those making this try)	EM (as % of all question attempts)	EM (as % of those making this try)
Correct at 1st try	43.0	43.0	34.3	34.3
Correct at 2nd try	12.9	22.7	12.2	18.5
Correct at 3rd try	5.5	12.4	5.2	9.8
Correct at 1st, 2nd or 3rd try	61.4		51.7	

Table 5 illustrates the extent to which the feedback provided in the hints appears to have been ignored, as implied by student responses being identical from one try to the next. This behavior is surprisingly common, and indeed a total of 20.8% of all student attempts at questions on *The Quantum World* and 28.7% of all student attempts at *Electromagnetism* questions offered three identical responses. Interestingly, while for *The Quantum World* most of these instances comprised situations where a response had been entered at first try but was not then subsequently altered, for *Electromagnetism* it was more common for students to enter no response at all.

Table 5. *Percentage of responses that were identical at first, second and third try, for all available questions on The Quantum World (QW) and Electromagnetism (EM).*

	QW (as % of all question attempts)	QW (as % of those making this try)	EM (as % of all question attempts)	EM (as % of those making this try)
Identical response at 1st & 2nd try.	21.7	38.1	30.4	46.3
Identical response at 2nd & 3rd try.	24.9	56.5	34.1	63.7
Three blank responses.	6.4	14.4	19.4	36.3
Three identical but non-blank responses.	14.4	32.7	9.3	17.4

4.3.2. *Repetition of questions*

We now report the extent to which students made repeated attempts at different versions of whole questions. For *The Quantum World*, over all questions and all presentations, the mean number of attempts at each question was 2.1, while for *Electromagnetism* the mean number of attempts was 2.5. However, these figures mask huge variation between different questions and different students. Unsurprisingly, questions which had a lower mean first attempt score were more likely to be repeated than higher scoring questions. It is also noteworthy that the mean number of attempts was very slightly higher for constructed-response questions, in which students construct their own answer (2.1 for *The Quantum World* and 2.6 for *Electromagnetism*) than for selected-response questions such as multiple-choice, in which students select from options provided in the question (2.0 for *The Quantum World* and 2.3 for *Electromagnetism*).

Figure 5 shows the overall distribution of question attempts for students on QW New 5 and EM New 5. Most students attempted most questions just once, but a small number of students attempted questions very many times. Note the very different pattern of Figure 5(a) (for *The Quantum World*) and Figure 5(b) (for *Electromagnetism*). For *Electromagnetism*, a larger number of students attempted the questions precisely four times than would be expected from the overall trend. Since all *Electromagnetism* questions have three versions, this may be as a result of students "gaming the system" i.e. writing down the correct answer to the first version they received then waiting for the same version to appear again. This point is further explored in the next section. *The Quantum World* questions had a variable number of versions.

Figure 5. *Distribution of the number of attempts made by all students on all questions for (a) QW New 5 and (b) EM New 5.*

(a)   (b)
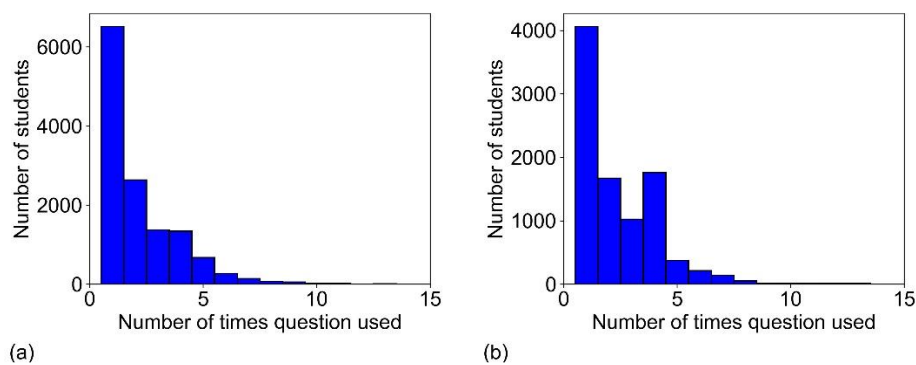
Figure 6 illustrates the impact of repeating questions on the final scores for each computer-marked assignment. Students on both modules appeared to have taken very seriously the advice that they should aim for at least 75%, and they repeated questions to achieve this. The first attempt score for a typical assignment had a mean of around 50% for *The Quantum World* and around 45% for *Electromagnetism* (shown by the blue bars in Figure 6); this increased to around 80% for both modules on the final attempt (shown by the red bars).

Figure 6. *The boost in marks achieved by repeating computer-marked questions in CMA1-CMA6 for (a) The Quantum World (data from QW New 2 to QW New 5 combined) and (b) Electromagnetism (data from EM New 2 to EM New 5 combined.) Blue bars show the first attempt, red bars the last.*

(a)   (b)
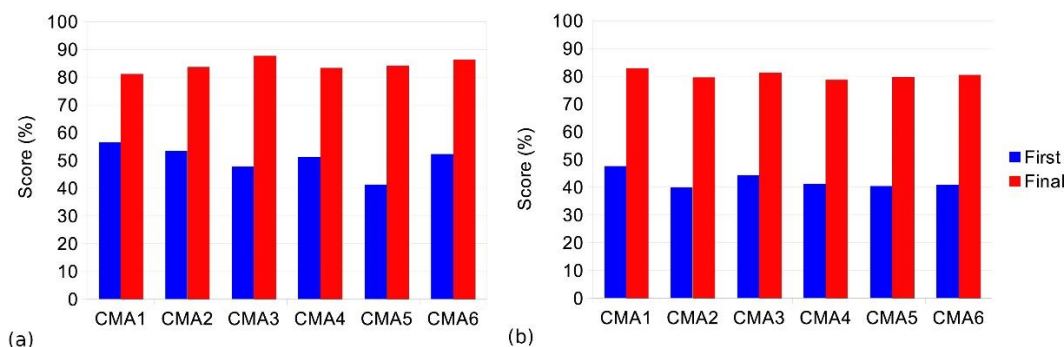
4.3.3. *Gaming the system*

From a baseline of 566 registered students on the QW New 3 and QW New 4 presentations of *The Quantum World*, 65.5% *never* exhibited the behavior of repeating questions for exactly the same number of times as there are versions and then getting the question correct immediately afterwards (a signature of "gaming the system"). The corresponding percentage from a baseline of 611 registered students on the equivalent presentations of *Electromagnetism* was 52.8%. A further 9.8% of the students on *The Quantum World* and 11.2% of those on *Electromagnetism* exhibited this behavior on just one of the 48 questions. A minority of students exhibited the behavior more widely. On *The Quantum World*, 12.9% did so for 5% of the questions, 6.9% for 10% of the questions and 1.8% for more than half of the questions. The corresponding percentages for the *Electromagnetism* module are slightly larger: 19.8%, 9.1% and 1.8%.

5. DISCUSSION and CONCLUSIONS

We were seeking to investigate the impact on student engagement of allowing students freedom in choosing a blend of tutor-marked and computer-marked assignments, subject to modest thresholds on grades. This compared with a more rigid previous system based on summative tutor-marked and formative computer-marked assignments. At the whole module level, a critical measure of student engagement is whether students are retained to the end of the

module. If the change in assessment strategy had resulted in a reduction in retention, this would have been a real cause for concern, while an increase in retention would have been pleasing. However, there was no significant change in retention at all.

In contrast, the introduction of computer-marked assignments in the first place coincided with a significant increase in retention for that presentation of *The Quantum World*, and the improved retention rate was maintained through the change in assessment strategy. This is particularly pleasing in the light of the Open University's open entry policy. Care must be taken in attributing causality for the improvement, because other changes were made at the same time as computer-marked assignments were introduced, but a student survey conducted at that time, to which 53 students responded, illustrated strong student satisfaction with the computer-marked assignments (Bolton, 2010b). 83% of respondents agreed that computer-marked assignments helped them to acquire skills and knowledge, while 79% agreed that computer-marked assignments helped them to prepared for the examination, and 75% agreed that computer-marked assignments helped them to understand what they needed to study further. Although feedback from humans, in this case the students' tutors, is inevitably more flexible than the automated hints provided by the computer-marked assignments, only a minority of respondents (42%) agreed that tutor-marked assignments taught them more than computer-marked assignments. The positive reaction to computer-marked assessment is in line with the findings of Holmes (2015) and Ghabraie (2020).

When a modest threshold was introduced, most students were observed to do considerably more than was required, both in terms of the number of assignments submitted and as measured by the marks obtained for each assignment (noting that, provided the threshold had been reached, the mark did not contribute to students' results). Most starkly, the computer-marked assignment submission rate increased markedly when the threshold was introduced, and students repeated computer-marked questions in order to achieve an improvement in their score, leading to a marked increase in the mean score for each assignment. By these measures it appears that the modest threshold, accompanied by advice to aim for a much higher score, was effective in encouraging engagement.

After considering other reasons for the correlation between the number of assignments submitted and examination score, it appears that submitting more assignments had a genuinely positive impact on final outcome for most students. This contrasts with Kibble's (2007) finding that a significant number of students who scored 100% on quizzes in response to incentives (credit of between 0.5% and 2% per quiz) did not subsequently perform well on summative examinations. The quizzes, like ours, had a primarily formative function and Kibble describes the student usage as "inappropriate", with a suspicion that students had copied from their peers, meaning that they were taking approaches in which "quiz points are scored, but which circumvent learning" (Kibble, 2007, p. 259).

Less positively, on both of the modules in our study, but in particular *Electromagnetism*, it seems that many students were not making as much use as had been hoped of the hints provided after one or two unsuccessful tries at computer-marked questions, relying rather more on the full answers provided after three unsuccessful tries. This result is in line with a survey in which students on *The Quantum World* were asked to rank eight features of the computer-marked assignments in terms of their helpfulness: "Being given a full solution" was reported as the most helpful aspect by 50% of the 53 respondents, compared with "Being given hints when my answer was incorrect" which was only considered to be the most helpful aspect by 23% of respondents (Bolton, 2010b, p. 85). These findings contrast with those from another Open University module, where hints were found to be more useful than the full answers (Jordan, 2011). However, that module used a very different assessment strategy to that considered here: the computer-marked assignments were summative and students were not able to repeat complete questions, meaning that students were not able to act on the final answers given (Gibbs

& Simpson, 2005). This serves to illustrate the importance of assessment strategy in determining the outcome of a change in assessment practice (Nguyen et al., 2017).

Some interesting insights can be obtained by comparing the use that was made of feedback on the two modules in the study. *Electromagnetism* students were less likely to get the correct answer after three tries, and considerably more likely to enter a blank response at each try, though students on *The Quantum World* were more likely to enter identical but not blank responses. *Electromagnetism* students were also more likely to make repeated question attempts, more likely to make precisely four attempts at questions, and more likely to repeat questions with zero score until one attempt more than the number of versions available, at which point they got a previously seen version of the question correct. There are several possible explanations for this differing behavior, each supported by some evidence, so all the explanations are worthy of consideration and of further investigation.

Firstly, *Electromagnetism* was generally considered to be a more difficult module than *The Quantum World*. We have already noted that questions with a lower mean first attempt score were more likely to be repeated than higher scoring questions, and students were plausibly repeating the difficult questions in order to learn from the full answers provided after three unsuccessful tries.

It has been noted elsewhere that students leave blank responses when they do not understand the question or the feedback provided (Jordan, 2014). Prior to the study reported here, student responses to the computer-marked assessment questions on *The Quantum World* had been analyzed in detail which had led to the removal of some questions, the rewording of others and additional tailored feedback for common incorrect responses. This detailed analysis had not taken place for *Electromagnetism*, and the difference in question behavior on the two modules points towards the importance of writing high quality questions, monitoring their use, and iterative development (Benson & Brack, 2010; Jordan, 2023).

It is to be expected that students will be more likely to enter a response of some sort rather than leaving a blank response when the answer is easier to guess. Therefore, we would expect more blank responses to constructed-response questions than to selected-response questions, in which students are provided with options from which to guess. Overall, 18.8% of the *Electromagnetism* questions are selected response, while 30.0% of *The Quantum World* questions are selected response.

Despite some concern that students were "gaming the system", for both *The Quantum World* and *Electromagnetism*, the majority of students never repeated questions for exactly the same number of times as there were different versions and then got the answer correct immediately afterwards. Where students did exhibit this behavior, it was usually just on a small number of questions, with only a handful of students doing so on most questions. It seems likely that the behavior was a "helpless reaction" because they did not know how to proceed by other means (Jordan, 2014). One of the factors that has been associated with increased incidence of cheating is a student's perceived inability to complete the assessment task themselves to the standard they feel is required (Sevnarayan & Bridget Maphoto, 2024). It is reassuring that, in general, students who occasionally in early assignments repeated questions unsuccessfully until they obtained a version that they had seen before did not then go on to do so regularly. This gives some optimism that most students appreciated that the assessment's formative focus meant that obtaining the right answer by this method was only cheating themselves (Ashworth et al., 1997). Although the problem is a relatively minor one, it is worth monitoring questions that seem to provoke this behavior and make changes where necessary. It is also advisable for questions to have a variable number of versions.

5.1. Limitations and Future Work to Address These

A known limitation of the study is its reliance on data gathered from a changing student population. We selected student cohorts that were as stable as possible, but we could not entirely remove the risk of impact. In addition, while we hypothesize plausible explanations for some of the different student behaviors on the two modules in the study, it has not been possible to test these hypotheses. The physics curriculum at the Open University is currently being redeveloped, providing an opportunity for further investigation, on a more stable student population, into the impact of assessment strategy on student engagement. However, as the educational community considers future assessment practice beyond the Covid-19 pandemic, we also encourage colleagues in other institutions to conduct similar investigations into the effectiveness of formative thresholded assessment and also to research the impact of a range of apparently minor factors, such as the level of the students and the guidance provided to them on assessment strategies.

5.2. Conclusions

In response to our first research question, we found no detrimental impact following the introduction of modest thresholds for tutor-marked and computer-marked assignments, compared with a previous situation in which some assignments had been summative and some had been purely formative. There was some evidence of improved engagement under the revised assessment strategy, but no change to overall retention on the modules included in the study. Computer-marked questions could be repeated, using a number of versions, as many times as a student wished prior to the deadline, and students were found to repeat questions to improve their score, well beyond what was required in order to reach the required threshold.

In answer to our second research question, students were found to make better use of the model answer provided after three tries than the hints provided after each unsuccessful try. There was evidence that on some occasions some students repeated a question until they obtained a version that they had seen previously, but most students exhibited this behavior not at all or on a very small number of occasions.

More generally, the study provided evidence of the need to address assessment strategy in addition to question type when moving toward online assessment. In addition, it has highlighted the importance of considering detail when designing assessment, as apparently minor factors can have a disproportionate effect on student engagement.

Acknowledgments

Declaration of Conflicting Interests and Ethics

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

Authorship Contribution Statement

Sally E. Jordan: Provided wider context within the Open University and externally, some data analysis, drafted paper. **John P.R. Bolton**: Led the investigation into student engagement and attitudes to assessment on the modules in the study, data collection and most of analysis.

Orcid

Sally E. Jordan https://orcid.org/0000-0003-0770-1443
John P.R. Bolton https://orcid.org/0009-0004-7191-6283

REFERENCES

Adedoyin, O.B., & Soykan, E. (2023). Covid-19 pandemic and online learning: The challenges and opportunities. *Interactive Learning Environments*, *31*(2), 863-875. https://doi.org/10.1080/10494820.2020.1813180

Ahsan, K., Akbar, S., & Kam, B. (2022). Contract cheating in higher education: A systematic literature review and future research agenda. *Assessment & Evaluation in Higher Education*, *47*(4), 523-539. https://doi.org/10.1080/02602938.2021.1931660

Ashworth, P., Bannister, P., & Thorne, P. (1997). Guilty in whose eyes? University students' perceptions of cheating and plagiarism in academic work and assessment. *Studies in Higher Education*, *22*(2), 187-203. https://doi.org/10.1080/03075079712331381034

Barthakur, A., Joksimovic, S., Kovanovic, V., Richey, M., & Pardo, A. (2022). Aligning objectives with assessment in online courses: Integrating learning analytics and measurement theory. *Computers & Education*, 190, 104603. https://doi.org/10.1016/j.compedu.2022.104603

Benson, R., & Brack, C. (2010). *Online learning and assessment in higher education: A planning guide*. Chandos Publishing.

Bolton, J. (2010a). CHASE: Prior course grades as indicators of success and failure in SM358. In J. Bolton (Ed.), *e-Tutorials and Learner Support* (pp. 124-140). The Physics Innovations Centre for Excellence in Teaching and Learning. https://www5.open.ac.uk/scholarship-and-innovation/esteem/resources/picetl-book-reports-e-tutorials-and-learner-support-2010

Bolton, J. (2010b). Using interactive computer-marked assignments in level 3 quantum mechanics. In S. Jordan (Ed.), e-Assessment (pp. 82-87). The Physics Innovations Centre for Excellence in Teaching and Learning. https://www5.open.ac.uk/scholarship-and-innovation/esteem/resources/picetl-book-reports-e-assessment-2010

Bond, M., & Bedenlier, S. (2019). Facilitating student engagement through educational technology: Towards a conceptual framework. *Journal of Interactive Media in Education*, 2019(1), 1-14. https://doi.org/10.5334/jime.528

Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. *International Journal of Educational Technology in Higher Education*, *17*(1), Article 2. https://doi.org/10.1186/s41239-019-0176-8

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, *22*(2), 151-167. https://doi.org/10.1080/713695728

Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-413. https://doi.org/10.1080/02602938.2015.1018133

Burrows, S., Gurevych, I., & Stein, B. (2015). The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education*, *25*, 60-117. https://doi.org/10.1007/s40593-014-0026-8

Chen, P.-S.D., Lambert, A.D., & Guidry, K.R. (2010). Engaging online learners: The impact of web-based learning technology on college student engagement. *Computers & Education*, *54*(4), 1222-1232. https://doi.org/10.1016/j.compedu.2009.11.008

del Gobbo, E., Guarino, A., Cafarelli, B., Grilli, L., & Limone, P. (2023). Automatic evaluation of open-ended questions for online learning. A systematic mapping. *Studies in Educational Evaluation*, *77*, 101258. https://doi.org/10.1016/j.stueduc.2023.101258

Englund, C., Olofsson, A.D., & Price, L. (2017). Teaching with technology in higher education: Understanding conceptual change and development in practice. *Higher Education Research and Development*, *36*(1), 73-87. https://doi.org/10.1080/07294360.2016.1171300

Farrelly, T., & Baker, N. (2023). Generative artificial intelligence: Implications and considerations for higher education practice. *Education Sciences*, *13*(11), 1109. https://doi.org/10.3390/educsci13111109

Finn, J.D., & Zimmer, K.S. (2012). Student engagement: What is it? Why does it matter?. In S. Christenson, A.L. Reschly, C. Wylie (Eds.), *Handbook of Research on Student Engagement* (pp. 97-131). Springer.

Ghabraie, K. (2020). Computer-marked assessments to enhance learning in engineering education. *International Journal on Innovations in Online Education*, *4*(1). https://doi.org/10.1615/IntJInnovOnlineEdu.2020033192

Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, *1*, 3-31. http://eprints.glos.ac.uk/3609/

Guo, Y., Liu, Z., Meng, X., & Yin, H. (2023). Unravelling the relationship between student engagement and learning outcomes in emergency online learning: A synthesis of quantitative and qualitative results. *Assessment & Evaluation in Higher Education*, *48*(8),1325-1338. https://doi.org/10.1080/02602938.2023.2214345

Holmes, N. (2015). Student perceptions of their learning and engagement in response to the use of a continuous e-assessment in an undergraduate module. *Assessment & Evaluation in Higher Education*, *40*(1), 1-14. https://doi.org/10.1080/02602938.2014.881978

Holmes, N. (2018). Engaging with assessment: Increasing student engagement through continuous assessment. *Active Learning in Higher Education*, *19*(1), 23-34. https://doi.org/10.1177/1469787417723230

Jordan, S. (2011). Using interactive computer-based assessment to support beginning distance learners of science. *Open Learning: The Journal of Open, Distance and e-Learning*, *26*(2), 147-164. https://doi.org/10.1080/02680513.2011.567754

Jordan, S. (2014). Using e-assessment to learn about students and learning. *International Journal of eAssessment*, *4*(1), 1-16. https://oro.open.ac.uk/40579

Jordan, S. (2023). Computer-marked assessment and concept inventories. In A.K. Wood (Ed.), *Effective Teaching in Large STEM Classes* (pp. 6-1 to 6-20). IOP Publishing. https://doi.org/10.1088/978-0-7503-5231-4ch6

Kahu, E.R. (2013). Framing student engagement in higher education. *Studies in Higher Education*, *38*(5), 758-773. https://doi.org/10.1080/03075079.2011.598505

Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: Effects of incentives on student participation and performance. *Advances in Physiology Education, 31*(3), 253-260. https://doi.org/10.1152/advan.00027.2007

Kirkwood, A. (2009). E-learning: You don't always get what you hope for. *Technology, Pedagogy and Education*, *18*(2), 107-121. https://doi.org/10.1080/14759390902992576

Kuh, G.D. (2009). The National Survey of Student Engagement: Conceptual and empirical foundations. *New Directions for Institutional Research*, *141*, 5-20. https://doi.org/10.1002/ir.283

Kuh, G.D., Cruce, T.M., Shoup, R., Kinzie, J., & Gonyea, R.M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education*, *79*(5), 540-563. https://doi.org/10.1080/00221546.2008.11772116

Kumar, D. & Sajja, R. (2020). Qualifying online assessment during COVID-19 pandemic: Reflecting on our experience under the cognitive lens of Miller's pyramid. *Research and Development in Medical Education*, *9*(1), 1-2. https://rdme.tbzmed.ac.ir/Article/rdme-31855

Lau, A.M.S. (2016). "Formative good, summative bad?": A review of the dichotomy in assessment literature. *Journal of Further and Higher Education, 40*(4), 509-525. https://doi.org/10.1080/0309877X.2014.984600

Lefevre, D., & Cox, B. (2017). Delayed instructional feedback may be more effective, but is this contrary to learners' preferences? *British Journal of Educational Technology*, *48*(6), 1357-1367. https://doi.org/10.1111/bjet.12495

Lipnevich, A.A., & Smith, J.K. (2009). "I really need feedback to learn": Students' perspectives on the effectiveness of the differential feedback messages. *Educational Assessment, Evaluation and Accountability*, *21*(4), 347-367. https://link.springer.com/content/pdf/10.1007/s11092-009-9082-2.pdf

Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *The Internet and Higher Education*, *24*, 26-34. https://doi.org/10.1016/j.iheduc.2014.09.005

McCallum, S., & Milner, M.M. (2021). The effectiveness of formative assessment: student views and staff reflections. *Assessment & Evaluation in Higher Education*, *46*(1), 1-16. https://doi.org/10.1080/02602938.2020.1754761

Miller, T. (2009). Formative computer-based assessment in higher education: The effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education*, *34*(2), 181-192. https://doi.org/10.1080/02602930801956075

Morris, R., Perry, T., & Wardle, L. (2021). Formative assessment and feedback for learning in higher education: A systematic review. *Review of Education*, *9*(3), e3292. https://doi.org/10.1002/rev3.3292

Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, R., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior*, *76*, 703-714. https://doi.org/10.1016/j.chb.2017.03.028

Nicol, D.J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, *31*(2), 199-218. https://doi.org/10.1080/03075070600572090

Ramesh, D., & Sanampudi, S.K. (2022). An automated essay scoring systems: A systematic literature review. *Artificial Intelligence Review*, *55*(3), 2495-2527. https://doi.org/10.1007/s10462-021-10068-2

Rashid, T., & Asghar, H.M. (2016). Technology use, self-directed learning, student engagement and academic performance: Examining the interrelations. *Computers in Human Behavior*, *63*, 604-612. https://doi.org/10.1016/j.chb.2016.05.084

Riegel, K., & Evans, T. (2021). Student achievement emotions: Examining the role of frequent online assessment. *Australasian Journal of Educational Technology*, *37*(6), 75-87. https://doi.org/10.14742/ajet.6516

Roediger, H.L., & Karpicke, J.D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181-210. https://doi.org/10.1111/j.1745-6916.2006.00012.x

Roediger, H.L., & Karpicke, J.D. (2018). Reflections on the resurgence of interest in the testing effect. *Perspectives on Psychological Science*, *13*(2), 236-241. https://doi.org/10.1177/1745691617718873

Rolim, C., & Isaias, P. (2019). Examining the use of e-assessment in higher education: Teachers' and students' viewpoints. *British Journal of Educational Technology*, *50*(4), 1785-1800. https://doi.org/10.1111/bjet.12669

Ross, S., Jordan, S., & Butcher, P. (2006). Online instantaneous and targeted feedback for remote learners. In C. Bryan, & K. Clegg (Eds.), *Innovative Assessment in Higher Education* (pp. 123-131). Routledge.

Sangwin, C.J. (2013). *Computer Aided Assessment of Mathematics*. Oxford University Press.

Sangwin, C.J., & Harjula, M. (2017). Online assessment of dimensional numerical answers using STACK in science. *European Journal of Physics*, *38*(3), 035701. https://iopscience.iop.org/article/10.1088/1361-6404/aa5e9d/

Şenel, S., & Şenel, H. (2021). Remote assessment in higher education during COVID-19 pandemic. *International Journal of Assessment Tools in Education*, *8*(2), 181-199. https://doi.org/10.21449/ijate.820140

Sevnarayan, K., & Bridget Maphoto, K. (2024). Exploring the dark side of online distance learning: Cheating behaviours, contributing factors, and strategies to enhance the integrity of online assessment. *Journal of Academic Ethics*, 1-20. https://doi.org/10.1007/s10805-023-09501-8

Sim, G., Holifield, P., & Brown, M. (2004). Implementation of computer assisted assessment: Lessons from the literature. *ALT-J*, *12*(3), 215-229. https://doi.org/10.3402/rlt.v12i3.11255

St-Onge, C., Ouellet, K., Lakhal, S., Dubé, T., & Marceau, M. (2022). COVID-19 as the tipping point for integrating e-assessment in higher education practices. *British Journal of Educational Technology*, *53*(2), 349-366. https://doi.org/10.1111/bjet.13169

Süzen, N., Gorban, A.N., Levesley, J., & Mirkes, E.M. (2020). Automatic short answer grading and feedback using text mining methods. *Procedia Computer Science*, *169*, 726-743. https://doi.org/10.1016/j.procs.2020.02.171

Winstone, N.E., & Boud, D. (2022). The need to disentangle assessment and feedback in higher education. *Studies in Higher Education*, *47*(3), 656-667. https://doi.org/10.1080/03075079.2020.1779687

Yang, B.W., Razo, J., & Persky, A.M. (2019). Using testing as a learning tool. *American Journal of Pharmaceutical Education*, *83*(9), Article 7324. https://doi.org/10.5688/ajpe7324

Yang, L.P., & Xin, T. (2022). Changing educational ssessments in the post-COVID-19 era: From assessment of learning (AoL) to assessment as learning (AaL). *Educational Measurement: Issues and Practice*, *41*(1), 54-60. https://doi.org/10.1111/emip.12492

Research Article

The effects of reverse items on psychometric properties and respondents' scale scores according to different item reversal strategies

Mustafa İlhan[1*], **Neşe Güler**[2], **Gülşen Taşdelen Teker**[3], **Ömer Ergenekon**[4]

[1]Dicle University, Ziya Gokalp Faculty of Education, Department of Mathematics and Science Education, Diyarbakir, Türkiye.
[2]İzmir Democracy University, Faculty of Education, Department of Educational Sciences, Izmir, Türkiye.
[3]Hacettepe University, Faculty of Medicine, Department of Medical Education and Informatics, Ankara, Türkiye.
[4]Nafiye Ömer Şevki Cizrelioğlu Kindergarten, Diyarbakir, Türkiye.

Abstract: This study aimed to examine the effects of reverse items created with different strategies on psychometric properties and respondents' scale scores. To this end, three versions of a 10-item scale in the research were developed: 10 positive items were integrated in the first form (Form-P) and five positive and five reverse items in the other two forms. The reverse items in the second and third forms were crafted using antonyms (Form-RA) and negations (Form-RN), respectively. Based on the research results, Form-P was unidimensional, while other forms were two-dimensional. Moreover, although reliability coefficients of all forms were obtained as above .80, the lowest one was acquired for Form-RN. There were strong-positive relationships between students' scores in the three scale forms. However, the lowest one was estimated between Form-P and Form-RN. Finally, there was a significant difference between the students' mean scores obtained from Form–RN and other two versions, but the effect size of the said difference was small. In conclusion, all these results indicate that different types of reverse items influence psychometric properties and respondents' scale scores differently.

1. INTRODUCTION

Likert-type scales, which are introduced by Rensis Likert, have been frequently used to measure the complex psychological constructs by many researchers from diverse disciplines since 1932. In such scales, items related to the construct to be measured and ordered response options are presented to the individuals and they respond to the items by selecting the category that best reflects them. That is, Likert scales are of self-report type and therefore are open to the response bias. Biased responses can occur in varied forms such as central tendency, extreme responding, social desirability, acquiescence, and dissent with Likert scales.

Central tendency bias happens because of the respondent's reluctance to select extreme response positions (Brace & Bolton, 2022). Contrary to the central tendency, in extreme response style, individuals tend to use the end points of the scales than middle response options (Kline, 2005). Social desirability stems from the respondent's willingness to portray herself or

---

*CONTACT: Mustafa İlhan ✉ mustafailhan21@gmail.com ⌨ Dicle University, Ziya Gokalp Faculty of Education, Department of Mathematics and Science Education, Diyarbakir, Türkiye.

himself favorably, regardless of his or her true characteristics; hence, social desirability bias is also called "faking good" (Furr, 2018). Acquiescence bias is the tendency to agree ("yea-saying") to all or majority of the items, regardless of their contents (Karandashev, 2021). On the other hand, dissent bias is the opposite tendency: it refers mainly to the respondent's propensity to disagree with statements (Stauesberg, 2002). These distortions in responses lead to construct irrelevant variance and threat the validity of measurements. Therefore, to obtain valid and reliable measures, precautions are needed to mitigate these biases.

Inclusion of reverse items (negative items) to the scale is a commonly proposed antidote to cope with response biases, especially acquiescence and dissent types (Ahlawat, 1985; Bandalos, 2018; Coolican, 2013; Bolt et al. 2020; Mayerl & Giehl, 2018). Reverse items function as cognitive "speed bumps" (Podsakoff et al., 2003) and can make respondents read the items more carefully, thereby reducing the probabilities of giving distracted and more generalized responses (Locker et al., 2007). Unfortunately, there is a price to pay for utilizing reverse items. For instance, reversals in item polarity may be confusing to respondents, especially when completing a long instrument (DeVellis, 2017). This situation can increase measurement error, reduce the validity and reliability of measures, and distort factor structure (Weijters & Baumgartner, 2012).

In general, two main strategies are available for item reversal. The first one involves adding negations to the item statements, namely, words like "not" or "no" and affixal morphemes like "un-," "non-," "dis," or "-less." In this case, the item's direction is changed without substantially modifying its wording (van Sonderen et al., 2013). The second strategy, on the other hand, is replacing the word or phrase in the original item with that of the antonym (Zhang et al., 2016). Reverse items created based on the first strategy are called *negatively worded negatively keyed*, and those formed based on the second one is named *positively worded negatively keyed* (Finney, 2001). Simply put, the items "*The conditions of my life are not good*" and "*The conditions of my life are bad*" are the two versions of the item "*The conditions of my life are good*" created based on the first and second strategies, respectively (Kam, 2023).

Notably, both strategies can bring some weak points. The reverse items formed with negations may cause erroneous data due to carelessness responses and more difficult judgmental process (Zhang et al., 2016). On the other side, in reverse items formed with antonyms, some respondents may not view an antonym intended as a reversal by the survey developer as being opposite in meaning to positive items (e.g., "relaxed" may not be perceived as an antonym for "stimulated") (Weijters & Baumgartner, 2012). Thereupon, respondents may give the same responses to a positive item and its reverse counterpart, and in this case, reversal ambiguity comes up. For instance, a respondent might agree with both the items *I like simple tasks* and *I like complex tasks* because *liking simple tasks* does not necessarily purport *disliking complex tasks* (Zhang et al., 2016). Swain et al. (2008) analyzed nearly 2000 Likert items in Bearden and Netemeyer's (1999) *Handbook of Marketing Scales* and ascertained that 81% of the reverse items were with negations, that is, items created based on the first strategy. The main point is that regardless of which strategy was created with, higher scores on the reverse items signify that the respondent has a low level of measured trait.

Although Likert scales have been utilized for about 90 years, researchers still experience some dilemmas about how they should act while developing these measurement tools. One of the foremost dilemmas encountered by researchers when constructing a scale concerns on either the integration of reverse items to the scale or its composition of items as being worded in the same direction or not. The dilemma about reverse items is not just about their incorporating into the scales or not. Another major dilemma is how they should be worded if reverse items are to be employed. Should the item reversal be achieved by using antonym expressions or negations (Weijters & Baumgartner, 2012).

A review of the literature disclosed that research about reverse items has a long history, and most studies focused on the first dilemma mentioned previously. The effects of reverse items on validity, internal consistency, factorial structure and factor loadings, item correlations, missing values, and respondents' ability estimation (mean scores on the scale) are principal subjects emphasized in these studies (e.g., Ahlawat, 1985; Bergstrom & Lunz, 1998; Boley et al., 2021; Bolt et al., 2020; Bulut & Bulut, 2022; Bulut, 2021; Chamberlain & Cummings, 1984; Conrad et al., 2004; Dooden, 2014; Dueber et al., 2021; Guyatt et al., 1999; Herche & England, 1996; Hooper et al., 2013, İlhan & Güler, 2017; Kula Kartal, 2021; Kula Kartal et al., 2022; Locker et al., 2013; Marsh, 1996; Schriesheim & Hill, 1981; Salazar, 2015; Schotte et al., 1996; Suárez-Alvarez et al., 2018; Vigil-Colet et al., 2020).

Conversely, the number of studies seeking answer to the second dilemma regarding reverse items is more limited. In simpler terms, there is a relatively small body of literature that deals with the comparison between the effect of different types of reverse items on the measurement qualities. With regard to this, Schriesheim et al. (1991) conducted a study on the effect of negation and polar opposite item reversals on reliability and validity and found out that the effect of reverse items on psychometric properties differed based on the strategy used for item reversal. Moreover, Salazar (2015) scrutinized the pros and the cons of combining reverse (negative) and regular (positive) items in scales in the Spanish context and concluded that individuals do not respond in the identical fashion to all types of reverse items. Similarly, Zhang et al. (2016) manipulated the types of reverse items (antonym vs. negation) while they investigated the effect of reverse items on the factor structure of the Need for Cognition Scale. As a result, they established that both the number and type of reverse items affect the factor structure of the scale.

In previous studies, variables that play a role in individuals' responses to reverse items have also been investigated. The results obtained demonstrate that the effects that arise from reverse items vary according to culture, age, linguistic features, and respondents' reading proficiency. For example, Marsh (1986) analyzed the bias of reverse items on a sample of preadolescent children and detected that younger children and children with low-reading proficiencies are clearly less able to respond to reverse items appropriately. Also, Williams and Swanson (2001) (cited in Weems et al., 2006) found a similar relationship for adults. Likewise, Bulut and Bulut (2022) revealed that the severity of item wording effect that emerges because of reverse items is related to the reading ability. Hooper et al. (2013) audited the behavior of reverse items on the Confidence in Mathematics Scale administered to students in TIMSS 2011. They proved that the effect of reverse items differs across grade levels and countries. Furthermore, Schmitt and Allik (2005) concluded that reverse items were interpreted differently across nations based on the data collected from 53 nations by means of Rosenberg Self-Esteem Scale. In the same vein, Wong et al. (2003) reported that the problems associated with reverse items do not occur in the same manner in all cultures and languages.

Considering the cultural backgrounds that affect the functioning of reverse items and the limited studies on the effect of types of reverse items (antonym vs. negation) on measurements, the authors believe that research conducted in different cultures on the subject would contribute to the literature. Indeed, researchers mentioned that it would be relevant to explore this phenomenon in diverse countries (e.g., Salazar, 2015). From this point of view, the effect of reverse items created with negations and with antonyms on the measurements in a Turkish-speaking sample was examined. Also, the current research was carried out on high school students unlike those studies in which the effect of different types of reverse items on psychometric qualities was tested on primary school students, undergraduates and older adults. In light of all these expositions, this research expectedly contributes to the literature and offers important insights into the debate about reverse items, which has a long history and is still a hot

topic today. As the scales are fundamental data collection tool in a wide range of scientific disciplines, it is thought that the present paper will appeal an extensive audience and expands the existed knowledge about the problems associated with reverse items.

Within the scope of this specific research, three scale forms were identified: the first one includes only 10 positive items (Form-P); the second one is the combination of five positive items and five reverse items achieved with antonyms (Form-RA); and the third one, on the other hand, comprises five positive items and five reverse items created with negations (Form-RN). By comparing the three scale forms, this study sought answers to the following research questions:

1) Do the three scale forms differ in terms of (a) factorial structure, (b) concurrent validity, and (c) internal consistencies?
2) Do the scores of the respondents vary from one scale form to another?

2. METHOD

2.1. Participants

This study is conducted in Diyarbakır, a province in the southeast of Türkiye. Because research in the relevant literature (e.g., Dagnew, 2017; Geddes et al., 2010; Verešová & Malá, 2016) reveals that students' attitude toward school is significantly related to their academic achievement, a study group with participants from different achievement levels was created. While deciding on the achievement levels of the schools, the placement scores of the national high-stake exams applied were taken in order to select students for high schools in Turkey as a reference. After randomly choosing a school from low, medium, and high achievement levels, a total of 1166 students, 666 girls and 500 boys, aged between 14 and 19 (M=15.06, SD=1.05) took part in the sample.

2.2. Instruments

The instrument to be used in the research must be convertible into reversed items created with negations and antonym expressions, without changing the meaning and wording severity of the regular items it contains. When the Turkish literature was reviewed, such a scale was not found. On that account, instead of drawing upon a scale with tested psychometric properties in Turkish culture, a new measurement tool was generated in the study. In line with the research purpose, the three versions of a School Attitude Scale (SAS) in the study were developed: Form-P, Form-RA, and Form-RN. While constructing the scale forms, 14 positive items were initially written to measure the attitude toward school. These items were composed in such a way that they can be converted into reverse items with antonyms and negations. Then, the draft form consisting of positive items was sent to three experts. While two of these experts were from the field of psychological counseling and guidance, the other one was from the field of measurement and evaluation in education. Relying on the experts' opinions, the content validity indices (CVI) of the items were computed and values ranging from –.33 to 1.00 were obtained. Two items with a CVI of –.33, that is, which two out of three experts deemed inessential, were removed from the scale. Thus, 12 items remained. The CVI for the entire scale based on these 12 items was detected as .83.

Next, seven of the 12 items in the scale were converted into reverse items to generate Form-RA and Form-RN. For instance, the item "*I try to attend school regularly*" in Form-P was transformed to "*I skip school whenever I get the chance*" in Form-RA and "*It is not important for me to attend school regularly*" in Form-RN. Then, the formed reverse counterparts of seven items were sent to six experts who evaluated their equivalence. Three of the experts were PhD candidates in the field of educational sciences, and their bachelor's degree was in Turkish language teaching. The fourth expert was an associate professor in the field of measurement

and evaluation in education, and her bachelor's degree was also in Turkish language teaching. The fifth and sixth experts were professors, one from the field of psychological counseling and guidance, and the other from the field of curriculum and instruction. Experts judged each positive item and its reverse counterparts as "equivalent" and "not equivalent", and also offered certain suggestions to enhance the equivalence of items in the three forms. The Fleiss' kappa statistic for the agreement between the experts was .506. Four experts reported that two of the items in the three scale forms were not equivalent in terms of the response category to be endorsed by any participants. Therefore, the two items were removed in question from the scale. Moreover, two experts pointed out that the items in the reversed form were relatively strict for the item "*I think the school supports my personal development.*" So, this item was rearranged, which was initially stated as "*The school does not contribute to my personal development*" in Form-RN, as "*I do not think that the school contributes to my personal development.*" Similarly, this item was converted, which was originally stated as "*School is just a waste of time for my personal development*" in Form-RA, into the sentence "*I see school as a waste of time for my personal development.*" Thus, three scale forms were drawn up, each consisting of 10 items. The items in these forms were presented in Table 1.

Table 1. *Items in three different forms of the SAS.*

Item Number	Form–P	Form–RA	Form–RN
1	I believe schools are important institutions for the progress of societies.	I believe schools are important institutions for the progress of societies.	I believe schools are important institutions for the progress of societies.
2	I look forward to the opening of schools while on holiday.	When I'm on holiday, I get depressed as the time for schools to open approaches.	I don't want schools to reopen while I'm on holiday.
3	I believe that school helps us to be responsible individuals.	I believe that school helps us to be responsible individuals.	I believe that school helps us to be responsible individuals.
4	I try to attend school regularly.	I skip school whenever I get the chance.	It is not important for me to attend school regularly.
5	Learning new things at school makes me happy.	Learning new things at school makes me happy.	Learning new things at school makes me happy.
6	I go to school willingly.	I go to school reluctantly.	I wouldn't go to school if I could.
7	I think I learned a lot of things in school that will benefit me.	I think I learned a lot of things in school that will benefit me.	I think I learned a lot of things in school that will benefit me.
8	I think school is an enjoyable place.	I think school is a boring place.	I don't see school as an enjoyable place.
9	I think that school supports my personal development.	I see school as a waste of time for my personal development.	I do not think that the school contributes to my personal development.
10	I believe that all children/young people should go to school.	I believe that all children/young people should go to school.	I believe that all children/young people should go to school.

The table shows the commonality of items 1, 3, 5, 7, and 10 in all three forms. On the other side, Items 2, 4, 6, 8, and 9 (gray shaded in Table 1) were positive in Form-P, reversed with opposite meanings in Form-R, and reversed with negations in Form-RN. Also, Table 1 showed that negative items were given randomly order in the scale. The items were anchored with a five-point rating of *strongly disagree* (1), *disagree* (2), *somewhat agree* (3), *agree* (4) and *strongly agree* (5) in all three versions of the scale. Considering the possible effects of response option orders to participants' responses, the order of response options in ascending format (i.e., *strongly disagree* to *strongly agree*) was arranged in all three scale versions.

Before using the scale forms that were developed to collect the research data, a pilot study on Form-P was performed. To this end, Form-P was administered to 394 high school students other than the main sample. The pilot data were randomly divided into two parts to conduct exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). EFA and CFA were performed after examining the relevant assumptions. Since the multivariate normality assumption was violated, principal axis analysis in EFA and Satorra-Bentler robust maximum likelihood (MLM) in CFA were utilized as the estimation methods. Based on the EFA results, a single-factor structure was acquired where factor loadings ranged from .55 to .72 and the extracted variance was 42.50%. As a result of CFA, on the other hand, the factor loadings belonging to single-factor model varied between .44 and .77, and the fit statistics were within acceptable limits (χ^2/df = 3.01, RMSEA = .10, [90% CI = .080, .124], SRMR = .058, TLI = .89, and CFI = .91). Furthermore, McDonald's ω was .878 (95% CI = .860, .896), and item-total correlations were within .47 and .66 for the dataset obtained by combining the data files from which EFA and CFA were conducted.

2.3. Procedure

The data collection process was fulfilled in three applications because of the three forms. A complete counterbalanced design was used to control for order effects that could arise from the sequence of administration of the scale forms, thus enhancing the internal validity of the research. The scale forms were applied to the students with 10-day intervals. Figure 1 summarizes the process followed for collecting the research data.

Figure 1. *The process followed in data collection.*
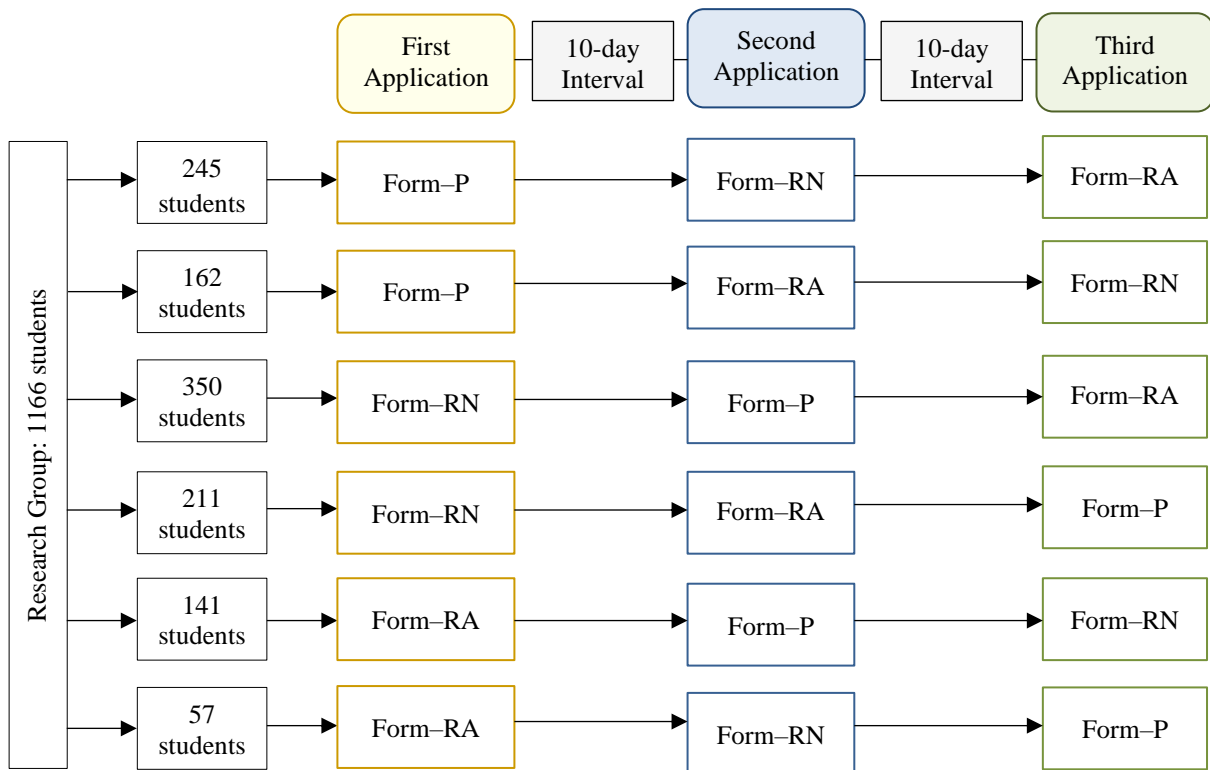
At the beginning of the scale forms, along with the demographics of gender and grade level, the students were asked the following: "*How many points would you give if you would rate your love of school between 1 and 10?*" The data obtained through this question was used to test and compare the concurrent validity of the three different versions of the SAS. The scales were administered to the students in pencil-and-paper format, in their actual classrooms, and in Turkish. Students' participation in the research was on a voluntary basis and they did not receive

any compensation. Furthermore, students were assured that their data would remain anonymous and would not be shared with any other person or institutions; however, they were required to use a nickname to match the three scale forms they responded. Fortunately, we did not encounter any students who refused to take part in the study in any of the classes where the administration was performed. In fact, data from 1329 participating students were collected. However, the 157 students who were absent in any of the three administrations and six students who had missing data in any scale forms although they participated in all the three applications were excluded. Consequently, a dataset of 1166 students was achieved. The compliance of the research with current ethical standards was approved by the Social and Humanities Ethics Committee of Dicle University, Diyarbakır, Türkiye.

2.4. Data Analysis

The analysis process started with the preparation of the data for analysis and the examination of the distribution characteristics. First, there was no missing value in the dataset. Because it is recommended in the literature to conduct EFA and CFA on different samples (Dawson, 2017; Fabrigar et al., 1999), the study sample was randomly split out into two halves before starting the analysis. Following this, the outliers was extracted from the datasets. As a result, 20 participants were removed from the sample used for EFA, leaving 563 students in this dataset. In addition, 26 participants were excluded from the dataset used for CFA, remaining 557 students in this dataset. In the next step, the skewness and kurtosis coefficients were examined to test the univariate normality and the Henze Zirkler tests for checking over multivariate normality. Notably, the skewness and kurtosis coefficients of the datasets were within ±1, and univariate normality was achieved. On the other hand, significant results of Henze Zirkler tests revealed that multivariate normality was violated. Hence, principal axis analysis in EFA and Satorra-Bentler robust maximum likelihood (MLM) in CFA were operated as the estimators.

In EFA, KMO values for Form-P, Form-RA, and Form-RN were .91, .88, and .86, respectively. Besides, Bartlett's test results were statistically significant for all three forms [χ^2_{Form-P} = 2040.49, $\chi^2_{Form-RA}$ = 2161.18, $\chi^2_{Form-RN}$ = 1579.70; df = 45; p < .001]. In conclusion, sample and data were satisfactory and continued with the factor analysis. Also, the parallel analysis method was employed to identify the number of factors to be extracted in EFA and the scree plots including parallel analysis results was presented in the Appendix.

Another evidence of validity that was investigated in the research was concurrent validity. Within this framework, the Pearson correlations between students' responses to the question of "*How many points would you give if you would rate your love of school between 1 and 10?*" and their mean scores from each scale version were calculated. To boot, the correlation coefficients obtained as a measure of concurrent validity were interpreted, and the variability of the three forms of the SAS with regard to concurrent validity was tested by analyzing the significance of the differences between the correlation coefficients.

To estimate the internal consistencies of the three scale versions and compare their reliability coefficients, the McDonald's ω was calculated for each form. Given that the presence of univariate normality, parametric tests were used to compare students' scores on three different versions of the SAS. Pearson correlation analysis was carried out to see the relationships between students' scores on three scale forms, and repeated measures ANOVA was executed to assess whether students' school attitude scores differ from one version of the scale to another. Mauchly's test of sphericity was checked before repeated measures ANOVA. Results revealed that the sphericity assumption had been met (Mauchly's W = .995, χ^2 = 5.15, df = 2, p > .005). Moreover, the eta squared value (η^2) was checked to appraise the effect size of the difference observed in ANOVA.

In the research, for detecting multivariate outliers and calculating the Henze Zirkler multivariate normality test, the web tool developed by Aybek (2021) with R software running in its background was employed. The significance of the differences between the correlation coefficients obtained for the concurrent validity of the three versions of the SAS was tested by means of the interface developed by Diedenhofen and Musch (2015). The functions in this web application are based on the tests implemented in the cocor package of the R programming language. All other analyses in the study were conducted in JASP 0.16.

3. RESULTS

The EFA outputs for the three versions of the SAS were primarily inspected. Table 2 summarizes the results of EFA.

Table 2. *EFA results for the three versions of the SAS.*

	Item Number	Form–P	Form–RA		Form–RN	
		Factor–1	Factor–1	Factor–2	Factor–1	Factor–2
Factor Loadings	1	.676	.740	.007	.693	.060
	2	.426	.106	.734	.098	.610
	3	.728	.807	.042	.704	.043
	4	.579	–.076	.413	–.196	.391
	5	.640	.628	–.063	.577	-.111
	6	.691	.068	.821	–.178	.610
	7	.792	.838	.023	.832	.048
	8	.518	.001	.801	.147	.662
	9	.665	–.302	.448	–.149	.468
	10	.604	.590	–.026	.608	.020
Variance Explained		40.90%	49.75%		41.20%	

Table 2 exhibits that Form-RA and Form-RN have a two-factor structure with positive items in one factor and reverse items in the other unlike the single-factor Form-P. Moreover, the variances explained are close to each other in Form-P and Form-RN. On the other hand, the extracted variance in Form-RA is clearly higher than that in the other two forms. Another remarkable result in Table 2 is as follows: Form-RA and Form-RN overlap in terms of the dispersion of the items to the factors, but the items' factor loadings are generally higher in Form-RA than those in Form-RN.

Under the CFA, the factor structures that emerged in the EFA for all three scale versions were tested. However, the unidimensional model in addition to the two-factor structure in Form-RA and Form-RN was also examined, because the SAS was prepared by forecasting a single-factor structure and Form-P was unidimensional. Table 3 comprises the fit indices reported in the CFA and shows the critical values for the fit indices in the second column.

Table 3. *Fit statistics obtained in CFA for the three versions of the SAS.*

Fit Statistics	Critical values	Form–P	Form–RA		Form–RN	
			Single-factor model	Two-factor model	Single-factor model	Two-factor model
χ^2/df	≤ 5[a]	3.819	13.654	2.510	7.419	3.105
RMSEA	$\leq .10$[b]	.071 90% CI (.059, .084)	.151 90% CI (.139, .163)	.052 90% CI (.038, .066)	.107 90% CI (.095, .120)	.061 90% CI (.048, .075)
SRMR	$\leq .10$[c]	.042	.095	.037	.068	.042
MFI	$\geq .90$[c]	.915	.672	.955	.817	.938
CFI	$\geq .90$[c]	.944	.784	.975	.851	.953
NNFI (TLI)	$\geq .90$[c]	.928	.722	.967	.808	.937

[a] Marsh & Hocevar (1985), [b] Meyers at al. (2006), [c] Pituch & Stevens (2016)

Table 3 illustrates that the fit indices of the single-factor model are out of the critical values for Form-RA and Form-RN. By contrast, all fit statistics belonging to two-factor model are within acceptable limits in both Form-RA and Form-RN. Apparently, model-data fit was provided only in Form-P for single-factor analysis. Figure 2 represents the measurement models in which the model-data fit was achieved for the three versions of the SAS. It presents that, just like in EFA, the factor loadings of the items are generally higher in Form-RA than they are in Form-RN. Additionally, the correlation between the factor with positive items and the factor with reverse items is higher in Form-RN than it is in Form-RA.

Figure 2. *Measurement models for (a) Form-P, (b) Form-RA, and (c) Form-RN of the SAS.*
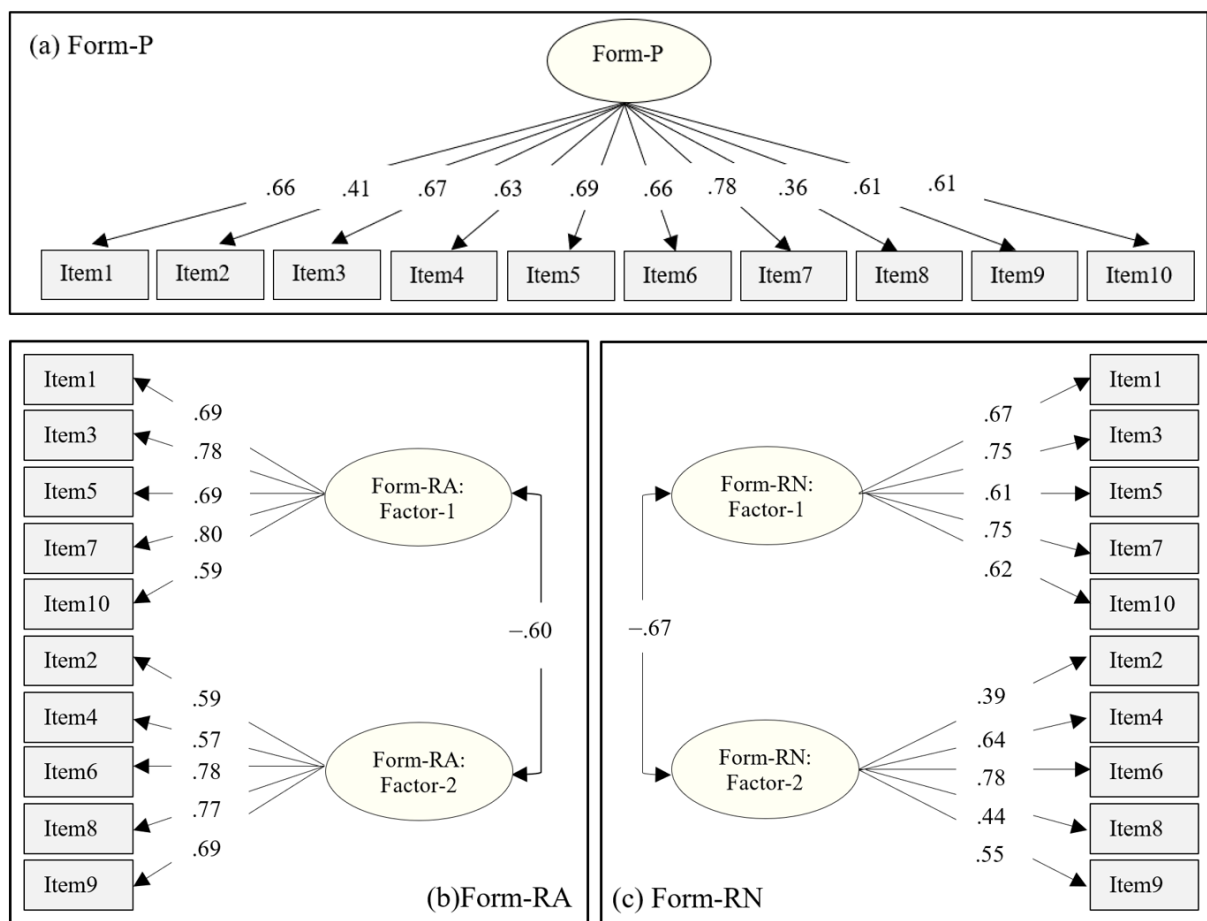
Having completed EFA and CFA, the concurrent validity results were analyzed. The correlation coefficients from concurrent validity of the three versions of the SAS were calculated as .592 (95% CI [.552, .628], $p < .001$), .644 (95% CI [.608, .677], $p < .001$), and .611 (95% CI [.573, .647], $p < .001$) for Form-P, Form-RA, and Form-RN, respectively. Salkind (2010) recommends a series of range for interpreting correlation coefficients as very weak ($<.20$), weak (.20–.40), moderate (.40–.60), strong (.60–.80), and very strong ($>.80$). Accordingly, the obtained correlation coefficients point to strong relationships for the concurrent validity of all three forms of the SAS. As a striking result, the correlation coefficient of Form-RA was calculated higher in proportion to the other two forms of the SAS. When the significance of the differences between acquired correlation coefficients was tested via Pearson and Filon's z statistic, there was a significant difference between Form-P and Form-RA ($z = -2.577$, 95% CI [$-.092$, $-.013$], $p < .01$). Conversely, there were no significant differences between Form-P and Form-RN ($z = -.893$, 95% CI [$-.061$, $.023$], $p > .05$) and between Form-RA and Form-RN ($z = 1.646$, 95% CI [$-.006$, $.073$], $p > .05$) in terms of their concurrent validity evidence.

Subsequent to comparing the three versions of the SAS in terms of validity proofs, the reliability estimations were performed. McDonald's ω coefficients of the measures were calculated as .861 (95% CI [.849, .874]), .855 (95% CI [.842, .868]), and .816 (95% CI [.800, .832]) for Form-P, Form-RA, and Form-RN, respectively. Evidently, the reliability values estimated for all three versions of the SAS are over .80. DeVellis (2017) proposed the following ranges when judging the reliability: below .60, unacceptable; between .60 and .65, undesirable; between .65 and .70, minimally acceptable; between .70 and .80, respectable; and between .80 and .90, very good. On the other hand, the coefficients above .90 signify redundant items and mean that the scale should be shortened. Relying on the intervals listed, all three scale forms yield quite reliable measurements. What stands out in reliability analysis results is that the internal consistency coefficient of Form-RN is lower than those of the other two forms.

After checking the three versions of the SAS against their psychometric properties, the respondents' scores on the different forms were compared. Based on the results of Pearson correlation analysis, there were significant relationships between students' scores in the three scale forms. To clarify, the correlation coefficients were estimated as .695 (95% CI [.633, .724]) between Form-P and Form-RA, .647 (95% CI [.611, .680]) between Form-P and Form-RN, and .702 (95% CI [.671, .731]) between Form-RA and Form-RN. Although there are strong positive relationships between students' scores in the three scale forms, the correlation between Form-P and Form-RN is relatively low. Eventually, repeated measures ANOVA was conducted to establish whether the students' mean scores differ across three scale forms, and the results are shown in Table 4.

Table 4 denotes that students' school attitude scores differ significantly from one scale form to the other ($F_{(2, 2238)} = 4,36$, $p < .05$, $\eta^2 = .004$). Although post-hoc test results revealed that the significant differences were observed between Form-RN and the other two versions, a closer inspection of the results suggests that the eta squared value, which is a measure of effect size, is quite low. In other words, Cohen (1988) proposed the following guidelines for interpreting the eta squared values: .01, small effect; .06, moderate effect; and .14, large effect (cited in Pallant, 2005). These benchmarks notify that the statistically significant differences detected are minor in practice.

Table 4. *Results of repeated measures ANOVA regarding the difference between the means across three versions of the SAS.*

Scale Version	Mean	SD	F(2, 2238)	Post-Hoc (LSD)	η^2
Form–P	3.47	.78			
Form–RA	3.46	.82	4.36[*]	Form–RN > Form–RA Form–RN > Form–P	.004
Form–RN	3.51	.76			

[*] $p<.05$

4. DISCUSSION and CONCLUSION

In the present research, the impacts of reverse items created with negations and with antonyms on the psychometric properties and respondents' scale scores were evaluated. The EFA results exposed that Form-P has a unidimensional structure, while Form-RN and Form-RA have a two-factor structure in which positive and reverse items are in separate factors. Furthermore, CFA outputs verified that the model where positive and reverse items were considered as distinct factors fit the data better than the single-factor solution for both Form-RA and Form-RN. More exactly, regardless of which reversal strategy is used, the reverse items caused an artificial factor that can also be called a method factor in addition to the trait factor. Consistent with this result, many studies in the literature revealed that reverse items in scales measuring a single unitary construct distort the factor structure of the scale by generating a separate dimension, which leads to an unintentionally multidimensional factor solution (Bulut & Bulut; 2022; DiStefano & Motl, 2006; Dunbar et al., 2000; Hazlett-Stevens et al., 2004; Herche & Engelland, 1996; Knight et al., 1998; Pilotte & Gable, 1990; Spector et al., 1997).

As a result of EFA, the factor loadings in most of the items of Form-RA and, in a parallel manner, its extracted variance were higher than those of the other scale forms. Besides, CFA results showed that Form-RA had higher factor loadings for many, if not all, items. The concurrent validity results were also in this direction and were higher for Form-RA. These results hint that the reverse items created with negations do not serve the purpose of increasing the validity of the measurements. The reverse items formed with antonyms, on the other hand, partially improve the validity even though they jeopardize the unidimensionality of the scale.

The research result regarding the variance ratio is contrary to that of Suárez-Alvarez et al. (2018). To sum up, the explained variance ratio in Form-RA was higher than that in Form-P. Conversely, Suárez-Alvarez et al. (2018) ascertained that the percentage of explained variance in the version containing both positive and reverse items with antonyms was lower than the form consisting of only positive items. This may be related to language properties, to the group in which the study was conducted, or to the differences between the scale forms used. As this research was conducted on a Turkish-speaking sample and that of Suárez-Alvarez et al. (2018) was carried out on a Spanish-speaking group, the aforementioned discrepancy could be attributed to language features. Indeed, Schmitt and Allik (2005) and Hooper et al. (2013) disclosed that reverse items were interpreted differently across countries. The mentioned contradiction may also be associated with the sample characteristics because the factors such as reading achievements (Bulut & Bulut, 2022; Michaelides, 2019), age levels (Bulut, 2021; Marsh 1986) and cognitive abilities (Gnambs & Schroeders, 2020) of the samples can affect their responses to the reverse items. Weems et al. (2003) analyzed profiles of individuals who respond inconsistently to positive and reverse items on Likert scales and remarked that certain characteristics of the participants made them more likely to generate differential patterns of responses to the positive and reverse items. In addition to the issues listed, the use of different scales in this paper and in Suárez-Alvarez et al.'s (2018) research might have caused discordant results concerning the explained variance ratio.

The results of the study indicated that internal consistency coefficient of Form-RN was lower than that of the other two versions. The facts that the reliability coefficient of Form-RN is lower than that of Form-P are coherent with those found by other researchers (Bourque & Shen, 2005; Carlson, 2011; Coleman, 2013; Ebesutani et al., 2012; Johnson et al., 2004; Roszkowski & Soven, 2010; Salazar, 2015; Suárez-Alvarez, 2018). Accordingly, we can infer that reverse items created with negations are more open to measurement errors. This is thought to reflect the fact that negatively worded phrases (i.e., reverse items) require additional cognitive efforts and usually cause confusion for respondents (Chyung et al., 2018). Justifiably, the foregoing situation is not true for reverse items built with antonyms as the reliability value of Form-RA

is similar to that of Form-P. The research results, in which Form-RA produces more valid and reliable measurements compared with Form-RN, accord with the results obtained in earlier studies. Weijters and Baumgartner (2012) executed a comprehensive literature review on reverse items and stated that negations are problematic at the judgment stage because they require additional mental steps during item processing. Further, they posited that negations should be employed sparingly and reverse items created with antonyms may be more beneficial than those with negations. Likewise, Zhang et al. (2016) criticized the reverse items derived from negations as follows: this type of reverse items might engender careless responding or judgmental complication for some individuals. Some respondents may fail to notice the presence of a negative particle in the item stem (e.g., misread I am not happy as I am happy), making errors due to carelessness. Moreover, a reverse item generated with negation makes it more difficult for the respondents to judge whether the item content is match with his or her own beliefs. Considering the results respecting "antonyms vs. negations" comparison in the light of the literature, arguably, using antonym is a better strategy for creating reverse items.

The analysis of comparing the scale scores of the participants in the three forms denoted that there were positive and strong correlations between the scores on the different forms. The correlation coefficients obtained reflect that the relative agreement between the scores of the participants in the three scale forms is high. Accordingly, the participants have, by and large, similar rankings in terms of their scores in the three forms, but there is no exact identical ranking. Moreover, the mean scores in Form-RN were significantly higher than those of the other two scale versions, and the difference between the scores of Form-P and Form-RA was not significant. The calculated effect size set out that the statistically significant difference was practically quite small. In simpler terms, even if there were not large differences between participants' mean scores on three scale forms, there was no full absolute agreement, either. On the basis of these results, it can be conceivably hypothesized that in cases where small score differences are important, the decision to be taken about the participants may change depending on the item wording effect. This inference matches with the ideas of Schotte et al. (1996) who stated that not only *what is asked* but also *how it is asked* influences the responses of participants in self-report instrument.

Comparing the study's results regarding the effect of reverse items on scale scores with the results of previous studies in the literature, a rather contradictory picture appears. First and foremost, similar to this research, Benson and Hocevar (1985), Weems et al. (2006), Hughes (2009), and Locker et al. (2013) established that reverse items change the scale scores of the respondents. Conversely, Greenberger et al. (2003) illustrated that item wording did not influence participants' mean scores, and Zhang et al. (2016) specified that the item means were similar across four scale versions with a different composition in terms of positive and reverse items. Second, a significantly higher mean score was detected in Form-RN than those of the other two scale versions. Parallel to this result, Taylor and Bowers (1972) (as cited in Schriesheim & Hill, 1981), Schotte et al. (1996), and Suárez-Alvarez et al. (2018) uncovered that reverse item generates a higher mean response than does the positive counterpart. On the other hand, Stewart and Frye (2004), Locker et al. (2013), Vigil-Colet et al. (2020), and Dueber et al. (2021) found that reverse items yield lower mean scores after coding them in the same direction of positive items. Hence, there is no consistency among the studies on whether the reverse items differentiate the scale scores, and if so, in which direction. These differences may originate because of the varieties between the samples of the studies and the measurement tools used. In particular, it seems possible that differences in mean scores due to reverse items are in opposite direction in scales where the measured attribute is negative in nature (e.g., depression) and in scales where it is positive (e.g., happiness).

The combination of the research results provides important suggestions for practice. Before anything else, an instrument developer should be cautious about the usage of a mix of positive and reverse items when constructing a scale. He or she must make up his or her mind whether reverse items are really necessary and avoid using them unless there is a clear justification. Considering that negated reverse items will attenuate the psychometric properties and elicit a difference, albeit small, in the scale scores of the participants, he or she should prefer items created with antonyms instead of items built with negations in cases where he or she utilizes reverse items. In addition, because item phrasing may differentiate scale scores, researchers should be careful when comparing scale scores that have diverse item word formats. Nonetheless, research limitations that restrict the generalizability of these implications must be inculcated. Corroborating studies are required to reach a more decisive conclusion on this issue.

4.1. Limitations and Future Avenues Research

The present study has certain limitations. First, this study was conducted on a Turkish-speaking sample, and the way reverse items are formed differs from one language to another. For example, in Turkish, the verb is at the end of the sentence, and suffixes in the form of "-me, -ma" are added to the verb to achieve reverse items with negations. In English, on the other hand, the verb is after the subject, and the words like "not or no" are used before the verb to form negated reverse items. In other languages the situation is likely to be different. In this sense, similar studies should be conducted in other languages regarding the impact of reverse items. Second, participants of this study were exclusively high school students whose age mean was approximately 15.06. Future research can focus on different age groups, as the effect of reverse items varies depending on the age and cognitive development of the respondents. Finally, the current study employed the SAS to investigate the influences of reverse items. Therefore, researchers must utilize other instruments when replicating this 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**: Dicle University, 27.12.2021-21445.

Authorship Contribution Statement

Mustafa İlhan: Investigation, creating of the instruments, data analysis, resources, visualization, and writing-original draft. **Neşe Güler**: Investigation, creating of the instruments, receiving experts' opinions and writing-original draft. **Gülşen Taşdelen Teker**: Creating of the instruments, resources and writing-original draft. **Ömer Ergenekon**: Data collection and writing-original draft.

Orcid

Mustafa İlhan 🆔 https://orcid.org/0000-0003-1804-002X
Neşe Güler 🆔 https://orcid.org/0000-0002-2836-3132
Gülşen Taşdelen Teker 🆔 https://orcid.org/0000-0003-3434-4373
Ömer Ergenekon 🆔 https://orcid.org/0000-0001-9964-5535

REFERENCES

Ahlawat, K.S. (1985). On the negative valence items in self-report measures. *The Journal of General Psychology*, *112*(1), 89–99. https://doi.org/10.1080/00221309.1985.9710992

Aybek, E.C. (2021). *Data preparation for factor analysis*. URL: https://shiny.eptlab.com/dp2fa/

Bandalos, D.L. (2018). *Measurement theory and applications for the social sciences*. The Guilford.

Benson, J., & Hocevar, D. (1985). The impact of item phrasing on the validity of attitude scales for elementary school children. *Journal of Educational Measurement, 22*(3), 231–240. https://doi.org/10.1111/j.1745-3984.1985.tb01061.x

Bergstrom, B.A., & Lunz, M.E. (1998, 13–17, April). *Rating scale analysis: Gauging the impact of positively and negatively worded items*. Annual Meeting of the American Educational Research Association, San Diego.

Boley, B.B., Jordan, E., & Woosnam, K.M. (2021). Reversed polarity items in tourism scales: Best practice or dimensional pitfall? *Current Issues in Tourism*, *24*(4), 1-13. https://doi.org/10.1080/13683500.2020.1774517

Bolt, D., Wang, Y.C., Meyer, R.H., & Pier, L. (2020). An IRT mixture model for rating scale confusion associated with negatively worded items in measures of social-emotional learning. *Applied Measurement in Education*, *33*(4), 1-18. https://doi.org/10.1080/08957347.2020.1789140

Brace, I., & Bolton, I. (2022). *Questionnaire design: How to plan, structure and write survey material for effective market research* (5th ed.). Kogan Page.

Bulut, H.Ç. (2021). Item wording effects in psychological measures: Do early literacy skills matter? *Journal of Measurement and Evaluation in Education and Psychology, 12*(3), 239-253. https://doi.org/10.21031/epod.944067

Bulut, H.Ç., & Bulut, O. (2022). Item wording effects in self-report measures and reading achievement: Does removing careless respondents help? *Studies in Educational Evaluation*, *72*(2). https://doi.org/10.1016/j.stueduc.2022.101126

Carlson, M., Wilcox, R., Chou, C.-P., Chang, M., Yang, F., Blanchard, J., Marterella, A., Kuo, A., & Clark, F. (2011). Psychometric properties of reverse-scored items on the CES-D in a sample of ethnically diverse older adults. *Psychological Assessment*, *23*(2), 558–562. https://doi.org/10.1037/a0022484

Chamberlain, V.M., & Cummings, M.N. (1984). Development of an instructor/course evaluation instrument. *College Student Journal*, *18*(3), 246–250.

Chyung, S.Y.Y., Barkin, J.R., & Shamsy, J.A. (2018). Evidence-based survey design: The use of negatively worded items in surveys. *Performance Improvement*, *57*(3), 16–25. https://doi.org/10.1002/pfi.21749

Coleman, C.M. (2013). *Effects of negative keying and wording in attitude measures: A mixed-methods study* [Unpublished doctoral dissertation, James Madison University]. https://commons.lib.jmu.edu/diss201019/73/

Conrad, K.J., Wright, B.D., McKnight P., McFall, M., Fontana, A., & Rosenheck, R. (2004). Comparing traditional and Rasch analyses of the Mississippi PTSD Scale: Revealing limitations of reverse-scored items. *Journal of Applied Measurement*, *5*(1), 15–30.

Coolican, H. (2013). *Research methods and statistics in psychology* (5th ed.). Routledge.

Dagnew, A. (2017). The relationship between students' attitudes towards school, values of education, achievement motivation and academic achievement in Gondar secondary schools, Ethiopia. *Research in Pedagogy*, *7*(1), 30–42. https://doi.org/10.17810/2015.46

Dawson, J. (2017). *Analysing quantitative survey data for business and management students*. Sage.

DeVellis, R. (2017). *Scale development: Theory and applications* (4th ed.). Sage.

Diedenhofen, B., & Musch, J. (2015). Cocor: A comprehensive solution for the statistical comparison of correlations. *PLoSONE*, *10*(4), 1-12. https://doi.org/10.1371/journal.pone.0121945

Dooden, H. (2014). The effects of positively and negatively worded items on the factor structure of the UCLA loneliness scale. *Journal of Psychoeducational Assessment*, *33*(3), 259–267. https://doi.org/10.1177/0734282914548325

Dueber, D.M., Toland, M.D., Lingat, J.E., Love, A.M.A., Qiu, C., Wu, R., & Brown, A.V. (2021). To Reverse item orientation or not to reverse item orientation, that is the question. *Assessment*, *29*(7), 1422–1440. https://doi.org/10.1177/10731911211017635

Dunbar, M., Ford, G., Hunt, K., & Der, G. (2000). Question wording effects in the assessment of global self-esteem. *European Journal of Psychological Assessment*, *16*(1), 13–19. https://doi.org/10.1027//1015-5759.16.1.13

Ebesutani, C., Drescher, C.F., Reise, S.P., Heiden, L., High, T.L., Damon, J.D., & Young, J. (2012). The loneliness questionnaire-short version: An evaluation of reverse- worded and non-reverse-worded items via item response theory. *Journal of Personality Assessment*, *94*(4), 427–437. https://doi.org/10.1080/00223891.2012.662188

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. http://dx.doi.org/10.1037/1082-989X.4.3.272

Finney, S.J. (2001). *A comparison of the psychometric properties of negatively and positively worded questionnaire items* (Publication No. 3009716) [Doctoral dissertation, University of Nebraska]. ProQuest Dissertations & Theses Global.

Furr, R.M. (2018). *Psychometrics: An introduction* (3rd ed.). Sage.

Geddes, J.D., Murrell, A.R., & Bauguss, J. (2010). Childhood learning: An examination of ability and attitudes toward school. *Creative Education*, *1*(3), 170-183. https://doi.org/10.4236/ce.2010.13027

Gnambs, T., & Schroeders, U. (2020). Cognitive abilities explain wording effects in the Rosenberg Self-Esteem Scale. *Assessment*, *27*(2), 404-418. https://doi.org/10.1177/1073191117746503

Greenberger, E., Chen, C., Dmitrieva, J., & Farruggia, S.P. (2003). Item-wording and the dimensionality of the Rosenberg Self-Esteem Scale: Do they matter? *Personality and Individual Differences*, *35*(6), 1241-1254. https://doi.org/10.1016/S0191-8869(02)00331-8

Guyatt, G.H., Cook, D.J., King, D., Norman, G.R., Kane, S.L., & Van Ineveld, C. (1999). Effect of the framing of questionnaire items regarding satisfaction with training on residents' responses. *Academic Medicine*, *74*(2), 192-194. https://doi.org/10.1097/00001888-199902000-00018

Hazlett-Stevens, H., Ullman, J.B., & Craske, M.G. (2004). Factor structure of the Penn State worry questionnaire: Examination of a method factor. *Assessment*, *11*(4), 361–370. https://doi.org/10.1177/1073191104269872

Herche, J., & England, B. (1996). Reversed-polarity items and scale unidimensionality. *Journal of the Academy of Marketing Science*, *24*(4), 366-374. https://doi.org/10.1177/0092070396244007

Hooper, M., Arora, A., Martin, M.O., & Mullis, I.V.S, (2013, 24–25 June). *Examining the behavior of "reverse directional" items in the TIMSS 2011 context questionnaire scales* [Conference presentation]. 5th IEA International Research Conference. National Institute of Education, Nanyang Technological University, Singapore. https://www.iea.nl/sites/default/files/2019-04/IRC-2013_Hooper_etal.pdf

Hughes, G.D. (2009). The impact of incorrect responses to reverse-coded survey items. *Research in the Schools*, *16*(2), 76–88.

Ilhan, M., & Guler, N. (2017). The number of response categories and the reverse scored item problem in Likert-type scales: A study with the Rasch model. *Journal of Measurement and Evaluation in Education and Psychology, 8*(3), 321-343. http://dx.doi.org/10.21031/epod.321057

JASP Team (2022). *JASP (Version 0.16.3)* [Computer software].

Johnson, J.M., Bristow, D.N., & Schneider, K.C. (2004). Did you not understand the question or not? An investigation of negatively worded questions in survey research. *Journal of Applied Business Research*, *20*(1), 75–86. https://doi.org/10.19030/jabr.v20i1.2197

Kam, C.C.S. (2023). Why do regular and reversed items load on separate factors? Response difficulty vs. item extremity. *Educational and Psychological Measurement*, *83*(6), 1085–1112. https://doi.org/10.1177/00131644221143972

Karandashev, V. (2021). *Cultural models of emotions*. Springer.

Kline, T.J.B. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage.

Knight, R.G., Chisholm, B.J., Marsh, N.V., & Godfrey, H.P.D. (1998). Some normative, reliability, and factor analytic data for the revised UCLA Loneliness Scale. *Journal of Clinical Psychology*, *44*(2), 203-206. https://doi.org/10.1002/1097-4679(198803)44:2<203::AID-JCLP2270440218>3.0.CO;2-5

Kula Kartal, S. (2021). Examining scale items in terms of method effects based on the bifactor item response theory model. *Kastamonu Education Journal*, *29*(1), 201–209. https://doi.org/10.24106/kefdergi.708968

Kula Kartal, S., Aybek, E.C., & Yaşar, M. (2022). Investigating the wording effect in scales based on the different dimension reduction techniques. *Journal of Uludag University Faculty of Education*, *35*(1), 44–67. https://doi.org/10.19171/uefad.1033284

Locker, D., Jokovic, A., & Allison, P. (2013). Direction of wording and responses to items in oral health related quality of life questionnaires for children and their parents. *Community Dentistry Oral Epidemiology*, *35*(4), 255–262. http://dx.doi.org/10.1111/j.1600-0528.2007.00320.x

Marsh, H.W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, *22*(1), 37-49. https://doi.org/10.1037/0012-1649.22.1.37

Marsh, H.W. (1996). Positive and negative global self- esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology*, *70*(4), 810–819. https://doi.org/10.1037/0022-3514.70.4.810

Marsh, H.W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, *97*(3), 562–582. https://doi.org/10.1037/0033-2909.97.3.562

Mayerl, J., & Giehl, C. (2018). A closer look at attitude scales with positive and negative items response latency perspectives on measurement quality. *Survey Research Methods*, *12*(3), 193–209. https://doi.org/10.18148/srm/2018.v12i3.7207

Meyers, L.S, Gamst, G., & Guarino, A.J. (2006). *Applied multivariate research: Design and interpretation*. Sage.

Michaelides, M.P. (2019). Negative keying effects in the factor structure of TIMSS 2011 motivation scales and associations with reading achievement. *Applied Measurement in Education*, *32*(4), 365–378. https://doi.org/10.1080/08957347.2019.1660349

Pallant, J. (2005). *SPSS survival manual* (2nd ed.). Allen & Unwin.

Pilotte, W.J., & Gable R.K. (1990). The impact of positive and negative item stems on the validity of a computer anxiety scale. *Educational and Psychological Measurement*, *50*(3), 603–610. http://dx.doi.org/10.1177/0013164490503016

Podsakoff, P.M., MacKenzie, S.B., Lee J.-Y., Podsakoff, N.P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879-903. https://doi.org/10.1037/0021-9010.88.5.879

Roszkowski, M.J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, *35*(1), 113-130. http://dx.doi.org/10.1080/02602930 802618344

Salazar, M.S. (2015). The dilemma of combining positive and negative items in scales. *Psicothema*, *27*(2), 192–200. https://doi.org/10.7334/psicothema2014.266

Pituch, K.A., & Stevens, J.P. (2016). *Applied multivariate statistics for the social sciences* (6th ed.). Routledge.

Schmitt, D.P., & Allik, J. (2005). Simultaneous administration of the Rosenberg self-esteem scale in 53 nations: Exploring the universal and culture-specific features of global self-esteem. *Journal of Personality and Social Psychology*, *89*(4), 623-642. https://doi.org/1 0.1037/0022-3514.89.4.623

Schotte, C.K.W., Maes, M., Cluydts, R., & Cosyns, P. (1996). Effects of affective-semantic mode of item presentation in balanced self-report scales: Biased construct validity of the Zung self-rating depression scale. *Psychological Medicine*, *26*(6), 1161–1168. http://dx.doi.org/10.1017/s0033291700035881

Schriesheim, C.A., Eisenbach, R.J., & Hill, K.D. (1991). The effect of negation and polar opposite item reversals on questionnaire reliability and validity: An experimental investigation. *Educational and Psychological Measurement*, *51*(1), 67-78. https://doi.or g/10.1177/0013164491511005

Schriesheim, C.A., & Hill, K.D. (1981). Controlling acquiescence response bias by item reversals: The effect on questionnaire validity. *Educational and Psychological Measurement*, *41*(4), 1101–1114. http://dx.doi.org/10.1177/001316448104100420

Spector, P.E., Van Katwyk, P.T., Brannick, M.T., & Chen, P.Y. (1997). When two factors don't reflect two constructs: How item characteristics can produce artifactual factors. *Journal of Management*, *23*(5), 659–677. https://doi.org/10.1016/S0149-2063(97)90020-9

Stauesberg, M. (2002). Survey and questionnaires. In S. Engler & Micheal Stausber (Eds,), *The Routledge handbook of research methods in the study of religion* (pp. 461–482). Routledge.

Stewart, T.J., & Frye, A.W. (2004). Investigating the use of negatively phrased survey items in medical education settings: Common wisdom or common mistake? *Academic Medicine*, *79*(10), 18–20. https://doi.org/10.1097/00001888-200410001-00006

Suárez-Alvarez, J., Pedrosa, I., Lozano, L.M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, *30*(2), 149–158. https://doi.org/10.7334/psicothema2018.33

Swain, S.D., Weather, D., & Niedrich, R.W. (2008). Assessing three sources of misresponse to reversed Likert items. *Journal of Marketing Research, 45*(1)*, 116-131. https://doi.org/1 0.1509/jmkr.45.1.116

van Sonderen, E., Sanderman, R., & Coyne, J.C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PLoS ONE*, *8*(7), 1-7. https://doi. org/10.1371/journal.pone.0068967

Verešová, M., & Malá, D. (2016, 11–15October). *Attitude toward school and learning and academic achievement of adolescents* [Conference Presentation]. 7th International Conference on Education and Educational Psychology. https://www.europeanproceedin gs.com/article/10.15405/epsbs.2016.11.90

Vigil-Colet, A., Navarro-González, D., & Morales-Vives, F. (2020). To reverse or to not reverse Likert-type items: That is the question. *Psicothema*, *32*(1), 108-114. https://doi.org/10.7 334/psicothema2019.286

Weems, G.H., Onwuegbuzie, A.J., & Lustig, D. (2003) Profiles of respondents who respond inconsistently to positively- and negatively-worded items on rating scales. *Evaluation & Research in Education*, *17*(1), 45–60. http://dx.doi.org/10.1080/14664200308668290

Weems, G.H., Onwuegbuzie, A.J., & Collins, K.M.T. (2006) The role of reading comprehension in responses to positively and negatively worded items on rating scales. *Evaluation & Research in Education*, *19*(1), 3-20. https://doi.org/10.1080/09500790608668322

Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research*, *49*(5), 737-747. https://doi.org/10.1509/jmr.11.0368

Wong, N., Rindfleisch, A., & Burroghs, J.E. (2003). Do reversed-worded items confound measures in cross-cultural consumer research? The case of material values scale. *Journal of Consumer Research, 30*(1), 72–91. https://doi.org/10.1086/374697

Zhang, X., Noor R., & Savalei, V. (2016) Examining the effect of reverse worded items on the factor structure of the need for cognition scale. *PLoS ONE*, *11*(6), 1-15. https://doi.org/10.1371/journal.pone.0157795

APPENDIX

Scale version	Scree plot with parallel analysis results

Form—P
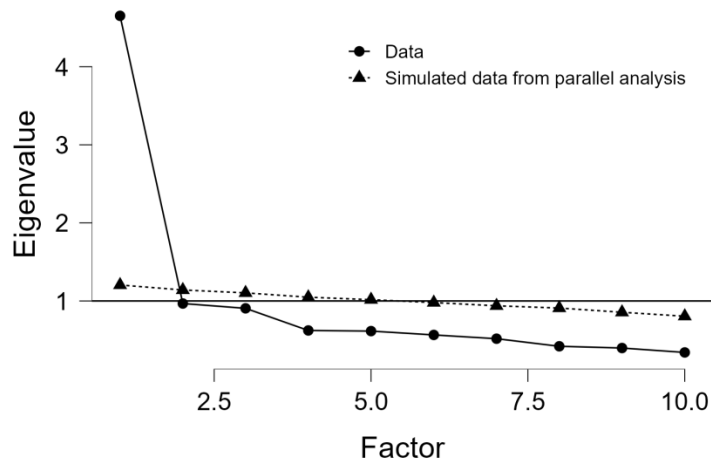
Form—RA
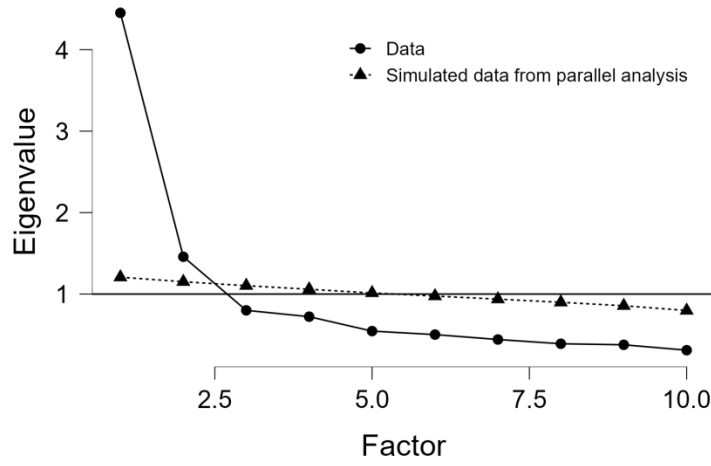
Form—RN
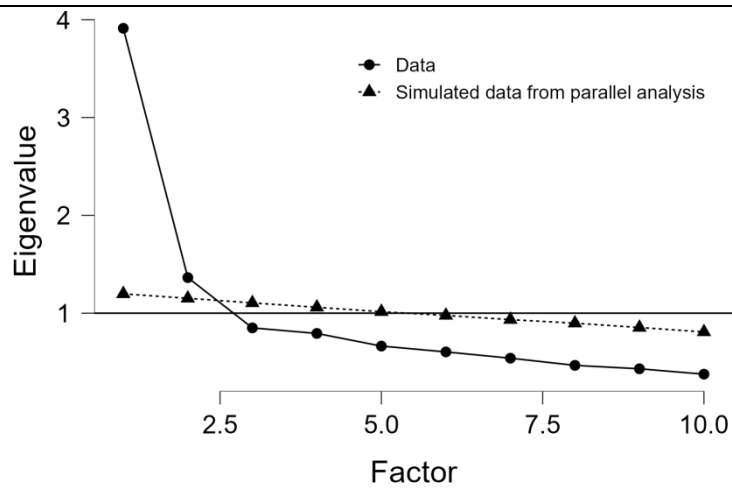
The examination of mediating role of distributed leadership in the relationship between school structure and accountability

Mustafa Orhan [1,*], **Tuncay Yavuz Özdemir** [2]

[1]Erzurum Directorate of National Education, Erzurum, Türkiye
[2]Fırat University, Faculty of Education, Department of Educational Science, Elazığ, Türkiye

Abstract: The aim of this study was to investigate the relationship between school structure, distributed leadership and accountability of school administrators. Relational survey design was adopted in the study. 444 elementary school teachers working in Aziziye, Palandöken and Yakutiye in Erzurum participated in the study. In sample selection, stratified sampling method was used. In data collection, the Enabling School Structure Scale, the Distributed Leadership Scale and the Accountability Behavior Scale for School Administrators were used. The data were analyzed using SPSS 23.0 for preliminary statistical analyses, LISREL 8.80 for Structural Equation Modeling (SEM), and the PROCESS macro for SPSS v3.3 for mediation analyses. In data analysis, Pearson Product-Moment Correlation Analysis, Bootstrap Analysis, and SEM were used. Furthermore, a mediation analysis was conducted to investigate whether distributed leadership played a mediating role in the relationship between school structure and accountability. The results revealed that coercive bureaucracy had a negative effect on accountability and distributed leadership while enabling bureaucracy had a positive effect on accountability and distributed leadership. Furthermore, the analysis revealed a positive effect of distributed leadership on accountability. This study revealed the mediating role of distributed leadership in the effect of coercive and enabling bureaucracy on accountability. In other words, it was found that the coercive and enabling bureaucracy had direct and indirect effects on accountability. Finally, a number of recommendations were made to educators, policy makers and researchers based on these findings.

1. INTRODUCTION

Organizational structure plays a remarkable role in decisions, communication, and performance of employees and managers. This structure enables managers to effectively allocate responsibilities, distribute decision-making authority, coordinate and control the organization's work, and ensure employees are accountable for their actions. By doing so, it supports the organization in achieving its goals. On the other hand, a poorly designed structure may lead to great waste, confusion and frustration for employees, resource providers and users (Bovée & Till, 2012, p.166). Similar to all organizations, schools also have a structure (Sinden et al., 2004,

p.464). Organizational structure is the formal features or continuing functioning of a school. Structure refers to the relationship between different roles established to achieve educational goals (Miskel, 1979, p.5; Miskel et al., 1979, p.100). It is necessary to establish a harmony between elements such as purpose, environment, technology, people and technology in the organization of the school structure. Some of the school-related problems derive from inappropriate organizational structures and systems (Şişman, 2019, p.207). One of the main elements of more effective schools is that they have a school structure that enables participants to do their work more creatively, collaboratively and professionally (Hoy & Sweetland, 2007, pp.362-363).

Schools need enabling structures rather than coercive ones. In an enabling structure, the authority hierarchy, rules, and procedures serve as mechanisms to support teachers, rather than tools to increase the principal's power. In a coercive structure, teacher behavior is closely monitored and strictly regulated. Accordingly, within a coercive structure, the authority hierarchy, rules, and regulations are employed to enforce compliance, maintain control, and penalize deviations (Hoy, 2003, p.90-91). Furthermore, schools typically possess bureaucratic structures characterized by a hierarchy of authority, division of labor, objective standards, and rules and regulation (Weber, 1947; cited in Hoy & Sweetland, 2001, p.296).

Adler and Borys (1996) examined bureaucracy in two dimensions as enabling and coercive. Hoy and Sweetland (2000, pp.530-531; 2001, s.297-300) dealt with two characteristics of bureaucratic organization: formalization and centralization. Coercive formalization refers to the implementation of rules and procedures that primarily focus on punishing subordinates instead of rewarding productive practices, and it emphasizes compliance rather than fostering organizational learning. Enabling formalization is the rules and procedures that enable employees in completing their tasks and resolving issues effectively. In addition, obstructive centralization refers to a structure and management style that produces obstacles for subordinates in completing their tasks and resolving issues. On the other hand, facilitating centralization encompasses a structure and management approach that enables subordinates to remove obstacles and resolve issues in an effective way (Hoy, 2003, p.89-90). Apparently, the organizational structure guides and helps the organization leaders. Thus, it is important for the managers to lead their organizations in order to reach the goals of the organizations.

Contemporary principals, who had mostly managerial responsibilities in the past, are expected to lead their schools, increase the academic success of students and help the professional development of the staff (Hermann, 2016, p.6). The constant and increasing demands of students, parents and the environment has paved the way for school administrators to question their leadership behaviors. This query resulted in the adoptatipn of different behaviors by school administrators (Cemaloğlu, 2013, p.158). Accordingly, school administrators are required to perform to meet social demands and to lead the school community, as well as being responsible for the accountability of schools.

Accountability refers to determining responsible public officers and holding them accountable for their actions (Kondo, 2002, p.7). In other words, accountability requires a willing or compulsory report given to other people and includes having a conscientious or moral responsibility for the action of a person (Maile, 2002, p.326). In the context of education, accountability pertains to the practice of holding education systems responsible for the quality of their outcomes, including students' knowledge, skills, and behaviors (Stecher & Kirby, 2004, p.1). The most widely accepted definition of accountability consists the implementation of administrative data-driven mechanisms to improve student achievement (Figlio & Loeb, 2011, p.384). The purposes of accountability are basically threefold: First, it controls the abuse of public authority. Second, accountability provides reassurance regarding the utilization of public resources and adherence to legal and public service values. Third, it fosters and promotes a

culture of learning aimed at achieving continuous improvement in public administration. (Aucoin & Heintzman, 2000, p.45).

Traditionally, teachers have been accountable for addressing specific content outlined in the curriculum, students for learning through grading systems, administrators for monitoring student test scores, graduation rates, and other performance indicators such as student attendance in schools and districts. As accountability systems have evolved, states and communities have started to reconsider how to hold students, teachers, and administrators accountable (Stapleman, 2000, p.1). Accountability mechanisms primarily involve holding individuals in schools accountable for their actions to a person in an official authority position, either within or outside the school, through a range of formal and informal methods (Abelmann et al., 1999, p.4). According to Behn (2003a, pp.60-61), education is a shared concern among parents, legislators, governors, supervisors, civil leaders, and ultimately, everyone. Although these factors can both contribute to and detract from children's education, they are not typically held accountable. Behn argued that this is because accountability tends to be focused on others, with schools, principals, and teachers being the easiest targets. Kalman and Gedikoğlu (2014, p.117) defined accountable school administrators as individuals who understand their responsibilities, provide clear information to school stakeholders, and can answer questions regarding school issues. By exhibiting such behaviors, accountable school administrators can cultivate good relationships with teachers and establish trust, contributing to a healthy organizational climate. However, there is currently no research examining the relationship between school structure and accountability. In Türkiye, where the education system is highly centralized, school administrators have limited decision-making opportunities and autonomy (Karataş, 2022, pp.10-12). The impact of centralized school structures on the accountability of school administrators is an area of interest, given the understanding that all decision-makers should be accountable for their actions and decisions (Çiçekli, 2016, p.62). Additionally, distributed leadership may also play a role in influencing the accountability of school administrators.

Distributed leadership is a leadership practice that emerges as a result of interactions between each person contributing to school life, such as teachers, administrators, classroom assistants, support staff, parents and students (Harris, 2005, p.8). Smylie et al. (2007, p.470) explain distributed leadership as the sharing, dissemination and distribution of leadership among individuals and roles within the school. Managing and leading a school alone can be a challenging task for school administrators, especially in the current complex and changing social environment. It is suggested that leadership should be distributed and transformed into a team behavior to overcome these challenges (Beycioğlu & Aslan, 2010, p.766). To the best of the researchers' knowledge, no study has specifically examined the relationship between distributed leadership and accountability. However, these concepts are known to be closely interconnected. According to Elmore (2005, p.141), accountable leadership is synonymous with distributed leadership. As schools establish internal accountability and develop improvement practices, leadership becomes distributed based on expertise. Some teachers possess more knowledge about the teaching challenges their schools face and the solutions to those challenges. Therefore, expertise is necessary to create successful practices across the organization. In distributed leadership, decision-making and problem-solving responsibilities should be distributed according to employees' areas of expertise. The decentralized approach to accountability assumes that school principals should not act as the sole decision-makers but rather involve parents and teachers whose voices are beginning to be heard. This accountability approach requires school leaders to empower these individuals and actively encourage them to share power previously held only by the principal. School principals are expected to act as team members rather than making decisions alone (Leithwood, 2001, p.3).

It is unlikely that anyone has all the knowledge, skills, and abilities to fulfill all leadership functions without distributing them among others, given the increasing complexity of the education system (Hulpia & Devos, 2010, p.565). Distributing leadership among school staff, rather than limiting it to principals or top-level administrators, has a positive effect on student learning outcomes, an important factor in school success (Bell et al., 2003, p.3). In the age of school-based accountability, focusing the responsibility for change and transformation on a single individual, such as a principal, due to workplace pressures and complexity, is no longer effective. Therefore, distributing or sharing leadership among a large number of colleagues can be presented as a solution to excessive role overload (Woods & Gronn, 2009, p.441). However, distributed leadership allows teachers with specific expertise to contribute to the school's decision-making processes. By allowing teachers and other school leaders to contribute to decision-making processes, school principals can provide important leadership experiences to the school's future leaders (Jacobs, 2010, p.13). Additionally, distributed leadership is considered a means of understanding leadership practice in schools, promoting democracy within schools, enhancing efficiency and effectiveness, and developing human capacity (Mayrowetz, 2008). Studies examining the relationship between distributed leadership and school structure indicate a positive relationship between enabling bureaucracy and distributed leadership, while there is a negative relationship between coercive bureaucracy and distributed leadership (Oldaç & Kondakçı, 2020). This relationship suggests a mutual interaction between school structure and distributed leadership. Accordingly, an increase in coercive bureaucracy in schools negatively affects distributed leadership behaviors, while an increase in enabling bureaucracy positively affects distributed leadership behaviors.

Based on the statements and definitions above, it can be investigated whether there is a relationship between the accountability of school administrators, school structure and distributed leadership in educational institutions in Turkey, where the centralist aspect is dominant.

A brief literature review shows that numerous studies have been carried out on school structure, distributed leadership and accountability.

In this context, the school structure has been examined in terms of academic optimism (Anderson, 2012; Çalık & Tepe, 2019; Messick, 2012; Özdemir & Kılınç, 2014), school climate (Jacob, 2003; Mayerson, 2010), school effectiveness (Çalık & Tepe, 2019; Koohi et al., 2019; Mayerson, 2010), organizational support (Çokyigit, 2020), organizational citizenship (Alev, 2019; Messick, 2012), teacher professionalism (Cerit, 2012), teacher self-efficacy (Kılınç et al., 2016), teachers' career satisfaction (Koohi et al., 2019), awareness and teacher empowerment (Watts, 2009), teachers' proactive behaviors (Cerit & Akgün, 2015), student success (Anderson, 2012), organizational trust (Çokyigit, 2020), collective student trust (Koster, 2016), organizational silence (Bozkuş et al., 2019; Demirtaş et al., 2016), school principals' leadership styles (Buluç, 2009), organizational cynicism (Demirtaş, et al., 2016), organizational health (Ordu & Tanrıöğen, 2013), job satisfaction (Bozkuş et al., 2019)

In addition, distributed leadership has been examined in terms of organizational commitment (Aboudahr & Jiali, 2019; Akdemir & Ayık, 2017; Hulpia et al., 2009; Yetim, 2015), organizational trust (Adıgüzelli, 2016; Çiçek, 2018; Yılmaz, 2014), teacher motivation (Wahab et al., 2013), teacher self-efficacy (Halim & Ahmad, 2016; Kurt, 2016), organizational citizenship (Çakır, 2019; Jofreh et al., 2012), organizational support (Yılmaz, 2014), school development and mathematics achievement (Heck & Hallinger, 2009), job satisfaction (Hulpia et al., 2009), job stress (Rabindarang et al., 2014), academic optimism (Ataş Akdemir, 2016; Cansoy & Parlar, 2018; Mascall et al., 2008), school effectiveness (Atılkan, 2019), organizational climate (Çomak, 2021)

Also accountability has been examined in terms of school climate and teacher stress (Nathaniel et al., 2016), organizational justice (Kalman & Gedikoğlu, 2014), organizational silence (Karagöz, 2020), academic freedom (Doğan, 2015), management styles of school administrators (Yağ, 2019), organizational commitment (Yiğit, 2017), organizational cynicism (Argon et al., 2015), servant leadership (Kandemir & Akgün, 2019).

The primary aim of this study is to uncover the mediating role of distributed leadership in the potential relationship between school structure and school administrators' accountability behaviors, based on teacher perceptions. If the hypotheses proposed in this study are confirmed, the effects of school structure on distributed leadership and, indirectly, on accountability would be discerned. Consequently, it is anticipated that interest in the organization of school structure will increase.

There is insufficient evidence to determine whether the school structure has an impact on the distributed leadership behaviors and accountability of school administrators. This study aims to address this gap in the literature and provide potential solutions to the aforementioned issues.

The findings of this study may serve as a valuable resource for improving the school structure to enhance the accountability of school administrators. Furthermore, it may inspire researchers to explore new hypotheses and contribute to expanding the literature on this topic.

The expected outputs stated above have been the motivational elements in the realization of the study. In this study, it is aimed to reveal the predictive and intermediary relationships between the variables of coercive/enabling bureaucracy, distributed leadership and accountability. In this context, the hypotheses developed in line with the general purpose of the research are presented below:

H1: Coercive bureaucracy significantly and negatively predicts accountability.
H2: Coercive bureaucracy significantly and negatively predicts distributed leadership.
H3: Distributed leadership significantly and positively predicts accountability.
H4: Coercive bureaucracy has an indirect influence on accountability through distributed leadership.
H5: Enabling bureaucracy significantly and positively predicts accountability.
H6: Enabling bureaucracy significantly and positively predicts distributed leadership.
H7: Enabling bureaucracy has an indirect influence on accountability through distributed leadership.

The conceptual diagram of Model 1 tested in the study is presented in Figure 1.

Figure 1. *The conceptual diagram of Model 1.*
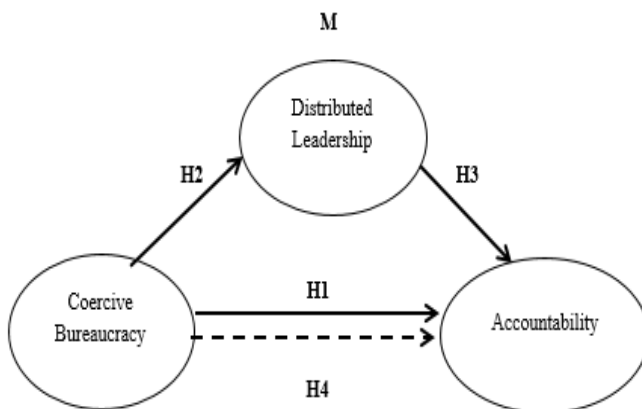
The conceptual diagram of Model 2 tested in the study is shown in Figure 2.

Figure 2. *The conceptual diagram of Model 2.*
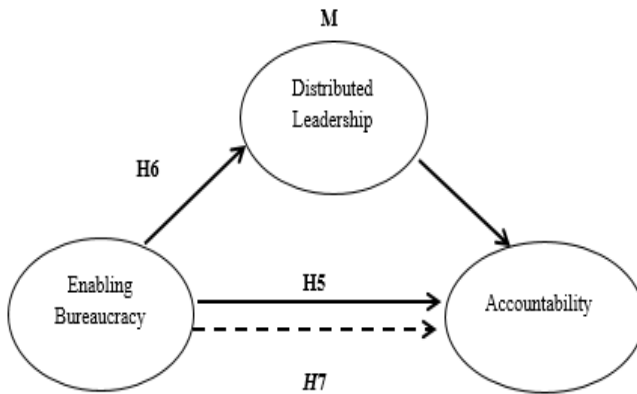
2. METHOD

2.1. Research Model

The primary objective of this study was to explore the predictive and mediating relationships between school structure and the variables of distributed leadership and school administrators' accountability behaviors. To accomplish this, a relational survey model was utilized. The purpose of this model is to evaluate the presence or degree of change between two or more variables (Karasar, 2009, s. 81). The study utilized two models to investigate the relationship between school structure, distributed leadership, and school administrators' accountability. In the first model, coercive bureaucracy was the independent variable, while distributed leadership was both the dependent and independent variable, and accountability was the dependent variable. The second model, on the other hand, had enabling bureaucracy as the independent variable, distributed leadership as the dependent and independent variable, and accountability as the dependent variable.

2.2. Population and Sample

The population of the study consisted of 3171 teachers working in elementary schools in the central districts of Erzurum (Aziziye, Palandöken and Yakutiye) in the 2021-2022 academic year. The proportional stratified sampling method, which is a form of random sampling, was employed for sample selection. Stratified sampling ensures that the subgroups in the universe are represented in the sample with the same proportions in the population (Özen & Gül, 2007, p. 402). Accordingly, the districts of Aziziye, Palandöken and Yakutiye was considered as separate strata. In sample selection, the ratio of the total number of teachers working in elemantary schools in each stratum to the total number of teachers in the population was considered. As a result, the sample consisted of 444 teachers, 34 from Aziziye, 201 from Palandöken, and 209 from Yakutiye.

The demographics of the participants are presented in Table 1. Table 1 revealed that of the participants, 273 (61.5%) were female and 171 (38.5%) were male. In addition, the participants' professional seniority was concentrated between 6-10 years (38.1%) and 11-15 years (25.0%), while their working time at the school was focused between 1-5 years (59.0%) and 6-10 years (29.3%). the greatest level of participation in the study, according to discipline, came from teachers in Turkish (17.8%), Social Studies (12.8%), Mathematics (12.6%), and Science (11.5%).

Table 1. *The demographic features of the participants.*

Demographic Characteristics		f	%
Gender	Female	273	61.5
	Male	171	38.5
Work Experience (years)	1-5	71	16.0
	6-10	169	38.1
	11-15	111	25.0
	16-20	54	12.2
	21 and above	39	8.8
Working Time at the School (years)	1-5	262	59.0
	6-10	130	29.3
	11-15	30	6.8
	16-20	16	3.6
	21 and above	6	1.4
Discipline	Turkish	79	17.8
	Mathematics	56	12.6
	Science	51	11.5
	Social Studies	57	12.8
	Religion	47	10.6
	English	45	10.1
	Information Technologies	15	3.4
	Physical Education	23	5.2
	Technology and Design	9	2.0
	Visual Arts	13	2.9
	Music	12	2.7
	Arabic	11	2.5
	Guidance and Psychological Counseling	26	5.9
Total		444	100

2.3. Data Collection Tools

Three different scales were used in data collection: Accountability Behavior Scale for School Administrators (Orhan, 2022), Enabling School Structure Scale (Özer & Dönmez, 2013), Distributed Leadership Scale (Özer & Beycioğlu, 2013). Information on data collection tools is presented below.

2.3.1. *Accountability behavior scale for school administrators*

The "Accountability Behavior Scale for School Administrators" developed by Orhan (2022) consists of 16 items and has four sub-dimensions: accountability behaviors towards students (5 Items), accountability behaviors towards teachers (3 Items), accountability behaviors towards parents (4 Items), accountability behaviors towards superiors (4 Items). The items were scaled from (1) never to (5) always. The Cronbach's Alpha, which shows the internal consistency of the scale, was calculated as .922 for the total scale, .862 for the accountability behaviors towards students, .721 for the accountability behaviors towards teachers, .834 for the accountability behaviors towards the parents and .853 for the accountability behaviors towards superiors. In the present study, the Cronbach's Alpha were found as .851 for the accountability behaviors

towards students, .704 for the accountability behaviors towards teachers, .855 for the accountability behaviors towards the parents, .850 for the accountability behaviors towards superiors, and .922 for the total scale.

2.3.2. *Enabling school structure scale*

The "Enabling School Structure" scale was originally developed by Hoy and Sweetland (2000) and later adapted into Turkish by Buluç (2009). Its psychometric properties were re-examined by Özer and Dönmez (2013). The scale comprises 12 items and underwent validity and reliability testing, which revealed that the Turkish version consisted of two factors: coercive bureaucracy and enabling bureaucracy. The scale utilizes a 5-Point Likert-type response format, ranging from "never" to "always." In the scale development study, the factor loadings of the items ranged between .557 and .832 for the coercive bureaucracy dimension, and between .485 and .785 for the enabling bureaucracy dimension. The items in the scale accounted for approximately 51% of the total variance. The Cronbach's Alpha values for the scale were .806 for the enabling bureaucracy dimension and .774 for the coercive bureaucracy dimension. In the present study, the internal consistency coefficients were calculated as .715 for the coercive bureaucracy dimension and .848 for the enabling bureaucracy dimension.

2.3.3. *Distributed Leadership Scale*

The "Distributed Leadership Scale," devised by Özer and Beycioğlu (2013), encompasses 10 items exhibiting a unidimensional structure. The scale employs a 5-Point Likert-type rating system, spanning from "never" to "always." The scale yields a minimum score of 10 and a maximum score of 50. The Cronbach's Alpha coefficient of the scale was determined to be .92, with a test-retest correlation coefficient of .82. In the context of this study, the calculated internal consistency coefficient was .900.

2.4. Data Collection

The data were collected by visiting schools during the first term of the 2021-2022 academic year. Information about the aim and scope of the study was provided, and care was taken to adhere to ethical values during the data collection process. Participation in the study was based on voluntariness. Furthermore, efforts were made to collect data in a manner that would not interrupt the educational process. The data were obtained during teachers' free periods, lunch breaks, and after school hours. Copies of research and application permission letters were provided to schools that requested them. Gratitude was expressed to the school administrators and teachers who took part in the research.

2.5. Data Analysis

The data were analyzed using SPSS 23.0, LISREL 8.80 and PROCESS macro for SPSS v3.3. Before the analyses, outliers, normality, multicollinearity problem and sample size, the prerequisites of the Structural Equation Modelling (SEM), were examined. In the first stage, outliers were examined and 11 outliers were removed from the data set (the absolute values of -3 and 3 were considered based on Z-Score).

In the second stage, the normality of the data was assessed by examining the skewness and kurtosis values. It was observed that the skewness and kurtosis values fell within the range of -1.5 to +1.5, indicating a normal distribution of the data (Tabachnick & Fidell, 2007). The skewness and kurtosis values of the variables are presented in Table 2.

Table 2. *The skewness and kurtosis values of the variables.*

Scales	Skewness	Kurtosis
Distributed Leadership	-.739	-.234
Accountability towards students	-1.145	.606
Accountability towards teachers	-1.048	-.506
Accountability towards parents	-.650	-.615
Accountability towards superiors	-1.117	.263
Accountability	-.857	-.256
Coercive Bureaucracy	1.154	.927
Enabling Bureaucracy	-.747	.371

Table 2 indicated that the data were normally distributed since the skewness and kurtosis coefficients of the variables were between -1.5 and +1.5.

In the third stage, the multicollinearity problem was examined. The fact that the correlation value between the variables is below .90 indicates that there is no multicollinearity problem (Çokluk et al., 2010, s.210). In order to investigate whether the multivariate normality was provided, the scatter diagram matrix was examined and it was found that the variable pairs formed diagrams similar to ellipse. The findings regarding the correlation between the variables are presented in Table 3.

Table 3. *Correlation values between the variables.*

	1	2	3	4	5	6	7	8
(1) Distributed Leadership	1	.586	.630	.562	.527	.683	-.369	.457
(2) Accountability towards students		1	.644	.568	.595	.848	-.286	.336
(3) Accountability towards teachers			1	.584	.605	.816	-.295	.345
(4) Accountability towards parents				1	.613	.840	-.274	.293
(5) Accountability towards superiors					1	.841	-.314	.370
(6) Accountability		,				1	-.348	.399
(7) Coercive Bureaucracy							1	-.346
(8) Enabling Bureaucracy								1

p<.001

Table 3 revealed that distributed leadership had a positive and moderate relationship with accountability towards students ($r=.586$, $p<0.01$), accountability towards teachers ($r=.630$, $p<0.01$), accountability towards parents ($r=.562$, $p<0.01$), accountability towards superiors ($r=.527$, $p<0.01$), the total accountability scale ($r=.683$, $p<0.01$) and enabling bureaucracy ($r=.457$, $p<0.01$) and a negative and moderate relationship with coercive bureaucracy ($r=-.369$, $p<0.01$). Accountability towards students correlated positively and moderately with accountability towards teachers ($r=.644$, $p<0.01$), accountability towards parents ($r=.568$, $p<0.01$) accountability towards superiors ($r=.595$, $p<0.01$) and enabling bureaucracy ($r=.336$, $p<0.01$), positively and strongly with the total accountability scale ($r=.848$, $p<0.01$), and negatively and weakly with coercive bureaucracy ($r=-.286$, $p<0.01$).

Accountability towards teachers had a positive and moderate relationship with accountability towards parents ($r=.584$, $p<0.01$), accountability towards superiors ($r=.605$, $p<0.01$), enabling bureaucracy ($r=.345$, $p<0.01$), a positive and strong relationship with the total accountability scale ($r=.816$, $p<0.01$), and negative and weak relationship with coercive bureaucracy ($r=-.295$, $p<0.01$).

Accountability towards parents had a positive and moderate relationship with accountability towards superiors (r=.613, p<0.01), a positive and strong relationship with the total accountability scale (r=.840, p<0.01) a positive and weak relationship with enabling bureaucracy (r=.293, p<0.01) and a negative and weak relationship with coercive bureaucracy (r=-.274, p<0.01). Accountability towards superiors correlated positively and strongly with the total accountability scale (r=.841, p<0.01), positively and moderately with enabling bureaucracy (r=.370, p<0.01), and negatively and moderately with coercive bureaucracy (r=-.314, p<0.01). The total accountability scale had a negative and moderate relationship with coercive bureaucracy (r=.348, p<0.01) and a positive and moderate relationship with enabling bureaucracy (r=.399, p<0.01). The coercive bureaucracy had a negative and moderate relationship with enabling bureaucracy (r=-.346, p<0.01).

In the fourth stage, the sufficiency of the sample size was examined. It was stated in the literature that the sample size should be at least five or even ten times the number of the items. (Bryman & Cramer, 2001; cited in Tavşancıl, 2002, p.17). Accordingly, the sample of the study, in which 444 participants were included, were sufficient since there were a total 38 items in the study.

Finally, SEM analysis was performed to test the mediating effect of distributed leadership on the effect of coercive/enabling bureaucracy on accountability. In order for the mediating effect of distributed leadership to be significant in the effect of coercive/enabling bureaucracy on accountability, the followings were required:

The effect of the independent variable of coercive/enabling bureaucracy on the mediating variable of distributed leadership (Path A) and the effect of the mediating variable of distributed leadership on accountability (Path B) should be significant. The effect of the independent variable of coercive/enabling bureaucracy on the dependent variable of accountability (Path C) should be significant.

Finally, when the model is run, the effect of the independent variable of coercive/enabling bureaucracy on the dependent variable of accountability (Path C) should lose statistical significance or there should be a significant decrease in the level of this effect (Baron & Kenny, 1986).

Paths in the model are shown in Figure 3.

Figure 3. *Paths in the model.*
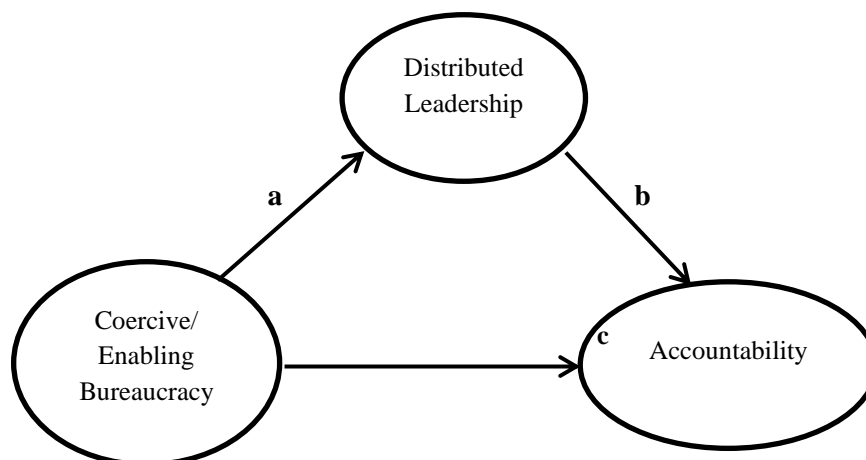
In the study, the figures presented for the structural equation model were generated using an online diagram software (diagrams.net).

2.5.1. *Testing the measurement model*

The measurement model for Model 1 in the study was tested (Figure 4), and the findings are presented in Table 4.

Figure 4. *The measurement model for Model 1.*
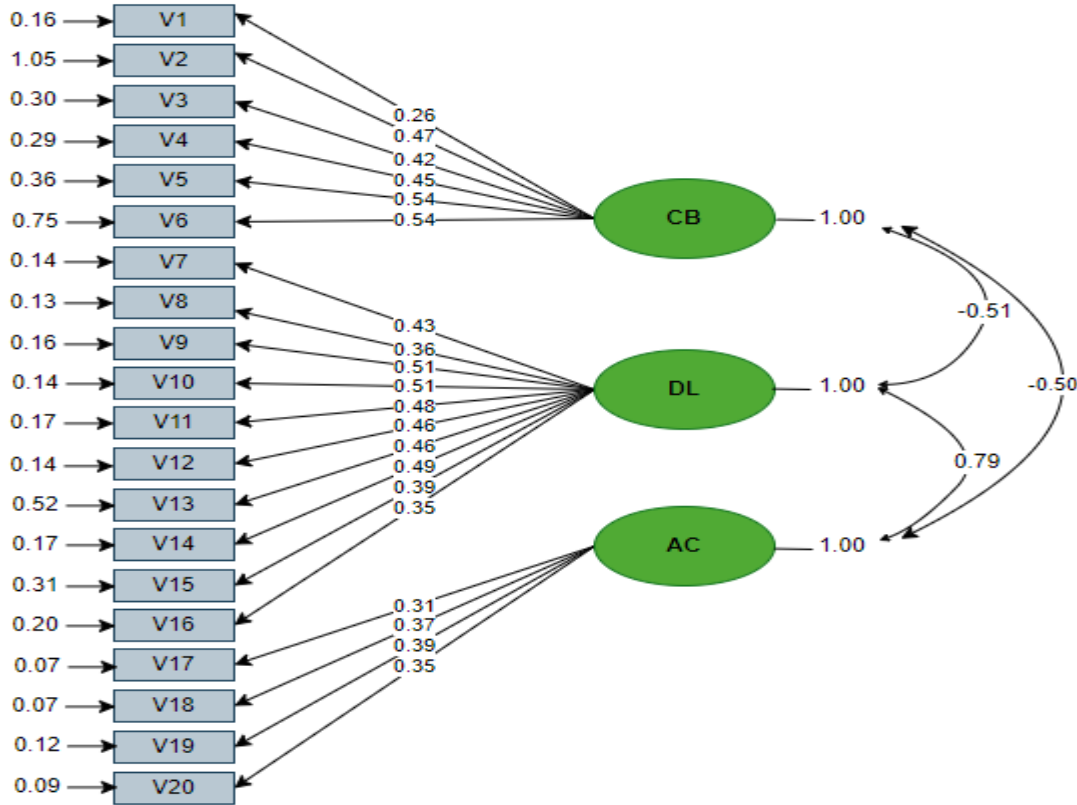
Findings on the testing Model 1 are presented in Table 4.

Table 4. *Fit indices for Model 1.*

χ²/sd	AGFI	GFI	SRMR	CFI	RMSEA	NFI	NNFI
3.53	.85	.88	0.053	0.97	0.076	0.95	0.96

It was found that Model 1 in the study had acceptable fit indices (Browne & Cudeck, 1993; Meydan & Şeşen, 2011; Schermelleh-Engel et al., 2003; Sümer, 2000). Thus, in the later stage the structural equation modelling was tested.

The measurement model for Model 2 in the study was tested (Figure 5), and the findings are presented in Table 5.

Figure 5. *The measurement model for Model 2.*
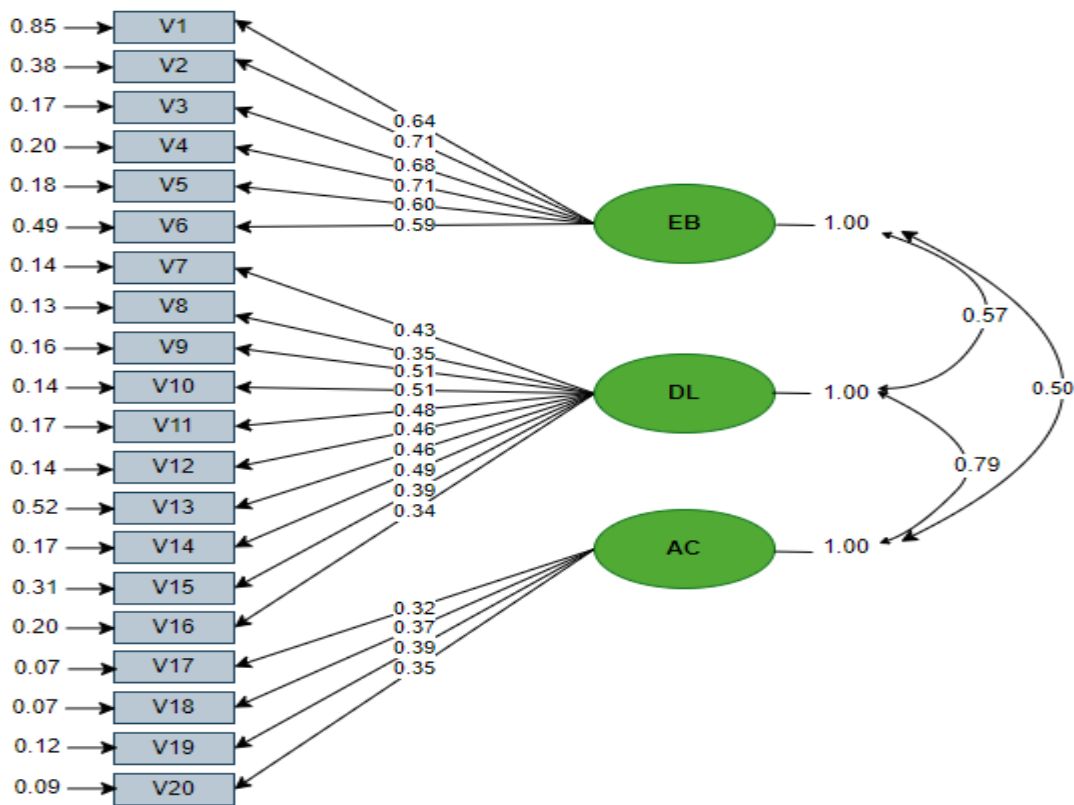
Findings on the testing the second measurement model are presented in Table 5.

Table 5. *Fit indices for Model 2.*

χ^2/sd	AGFI	GFI	SRMR	CFI	RMSEA	NFI	NNFI
3.70	0.85	0.88	0.052	0.97	0.078	0.96	0.97

It was found that the second measurement model in the study had acceptable fit indices (Browne & Cudeck, 1993; Meydan & Şeşen, 2011; Schermelleh-Engel et al., 2003; Sümer, 2000). Thus, in the later stage the structural equation modelling was tested.

3. FINDINGS

3.1. Findings on Model 1

3.1.1. *The mediating role of distributed leadership in the relationship between coercive bureaucracy and accountability*

First, it was found that the coercive bureaucracy had a negative and significant (t=-11.68, p<0.01) effect (β=-0.69) on accountability. Then, the coercive bureaucracy was found to have a negative and significant (t=-9.21, p<0.01) effect (β=-0.52) on distributed leadership. Similarly, it was found that distributed leadership had a positive and significant (t=13.46, p<0.01) effect (β= 0.80) on accountability. These findings indicated that the model was suitable for the mediation test. Accordingly, the mediating role of distributed leadership in the relationship between coercive bureaucracy and accountability was examined. The figures presented for the structural equation model in the study were created using online diagram software (diagrams.net). Findings on Model 1 are presented in Figure 6.

Figure 6. *Findings on Model 1.*
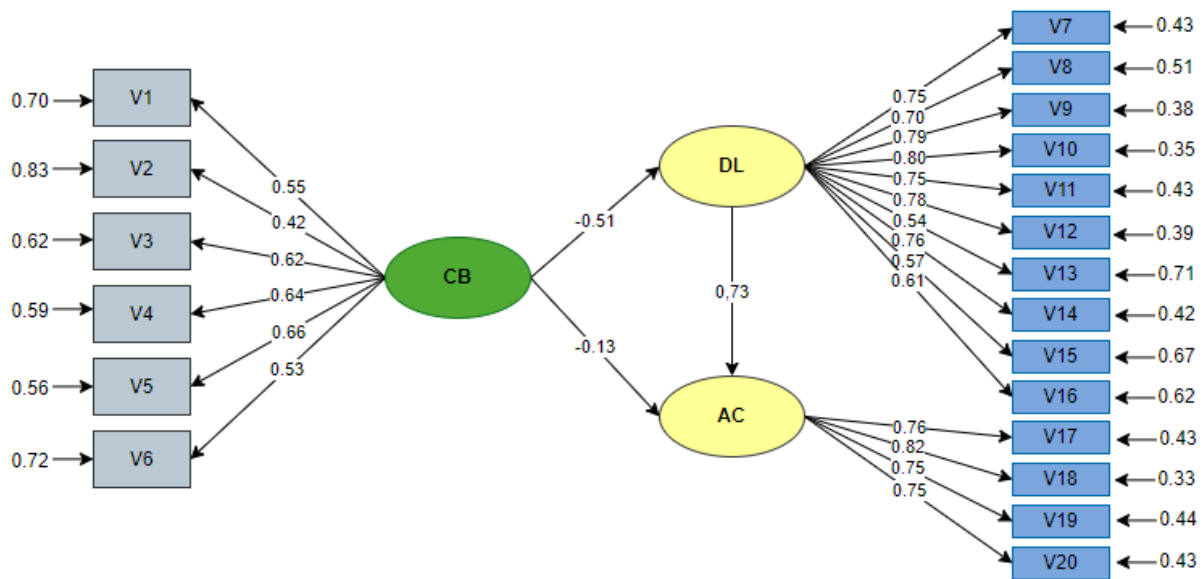
Figure 6 showed that before the model was run, there was an almost strong, negative and significant relationship between coercive bureaucracy and accountability (β=-0.69, $p<0.01$), and this relationship decreased after the model (β=-0.13, $p<0.05$); however this and other predictive relationships did not lose statistical significance. This finding revealed that distributed leadership had a mediating role in the relationship between coercive bureaucracy and accountability. Accordingly, the first four hypotheses of the study were confirmed. The fit indices of the model were found to be χ^2/sd =3.53, AGFI=0.85, GFI=0.88, SRMR=0.053, CFI=0.97, RMSEA=0.076, NFI=0.95, NNFI=0.96, which was at an acceptable level. Figure 7 shows t values of the model.

Figure 7. *T-values of Model 1.*
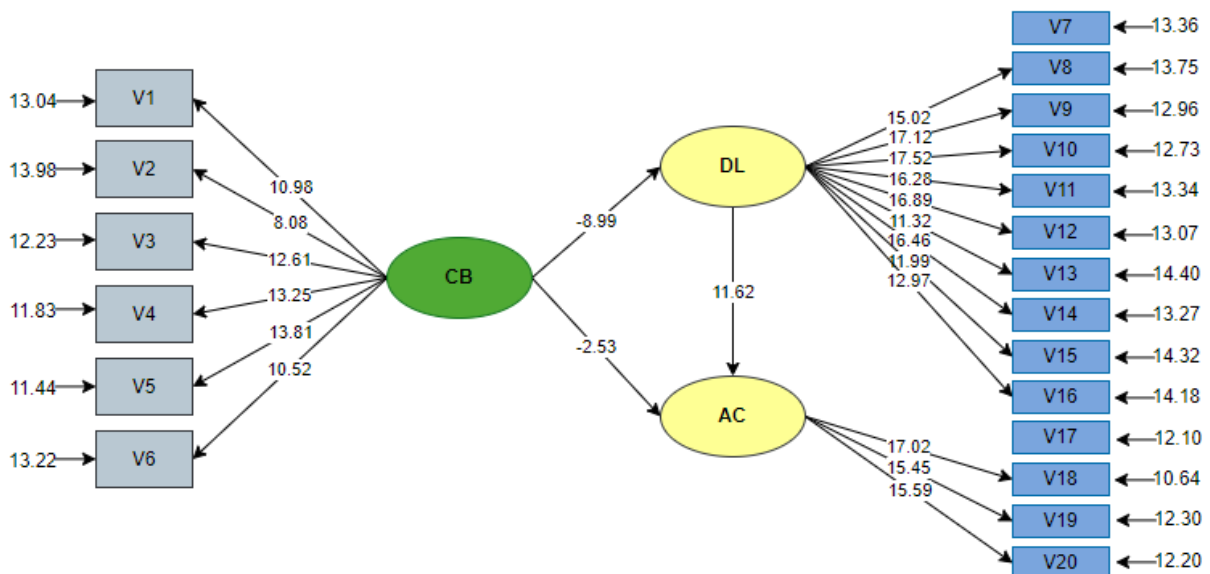
To investigate the significance of the mediating role of distributed leadership (H4) in the relationship between coercive bureaucracy and accountability, bootstrap analysis was conducted using PROCESS v3.3 Model 4. A sample size of 5,000 participants was utilized to examine the direct and indirect effects (Preacher & Hayes, 2008). The results and corresponding confidence intervals (CI) are presented in Table 6.

Table 6. *Bootstrap analysis results on predictive effects in Model 1.*

Paths	Prediction Coefficient	*t*	*p*	R²	95 % CI Lower	95 % CI Upper
Coercive Bureaucracy ▼ Distributed Leadership	.-.3283	-8.34	.0000	.14	.-.4056	-.2510
Coercive Bureaucracy ▼ Accountability	-.0822	-2.98	.0030	.48	-.1364	-,0281
Distributed Leadership ▼ Accountability	.5377	17.35	.0000		.4768	.5985
Diret Effects	**Effect Coefficient**				**95 % CI Lower**	**Upper**
Coercive Bureaucracy ▼ Accountability	-.0822	-2.98	.0030		-.1364	-.0281
Indirect Effects						
C. B. ▼D. L.▼ Accountability.	-.1765				-.2260	-.1316
Total Effect						
Coercive Bureaucracy ▼ Accountability	-.2587	-7.79	.0000		-.3240	-.1935

As shown in Table 6, coercive bureaucracy predicted distributed leadership (β=-.3283, *p*<.01) and explained 14% of the variance in distributed leadership. It was also found that coercive bureaucracy significantly predicted accountability (β=-.0822, *p*<.01), and distributed leadership predicted accountability statistically (β= .5377, *p*<.01). Coercive bureaucracy and distributed leadership explained approximately 48% of the variance in accountability.

The findings revealed that the direct effect of coercive bureaucracy on accountability (-.0822) was significant since the 95% confidence interval for the observed values did not encompass zero, which revealed statistical significance (-.1364<-.0281, *t*=-2.98, *p*>0.01).

The indirect effect of coercive bureaucracy on accountability through distributed leadership (-.1765) was significant since the 95% confidence interval for the observed values did not encompass zero (-.2260 < -.1316). The sum of the direct and indirect effects of the coercive bureaucracy on accountability (-.2587) was also significant since the 95% confidence interval for the observed values did not encompass zero (-.3240 < -.1935, *t*=-7.79, *p*<.0.01)

3.2. Findings on Model 2

3.2.1. *The mediating role of distributed leadership in the relationship between enabling bureaucracy and accountability*
First, it was found that the enabling bureaucracy had a positive and significant (*t*= 9.92, *p*<0.01) effect (β= 0.54) on accountability. The, enabling bureaucracy was found to have a positive and significant (*t*= 10.83, *p*<0.01) effect (β= 0.57) on distributed leadership. It was also found that distributed leadership had a positive and significant (*t*=13.46, *p*<0.01) effect (β= 0.80) on accountability. These findings indicated that the model was suitable for the mediation test. Accordingly, the mediating role of distributed leadership in the relationship between enabling bureaucracy and accountability was examined. The findings on Model 2 is presented in Figure 8.

Figure 8. *The findings on Model 2.*
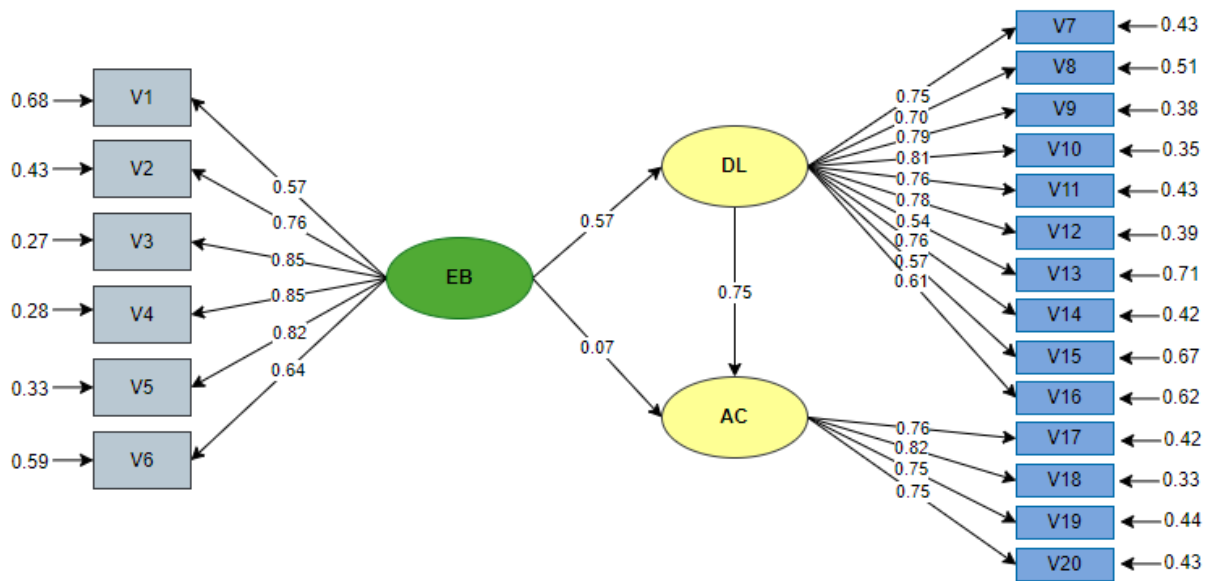
Figure 8 showed that there was a moderate, positive, and statistically significant relationship between enabling bureaucracy and accountability prior to running the model (β= 0.54, *p*<0.01), and this relationship lost statistical significance after the model (β= 0.07, *p*>0.05). Other paths related to the model maintained their significance. This finding shows that distributed leadership has a full mediating role in the relationship between enabling bureaucracy and accountability. Accordingly, the last three hypotheses of the study were also confirmed. The fit indices for the model were as follows: χ^2/sd=3.70, AGFI=0.85, GFI=0.88, SRMR=0.052, CFI=0.97, RMSEA=0.078, NFI=0.96, NNFI=0.97, which was at an acceptable level. The *t* values of the Model 2 are presented in Figure 9.

Figure 9. *t-values of Model 2.*
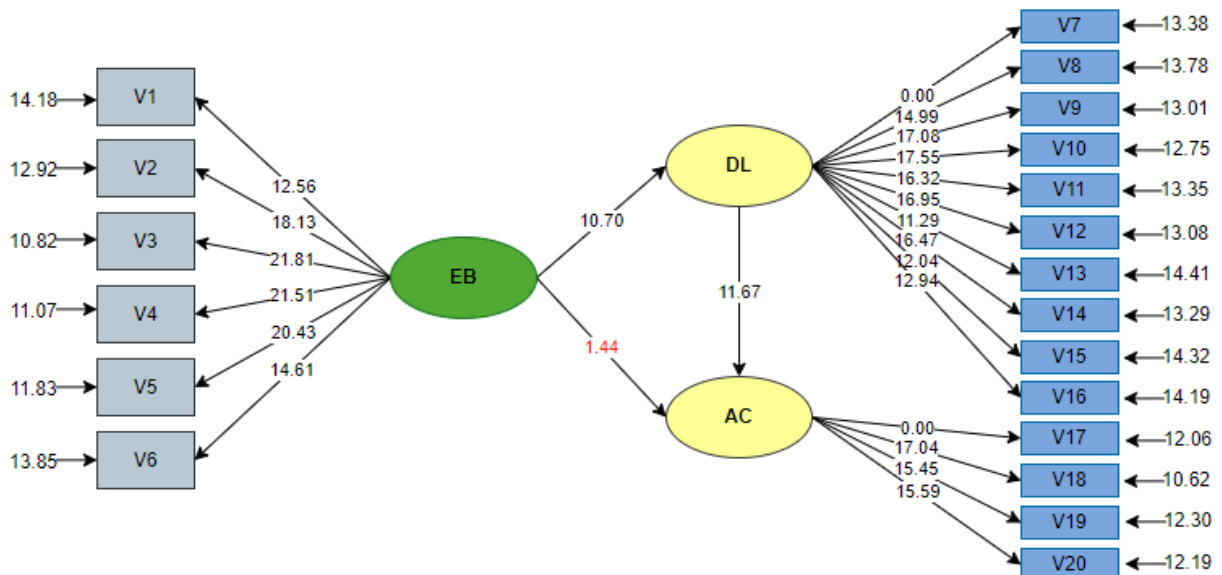
Bootstrap analysis was performed using PROCESS v3.3 Model 4 in order to examine whether the mediating role of distributed leadership (H7) was significant in the relationship between enabling bureaucracy and accountability in a sample of 5,000 people, and to examine its direct and indirect effects (Preacher & Hayes, 2008). The findings and confidence intervals (CI) are presented in Table 7.

Table 7. *Bootstrap analysis results on predictive effects in Model 2.*

Paths	Prediction Coefficient	t	p	R²	95 % CI Lower	Upper
Enabling Bureaucracy ▼ Distributed Leadership	.3117	10.79	.0000	.21	2550	3684
Enabling Bureaucracy. ▼ Accountability	.0626	2.83	.0048	.48	.0192	.1061
Distributed Leadership ▼ Accountability	.5298	16.35	.0000		.4662	.5935
Direct Effects	**Effect Coefficient**				**95 % CI Lower**	**Upper**
E. B. ▼ Accountability	.0626	2.83	.0048		.0192	.1061
Indirect Effects						
E. B. ▼ D. L. ▼ Accountability.	.1651				.1271	.2064
Total Effect						
E. B. ▼ Accountability	.2278	9.15	.0000		.1789	.2767

Table 7 revealed that the enabling bureaucracy predicted the distributed leadership (β= .3117, $p<.01$), and explained 21% of the variance in the distributed leadership.

It was also found that enabling bureaucracy significantly predicted accountability (β= .0626, $p<.01$) and distributed leadership significantly predicted accountability (β= .5298, $p<.01$). Enabling bureaucracy and distributed leadership explained approximately 48% of the variance in accountability.

The findings revealed that the direct effect of enabling bureaucracy on accountability (.0626) %95 was significant since the 95% confidence interval for the observed values did not encompass zero (.0192<.1061, $t=2.83$, $p>0.01$),

The indirect effect of enabling bureaucracy on accountability through distributed leadership (.1651) was significant since the 95% confidence interval for the observed values did not encompass zero (.1271 < .2064). The sum of the direct and indirect effects of the enabling bureaucracy on accountability (.2278) was also significant since the 95% confidence interval for the observed values did not encompass zero (.1789 < .2767, $t=9.15$, $p<.0.01$).

4. DISCUSSION and CONCLUSION

In this study, the relationships between coercive/enabling bureaucracy, distributed leadership and accountability were examined and the mediating role of distributed leadership in this relationship was analyzed. The findings revealed that all 7 hypotheses proposed in the study were confirmed.

First, it was found that coercive bureaucracy had a negative effect on accountability, while enabling bureaucracy affected accountability positively. A significant negative relationship was observed between coercive bureaucracy and accountability, while a strong and statistically meaningful association was found between enabling bureaucracy and accountability. In other words, an increase in coercive bureaucracy decreased accountability, whereas an increase in enabling bureaucracy increased accountability. Although there is no study on the relationship between school structure and accountability, there are some studies investigating common variables. For example, Kim (2005), Turner (2018) and Arik (2021) reported a strong and statistically meaningful association between organizational justice and school structure. In addition, Kalman and Gedikoğlu (2014) found that there was a strong and statistically meaningful association between organizational justice and the accountability of school

administrators. These findings suggest that the increase in school structure and accountability increases organizational justice.

Another common variable between school structure and accountability is school climate. While Jacob (2003) found that there was not a significant relationship between school climate and school structure, Nathaniel et al. (2016) reported a negative relationship between school climate and exam-based accountability policies. This finding indicates that the increase in exam-based accountability policies increases negative perceptions about the school climate.

Organizational cynicism is also a common variable between school structure and accountability. Demirtaş et al. (2016) and Karaoğlan (2019) found that there was a positive association between organizational cynicism and bureaucratic school structure. However, Argon et al. (2015) reported a negative relationship between organizational cynicism and accountability practices in the Turkish National Education System. Accordingly, it can be argued that the increase in the bureaucratic school structure increases organizational cynicism, while the increase in accountability practices reduces organizational cynicism.

Furthermore, the analysis revealed that coercive bureaucracy had a negative impact on distributed leadership, while enabling bureaucracy exerted a positive influence on distributed leadership. Consequently, a negative and significant relationship between coercive bureaucracy and distributed leadership was established, along with a positive and significant relationship between enabling bureaucracy and distributed leadership. In other words, the increase in the coercive bureaucracy decreased the distributed leadership behaviors, whereas the increase in the enabling bureaucracy increased the distributed leadership behaviors. Accordingly, coercive bureaucracy had a negative effect on distributed leadership, while enabling bureaucracy affected distributed leadership positively. Similar to the present study, Oldaç and Kondakçı (2020) found a positive association between enabling bureaucracy and distributed leadership, and a negative relationship between coercive bureaucracy and distributed leadership.

While limited research exists specifically on the relationship between school structure and distributed leadership, several studies have explored these variables independently. For example, Messick (2012), Mitchell (2019) and Alev (2019) found a strong and statistically meaningful association between organizational citizenship and school structure. Furthermore, Jofreh et al. (2012) and Çakır (2019) revealed a strong and statistically meaningful association between organizational citizenship and distributed leadership. Tese findings indicates that the increase in school structure and distributed leadership increases organizational citizenship.

A common variable related to school structure and distributed leadership is academic optimism. McGuigan and Hoy (2006), Messick (2012), Özdemir and Kılınç (2014), Anderson et al. (2018), and Çalık and Tepe (2019) reported a positive and significant relationship between academic optimism and school structure. In addition, Mascall et al. (2008), Ataş Akdemir (2016), and Cansoy and Parlar (2018) found a strong and statistically meaningful association between academic optimism and distributed leadership. These findings reveal that the increase in school structure and distributed leadership increases academic optimism.

Job satisfaction is another common variable related to school structure and distributed leadership. In this sense, Bozkuş et al. (2019) found a weak and positive relationship between job satisfaction and enabling school structure. In addition, Ereş and Akyürek (2016) also reported a positive association between job satisfaction and distributed leadership characteristics of school principals. These findings indicate that the increase in school structure and distributed leadership increases job satisfaction.

Another common variable related to school structure and distributed leadership is organizational citizenship. In this sense, Mitchell (2019) found a strong and statistically meaningful association between organizational citizenship and enabling school structure.

However, Jofreh et al. (2012) and Çakır (2019) revealed a strong and statistically meaningful association between distributed leadership and organizational citizenship. These findings suggest that the increase in school structure and distributed leadership increases organizational citizenship.

Another common variable related to school structure and distributed leadership is teacher self-efficacy. Kılınç et al. (2016) found a strong and statistically meaningful association between teacher self-efficacy and school structure. Similarly, Kurt (2016) and Halim and Ahmad (2016) reported a strong and statistically meaningful association between teacher self-efficacy and distributed leadership. These findings show that the increase in school structure and distributed leadership increases teacher self-efficacy.

Organizational trust is another variable related to school structure and distributed leadership. In this sense, Geist (2002) and Çokyiğit (2020) found a strong and statistically meaningful association between organizational trust and school structure. In addition, Adıgüzelli (2016), Çiçek (2018) and Yılmaz (2014) revealed a strong and statistically meaningful association between organizational trust and distributed leadership. These findings indicate that the increase in school structure and distributed leadership increases organizational trust.

Organizational support is another variable related to school structure and distributed leadership. Çokyiğit (2020) found a strong and statistically meaningful association between organizational support and school structure. Similarly, Yılmaz (2014) revealed a strong and statistically meaningful association between organizational support and distributed leadership. These findings show that the increase in school structure and distributed leadership increases organizational support.

School effectiveness is another variable related to school structure and distributed leadership. In this issue, Gray (2016) and Çalık and Tepe (2019) reported a strong and statistically meaningful association between school effectiveness and school structure. Also, Atılkan (2019) found a strong and statistically meaningful association between school effectiveness and distributed leadership. These findings indicate that the increase in school structure and distributed leadership increases school effectiveness.

Who will be held accountable in schools, which issues will be accounted for, and who will be accountable have been the most debated issues. Argon (2015) found that teachers and administrators attributed the same meaning to accountability stating that everyone working in schools should be accountable. In addition, Koçak and Nartgün (2018) stated that besides teachers, school administration, parents and students should be accountable at schools. Based on these findings, it can be argued that all stakeholders should be held responsible for school accountability, which is actually related to distributed leadership. It is known that the effective performance of school leadership, which has become quite complex and difficult nowadays, is possible by distributing leadership authorities. Therefore, school administrators can distribute their authority to employees in order to spread accountability throughout the school and in turn position them as stakeholders of accountability. In other words, it can be said that it is consistent to hold the employees, who can participate in the decisions taken or take the initiative, accountable for their actions. In this sense, Behn (2003b, p.68) stated that if organizational managers do not distribute some of their authority to their assistants in managerial issues and if they want that every action should be asked to them, they cannot hold their assistants accountable.

Third, it was found that distributed leadership positively affected accountability. In other words, the increase in distributed leadership increased accountability. It was concluded in this study that there was a significant relationship between school administrators' distributed leadership behaviors and their accountability behaviors. Elmore (2005, p.141) states that accountable

leadership is actually distributed leadership and as schools are successful in ensuring internal accountability and schools have become places where leadership is distributed based on expertise, if schools ensures internal accountability and developing practices to improve it. Performance-based accountability and good management practices in schools generally require that particular individuals are held accountable for the guidance, management and ultimately performance of the organization. In this sense, administration leaders should aim primarily to develop the knowledge and skills of the people in the organization, create a culture of common expectation around the use of these skills and knowledge, keep the organization as a whole to be productive, and hold individuals accountable for contributing to the collective result (Elmore, 2000, p.15). Accordingly, school administrators should provide employees the opportunity to use their knowledge and skills in the school environment and empower them, and in turn position them as the addressee of school accountability. The decentralized approach to accountability proposes that the role of school leaders is to empower parents and teachers and encourage them to share the power previously used by the school principal (Leithwood, 2001, p.3).

Although there is no study examining the relationship between distributed leadership and accountability of school administrators, studies, in which related variables were examined, support the positive relationship between distributed leadership and accountability. In this sense, Yetim (2015), Akdemir and Ayık (2017) and Aboudahr and Jiali (2019) showed that that there was a strong and statistically meaningful association between organizational commitment and distributed leadership. Besides, Yiğit (2017) found a strong and statistically meaningful association between organizational commitment and accountability. The findings unveiled a noteworthy positive association between organizational commitment and both distributed leadership and accountability, which supports the significant positive relationship between distributed leadership and accountability behaviors of school administrators.

Another variable related to distributed leadership and accountability is school improvement/academic quality and mathematics achievement. In this sense, Heck and Hallinger (2009) concluded that distributed leadership had a direct effect on the change in the academic capacity of schools and indirect effects on students' mathematical development. Furthermore, Chang's (2011) found that distributed leadership had an indirect effect on student achievement. In addition, Rockoff and Turner (2010) showed that the pressure of accountability significantly increased student achievement in English and especially mathematics courses in low-performing schools. Also, it was found in the same study that the satisfaction of the parents of the children in these schools with the academic quality increased significantly. Based on these findings, the significant relationship between school development/academic quality and mathematics achievement and distributed leadership and accountability supports the significant positive relationship between distributed leadership and accountability behaviors of school administrators.

Another variable related to distributed leadership and accountability is job stress. In this sense, Rabindarang et al. (2014) reported a significant and negative relationship between job stress and distributed leadership. Accordingly, it can be said that the increase in distributed leadership behaviors reduces job stress. In addition, Nathaniel et al. (2016) found that there was a positive association between job stress and exam-based accountability policies. Based on this finding, it can be said that the increase in exam-based accountability policies increases job stress. Similarly, Mitani's (2018) found that No Child Left Behind sanctions were positively correlated with job stress, turnover rates, and transfer rates of school principals. Based on these results, the significant and negative relationship between job stress and distributed leadership, and the significant positive relationship between job stress and accountability do not support the

significant positive relationship between distributed leadership and accountability behaviors of school administrators found in the present study.

Another variable related to distributed leadership and accountability is school climate. Harrison (2005) stated that distributed leadership led to a positive learning environment and paved the way for teachers and students to be satisfied with the school. In addition, Çomak (2021) found that there was a positive association between school climate and distributed leadership. These findings suggest that the increase in distributed leadership behaviors increases the positive perceptions of the school climate. On the other hand, Nathaniel et al. (2016) found that there was a negative relationship between school climate and exam-based accountability policies. Accordingly, it can be said that the increase in exam-based accountability policies plays a role in the negative perceptions regarding the school climate. These studies revealed that the significant positive relationship between school climate and distributed leadership, and the significant and negative relationship between school climate and accountability are not lime with the significant positive relationship between distributed leadership and accountability behaviors of school administrators.

5. LIMITATIONS and SUGGESTIONS

Based on the results of this study, the following suggestions were made to educators, policy makers and researchers:

While this study provides significant insights, it only includes a sample of teachers from a single province in eastern Turkey. Therefore, the findings cannot be generalized to other provinces or regions. It is necessary to conduct further studies in different educational contexts to support and replicate the current findings. Additionally, more research is required to clarify the relationships between the variables discussed in the study. Researchers are advised to carry out similar studies in diverse educational contexts using different quantitative methods to uncover the relationships between school structure, distributed leadership, and accountability behaviors of school administrators more precisely. Conducting comparable studies in diverse cultural settings may also help to extend or limit research findings to wider contexts. In addition, this study involved teachers, and the findings were limited to the items in the data collection tools. Another limitation of this study is its cross-sectional design, which prevents it from establishing cause-effect relationships and limits it to being purely descriptive. Therefore, longitudinal studies are needed for more reliable results. However, based on the study's findings, it is crucial for school administrations to encourage teachers to take on responsibilities beyond their classrooms, alongside their teaching roles, and operate within the limits of their expertise. Potential stumbling blocks to a more democratic, participatory and free administrative structure of schools should be removed and professional development programs should be organized for school administrators. An education policy that will encourage school administrators to distribute their duties and responsibilities with teachers based on their knowledge, experience and expertise should be implemented. A performance evaluation approach in which all school stakeholders are responsible and accountable for the success performance of schools should be adopted. In this sense, a performance evaluation approach should be adopted in which teachers can have a say in the decision-making mechanisms and take responsibility in schools. Accordingly, legal measures should be taken.

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**: Firat University, Social Sciences and Humanities Research Ethics Committee, 29/07/2019-340737.

Authorship Contribution Statement

Mustafa Orhan: Investigation, Resources, Methodology, Software, Formal Analysis, and Writing-original draft. **Tuncay Yavuz Özdemir:** Supervision, Visualization, and Validation.

Orcid

Mustafa Orhan https://orcid.org/0000-0002-6254-6733
Tuncay Yavuz Özdemir https://orcid.org/0000-0002-5361-7261

REFERENCES

Abelmann, C., Elmore, R., Even, J., Kenyon, S., & Marshall, J. (1999). *When accountability knocks, will anyone answer?*. CPRE Policy Briefs.

Aboudahr, S.M.F.M., & Jiali, L. (2019). The relationship between distributed leadership and teachers' organizational commitment in Guangzhou city, China. *J. Educ. Manage. Stud., 9*(1), 1-5. http://jems.science-line.com

Adıgüzelli, Y. (2016). Examining the relationship between distributed leadership and organizational trust according to opinions of teachers. *Education & Science, 41*(185), 269–280. https://doi.org/10.15390/EB.2016.6218

Adler, P.S., & Borys, B. (1996). Two types of bureaucracy: Enabling and coercive. *Administrative Science Quarterly, 41*(1), 61-89. http://www.jstor.org/stable/2393986

Akdemir, Ö.A., & Ayık, A. (2017). The impact of distributed leadership behaviors of school principals on the organizational commitment of teachers. *Universal Journal of Educational Research, 5*(12), 18–26. https://doi.org/10.13189/ujer.2017.051402

Alev, S. (2019). Kolaylaştırıcı okul yapısı ve örgütsel vatandaşlık davranışları arasındaki ilişkinin incelenmesi [Examination of relationship between enabling school structure and organizational citizenship behaviors]. *Uluslararası Toplum Araştırmaları Dergisi, 10*(17), 420-443. https://doi.org/10.26466/opus.518009

Anderson, K. (2012). *Examining relationships between enabling structures, academic optimism and student achievement* [Unpublished doctoral dissertation]. Auburn University.

Anderson, K., Kochan, F., Kensler, L.A., & Reames, E.H. (2018). Academic optimism, enabling structures, and student achievement: Delving into relationships. *Journal of school leadership*, *28*(4), 434-461.

Argon, T. (2015). Teacher and administrator views on school principals' accountability. *Educational Sciences: Theory and Practice*, *15*(4), 925-944. https://doi.org/10.12738/estp.2015.4.2467

Argon, T., Uylas, S.D., & Yerlikaya, S. (2015). Perceptions of teachers related accountability practices ın Turkish national education sistem and organizational cynicism. *International Online Journal of Educational Sciences*, *7*(2), 204-219. http://dx.doi.org/10.15345/iojes.2015.02.016

Arik, S. (2021). The predictive powers of organizational trust and justice on enabling school structure: A structural equation modelling approach. *African Educational Research Journal*, *9*(2), 449-460. https://doi.org/10.30918/AERJ.92.21.044

Ataş Akdemir, Ö. (2016). *Öğretmen algılarına göre okul yöneticilerinin dağıtımcı liderlik davranışlarının okulların akademik iyimserlikleri ile örgütsel bağlılık ve okul başarısıyla ilişkisi [The relationship of distributed leadership behaviours of school principals with*

schools' academic optimism and organizational commitment and school achievement according to teacher perceptions] [Unpublished doctoral thesis]. Atatürk University.

Atılkan, N. (2019). *Okul müdürünün dağıtımcı liderlik davranışları ile okulların etkililiği arasındaki ilişkinin incelenmesi [Relationship between school rincipals's distributed leadership behaviors and school effectiveness]* [Unpublished master thesis]. Karabük University.

Aucoin, P., & Heintzman, R. (2000). The Dialectics of Accountability for Performance in Public Management Reform. *International Review of Administrative Sciences, 66*(1), 45–55. https://doi.org/10.1177/0020852300661005

Baron, R.M., & Kenny, D.A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173-1182. https://doi.org/10.1037/0022-3514.51.6.1173

Behn, R.D. (2003a). Rethinking accountability in education: How should who hold whom accountable for what? *International Public Management Journal, 6*(1), 43-73.

Behn, R.D. (2003b). *Demokratik hesap verme sorumluluğu [Rethinking democratic accountability].* (Çev: Paradigma Tercüme Ltd. Şti). Arı Yayıncılık.

Bell, L., Bolam, R., & Cubillo, L. (2003). *A systematic review of the impact of school headteachers and principals on student outcomes.* EPPI-Centre.

Beycioğlu, K., & Aslan, B. (2010). Öğretmen liderliği ölçeği: Geçerlik ve güvenirlik çalışması [Teacher leadership scale: A validity and reliability Study]. *İlköğretim Online, 9*(2), 764-775.

Bovée, C.L., & Till, J.V. (2012). *Business in action* (6th ed.). Pearson Education.

Bozkuş, K., Karacabey, M.F., & Özdere, M. (2019) Okulun kolaylaştırıcı örgütsel yapısı ile örgütsel sessizlik ve iş tatmini arasındaki ilişkiler [The relationships among enabling school structure, organizational silence and job satisfaction]. *Eğitim ve İnsani Bilimler Dergisi: Teori ve Uygulama*, *10*(19), 107-129.

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.

Buluç, B. (2009). İlköğretim okullarında bürokratik okul yapısı ile okul müdürlerinin liderlik stilleri arasındaki ilişki [The relationship between bureaucratic school structure and leadership styles of school principals in primary schools]. *Eğitim ve Bilim*, *34*(152), 71-86.

Cansoy, R., & Parlar, H. (2018). Examining the relationships among trust in administrator, distributed leadership and school academic optimism. *Kuram ve Uygulamada Eğitim Yönetimi, 24*(1), 1-44. https://doi.org/10.14527/kuey.2018.001

Cemaloğlu, N. (2013). Liderlik [Leadership]. In S. Özdemir (Ed.), *Eğitim yönetiminde kuram ve uygulama* (p.131-183). Pegem Akademi.

Cerit, Y. (2012). Okulun bürokratik yapısı ile sınıf öğretmenlerinin profesyonel davranışları arasındaki ilişki [The relationship between bureaucratic school structure and classroom teachers' professional behaviour]. *Kuram ve Uygulamada Eğitim Yönetimi, 4*(4), 497-521.

Cerit, Y., & Akgün, N. (2015). Okulun örgüt yapısı ile sınıf öğretmenlerinin proaktif davranışları arasındaki ilişki [The relationship between school organizational structure and classroom teachers' proactive behaviors]. *Kastamonu Eğitim Dergisi*, *23*(4), 1789-1802.

Chang, I.H. (2011). A study of the relationships between distributed leadership, teacher academic optimism and student achievement in Taiwanese elementary schools. *School Leadership & Management, 31*(5), 491-515. https://doi.org/10.1080/13632434.2011.614945

Çakır, Ç. (2019). Dağıtılmış liderlik ile örgütsel vatandaşlık davranışı arasındaki ilişkide örgütsel kimliğin rolü [The role of organizational identity on the relationship between distributed leadership and organizational citizenship behaviour]. *Turkish Studies-Educational Sciences*, *14*(4), 1189-1209. https://doi.org/10.29228/TurkishStudies.23482

Çalık, T., & Tepe, N. (2019). Ortaokul öğretmenlerinin algılarına göre okul etkiliği ile kolaylaştırıcı okul yapısı ve akademik iyimserlik arasındaki ilişki [The relationship between school effectiveness and enabling school structure and academic optimism according to the middle school teachers perceptions]. *Kastamonu Eğitim Dergisi*, *27*(4), 1471-1480. https://doi.org/10.24106/kefdergi.3008

Çiçek, E. (2018). *Eğitim kurumlarında dağıtımcı liderlik ve örgütsel güven arasındaki ilişki: Kayseri il merkezi örneği [The relationship between distributed leadership and organizational confidence in educational institutions: Kayseri province case]* [Unpublished master thesis]. Nevşehir Hacı Bektaş Veli University.

Çiçekli, A. (2016). *Kamu yönetiminde hesap verebilirlik ve BİMER'in hesap verebilir yönetime etkisi: Van sağlık hizmetleri örneği [Accountability in public administration, the effect of BİMER on accountable goverment; the example of Van health service managements]* [Unpublished master thesis]. Yüzüncü Yıl University.

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamalı [Multivariate statistics for the social sciences: SPSS and LISREL applications]*. Pegem Yayınları.

Çokyiğit, M. (2020). *Ortaokulların bürokratik yapısının etkiliği örgütsel güven ve örgütsel destek düzeyi arasındaki ilişki [The relationship between effectiveness of the bureaucratic structure of secondary schools organizational trust and organizational support level]* [Unpublished doctoral thesis]. Gazi University.

Çomak, M. (2021). *Okul yöneticilerinin dağıtımcı liderlik uygulama becerileri ile örgüt iklimi arasındaki ilişkiler [Relationships between distributed leadership practice skills of school administrators and organizational climate]* [Unpublished master thesis]. Kırşehir Ahi Evran University.

Demirtaş, Z., Özdemir, T.Y., & Küçük, Ö. (2016). Okulların bürokratik yapısı, örgütsel sessizlik ve örgütsel sinizm arasındaki ilişki [Relationships between bureaucratic structure of schools, organizational silence and organizational cynicism]. *Kuram ve Uygulamada Eğitim Yönetimi Dergisi, 22*(2), 193-216. https://doi.org/10.14527/kuey.2016.008

Doğan, D. (2015). *Türkiye'deki yükseköğretim kurumlarında hesap verebilirlik ve akademik özgürlük [Accountability and academic freedom in higher education institutions in Turkey]* [Unpublished doctoral thesis]. Eskişehir Osmangazi University.

Elmore, R.F. (2000). *Building a new structure for school leadership*. Albert Shanker Institute.

Elmore, R.F. (2005). Accountable leadership. *The Educational Forum*, *69*(2), 134-142.

Ereş, F., & Akyürek, M.İ. (2016). İlkokul müdürlerinin dağıtılmış liderlik davranışları ile öğretmenlerin iş doyumu algıları arasındaki ilişki [Relationship levels between distributed leadership behaviors of principals and the teachers' perception of job satisfaction]. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, *36*(3), 427-449.

Figlio, D., & Loeb, S. (2011). School accountability. In Hanushek, E.A., Machin, S., & Woessmann, L. (Eds.), *Handbooks in economics: Economics of education*, (Vol. 3, pp. 383–421). Elsevier.

Flowchart Maker Online Diagram Software. https://www.draw.io/

Geist, J.R. (2002). *Predictors of faculty trust in elementary schools: Enabling bureaucracy, teacher professionalism, and academic press* [Unpublished doctoral dissertation]. The Ohio State University.

Gray, J. (2016). Investigating the Role of Collective Trust, Collective Efficacy, and Enabling School Structures on Overall School Effectiveness. *Education Leadership Review*, *17*(1), 114-128.

Halim, R.A., & Ahmad, H. (2016). Distributed leadership, contextual factor and teachers' self-efficacy in Malaysia. *Educational Leader (Pemimpin Pendidikan)*, *4*, 31-38.

Harris, A. (2005). *Crossing boundaries and breaking barriers: Distributing leadership in schools.* Specialist Schools Trust.

Harrison, N.L.T. (2005). *The impact of distributed leadership on teachers* [Unpublished doctoral dissertation]. Indiana University.

Heck, R.H., & Hallinger, P. (2009). Assessing the contribution of distributed leadership to school improvement and growth in math achievement. *American Educational Research Journal, 46*(3), 659–689. https://doi.org/10.3102/0002831209340042

Hermann, K.R. (2016). *The Principal's role; distributed leadership* [Unpublished doctoral dissertation]. Old Dominion University.

Hoy, W.K. (2003). An analysis of enabling and mindful school structures: Some theoretical, research and practical considerations. *Journal of Educational Administration*, *41*(1), 87-109. https://doi.org/10.1108/09578230310457457

Hoy, W.K., & Sweetland, S.R. (2000). School bureaucracies that work: Enabling, not coercive. *Journal of School Leadership*, *10*(6), 525-541. https://doi.org/10.1177/105268460001000603

Hoy, W.K., & Sweetland, S.R. (2001). Designing better schools: The meaning and measure of enabling school structures. *Educational Administration Quarterly, 37*(3), 296–321.

Hoy, W.K., & Sweetland, S.R. (2007). Designing better schools: The meaning and nature of enabling school structures. In W. Hoy, & M. DiPaola (Eds.), *Essential ideas for the reform of American schools* (pp. 339-366). Information Age Publishing.

Hulpia, H., & Devos, G. (2010). How distributed leadership can make a difference in teachers' organizational commitment? A qualitative study. *Teaching and Teacher Education, 26*(3), 565–575.

Hulpia, H., Devos, G., & Rosseel, Y. (2009). The relationship between the perception of distributed leadership in secondary schools and teachers' and teacher leaders' job satisfaction and organizational commitment. *School Effectiveness and School Improvement*, *20*(3), 291-317.

Hulpia, H., Devos, G., & Van Keer, H. (2009). The influence of distributed leadership on teachers' organizational commitment: A multilevel approach. *The Journal of Educational Research*, *103*(1), 40-52.

Jacob, J.A. (2003). *A study of school climate and enabling bureaucracy in select New York City public elementary schools* [Unpublished doctoral dissertation]. St. John's University.

Jacobs, G.E. (2010). *The relationship between distributed leadership as practiced by principals and the organizational commitment of the teachers* [Unpublished doctoral dissertation]. Georgia Southern University.

Jofreh, M., Mohammadi, F., & Yasini, A. (2012). Leadership distribution consequences in schools: A particular look at organizational citizenship behavior of teacher. *Australian Journal of Basic and Applied Sciences*, *6*(12), 259-268.

Kalman, M., & Gedikoğlu, T. (2014). Okul yöneticilerinin hesap verebilirliği ile örgütsel adalet arasındaki ilişkinin incelenmesi [An investigation of the relationship between school administrators' accountability and organizational justice]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *29*(2), 115-128.

Kandemir, A., & Akgün, N. (2019). Ortaokul öğretmenlerinin hesap verebilirlik algıları ile yöneticilerinin hizmetkâr liderlik davranış düzeyleri arasındaki ilişki [Relationship between the accountability perceptions of secondary school teachers and their managers'

level of servant leadership behavior]. *Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 19*(4), 1637-1653. https://dx.doi.org/10.17240/aibuefd.2019..-541661

Karataş, İ.H. (2022). Türkiye'de okul yöneticiliğinin açmazları [The deadlocks of school management in Turkey]. *Okul Yönetimi, 1*(1), 9-18.

Karagöz, S. (2020). *Anaokulu öğretmenlerinin hesap verebilirlik ve örgütsel sessizlik davranışları arasındaki ilişki [The relationship between accountability and organizational silence behaviors of kindergarten teachers]* [Unpublished master thesis]. İstanbul Okan University.

Karaoğlan, Ö. (2019). *Okulların bürokratik yapısı, örgütsel sessizlik ve örgütsel sinizm arasındaki ilişki [The relationship between the bureaucratic structure of the schools, organizational silence and organizational cynicism]* [Unpublished master thesis]. Pamukkale University.

Karasar, N. (2009). *Bilimsel araştırma yöntemi [Scientific research method].* Nobel Yayın Dağıtım.

Kılınç, A.Ç., Koşar, S., Er, E., & Öğdem, Z. (2016). The relationship between bureaucratic school structures and teacher self-efficacy. *McGill Journal of Education/Revue des sciences de l'éducation de McGill*, *51*(1), 615-634. https://doi.org/10.7202/1037362ar

Kim, H.S. (2005). *Organizational structure and internal communication as antecedents of employeeorganization relationships in the context of organizational justice: A multilevel analysis* [Unpublished doctoral dissertation]. University of Maryland.

Koçak, S., & Nartgün, Ş.S. (2018). Öğretmenlerin okul hesap verebilirliğine ilişkin görüşlerini belirlemeye yönelik nitel çalışma [Qualitative study to determine teachers' views on school accountability]. In M. Koçali (Ed.), *9. Uluslararası Eğitim Yönetimi Forumu* (p. 235-254). Türkiye: Antalya.

Kondo, S. (2002). Fostering dialogue to strengthen good governance. *Public sector transparency and accountability: Making it happen*, 7-11.

Koohi, A., Hashemi, A., Minaei, A., & Dehghan, A. (2019). The relation between empowering structure in schools with teachers career satisfaction and schools efficiency in elementary schools in Qarchak. *Career and Organizational Counseling*, *11*(40), 99-114.

Koster, G.L. (2016). *Relationship between school structure and collective student trust in teachers* [Unpublished doctoral dissertation]. Oklahoma State University.

Kurt, T. (2016). Öğretmen liderliğini açıklamaya yönelik bir model: Dağıtımcı liderlik, örgütsel öğrenme ve öğretmenlerin öz yeterlik algısının öğretmen liderliğine etkisi [A model to explain teacher leadership: The effects of distributed leadership, organizational learning and teachers' sense of self-efficacy on teacher leadership]. *Eğitim ve Bilim*, *41*(183), 1-28. https://doi.org/10.15390/EB.2016.5081

Leithwood, K. (2001). Educational accountability and school leadership. *National College of School Leadership.* http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.495.5426&rep=rep1&type=pdf.

Maile, S. (2002). Accountability: An essential aspect of school governance. *South African Journal of Education*, *22*(4), 326-331.

Mascall, B., Leithwood, K., Straus, T., & Sacks, R. (2008). The relationship between distributed leadership and teachers' academic optimism. *Journal of Educational Administration, 46*(2), 214-228. https://doi.org/10.1108/09578230810863271

Mayerson, D.R. (2010). *The relationship between school climate, trust, enabling structures, and perceived school effectiveness* [Unpublished doctoral dissertation]. St. John's University.

Mayrowetz, D. (2008). Making sense of distributed leadership: Exploring the multiple usages of the concept in the field. *Educational Administration Quarterly, 44*(3), 424-435. https://doi.org/10.1177/0013161X07309480

McGuigan, L., & Hoy, W.K. (2006). Principal leadership: Creating a culture of academic optimism to improve achievement for all students. *Leadership and Policy in Schools, 5*(3), 203–229. https://doi.org/10.1080/15700760600805816

Messick, P.P. (2012). *Examining relationships among enabling school structures, academic optimism, and organizational citizenship behaviors* [Unpublished doctoral dissertation]. Auburn University.

Meydan, C.H., & Şeşen, H. (2011). *Yapısal eşitlik modellemesi AMOS uygulamaları* [Structural equation modeling AMOS applications]. Detay Yayıncılık.

Miskel, C. (1979). Demographic characteristics, faculty attitudes and school structure. *Paper presented at The Annual Meeting of The American Educational Research Association*. San Francisco, California, April 8-12.

Miskel, C.G., Fevurly, R., & Stewart, J. (1979). Organizational Structures and Processes, Perceived School Effectiveness, Loyalty, and Job Satisfaction. *Educational Administration Quarterly, 15*(3), 97–118. https://doi.org/10.1177/0013131X7901500308

Mitani, H. (2018). Principals' working conditions, job stress, and turnover behaviors under NCLB accountability pressure. *Educational Administration Quarterly*, *54*(5), 822-862. https://doi.org/10.1177/0013161X1878587

Mitchell, R.M. (2019). Enabling school structure & transformational school leadership: Promoting increased organizational citizenship and professional teacher behavior. *Leadership and Policy in Schools*, *18*(4), 614-627. https://doi.org/10.1080/15700763.2018.1475577

Nathaniel, P., Pendergast, L.L., Segool, N., Saeki, E., & Ryan, S. (2016). The influence of test-based accountability policies on school climate and teacher stress across four states. *Teaching and Teacher Education*, *59*, 492-502. https://doi.org/10.1016/j.tate.2016.07.013

Oldaç, Y.İ., & Kondakçı, Y. (2020). Multilevel analysis of the relationship between school-level variables and student achievement. *Educational Management, Administration & Leadership, 48*(4), 762-780. https://doi.org/10.1177/1741143219827303

Ordu, A., & Tanrıöğen, A. (2013). İlköğretim okullarında örgütsel yapı ile örgüt sağlığı arasındaki ilişkiler [The relationships between organizational structure and organizational health in primary schools]. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, *33*(33), 93-106.

Orhan, M. (2022). *Okul yapısı ile hesap verebilirlik arasındaki ilişkide dağıtımcı liderliğin aracılık rolünün incelenmesi [The examination of mediating role of distributed leadership in the relationship between school structure and accountability]* [Unpublished doctoral thesis]. Fırat University.

Özdemir, S., & Kılınç, A.Ç. (2014). The relationship between bureaucratic school structure and teachers' level of academic optimism/Bürokratik okul yapısı ile öğretmenlerin akademik iyimserlik düzeyleri arasındaki ilişki [The relationship between bureaucratic school structure and teachers' level of academic optimism]. *Eğitimde Kuram ve Uygulama*, *10*(1), 1-23.

Özen, Y., & Gül, A. (2007). Sosyal ve eğitim bilimleri araştırmalarında evren-örneklem sorunu [Populatıon-sampling issue on social and educational research studies]. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi*, (15), 394-422.

Özer, N., & Beycioğlu, K. (2013). Paylaşılan liderlik ölçeğinin geliştirilmesi: Geçerlik ve güvenirlik çalışmaları [The development, validity and reliability study of distributed leadership scale]. *İlköğretim Online*, *12*(1), 77-86.

Özer, N., & Dönmez, B. (2013). Kolaylaştırıcı okul yapısı ölçeğinin Türkçe formunun psikometrik özelliklerinin yeniden değerlendirilmesi [Reassesing the psychometric

properties of the turkish version of enabling school structure scale]. *Pegem Eğitim ve Öğretim Dergisi, 3*(4), 57-68.

Preacher, K.J., & Hayes, A.F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, *40*(3), 879-891. https://doi.org/10.3758/BRM.40.3.879

Rabindarang, S., Bing, K.W., & Yin, K.Y. (2014). The influence of distributed leadership on job stress in technical and vocational education. *International Journal of Academic Research in Business and Social Sciences*, *4*(1), 490-499. https://doi.org/10.6007/IJARBSS/v4-i1/555

Rockoff, J.E., & Turner, L.J. (2010). Short-run impacts of accountability on school quality. *American Economic Journal: Economic Policy*, *2*(4), 119-147.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluatingthe fit of structural equation models: Tests of significance and descriptive goodnessof-fit measures. *Methods of Psychological Research Online, 8*(2), 23-74.

Sinden, J.E., Hoy, W.K., & Sweetland, S.R. (2004). An analysis of enabling school structure: Theoretical, empirical, and research considerations. *Journal of Educational Administration, 42*(4), 462-478. https://doi.org/10.1108/09578230410544071

Smylie, M.A., Mayrowetz, D., Murphy, J., & Louis, K.S. (2007). Trust and the Development of Distributed Leadership. *Journal of School Leadership, 17*(4), 469–503.

Stapleman, J. (2000). *Standards-based accountability systems* [policy brief]. Mid-continent Research for Education and Leaming.

Stecher, B., & Kirby, S.N. (Eds.). (2004). *Organizational improvement and accountability: Lessons for education from other sectors.* Rand Publication.

Sümer, N. (2000). Yapısal eşitlik modelleri: Temel kavramlar ve örnek uygulamalar [Structural equation modeling: Basic concepts and applications]. *Türk Psikoloji Yazıları, 3*(6), 49-74.

Şişman, M. (2019). *Türk eğitim sistemi ve okul yönetimi [Turkish education system and school administration]* (11. Baskı). Pegem Akademi.

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th Edition). Pearson Education.

Tavşancıl, E. (2002). *Tutumların ölçülmesi ve SPSS ile veri analizi [Measurement of attitudes and data analysis with* SPSS*].* Nobel Yayıncılık.

Turner, J.B. (2018). *Organizational justıce and enabling school structure as predictors of organizational citizenship behavior* [Unpublished doctoral dissertation]. The University of Alabama.

Wahab, J.A., Hamid, A.H.A., Zainal, S., & Rafik, M.F.M. (2013). The relationship between headteachers' distributed leadership practices and teachers' motivation in national primary schools. *Asian Social Science*, *9*(16), 161-167. https://doi.org/10.5539/ass.v9n16p161

Watts, D.M. (2009). *Enabling school structure, mindfulness, and teacher empowerment: Test of a theory* [Unpublished doctoral dissertation]. The University of Alabama.

Woods, P.A., & Gronn, P. (2009). Nurturing democracy: The contribution of distributed leadership to a democratic organizational landscape. *Educational Management Administration & Leadership, 37*, 430–451. https://doi.org/10.1177/1741143209334597

Yağ, T. (2019). *Okul yöneticilerinin yönetim tarzları ile hesap verebilirlik yönelimleri arasındaki ilişki [The relationship between management styles of school administrators' and accountability orientations]* [Unpublished master thesis]. Kırşehir Ahi Evran University.

Yetim, Z. (2015). *İlköğretim kurumlarında dağıtımcı liderlik ile örgütsel bağlılık arasındaki ilişki (Seydikemer örneği) [The relation between the distributed leadership and*

organizational commitment in the elementary educational institutions (The case of Seydikemer)] [Unpublished master thesis]. İstanbul Aydın University.

Yılmaz, D. (2014). *Dağıtılmış liderliğin örgütsel güven ile algılanan örgütsel destek ve okul başarısıyla ilişkisi [The relationship between distributed leadership, organizational trust, perceived organizational support and school success]* [Unpublished doctoral thesis]. Eskişehir Osmangazi University.

Yiğit, B. (2017). *Yöneticilerin hesap verebilirlik uygulamalarının çalışanların örgütsel bağlılık davranışlarıyla ilişkisinin incelenmesi- Eğitim kurumları örneği [Examining the relation between accountability practices of directors and the organizational commitment behaviors of the employees: A case study on educational institutions]* [Unpublished master thesis]. Türk Hava Kurumu University.

Research Article

Adapting the teacher formative assessment literacy scale into Turkish: Validation and reliability study

Kemal İzci [ID][1*], **İlhan İlter** [ID][2], **Gökhan Izgar** [ID][1]

[1]Necmettin Erbakan University, Faculty of Education, Department of Educational Sciences, Konya, Türkiye
[2]Kahramanmaraş Sütcü Imam University, Faculty of Education, Department of Social Studies, Kahramanmaraş, Türkiye

Abstract: In the present study, the Teacher Formative Assessment Literacy Scale (TFALS), developed by Yan and Pastore (2022a), was translated into Turkish to examine the factor structure and psychometric characteristics of the scale in a Turkish sample. Data was collected from 318 teachers, of whom 168 were males and 150 were females, working in different state schools in a city in the Central Anatolia Region of Türkiye. As the first step, the scale items were translated and back-translated by experts in both English and Turkish. Afterward, experts were presented with the scale to check for consistency and accuracy based on the feedback received. A pilot study was carried out to establish the linguistic equivalence of the scale, followed by an examination of its structural validity in a Turkish sample. To demonstrate evidence of the confirmed factor structure of TFALS with 22 items, the confirmatory factor analysis (CFA) was conducted. The CFA results demonstrated that the three-factor model of the TFALS-Turkish-version had adequate fit indices. The Cronbach's alpha coefficient of the total scale was found to be .93. The Cronbach's alpha for each of the dimensions of TFALS was .90 for the conceptual dimension (7 items), .86 for the practical dimension (8 items), and .88 for the socio-emotional dimension (7 items). Composite reliability coefficients of the dimensions ranged from .78 to .84. The results of this study provide important evidence for the validity and reliability of the Turkish version of the TFALS, confirming that it has good psychometric properties in a Turkish sample

1. INTRODUCTION

Assessment and evaluation are the essential elements of an effective instructional process. Assessment and evaluation provide data both on the impact of education policies at large and on the quality and level of learning that takes place at the classroom level. In particular, studies conducted in recent years have shown that teachers' data-based decisions in the teaching process support student success (Bennet, 2011; Brookhart, 2018; Lee et al., 2020; Li, 2016). In this context, it is important that teachers both collect data showing their students' learning as a process and result and use these data to support teaching and learning (Yan & Pastore, 2022a). Assessment and evaluation can be used for diverse purposes in the educational process.

*CONTACT: Kemal İzci ✉ kemalizci@gmail.com 🖃 Necmettin Erbakan University, Faculty of Education, Department of Educational Sciences, Konya, Türkiye

Researchers classify the purposes of assessment and evaluation in education into three main categories: accountability, certification, and instructional purposes (Archer, 2017; Black & Wiliam, 2009; Brookhart & McMillan, 2020). Firstly, each country needs to monitor the success of its education policies at a broad level. Thus, educational institutions use national and international comparative assessments to illustrate the results of educational efforts for accountability purposes. Secondly, educational institutions need to provide certifications or diplomas to illustrate that an individual student has the knowledge and skills required to progress to the next level of education, to transfer to new types of schools, and to apply for a job requiring the certification. Thus, institutions use assessment and evaluation to provide certification.

Finally, at the classroom level, assessment and evaluation can be used for instructional purposes, including diagnosing students' prior knowledge and misconceptions, monitoring learning, providing feedback, and demonstrating the success of teaching and learning. When we look at the instructional purposes of assessment and evaluation, we see that they are generally classified as summative and formative (Black & Wiliam, 1998). Summative assessment focuses on determining students' level of learning in order to award them grades that certify their learning. Therefore, it is believed that the influence of summative assessment on shaping instruction and improving learning is limited. Formative assessment, on the other hand, is thought to support learning by enabling the teacher to monitor the learning process and the students to organize their own learning. For this reason, it is important that teachers understand formative assessment and use it in their teaching to facilitate effective and high-quality learning (Bennet, 2011; Black & Wiliam, 1998; Brookhart & McMillan, 2020; Izci, 2016; Shepard et al., 2017).

1.1. Theoretical Framework

1.1.1. *Formative assessment*

Researchers define formative assessment as the use of assessment processes to gather data about students' learning in order to aid their learning (Black & Wiliam, 2009; Haritage, 2010; Wylie, 2020). Teachers can use a variety of methods, such as observations, quizzes, and class discussions, in order to gain knowledge about their students' learning progress. However, the critical point in formative assessment is to use the collected information in a way to support learning. Providing effective feedback and adapting teaching to meet students' needs are some ways to use formative assessment results to facilitate students' learning. Formative assessment stands out as an important instructional approach because of its role in regulating instruction and enhancing learning (Black & Wiliam, 2009; Wylie, 2020).

When looking at the relevant literature, two approaches to formative assessment come to the fore. The first of these approaches is data-based decision-making, and the second is assessment for learning. The data-based decision-making approach involves collecting, interpreting, and using data from different sources, formally or informally, in order for students to achieve a specific learning goal (Schildkamp & Kuiper, 2010; Schildkamp et al., 2020). The students' learning levels are determined by analyzing the collected data, and tailored teaching activities are then provided to help them achieve the desired level. In this approach, the goal is the realization of learning. On the other hand, the assessment for learning approach aims to improve the quality of the learning process. Klenowski (2009) states that the purpose of assessment for learning is to aid and encourage the ongoing learning process.

What is important here is the interpretation and use of data collected through teaching activities such as dialogues, demonstrations, and observations made by students, teachers, and others. Based on this definition, the information collected about the learning process generally includes the information collected informally in the process. The effective use of this approach enhances the effectiveness of teaching and the success of students, as shown in previous studies (Pinger

et al., 2018; Shavelson, 2013). According to Bennet (2011), the success of assessment for learning depends on teachers' obtaining useful information about their students, interpreting this information, and using this interpretation for instructional decisions and feedback to students.

Although formative assessment offers significant potential, the desired effect of formative assessment depends on its comprehensive understanding and application by teachers. Previous research studies have yielded mixed results on the effect of formative assessment (Furtak et al., 2016; Kepek & Izci, 2021). Some researchers argue that the reason why formative assessment does not produce the desired effect is that teachers do not use assessment information to make data-supported instructional decisions (Bennet, 2011; Schildkamp et al., 2020). However, it has been concluded in some studies in the literature that how teachers use formative assessment is determinative of increasing learning (Black & Wiliam, 1998; Torrence, 2012). Hence, it can be posited that teachers hold a pivotal role in the achievement of formative assessment. In an investigation conducted by Cañadas and colleagues (2021), the potential gains and benefits of formative assessment were explored. This study's findings showed that implementing formative assessment techniques in educational environments could enhance students' learning outcomes, cultivate self-directed learning abilities, facilitate significant and contextual learning experiences, and encourage metacognitive processes, especially through collaborative and self-evaluation. If the competencies that teachers must have for a successful formative assessment are known, then teachers can be trained within these competencies. Upon examination of the pertinent literature, it becomes apparent that various concepts, including assessment literacy, assessment identity, and assessment expertise, are utilized to describe teachers' assessment competencies (Abell & Siegel, 2011).

These models also cover cognitive, affective, and practical competencies related to teacher assessment competencies. For example, in a literature review conducted by Gotch and French (2014), it was determined that 36 scales existed in the literature on assessment literacy, and most of these scales had low validity and reliability evidence. Assessment literacy pertains to the knowledge, skills, and techniques required by teachers to appraise the comprehension and competencies of their students, analyze the outcomes of these assessments, and apply the findings to provide constructive feedback or adapt their pedagogical strategies (Abell & Siegel, 2011; Xu & Brown, 2016). Yan and Pastore (2022a) defined the concept of formative assessment literacy based on the concept of assessment literacy. Based on the definition provided, the concept of formative assessment literacy includes the fundamental knowledge, skills, and attitudes that educators need to have in order to proficiently use the assessment process. Such a process must improve the learning experience and refine pedagogical practice.

However, it has been observed that limited studies have been conducted on teacher competencies in formative assessment literacy. In one of these studies, Schildkapm et al. (2020) analyzed 54 studies that were conducted within the scope of formative assessment. The examination led to the recommendation of a comprehensive set of skills that encompass three dimensions. The first element comprises the knowledge and skills necessary for teachers to possess in relation to formative assessment. The second element consists of social factors that relate to teachers. The last element comprises psychological factors that relate to formative assessment. According to the framework, the dimension of knowledge and skills includes various competencies, such as but not limited to data and assessment literacy mastery, proficiency in pedagogical content knowledge, effectiveness in goal-setting, provision of valuable feedback, leading meaningful in-class discussions, and expertise in utilizing information and communication technologies. The social factors dimension, on the other hand, involves the competencies of teachers to cooperate with their colleagues and their ability to involve students in the process. The psychological factors dimension consists of the competencies of attitude or belief, ownership, social pressure, and perception of control.

Although a model was proposed by Schildkapm et al. (2020) within the scope of formative assessment, only one scale that could be used to determine and improve teachers' formative assessment literacy was reached (Yan & Pastore, 2022a).

1.1.2. *Formative assessment in Türkiye*

In recent years, there has been an increased emphasis on formative assessment in Türkiye's learning and teaching process, which aligns with international developments in this area (Cañadas, 2023; Menéndez et al., 2019; Schildkamp et al., 2020). This is to ensure that student needs are met effectively. It is possible to state that formative assessment is deemed important for evaluating the successful implementation of the curriculum. In this sense, teachers can guide their students to develop meaningful learning by choosing formative assessment tasks and methods (Biggs & Tang, 2011). In addition, within the scope of teacher competencies in Türkiye, the formative assessment is also highlighted by the Ministry of National Education (MoNE). While the MoNE does not provide a distinct definition for formative assessment, the MoNE (2013) explained the assessment vision of the 2013 curriculum as "it has adopted an assessment approach aimed at identifying learning difficulties and providing continuous feedback in order to monitor and guide instruction to support meaningful learning." (MoNE, 2013, p. IV).

Similarly, the "General Competencies for the Teaching Profession" document, which was issued by the MoNE in 2017, constitutes the Turkish teacher competencies and places great emphasis on formative assessment by requiring teachers to meet the following two competencies: "1) Giving correct and constructive feedback to students and other stakeholders by taking into account the results of assessment; and 2) Undertaking a reorganization of the procedures utilized for both teaching and learning, taking into account the outcomes obtained from assessments." (MoNE, 2017, p. 15). This shows that the MoNE of Türkiye values the formative function of assessment as it emphasizes the use of assessment to elicit student learning, provide feedback, and adjust instruction within the two documents. While the beneficial impact of formative assessment on the learning and teaching process is recognized, its potential practical value depends on teachers' application of the approach (Bennet, 2011; Black & Wiliam, 2009; Schildkamp et al., 2020). Thus, it is important to focus on teachers' formative assessment literacies.

A limited number of studies that focus directly on formative assessment literacy are noticeable in Türkiye (Aras, 2019; Bayrak et al., 2019; Buldur & Hasbek, 2020; Karaman, 2017; Karaman & Karaman, 2017; Kaya et al., 2021; Yasar, 2017). One of the reasons for this is that alternative assessment and evaluation are more emphasized in MoNE documents, and, therefore, researchers focus more on alternative assessment and evaluation in their studies. However, alternative assessment and evaluation refers to the use of different and flexible techniques that are different from the traditional techniques that can be used in the assessment of students' learning as a process and a product (Şahin & Kaya, 2020). Therefore, the studies conducted in this area in Turkey mostly focused on teachers' opinions, knowledge, and frequency of using alternative assessment and evaluation techniques (Ayan & Erdemir, 2023; Şahin & Kaya, 2020). When we look at the limited studies that focus on teachers' formative assessment literacies, it is seen that these studies mostly focus on any of the social-emotional, conceptual, and practical parts of formative assessment rather than focusing on teachers' formative assessment literacies as a whole. In addition, it is seen that the related studies mostly examined the formative assessment of the teachers by collecting qualitative data.

Aras (2019) conducted a study to support the development of formative assessment practices among three preschool teachers and obtained research data through semi-structured interviews and observations. The results of the study indicate that teachers made improvements, especially in the areas of collecting data about students' learning in a planned manner, developing lesson plans according to the collected data, and using portfolios in a way that supports student

participation. In another study conducted with preschool teachers, Karaman and Karaman (2017) examined the formative assessment practices of 12 service preschool teachers. In the study, in which the data was obtained through the notes kept by the participants and open-ended questions posed by the researcher, the findings showed that the participants used observation and follow-up as an assessment tool, integrated assessment and learning processes, and used process-oriented assessments.

In their study, Bayrak et al. (2019) focused on biology teachers' ability to identify students' learning deficiencies and their plans for addressing the deficiencies. According to the data obtained through semi-structured interviews, the authors concluded that teachers mostly identified students' learning deficiencies by using verbal questions and discussions during the lesson and tried to overcome learning deficiencies by giving short repetitions of the subject and different examples. Buldur and Hasbek (2020) used the metaphor test as a data collection tool in their study, in which they examined pre-service teachers' perceptions about formative assessment through metaphors. The findings of the study showed that there were 42 metaphors produced by 127 pre-service teachers, and these metaphors were classified in the categories of "improving student learning; identifying the learning gap; re-planning the teaching process; and assessment throughout the process." It was determined that the most metaphors were in the "Improving student learning" category. Kaya et al. (2021) aimed to examine science teachers' formative assessment awareness and the effect of this awareness on formative assessment practices. The authors collected data through a form consisting of open-ended questions and classroom observations. The findings showed that the participants exhibited three different levels of formative assessment awareness and practices: naive, eclectic, and conscious. It was found that the majority of the participants had an eclectic level, which meant that they had a high level of awareness but could not reflect this awareness into practice. Yaşar (2017) examined the perceptions of pre-service science teachers about formative assessment and obtained data from 17 participants through semi-structured interviews. The findings showed that pre-service science teachers mostly knew the definition and function of formative assessment, but they were very inadequate in terms of the purposes and uses of formative assessment and adapting instruction according to the results of formative assessment. Lastly, Karaman (2017) adapted a scale for determining pre-service teachers' attitudes and intentions towards formative assessment into Turkish and collected data from 301 pre-service teachers. The results showed that pre-service teachers' attitudes towards formative assessment and behavioral control levels affect their intentions to implement formative assessment.

As mentioned above, Turkish teachers' formative assessment literacy areas were mostly addressed by using qualitative research methods and data. Only Karaman (2017) translated a quantitative measurement tool that addressed teachers' perceptions and goals of formative assessment into Turkish. Thus, there is no quantitative tool to measure Turkish teachers' formative assessment literacy. Teachers' formative assessment abilities can be measured accurately and reliably through a valid tool, which in turn can help focus on learning-related elements. This can contribute to the development of teachers' teaching competencies and aid in the creation of effective strategies for instructional planning. In this context, the present study aims to adapt the TFALS developed by Yan and Pastore (2022a) into Turkish in order to measure the formative assessment literacy of teachers.

1.2. Aim of the Present Study

Yan and Pastore (2022a) have created a measuring tool called "The Teacher Formative Assessment Literacy Scale [TFALS]," which measures the formative assessment literacy of teachers. In the process of scaling relevant factors, the studies on formative assessment and the models and scales proposed concerning assessment literacy have been employed. The TFALS comprises three dimensions: conceptual, practical, and socio-emotional. It measures teachers' formative assessment literacy based on their statements. A study carried out by Yan and Pastore

(2022a) on primary and secondary school teachers showed that the TFALS had acceptable psychometric properties. One of the reasons for conducting the present research is that formative assessment has not been adequately addressed nationally, as mentioned above. Furthermore, just one study (Karaman, 2017) employed a quantitative approach by adapting a scale to investigate preservice teachers' aims and attitudes toward formative assessment. Given the constraints of time and cost, a decision was made to forego the development of a new scale in favor of translating an established, reliable scale that has been introduced to the international literature into Turkish. Besides, adapting a scale to a new group helps researchers investigate the validity and reliability of the scale for different groups of participants, which allows them to use the scale to compare groups. Furthermore, the adaptation of a pre-existing scale to a new cultural context has the potential to enrich the literature on the cultural validity of said scale.

Formative assessment undeniably impacts teaching and learning processes in the classroom (Schildkamp et al., 2020). Formative assessment has become a policy pillar with educational significance (Van der Kleij et al., 2018) due to its potential to improve student learning (Black & Wiliam, 1998). The fact that the scale has been used in different cultures, such as Hong Kong and Italy (Yan & Pastore, 2022b), and has high psychometric properties, created a motivation for its adaptation to Turkish culture. Adaptation of the TFALS into Turkish will be beneficial in terms of determining the levels of formative assessment literacy of our teachers, meeting their formative assessment needs by identifying them, and allowing for different comparisons. The lack of any quantitative studies on teachers' formative assessment literacy, especially at the national level, is attributed to the absence of a data collection instrument that can measure teachers' formative assessment literacy levels. Therefore, adapting the scale is expected to enhance teaching competencies and facilitate the implementation of formative assessment processes in education. In addition, the use of an instrument that measures teachers' formative assessment literacy can give future teachers an idea about making decisions concerning their instructional steps and how to improve their assessment practices. This research is expected to benefit teachers, students, and researchers.

2. METHOD

2.1. Research Method

This study aimed to examine the psychometric properties of the Teacher Formative Assessment Literacy Scale (TFALS) in a sample of Turkish teachers. The study employs a survey model, which is one of the quantitative research approaches, in order to achieve its aim. For the purpose of this study, the investigation was executed through the implementation of the stages involved in the process of scale adaptation.

2.2. Study Group

The sample consists of elementary and secondary school teachers who registered at the non-thesis master's degree in educational sciences at a higher education institution in Türkiye. The sample was composed of 318 teachers, of whom 150 were females (47.16%), and 168 were males (52.83%). Participants' ages ranged from 24 to 46, with a mean age of 34 years. Convenience sampling was used to recruit participant respondents for data collection. When we examined the teaching experience of the sample, 10 (3.1%) participants reported less than five years of experience (0- 5 years), 66 (20.8%) were between 6 and 10 years, 60 (18.9%) between 11 and 15 years, 70 (22%) between 16 and 20 years, and 112 (35.2%) between 21 years or more. Participants differed according to their teaching profession fields and worked at different school levels, such as elementary and secondary schools. The distribution of participants according to their fields of expertise is shown in Table 1.

Table 1. *Distribution of participant teachers by their teaching profession fields*.

Teaching profession fields	N	%	Teaching profession fields	N	%
Elementary school teachers	38	11.95	English teachers	24	7.55
Social studies teachers	35	11.01	Physical education & sports teachers	15	4.72
Turkish teachers	34	10.69	Special education teachers	13	4.09
Science teachers	34	10.69	Religious culture & moral teachers	12	3.77
Math teachers	38	11.95	Information technology teachers	15	7.72
History teachers	15	4.72	Biology teachers	9	2.83
Geography teachers	18	5.66	Chemistry teachers	8	2.52
Physics teachers	10	3.14			

Additionally, in this study, to ensure the language validity of the scale, data were collected from an independent sample group of 25 pre-service teachers enrolled in the final year of the English language-teaching department at a university. Most of these participants were female ($N = 18$).

2.3. Data Collection Tool

2.3.1. *Teacher formative assessment literacy scale*

The TFALS was developed by Yan and Pastore (2022a) to measure the formative assessment literacy profile of teachers. The scale was developed with two randomly selected sample groups from Hong Kong and Italy. A total of 449 teachers, 295 females and 151 males, working in 12 schools in Hong Kong, participated in the administration of the scale. Of these teachers, 263 worked in elementary schools and 186 in high schools. In the Italian sample, data were collected from 309 teachers, most of whom ($N = 278$) were female, in ten selected schools in the Apulian region. Out of 309 teachers, 134 work in primary schools and 175 in high schools. It was designed based on the theoretical model of assessment literacy introduced by Pastore and Andrade (2019). Recognizing the critical role of teachers' assessment literacy, this model was developed to assess the conceptual, practical, and socio-emotional aspects of teachers' formative assessment literacy. Therefore, the original scale has three factors (dimensions) and twenty-two items based on a deductive approach. The items are designed to provide a comprehensive picture of the formative assessment literacy of teachers.

The main factor in the development of this TFALS was to develop a self-report tool that would comprehensively assess teachers' formative assessment literacy. The conceptual dimension has to do with formative assessment principles and content knowledge. The practical dimension aims to assess a teacher's formative assessment practices in order to promote learning and teaching. The socio-emotional dimension assesses a teacher's understanding of the social and emotional aspects of formative assessment. The conceptual dimension consists of seven items (items 1, 2, 3, 4, 5, 6, and 7); the practical dimension consists of eight items (items 8, 9, 10, 11, 12, 13, 14, and 15); and the socio-emotional dimension includes seven items (items 16, 17, 18, 19, 20, 21, and 22). For all items, a six-point Likert-type response scale was used (1 = strongly disagree, 6 = strongly agree). The factor structure of the original TFALS was tested with exploratory factor analysis (EFA) by Yan and Pastore (2022a). The EFA results indicated that three factors had eigenvalues greater than one that accounted. In addition, confirmatory factor analyses (CFA) were performed by the authors to determine the construct validity of the TFALS with 22 items. The CFA results showed that the three-factor structure of the original scale had good model fit [(χ^2/df = 2.370; TLI = .910, CFI =.921; RMSEA = .069). Cronbach's alpha (α) reliability coefficients was found to be for the dimensions of conceptual, practical, and socio-emotional .88, .88, and .89, respectively (Yan & Pastore, 2022a).

2.4. Adaptation Processes

In this study, language equivalence (language and cultural adaptation), construct validity, item discrimination, and Cronbach's α coefficients were examined within the framework of the psychometric properties of TFALS. To translate the TFALS, a six-step process was followed consistent with recommendations in the literature (Eremenco et al., 2005; World Health Organization [WHO], 2017).

2.4.1. *Language and culture adaptation*

In the literature, it is apparent that the steps employed for cultural adaptation are intricately described. In this study, the adaptation process was refined by the WHO through numerous studies and then reduced to the steps shown in Figure 1 (WHO, 2017).

Figure 1. *Steps to be followed in the adaptation process*.
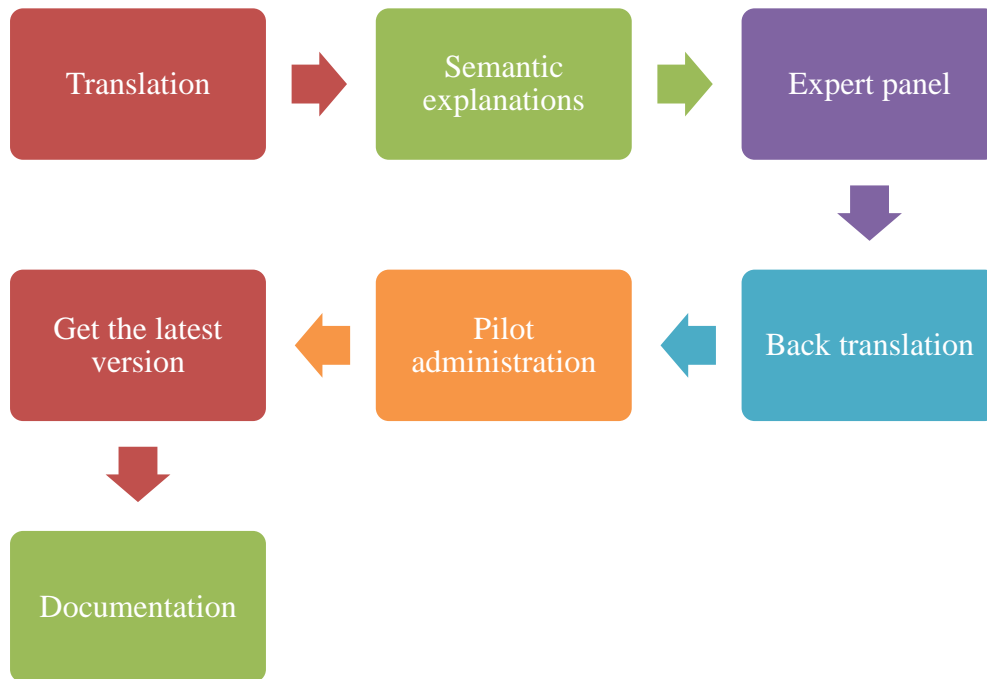
In order to adapt the original TFALS into Turkish, the authors of the scale were contacted via email to seek permission. Following the permission process, at the first stage, four faculty members who have mastery of both Turkish and English translated all of the items in the TFALS into Turkish. The experts are familiar with the culture of the original scale. During translation, they aimed for conceptual equivalence by independently translating the words or phrases in the items in the scale while staying very faithful to the original text. These experts avoided the use of any jargon when translating articles, such as technical terms, phrases, or colloquial terms that are not clearly understood. They determined whether items had changed meaning culturally (idiomatic equivalence). In particular, they examined whether it was culturally appropriate to apply each item to the target group (experimental equivalence). Thus, they evaluated whether the scale items questioned the same concepts in the new culture (conceptual equivalence) (Borsa et al., 2012).

The quality of the translation and the cultural relevance and intelligibility of the items on the scale were examined. By comparing the scale items obtained from the experts, a common Turkish form was created. In the second step, the TFALS-Turkish form was given to two experts who obtained their master's and doctoral degrees in an English-speaking country, and the experts were asked to translate the items in the Turkish form of the scale back into English. According to WHO (2017), this method, the back-translation of the scale, which passed the expert panel, should be done by an independent translator whose native language is English but

who also knows the translated language and has no knowledge of the scale. In the third step, a one-to-one comparison was made between each item's original expression and the expression that resulted from this translation. In the fourth step, after the back translation of the scale, the authors compare the two versions to identify differences between the back-translated form of the scale and the original scale. All experts reviewed all items in Turkish and English and investigated any semantic differences. In the fifth step, the TFALS-Turkish form obtained as a result of the comparison and the original scale were found to be generally equivalent, and thus the translation process was completed. In the sixth phase, the items in the scale translated into Turkish were individually evaluated by two experts who have studied Turkish language education in terms of language, intelligibility, significance, and clarity, and the scale was given its final form. According to WHO (2017), a pilot study should be conducted before claiming that a new scale is ready for data collection. Pilot implementation actually constitutes the last stage of the adaptation process. At this stage, a pilot application was made to 30–40 students for the clarity of the questions. In the selection of the sample, the aim was to reach the group that best reflected the target group. To ensure its linguistic equivalence, the TFALS-English version and the TFALS-Turkish version were administered at two-week intervals to a group of students (N = 25) from a higher education institution's English language education program. After passing all the steps described above, the authors achieved the final version of the scale. The results of the correlational analysis are presented in Table 2.

Table 2. *Correlation coefficient between the TFALS-Turkish and English versions.*

Factor	r
Factor 1	.91**
Factor 2	.85**
Factor 3	.83**
Total	.87**

**$p < .01$, Factor 1 = Conceptual, Factor 2 = Practical, Factor 3 = Socio-emotional

As seen in Table 2, the results revealed that all dimensions in both the TFALS-Turkish and TFALS-English versions were significantly correlated with each other (conceptual dimension $r= .91$, $p<. 01$, practical dimension $r= .85$, $p<.01$, and socio-emotional dimension $r= .83$, $p<. 01$). According to Büyüköztürk (2014), correlation coefficients between .30 and .70 indicate a moderate correlation, whereas those greater than .70 indicate a high correlation. The result indicated that linguistic equivalence was achieved between the TFALS-English and TFALS-Turkish versions.

2.5. Normal Distribution Analysis

Before conducting factor analyses, this study examined the assumption of normality using the AMOS tests for normality and outliers. To verify the existence of outliers, the Mahalanobis distance was used in the data set. Mahalanobis distance is a value used to detect the presence of extreme values that make it difficult to meet linearity and normality assumptions in regression analysis (Çokluk et al., 2012; Kline, 2011). There was no outlier that had a Mahalanobis distance score greater than the critical value, $\chi^2(10) = 29.558$, $p < .05$. In order to determine whether the univariate normal distribution assumption is met, the skewness and kurtosis coefficients for each item were examined. The normality results for the study data are shown in Table 3.

Table 3. *Normality results of the study data.*

Item/Factor	N	Skewness	Kurtosis
Factor 1: Conceptual	318	-1.548	-1.777
Item 1	318	-1.219	1.986
Item 2	318	-1.548	-1.778
Item 3	318	-1.496	-.571
Item 4	318	-.948	-1.501
Item 5	318	-1.793	-889
Item 6	318	-1.417	-1.369
Item 7	318	-1.239	-.413
Factor 2: Practical	318	-1.150	2.159
Item 8	318	-.696	-.485
Item 9	318	-.778	-1.552
Item 10	318	-1.359	1.023
Item 11	318	-1.589	-.548
Item 12	318	-1.423	.901
Item 13	318	-.974	-1.842
Item 14	318	-1.716	.844
Item 15	318	-1.158	-.548
Factor 3: Socio-emotional	318	-1.539	-1.457
Item 16	318	-1.896	.548
Item 17	318	-1.853	-1.471
Item 18	318	-.959	.963
Item 19	318	-1.112	-1.785
Item 20	318	-1.264	.875
Item 21	318	-1.489	-1.916
Item 22	318	-.895	1.750

Furthermore, normality was tested by using skewness and kurtosis coefficients. Results indicated that the skewness coefficients of the data ranged from -1.896 to -.696, and the kurtosis coefficients ranged from 1.916 to 2.159 from -2.0 to +2.0, suggesting that the data had a normal distribution (Tabachnick & Fidell, 2013; Trochim & Donnelly, 2006). In addition, the distribution of the data on a 45-degree line in the Q-Q graphs is another indicator that indicates a univariate normal distribution (Tabachnick & Fidell, 2013). The correlation coefficients calculated to determine the status of multicollinearity between items were below .90 coefficient specified by Kline (2011), indicating no multicollinearity problems.

2.6. Data Analysis

A confirmatory factor analysis (CFA) was conducted for the TFALS-Turkish version in the study sample using the AMOS 23 version. A three-factor model was created based on Yan and Pastore (2022a) with the scales used in the study. Model fit was assessed using Chi-square (χ^2), Tucker-Lewis Index (TLI), Incremental Fit Index (IFI), Comparative Fit Index (CFI), and Root-Mean-Square Error of Approximation (RMSEA (Brown, 2006; Kline, 2011). In order for the model constructed in CFA to have an acceptable fit, it should have values of $2 \le \chi^2/df \le 3$, $.05 \le RMSEA \le .08$, $.90 \le CFI \le .95$, $.90 \le TLI \le .95$ and $.90 \le IFI \le 95$ (Kline, 2011). The reliability of the scale was tested using Cronbach's alpha and composite reliability (CR), where a value

of α above 0.7 is considered acceptable (Büyüköztürk, 2014). Furthermore, the discriminatory power of the items was examined by calculating adjusted item-total correlations and comparing the upper 27% of the participants with the lower 27% (Erkuş, 2014; Tekindal, 2015). Therefore, item analyses were performed to estimate item-total correlation values. SPSS 22.0 program was used for composite reliability, Cronbach's alpha, construct validity and item analysis of TFALS.

3. FINDINGS

3.1. Factor structure of the TFALS

Researchers have suggested that when adapting an instrument developed in a culture, whose factor structure has been determined, to another culture, the adaptation should be done using CFA. It is also suggested that the convergent validity of the scale should be tested with CFA (Çokluk et al., 212; Seçer, 2015). Construct validity is a crucial aspect of data analysis that examines the extent to which a measure purports to measure or whether the factorial structure of the measurement tool is valid (Byrne, 2001). To examine the construct, the three-factor, 22-item structure of the TFALS-Turkish version, a CFA was performed by using the maximum likelihood estimation. The initial model included three latent variables with 25 items. Results of CFA showed that the three-factor model was not an acceptable level for close model fit χ^2 value (2.771, $p < .05$), but other fit indices (TLI = .88, IFI = .89, CFI = .89, RMSEA = .078) were within the acceptable range (Kline, 2011).

In order to improve the model, modification indices were tested among the items on the scale thought to contribute significantly to model fit. According to the experts, when conducting modifications, attention should be paid to the theoretical rationale and the items with associated error terms under the same factor (Çokluk et al., 2016; Karagöz, 2016). In this respect, modifications were conducted between the error terms (item1 & item2) of two items in the conceptual subscale of the scale, between which a latent relationship could be accepted and which was thought to contribute significantly to the model fit. Both modifications were made on items of the same size and assumed to measure similar phenomena. As a result of the renewed CFA after the modification suggestions that were provided to contribute to the model were processed, it was found that there were significant improvements in the goodness of fit indices (see Table 4) of the three-factor model of the Turkish version of the TFALS that met the adequate data fit values and were confirmed (χ^2/df = 2.379; IFI= .923; TLI= .912, CFI= .922; RMSEA= .066; SRMR= .036). The standardized factor loadings for the constructs of conceptual dimension were between .61 and .77, for practical dimension between .75 and .76 and for socio-emotional dimension between .77, and .82, which were similar to TFALS–English version (.78, .79, and .88, respectively).

Table 4. *The goodness-of fit indices and CFA results.*

Fit Indices	Perfect Fit	Acceptable Fit	Model fit indices (Yan & Pastore (2022a)	Model fit indices (Turkish version)
χ^2/df	$0 \leq \chi^2/df \leq 2$	$2 \leq \chi^2/df \leq 3$	2.370	2.379
RMSEA	$0 \leq RMSEA \leq .05$	$.05 \leq RMSEA \leq .08$	.069	.066
CFI	$.95 \leq CFI \leq 1$	$.90 \leq CFI \leq .95$	.92	.92
TLI	$.95 \leq TLI \leq 1$	$.90 \leq TLI \leq .95$	.91	.91
IFI	$.95 \leq IFI \leq 1$	$.90 \leq IFI \leq .95$	--	.92

3.2. Correlations between dimensions of the TFALS

Pearson correlation analysis was used to determine the relationship between the dimensions of the TFALS. The results of the descriptive and correlational analyses are presented in Table 5.

Table 5. *Correlations between dimensions of TFALS-Turkish version.*

	M	*SD*	Factor Correlation		
			1	2	3
Conceptual	4.96	.795	1.00		
Practical	5.02	.778	.780**	1.00	
Socio-Emotional	5.11	.819	.762**	.840**	1.00

**p< .01

As seen in Table 5, all correlations between variables were positively significant (*p*<.01). The correlations between the three dimensions of the TFALS ranged from .762 to .840. Büyüköztürk (2014) stated that the correlation between .00-.30 is low, between .30-.70 is moderate, and between .70-1.0 is high. Based on these parameters, it can be said that there were high correlations between the dimensions of TFALS. Furthermore, according to the correlation matrix for the items TFALS, all 22 items showed positive, significant, and above .30 correlations with the total score of the scale, and the level of correlation varies between medium and high.

3.3. Reliability Analysis

Cronbach's alpha (α) and composite reliability (CR) coefficients for each dimension of TFALS were used to assess the reliability of the TFALS-Turkish version. The results are shown in Table 6.

Table 6. *Reliability coefficients of the scale for the 22-item TFALS-Turkish version.*

	Number of items	Cronbach's α	Composite Reliability
Conceptual	7	.90	.84
Practical	8	.86	.81
Socio-emotional	7	.88	.78
Total	22	.93	.85

The Cronbach's alpha coefficients found for each of the dimensions of the TFALS were .90 for the conceptual (7 items), .86 for the practical (8 items), .88 for the socio-emotional (7 items), and .93 for the total scale. The CR coefficients of conceptual, practical, and socio-emotional dimensions were .84, .81, and .78, respectively, and .85 for the total scale. Fraenkel et al. (2012) and Hair et al. (2010) noted that measurements of reliability coefficients of .70 and above are generally agreed upon as an acceptable value. Additionally, George and Mallery (2003) suggested a tiered approach for the range of Cronbach's alpha consisting of the following: ≥0.9-excellent, ≥0.8-good, ≥0.7-acceptable, ≥0.6. All reliability coefficients are above .72 provides additional evidence for the reliability of the TFALS-Turkish version. Figure 2 shows the DFA results for the Turkish version of TFALS.

Figure 2. *The confirmatory factor analysis model tested with the TFALS-Turkish version.*
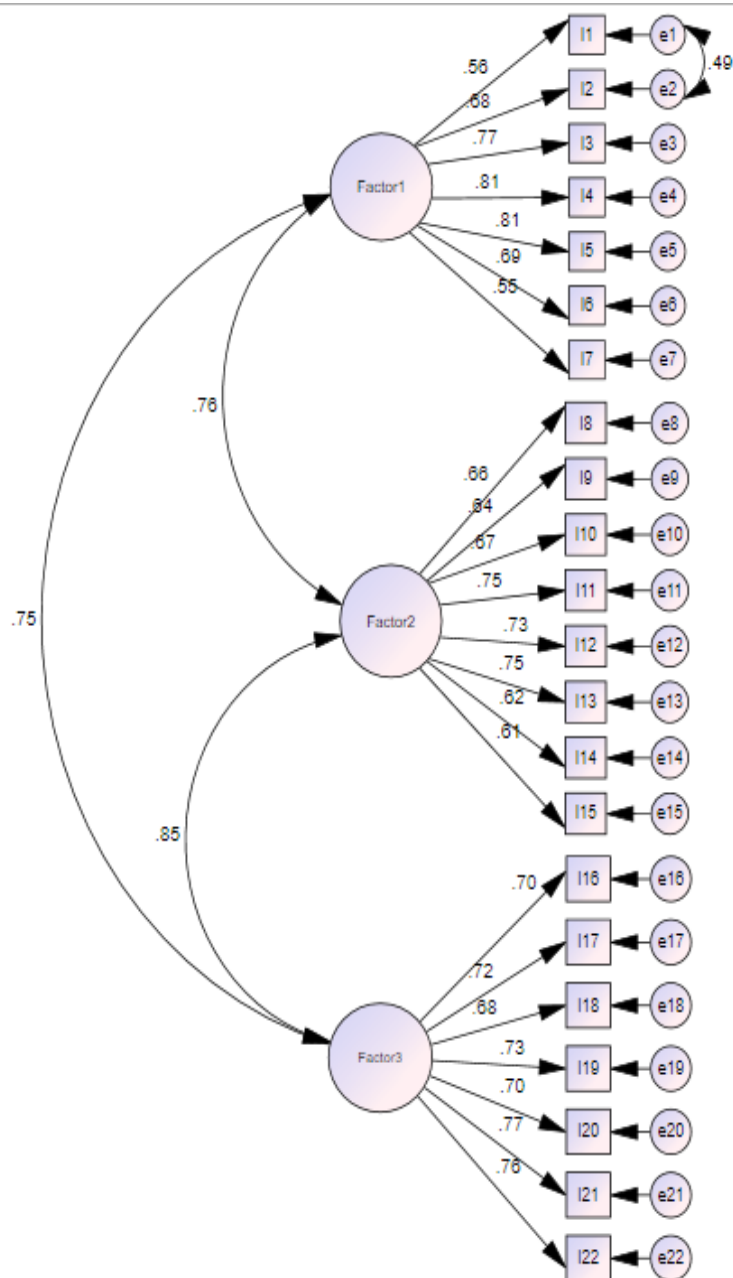
Note. TFALS= Teacher Formative Assessment Literacy Scale (TFALS) Factor 1: Conceptual, Factor 2: Practical, Factor 3: Socio-Emotional

3.4. Item Analysis

In this study, the discriminatory power of the items was examined by comparing the participants' lower 27% and upper 27% and calculating adjusted item-total correlations. The lower and upper groups for each subscale of the TFALS must be determined by the scales formed by the multidimensional structure (Büyüköztürk, 2014). A t-test was used to determine whether there is a significant difference between the upper 27% and lower 27% of the group in terms of the subscale of the TFALS-Turkish version. Table 7 displays the item-total correlations and the lower-upper 27% t-test results.

Table 7. *Results of analysis for the 22-item TFALS.*

Items	Corrected Item Total Correlations (r)	M	SD	Lower -Upper 27% t-Test
Factor 1: Conceptual				
1	.78	4.62	1.054	15.48[*]
2	.72	4.81	.978	13.95[*]
3	.65	5.18	.962	11.78[*]
4	.52	5.13	.993	118.41[*]
5	.49	5.11	.979	17.74[*]
6	.79	5.02	.981	10.12[*]
7	.68	4.89	.908	13.89[*]
Factor 2: Practical				
8	.67	4.82	.977	14.75[*]
9	.77	4.85	1.021	5.69[*]
10	.58	5.03	.972	17.26[*]
11	.70	5.04	.977	18.47[*]
12	.78	5.18	.969	21.34[*]
13	.65	4.99	.884	19.47[*]
14	.74	5.09	.961	15.03[*]
15	.70	5.16	.978	17.85[*]
Factor 3: Socio-emotional				
16	.63	5.04	.933	20.24[*]
17	.57	5.04	1.068	14.96[*]
18	.75	4.96	1.037	22.23[*]
19	.71	5.13	.937	16.45[*]
20	.78	5.24	1.007	10.99[*]
21	.74	5.19	.933	20.45[*]
22	.62	5.16	.906	17.36[*]

[*]$p < .01$

Taking 27% as a cutoff value (lower and upper groups), the results showed t values for the difference between the upper 27% and lower 27% of the participants ranged from 10.54 to 14.55 for the conceptual dimension, between 14.85 and 14.25 for the practical dimension, and between 13.25 and 12.45 for the socio-emotional dimension. T-test values were significant for all items according to the comparison result between the participants' lower 27% and the upper 27%. Significant t-values in comparisons between the lower and upper groups of the participants were accepted as evidence of the items' discriminatory power (Erkuş, 2014). Table 7 also presents that item-total correlations ranged from .47 to .57 for the conceptual dimension, .54 to .62 practical dimension, and .37 to .50 socio-emotional dimension. Items with .30 and above coefficients are considered to have sufficient discriminatory power when interpreting item-total correlations (Büyüköztürk, 2014; Erkuş, 2014). The results indicate that all items met this requirement. Therefore, all items in the TFALS had discriminatory power, according to these findings.

3.5. Scoring of the Scale

The TFALS-Turkish version consists of 22 items total with three dimensions. It included 7 items in the conceptual dimension, 8 items in the practical dimension, and 7 items in the socio-emotional dimension. There are no reverse items among the items that constitute the scale. The

scores that can be obtained from the scale range from 22 to 132 points. Higher scores indicate higher levels of teachers' formative assessment literacy.

4. DISCUSSION and CONCLUSION

Assessment is an important element of the teaching process in terms of the quality of education and training services. Accurate and effective feedback can be given to students during the assessment and evaluation process, and more effective guidance can be provided. By doing this, variables that result from teaching methods, learning environments, course contents, and other factors that negatively affect learning can be identified and corrected (Yılmaz, 1998). In this context, the effectiveness of assessment activities today is related to whether they aim to improve the learning of students rather than reach a conclusion about the success or failure of students by using letter or numerical grades (National Council of Teachers of Mathematics [NCTM], 2023). At this point, the issue of formative assessment, which aims to support the learning of the students, gains importance in increasing the quality of educational services. Although early studies aimed to determine the features of formative assessments compared to other assessment activities, recent research has focused on how frequently teachers utilize formative assessment and its impact on students' learning outcomes (Furtak et al., 2016; Pinger et al., 2018; Yan & Pastore, 2022a). Furthermore, there is a requirement for research into scale development aimed at evaluating teachers' formative assessment literacy. Therefore, Yan and Pastore (2022a) devised a self-reported scale for evaluating teachers' formative assessment literacy in response to this demand. The purpose of the study was to translate and evaluate the psychometric properties of the TFALS developed by Yan and Pastore (2022a). When an instrument is translated from one language or dialect into another, reliability and validity studies should be conducted for the intended use across linguistic groups (Geisinger, 1994). Accordingly, in this study, initially, a language equivalence study was carried out while adapting the original scale into Turkish. Experts in educational sciences and the English language translated the scale into Turkish.

The resulting Turkish form was translated back into English using the back-translation method, and expert advice was provided. Adaptation studies have shown that the items in the English and Turkish versions of the TFALS are highly correlated with each other. A common source of validation information on scale structure is the factor analysis technique applied to test data. Factor analysis techniques are the most frequently used procedures for evaluating tests adapted to linguistically diverse populations (Geisinger, 1994). In this study, after the linguistic equivalence study of the scale, CFA was used to evaluate the three-factor model of the TFALS-Turkish version with 22 items for a Turkish teacher sample. The CFA results indicated that fit indices showed that the model was not within acceptable ranges. In order to improve the model, modification indices were examined among the items that were thought to contribute significantly to the model fit, and modifications were made between the error terms of two items (item 1 and item 2) in the conceptual sub-dimension of the scale. The results of the CFA repeated after the modifications showed that the model fit indices were within the acceptable limits of goodness of fit (Kline, 2011). The CFA results support the three-factor structure of the TFALS-Turkish version. The factor loadings of the items in the scale ranged from .61 to .82. According to the experts, factor loadings of .45 or more are a good measure of selection. (Büyüköztürk, 2014; Tabachnick & Fidell, 2013). Accordingly, it can be said that the three-factor structure and item factor loadings of the TFALS-Turkish version were found to be sufficient. Therefore, the results suggest that the Turkish version of the scale is structurally reliable and can be used in a Turkish sample.

The reliability of the Turkish-adapted scale was determined by calculating Cronbach's alpha and the composite reliability coefficient. For the Turkish version of the scale, the Cronbach's alpha for each of the dimensions of TFALS were .90 for the conceptual dimension (7 items), .86 for the practical dimension (8 items), and .88 for the socio-emotional dimension (7 items). In

addition, the composite reliability values of the dimensions in the adapted scale were found to be greater than .70. In the original form of the scale, Cronbach's alpha was .88 for the conceptual dimension, .88 for the practical dimension, .89 for the socio-emotional dimension, and .86 for the whole scale (Yan & Pastore, 2022a). The internal consistency coefficient obtained is similar to the results of previous research on the reliability of TFALS (Yan & Pastore, 2022a; Yan & King, 2023; Yan et al., 2022). At this point, it is seen that the reliability coefficients of the original scale and the Turkish-adapted scale are close to each other. According to Geisinger (1994), there may be differences in the understanding of the scale due to cultural differences in scale adaptation studies. Therefore, there may be differences in reliability values. Previous research has recommended that for a measure to be reliable in scale development or adaptation studies, the internal consistency and reliability coefficients should exceed .70 (Büyüköztürk, 2014; Fraenkel & Wallend, 2006; Robinson et al., 1999; Tezbaşaran, 1997).

The findings suggest that the internal consistency and composite reliability coefficients of the Turkish version of the scale are satisfactory. Currently, based on the evidence of validity and reliability, it is appropriate to claim that the Turkish-adapted version of the scale can be employed to assess teachers' formative assessment literacy. To evaluate whether the items on the scale measure the targeted attributes, item-total score correlations and mean scores collected from the upper and lower 27% groups were computed and compared according to Büyüköztürk (2014). On the scale, the differences observed in the item averages of the lower and upper 27% groups were found to be significant at the $p<.001$ level. These results provided evidence for the discrimination of the items in the scale (Erkuş, 2014). Item-total score correlations were found to vary, ranging from .55 to .75. The correlation values between the dimensions of the scale ranged from .32 to .60 and there were significant positive relationships ($p<.05$) between the dimensions of the scale (Erkuş, 2014). On the scale, the differences observed in the item averages of the lower and upper 27% groups were found to be significant at the $p<.001$ level. This result is evidence for the discrimination of the items in the scale (Erkuş, 2014).

4.1. Recommendations and Contributions to Education

All results showed that the multidimensional structure of the original TFALS had appropriate psychometric properties. The validity and reliability analyses of this scale adapted into Turkish were carried out with a systematic approach. However, there is a limitation in terms of research results that must be taken into consideration for future research. This limitation is that the participants are teachers who have a non-thesis master's degree in the field of education. In terms of the generalizability of the results of the study, validity and reliability analyses can be conducted on the data to be collected from a sample group consisting of teachers who have not undertaken postgraduate education in Türkiye. Although there is such a limitation in the study, the findings provided evidence that the TFALS adapted into Turkish has good reliability and validity. Consequently, this study translated the TFALS into Turkish and tested its reliability and validity among Turkish teachers. It can be said that the adapted scale is suitable for measuring the literacy levels of teachers from different teaching fields in Türkiye for the practice of formative assessment. Examining teachers' assessment literacy is one of the most important issues in education and teaching practice today. There is convincing evidence in the literature that assessment, especially formative assessment, is an important leverage point for improving students' academic outcomes (Andrade & Heritage, 2018; De Simone, 2009). In this direction, the adaptation of TFALS into Turkish may be informative in terms of teacher education and teacher practice. In addition, as far as we know, this study presents the first attempt to translate the TFALS into Turkish and test its reliability and validity among a group of teachers working in Türkiye. In other words, to date, there is no inductively developed and psychometrically tested instrument that measures the multidimensional construct of teachers' formative assessment literacy in Türkiye. The Turkish version of the scale is presented in the Appendix.

The scale adapted into Turkish can actually provide researchers, teacher educators, school administrators, and policymakers with important ideas about teachers' profiles (strengths and weaknesses) in formative assessment. In this sense, it may be possible to support teachers in developing their formative assessment literacies through the data to be collected through the scale. The adapted scale can also be used to evaluate the effectiveness of teacher education programs or interventions to increase teachers' formative assessment literacy (Yan & Pastore, 2022a). Furthermore, the three-dimensional formative assessment literacy model underpinning TFALS serves critical dimensions for the effectiveness of teaching practice and the improvement of student development. In future studies, teachers' formative assessment literacy can be measured using TFALS. Again, by using TFALS, researchers can get an idea about teachers' competencies related to formative assessment in teaching processes and how they use formative assessment. This is because the scale provides a multidimensional perspective and enables the complex structure of formative assessment literacy to be seen. In this context, the TFALS, adapted into Turkish, is expected to fill an important gap in the existing literature. Both the original form of the scale and the adaptation study, which were applied to teachers, can give researchers an idea about the efficiency and excellence of teaching services. The use of this scale in different studies and with different samples will also help to better understand its psychometric properties.

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**: Bartın University, E-23688910-050.01.04-2300023164

Contribution of Authors

Kemal İzci: Literature review**,** Investigation, Data collection, Resources, and Writing-original draft. **İlhan İlter**: Methodology, Supervision, Statistical analysis, Validation, and Writing-original draft. **Gökhan Izgar**: Data collection, Data Interpretation, Resources, Writing-original draf.

Orcid

Kemal İzci https://orcid.org/0000-0002-4228-8845
İlhan İlter https://orcid.org/0000-0002-1473-7172
Gökhan Izgar https://orcid.org/0000-0002-6835-9701

REFERENCES

Andrade, H.L., & Heritage, M. (2018). *Using formative assessment to enhance learning. achievement and academic self-regulation*. Routledge.

Aras, S. (2019). Improving early childhood teachers' formative assessment practices: Transformative role of collaborative action research. *International Journal of Curriculum and Instructional Studies, 9*(2), 221-240.

Archer, E. (2017). The assessment purpose triangle: Balancing the purposes of educational assessment. *Frontiers in Education, 2,* 1-17.

Ayan, A.D., & Erdemir, N. (2023). EFL teachers' perceptions of automated written corrective feedback and Grammarly. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi (AKEF) Dergisi, 5*(3), 1183-1198.

Bayrak, N., Çalık, M., & Doğan, S. (2019). Biology teachers' practice of formative assessment: A case of the identifying learning gap element. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 46*, 132-149.

Bennett, R.E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice, 18*(1), 5-25.

Biggs, J., & Tang, C. (2011). *Teaching for quality learning at university.* Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5(*1), 7–74.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability, 21*(1), 5–31.

Borsa, J.C., Damásio, B.F., & Bandeira, D.R. (2012). Cross-cultural adaptation and validation of psychological instruments: Some considerations. *Paidéia, 22*(53), 423–432.

Brookhart, S.M., & McMillan, J.H. (2020). *Classroom assessment and educational assessment. (p. 1-9).* Routledge.

Buldur, S., & Hasbek, G. (2020). Pre-service teachers' perceptions about formative assessment: A metaphorical study. *Inonu University Journal of the Faculty of Education, 21*(2), 595-615.

Büyüköztürk, Ş. (2014). *Sosyal bilimler için veri analizi el kitabı [Data analysis notebook for social sciences].* Pegem Akademik Yayın.

Byrne, B.M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing, 1*(1), 55–86.

Brookhart, S.M. (2018). Summative and formative feedback. In A. Lipnevich & J. Smith (Eds.), *The Cambridge handbook of instructional feedback (pp. 52–78).* Cambridge University Press. https://doi.org/10.1017/9781316832134.005

Cañadas, L. (2023). Contribution of formative assessment for developing teaching competences in teacher education. *European Journal of Teacher Education, 46*(3), 516-532.

Cañadas, L., Santos-Pastor, M.L., & Ruíz, P. (2021). Percepción del impacto de la evaluación formativa en las competencias profesionales durante la formación inicial del profesorado [Perception of impact of formative assessment on professional competences during initial teacher training]. *Revista Electrónica De Investigación Educativa 23*: 1–14. e03.

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 Akademi Yay.

Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199.

Eremenco, S.L., Cella, D., & Arnold, B.J. (2005). A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Evaluation & the Health Professions, 28*(2), 212–232.

Erkuş, A. (2014), *Psikolojide ölçme ve ölçek geliştirme-I: Temel kavramlar ve işlemler [Measurement and scale development in psychology-I: Basic concepts and processes*]. Pegem Akademi Yay.

Fraenkel, J.R., Wallen, N.E., & Hyun, H.H. (2012). *How to design and evaluate research in education* (Vol. 7, p. 429). McGraw-hill.

Fraenkel, J.R., & Wallen, N. (2006). *How to design and evaluate research in education.* McGraw-Hill.

Furtak, E.M., Kiemer, K., Circi, R.K., Swanson, R., de León, V., Morrison, D., & Heredia, S. C. (2016). Teachers' formative assessment abilities and their relationship to student learning: Findings from a four-year intervention study. *Instructional Science, 44*, 267-291.

Geisinger, K.F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment, 6*(4), 304-312. https://doi.org/10.1037/1040-3590.6.4.304

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Allyn & Bacon

Gotch, C.M., & French, B.F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice, 33*(2), 14–18.

Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2010). *Multivariate data analysis* (7th Ed). Pearson.

Heritage, M. (2010). *Formative assessment: Making it happen in the classroom.* Corwin Press.

Izci, K. (2016). Internal and external factors affecting teachers' adoption of formative assessment to support learning. *International Journal of Educational and Pedagogical Sciences, 10*(8), 2800–2807.

Karagöz, Y. (2016). *SPSS 23 ve AMOS 23 uygulamalı istatistiksel analizler* [SPSS 23 and AMOS 23 applied statistical analysis]. Nobel Akademik Yay.

Karaman, O. (2017). Adapting of formative assessment attitude and intention scale for pre-service teachers and a structural equation modelling. *Inonu University Journal of the Faculty of Education, 18*(3), 181-131.

Karaman, P., & Karaman, A. (2017). Investigating teacher candidates' use of formative assessment: Action research model. *Kastamonu Eğitim Dergisi, 25*(6), 2385-2400.

Kaya, G., Atasoy, V., Candan-Helvacı, S., & Pektaş, M. (2021). The role of science teachers' awareness in their classroom practice of formative assessment. *Education and Science, 46*(205), 335-357.

Kepek, T., & Izci, K. (2021). Examining the effect of alternative assessment methods on academic achievement and student opinions about these methods. *Necmettin Erbakan Üniversitesi Ereğli Eğitim Fakültesi Dergisi, 3*(1), 34-50.

Klenowski, V. (2009). Assessment for learning revisited: An Asian-Pacific perspective. *Assessment in Education: Principles, Policy and Practice, 16*(3), 263-268.

Kline, R.B. (2011). *Principles and practice of structural equation modeling.* The Guilford Press.

Li, H. (2016). How is formative assessment related to students' reading achievement? Findings from PISA 2009. *Assessment in Education: Principles, Policy & Practice, 23*(4), 473–494. https://doi.org/10.1080/0969594X.2016.1139543

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.

MoNE (2013). İlköğretim fen ve teknoloji dersi öğretim programi [Primary school science and technology curricula). Milli Eğitim Bakanlığı. Ankara.

MoNE (2017). Öğretmenlik mesleği genel yeterlikleri. [General competencies of the teaching profession] Öğretmen Yetiştirme ve Geliştirme Genel Müdürlüğü. Ankara. https://oygm.meb.gov.tr/dosyalar/StPrg/Ogretmenlik_Meslegi_Genel Yeterlikleri.pdf

Menéndez, I.Y.C., Napa, M.A.C., Moreira, M.L.M., & Zambrano, G.G.V. (2019). The importance of formative assessment in the learning teaching process. *International Journal of Social Sciences and Humanities, 3*(2), 238-249.

NCTM (2023). Using classroom assessment to improve student learning: Math problems aligned with nctm and common core state standards. https://www.nctm.org/store/Product s/Using-Classroom-Assessment-to-Improve-Student-learning--Math-Problems-Aligned-with-NCTM-and-Common-Core-State-Standards/

Pastore, S., & Andrade, H.L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education, 84*, 128-138

Pinger, P., Rakoczy, K., Besser, M., & Klieme, E. (2018). Implementation of formative assessment–effects of quality of programme delivery on students' mathematics achievement and interest. *Assessment in Education: Principles, Policy & Practice, 25*(2), 160-182.

Robinson, J.P., Shaver, P.R., & Wrightsman, L.S. (1999). *Measures of political attitudes.* Academic Press.

Seçer İ. (2015). *Psikolojik test geliştirme ve uyarlama süreci: SPSS ve LISREL uygulamaları [Psychological test development and adaptation process: SPSS and LISREL applications].* Anı Yayıncılık.

Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education, 26*(3), 482-496.

Schildkamp, K., van der Kleij, F.M., Heitink, M.C., Kippers, W.B., & Veldkamp, B.P. (2020). Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *International Journal of Educational Research, 103*, 101602. https://doi.org/10.1016/j.ijer.2020.101602

Shavelson, R.J. (2013) On an approach to testing and modeling competence. *Educational Psychologist, 48*(2), 73-86. https://doi.org/10.1080/00461520.2013.779483

Shepard, L.A., Penuel, W.R., & Davidson, K.L. (2017). Design principles for new systems of assessment. *Phi Delta Kappan, 98(6), 47-52.*

Şahin, Ç., & Kaya, G. (2020). Alternatif ölçme değerlendirme ile ilgili yapılan araştırmaların incelenmesi: Bir içerik analizi [A review of the research on alternative assessment evaluation: A content analysis]. *Nevşehir Hacı Bektaş Veli University Journal of ISS, 10*(2), 798-812.

Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics.* (6th Ed). Pearson.

Tekindal, S. (2015). *Duyuşsal özelliklerin ölçülmesi için araç oluşturma [Creating a tool for measuring affective traits].* Pegem Akademi Yayın.

Tezbaşaran, A. (1997). *Likert tipi ölçek geliştirme kılavuzu.* Türk Psikologlar Derneği Yayın.

Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformative and transformative assessment. *Oxford Review of Education, 38*(3), 323-342.

Trochim, W.M., & Donnelly, J.P. (2006). *The research methods knowledge base* (3rd ed). Atomic Dog.

Van der Kleij, F.M., Cumming, J.J., & Looney, A. (2018). Policy expectations and support for teacher formative assessment in Australian education reform. *Assessment in Education: Principles, Policy & Practice, 25*, 620–637.

World Health Organization (WHO) (2017). Process of translation and adaptation of instruments. Retrieved from http://www.who.int/substance_abuse/research_tools/translation/en/

Wylie, E.C. (2020). Observing formative assessment practice: Learning lessons through alidation. *Educational Assessment, 25*(4), 251–258.

Xu, Y., & Brown, G.T. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education, 58*(1), 149-162.

Yan, Z., Chiu, M.M., & Cheng, E.C.K. (2022). Predicting teachers' formative assessment practices: Teacher personal and contextual factors. *Teaching and Teacher Education, 114,* 103718. https://doi.org/10.1016/j.tate.2022.103718

Yan, Z., & King, R.B. (2023). Assessment is contagious: The social contagion of formative assessment practices and self-efficacy among teachers. *Assessment in Education: Principles, Policy & Practice, 30*(2), 130–150.

Yan, Z., & Pastore, S. (2022a). Are teachers literate in formative assessment? The development and validation of the Teacher Formative Assessment Literacy Scale. *Studies in Educational Evaluation, 74*, 101183. https://doi.org/10.1016/j.stueduc.2022.101183.

Yan, Z., & Pastore, S. (2022b). Assessing teachers' strategies in formative assessment: The teacher formative assessment practice scale. *Journal of Psychoeducational Assessment, 40*(5), 592-604. https://doi.org/10.1177/0734282922107.

Yaşar, M.D. (2017). Prospective science teachers' perception related to formative assessment approaches in Turkey. *Journal of Education and Training Studies, 5*(4), 29-43.

Yılmaz, H. (1998). *Eğitimde ölçme ve değerlendirme [Assessment in education].* Mikro Yayın.

APPENDIX: **The Teacher Formative Assessment Literacy Scale-Turkish Version**

Öğretmen İzlemeye Dayalı Değerlendirme (Formative Assessment) Okuryazarlığı Ölçeği Türkçe Versiyonu

Aşağıdaki ifadelere ne derece katılıp/katılmadığınızı belirtiniz.	Kesinlikle katılmıyorum	Katılmıyorum	Kısmen katılmıyorum	Kısmen katılıyorum	Katılıyorum	Kesinlikle katılıyorum
1. İzlemeye dayalı değerlendirmenin mantığını/ gerekçelerini açıklayabilirim.						
2. Öğrencilerin öğrenme ihtiyaçlarının izlemeye dayalı değerlendirme yoluyla belirlenebileceğini biliyorum.						
3. Değerlendirme etkinliklerinin öğrenme hedefleriyle örtüşmesi gerektiğini düşünüyorum.						
4. İzlemeye dayalı değerlendirme etkinliklerinin öğrencilerin öğrenme düzeylerine yönelik geri bildirim sunması gerektiğini düşünüyorum.						
5. Öğrencilerin öğrenme ihtiyaçlarını belirlemek adına izlemeye dayalı değerlendirme sonuçlarının öğretmenler için yararlı olduğunu biliyorum.						
6. Öğrenmenin arttırılması için öğrencilerin izlemeye dayalı değerlendirmeye katılmaları gerektiğini düşünüyorum.						
7. Öğrencilerin öğrendiklerini göstermelerine olanak sağlayacak çeşitli değerlendirme yöntemlerini biliyorum.						
8. Öğrencilerin öğrendiklerini göstermelerine olanak tanıyan çeşitli değerlendirme yöntemlerini kullanırım.						
9. Öğrencilere akran değerlendirme süreçlerine katılmayı öğretiyorum.						
10. Öğrencilerin öz değerlendirme becerilerini geliştirmelerine yardımcı olurum.						
11. Öğrencilere sunmuş olduğum geri bildirim bilgilerini kullanmalarını sağlarım.						
12. Değerlendirme sonuçlarına göre, öğrencilere o konuyu nasıl daha iyi öğrenebileceklerini gösteririm.						
13. Öğrencilere, öğrenmelerini geliştirmek amacıyla değerlendirme sonuçlarını kullanmalarını öğretirim.						
14. Kafa karışıklığı oluştuğunda, öğrenciler için değerlendirme amaçlarını netleştiririm.						
15. Değerlendirme kriterlerini /ölçütlerini öğrencilerle paylaşırım.						
16. Öğretmenler ve öğrenciler arasında izlemeye dayalı değerlendirmeye ilişkin ortak bir anlayış oluşturma ihtiyacının farkındayım.						
17. Öğrencilerin değerlendirmelere verdikleri duygusal tepkileri dikkate alırım.						
18. Öğrencilerin değerlerinin, inançlarının ve tutumlarının izlemeye dayalı değerlendirmeyi kullanmalarını etkilediğinin farkındayım.						
19. Değerlendirme sürecindeki geri bildirim bilgilerinin öğrencilerin öğrenme motivasyonunu nasıl etkilediğinin farkındayım.						
20. İzlemeye dayalı değerlendirmenin adalet ve öğrenci mahremiyeti gibi etik yönlerine duyarlıyım.						
21. İzlemeye dayalı değerlendirme sürecinde öğrencilerin iyi oluşlarını sağlamak için sorumluluklarımın farkındayım.						
22. Öğrencilerin izlemeye dayalı değerlendirme uygulamalarından yararlanma hakkına sahip olduğunun bilincindeyim.						

Research Article

Developing a game-based test to assess middle school sixth-grade students' algorithmic thinking skills

Emre Zengin[ID][1*], **Yasemin Karal**[ID][2]

[1]80th Year Republic Middle School, Gümüşhane, Türkiye
[2]Trabzon University, Faculty of Education, Department of Computer and Instructional Technologies Education, Trabzon, Türkiye

Abstract: This study was carried out to develop a test to assess algorithmic thinking skills. To this end, the twelve steps suggested by Downing (2006) were adopted. Throughout the test development, 24 middle school sixth-grade students and eight experts in different areas took part as needed in the tasks on the project. The test was given to 252 students attending the sixth grade who were selected through purposeful sampling. The content validity of the test was ensured by means of obtaining expert opinion, whereas the construct validity was ensured by performing an independent sample t-test on the difference between the lower and upper groups. As a result, the algorithmic thinking skills assessment test was finalized with 22 main items and 2 sample items, totalling 24 items. The KR-20 reliability analysis proved a quite reliable test based on the reliability coefficient of 0.83. As mentioned earlier, the independent sample t-test was applied to the difference of lower and upper groups for construct validation of the test. It was seen that the test items are significant in discriminating the students in the lower and upper groups ($p<0.01$).

1. INTRODUCTION

Technology is developing and affecting social life. These effects are evident in a variety of spheres including education and health and social, commercial, and professional areas. This necessitates individuals to acquire different skills and maintain their individual development. With the acquisition of these skills, which are known as 21st-century skills, individuals are expected to be productive, problem-solving, critical, entrepreneurial, and able to think creatively (Geisinger, 2016; Keane, 2012; Khanlari, 2013). Under these skills lies algorithmic thinking skills, which are common skills that should be possessed by all individuals today (Futschek & Moschitz, 2010; Mumcu & Yıldız, 2018; Sarı et al., 2022). Individuals who have acquired algorithmic thinking skills can produce solutions by applying the algorithm structures in their minds to the problems they encounter in daily life (Thomas et al., 2015; Thomas et al., 2017). Bearing in mind the possibility of facing problems at any moment in life, individuals should also be prepared for these situations (Aytekin et al., 2018; Lertlapnon et al., 2022; Turchi et al., 2019). It is emphasized that algorithmic thinking skills should be taught in a planned and programmed way (Erümit et al., 2018). Several researchers have put forward steps to be followed in the process of developing algorithmic thinking skills (Zsako & Slavi, 2012, p. 55,

---

*CONTACT: Emre ZENGİN ✉ emre_zengin20@gmail.com 🖥 80th Republic Middle School, Gümüşhane, Türkiye

as cited in Szanto, 2002; Vasconcelos, 2007; Zsako & Slavi, 2012).

There are many studies on algorithmic thinking skills, and they include games (Lee et al., 2014; Paspallis et al., 2022; Sungkaew et al., 2022). According to research (Evripidou et al., 2021; Kiss & Arki, 2017; Pivec & Kearney, 2007), games have positive effects on students' attention and motivation (Apostolellis et al., 2014; Debabi & Bensebaa, 2016; Shang et al., 2019) and they boost skills such as problem-solving, flexibility, adaptability and creativity. In addition, it is stated that the actions and rules in games progress in parallel with the logic of algorithmic thinking (Karakasis & Xinogalos, 2020, Wangenheim et al., 2019; Yılmaz, 2020).

A literature review was conducted on studies dealing with the development of algorithmic thinking skills in connection with computer-based games and unplugged games. It was seen that data were collected during the students' achievement of the tasks given in games and assessment was predominantly carried out based on such data (Chuechote et al., 2020; Czakoova, 2020; Kazimoglu, 2020). Some other studies featuring computer-based games used achievement tests and questionnaires (Elshahawy et al., 2020; Hsu & Wang, 2018; Li et al., 2020; Tsukamoto et al., 2017). As for the studies using unplugged games, inferences were made from the data obtained from session recordings (Lin et al., 2020; Scharf et al., 2020) or interviews and observations (Chen & Chi, 2020). Other studies applied hybrid games, and it was seen that the assessment was done by using the data obtained through the think-aloud technique (Lee et al., 2014).

1.1. Algorithm and Algorithmic Thinking

The algorithm is a procedure of sequenced instructions fulfilled to complete a specific task (Borkulo et al., 2021). Algorithmic thinking is the ability to comprehend, execute, create, and evaluate algorithms (Brown, 2015). People attempt to solve every problem they encounter in their daily lives (Doleck et al., 2017; Kanaki & Kalogiannakis, 2022; Yadav et al., 2017). In the problem-solving process, it is regarded as important to reveal the most efficient solution by considering all reasonable possibilities (Hu, 2011; Jancec & Vujicic, 2021; Katai, 2015). Therefore, it is recommended that individuals be trained as good algorithmic thinkers (Czakoova & Udvaros, 2021; Figueiredo et al., 2021; Mezak & Papak, 2018).

1.2. Developing Algorithmic Thinking Skills

Researchers examining the development process of algorithmic thinking skills have suggested some models as seen in Table 1.

Table 1. *Models of algorithmic thinking skill development.*

Zsako and Slavi (2012)	Vascencolos (2007)	Szanto (2002)	Erümit et al. (2018)
1. Recognizing and Understanding Algorithms 2. Implementing Algorithms 3. Analyzing Algorithms 4. Making Algorithms 5. Realizing Algorithms 6. Modifying and Changing Algorithms 7. Designing Complex Algorithms	1. Read and Comprehend the Problem Statement 2. Select Theoretical Concepts That May Be Applied 3. Qualitative Description of the Problem 4. Formulization of a Solution Strategy 5. Test and Description of the Solution	1. Application/Coding 2. Algorithm Writing 3. Analogic Thinking 4. Being Able to Change Algorithm and Adapt to Current Situation 5. Production/ Derivation	1. Understand the Problem 2. Devise A Plan 3. Compare the Strategies 4. Devise and Algorithm 5. Code the Algorithm 6. Identify and Correct the Error in A Different Code 7. Prepare and Code New Algorithms

1.3. Games and Game-Based Learning

According to Huizinga (1955), a game is a sequence of activities that flow according to a set of pre-determined rules within a certain time and place. Yılmaz (2020) defines games as a type of behavior that children perform in order to adapt to the real world. Game-based learning is defined as a learning environment that enables students to achieve their learning goals and solve problems that they may encounter in daily life by providing a sense of achievement through gaming activities (Kim et al., 2009; McFarlane et al., 2002; Prensky, 2001).

1.4. Assessing Algorithmic Thinking

A review was conducted on the previous studies looking into the effect of game-based activities on algorithmic thinking skills. As a result, several data collection tools were found. Assessment was performed by using the data obtained from tests, questionnaires, interviews, observations, and session recordings. There were seven studies that assessed students' algorithmic thinking skills by using tests. These studies are listed in Table 2 below with details including the research title, year, author, level, and scope of the assessment tool.

Table 2. *Studies using multiple-choice tests for assessing algorithmic thinking skills.*

Author	Research Title	Year	Level	Test Scope
Gürbüz et al.	*"What's the weather like today?": A computer game to develop algorithmic thinking and problem-solving skills of primary school pupils*	2017	8-10 year-olds	In a computer-based educational game, students are provided with known variables of sun, temperature, humidity, and wind. Students try to reach the correct answer among 144 possible answers by developing and executing algorithms by using the given values.
Tsukamoto et al.	*Evaluating Algorithmic Thinking Ability of Primary Schoolchildren Who Learn Computer Programming*	2017	3rd through 6th graders	There is a test which is comprised of 3 multiple-choice items on which students perform sequential tasks. Students use the basic structures and elements of the algorithm while completing this task.
Hsu and Wang	*Applying game mechanics and student-generated questions to an online puzzle-based game learning system to promote algorithmic thinking skills*	2018	4th graders	There are 3 test items in which students give commands containing the basic algorithm structures and elements such as conditions, loops, variables, etc. so that they can make sure the aircraft object arrives at the right position.
Elshahawy et al.	*Codaroutine: A serious game for introducing sequential programming concepts to children with autism*	2020	14, 12, 8 7 year-olds	In a three-phased game related to daily life tasks, students are asked questions about the tasks. The students have to use the basic algorithm structures and elements while trying to find the answer.
Li et al.	*Socially shared regulation of learning in game-based collaborative learning environments promotes algorithmic thinking, learning participation and positive learning attitudes*	2020	Middle schoolers	Students attend a six-week Kodu Game Lab training. Next, they try to answer 5 multiple-choice items and 1 open-ended item regarding the software. The questions are targeted at concept understanding, algorithm creation and complex game design. Kodu Game Lab is a specific software for designing 3D games.

Oluk and Cakir	*The Effect of Code.Org Activities on Computational Thinking and Algorithm Development Skills*	2021	6th Graders	In order to examine the effect of the applications on the development of algorithm development skills of students, the algorithm development achievement test was applied as an evaluation tool. The test includes questions within the framework of learning algorithm logic, choosing the best algorithm and editing faulty algorithms.
Dag et al.	*The effect of an unplugged coding course on primary school students' improvement in their computational thinking skills*	2023	3rd Graders and 4th Graders	In order to examine the effect of computer-free coding courses on students' computational thinking skills, evaluation was carried out with multiple-choice tests. While computational thinking is evaluated by covering different dimensions, 3 items in the test are related to algorithm design.

When we look at the current literature, it is pointed out that the brain is a part of the body and it shows the importance of the body in learning. The embodied cognition theory emphasizes the inseparable link between the brain, body, and the world. Advocates of the theory claim that the brain must be understood in the context of its physical body whilst, reciprocally, the active body can alter the function of the brain. Implications of embodied cognition theory in education have become a significant part of contemporary teaching and learning practices, under the umbrella of embodied learning (Anderson, 2003; Ayala et al., 2013; Paloma, 2017; Wilson, 2002). Participating in games through physical activities is seen as an important way to ensure embodied learning (Altakrouri & Schrader 2012; Iacolina et al. 2010). It has been demonstrated that by integrating this path into the educational environment, students' motivation and desires increase, their active participation is ensured, and their cognitive and academic performances are positively affected (Kosmas et al., 2018; Kosmas & Zaphiris, 2023). Kosmas et al. (2018) evaluated the use of physically based games on 31 primary school students from the framework of embodied learning. The results showed that the games had a positive impact on students' short-term memory and emotional states. In context-based evaluation, which is one of the types of educational evaluation, evaluations are carried out by focusing on the contexts in individuals' lives (La Belle et al., 1979; Taasoobshirazi & Carr, 2008). Based on this, it is stated that in evaluations carried out by selecting the situations in individuals' lives as context, individuals' mastery and active participation in the process will increase and the validity of the evaluation results will be high (Bellochi et al., 2016; Fensham & Rennie, 2013). Traditional games which are one of the cultural elements, are played by children in a fun way in natural environments. In the course of time, children transferred many features of the traditional game into their lives (Sümbüllü & Altınışık, 2016; Yılmaz, 2020). Departing from this, the present study aimed to develop a test in line to assess the algorithmic thinking skills of middle school sixth-grade students.

2. METHOD

This is a research study conducted to develop a test to assess algorithmic thinking skills through traditional games. In line with this purpose, the research questions to be answered in the study are as follows:

1. How to scope a test to evaluate algorithmic thinking skills?
2. What is the validity and reliability of the test developed to evaluate algorithmic thinking skills?

3. How is the criterion score of the test developed to evaluate algorithmic thinking skills determined?

The development process was implemented in compliance with the steps suggested by Downing (2006). The model developed by Downing (2006) consists of 12 steps that progress gradually. While different tasks are performed at each step, there are relational tasks that affect each other between different steps. Detailed explanations about the steps are given below by headings.

2.1. Participants

Various groups of participants took part in the study until the test development was completed. For each stage of the process, a different sample selection method was used. For example, the participants involved in the validity and reliability checks were selected with purposeful sampling. This method allows for choosing the most eligible individuals or groups who share a similar experimental background concerning the project objectives (Yıldırım & Şimşek, 2006). The inclusion criterion was that 24 6th-grade students who participated in the development phase of algorithmic thinking skill test items and 252 6th-grade students who participated in the evaluation phase completed the "Problem Solving and Coding" unit of the Ministry of National Education 6th Grade Information Technologies and Software Course Curriculum. It was explained that the aim was to evaluate algorithmic thinking skills with the items or tests presented to the students at the beginning of both stages. To ensure the content validity of the test, opinions of experts in the fields of Information Technologies, Turkish Language Literature and Mathematics Education were taken. A Turkish Language and Literature expert examined the grammatical and semantic suitability of the test items. The Information Technologies and Mathematics Education expert evaluated the suitability of the test items with the curriculum outcomes. The objectives of the 6th Grade Information Technologies and Software Curriculum Unit 5 Problem Solving and Coding Unit and four of the objectives of the 6th Grade Mathematics Unit 1 Numbers and Operations overlap with each other. The numbers of all the participants and their descriptive features are given along with the particular stage of assignment they were assigned in Table 3.

Table 3. *Study participants by stages of test development.*

	Stage	No of participants	Descriptive Features
1	Determining the game pool and checking the grammaticality and semantics of the items	1	Expert in Turkish Language and Literature
2	Determining the relationship between traditional games and suggested ways to develop algorithmic thinking skills	4	Expert in Information Technologies
3	Determining the content validity of the test	3	Expert in Information Technologies and Mathematics
		24	Middle School Sixth-Grade Students
4	Performing validity and reliability studies of the test	252	Middle School Sixth Grade Students
5	Determining the passing score of the test	3	Information Technologies Teacher

2.2. Data Collection Tools

2.2.1. *Document*

Models proposed by researchers such as Szanto (2002), Garner (2003), Futschek (2006), Vascencolos (2007), and Committee on Logic Education for the development of algorithmic

thinking skills were examined. In order to analyze the test items in detail, during the document creation process, a form containing the game pool and models for model selection was created and the opinions of 2 information technologies field experts were taken. Field experts expressed a common opinion about the suitability of the model proposed by Vascencolos. Therefore, the test items created within the scope of this research were divided into weekly worksheets in accordance with the algorithmic thinking skill teaching method suggested by Vascencolos (2007) and applied during the course, by consulting the experts. Question items were added, and the students were expected to answer the following questions: a) What exactly is the problem you are expected to solve? State the problem in your own words; b) Clarify the problem situation by drawing a picture or diagram of the problem; c) What are the known algorithmic concepts or elements? d) What is/are the unknown or requested situation(s)? e) Illustrate your solution with a row algorithm or flowchart; f) Test your solution and explain your answer. Thus, information was collected about the students' response status to the test items and their ability to identify and use the basic structure and elements of the algorithm in the items.

2.2.2. *Interview*

The interview is preferred in order to reveal the thoughts, attitudes, interests and beliefs of the source person or group for the research (Ocak, 2010). In this research, an unstructured interview, one of the interview types, was used. Unstructured interview: It enables the researcher to obtain detailed information about the situation by constantly asking different questions according to the answers he receives from the interviewee. (Gall et al., 1996; Türnüklü, 2000). Students and science experts were interviewed as needed at certain stages of the study. Sample questions are as follows: Can you give examples of traditional games that can be used in relation to algorithmic thinking skills during the research process for "Determining the Game Pool and Grammatics and Semantics of the Items" for Turkish Language Literature Field Expert? Are there any grammatical or semantic errors in the test items? How? To the Information Technologies Field Expert for "Determining the content validity", which of the basic structure and elements of the algorithm do you think are included in the relevant article? Information showing which topics were carried out and with whom the interview process was carried out is given in Table 3 above.

2.3. The Procedure

The test development process was based on Downing's (2006) test development model. The schematic depiction of the model is shown in Figure 1.

Figure 1. *Test development procedure (Downing, 2006).*
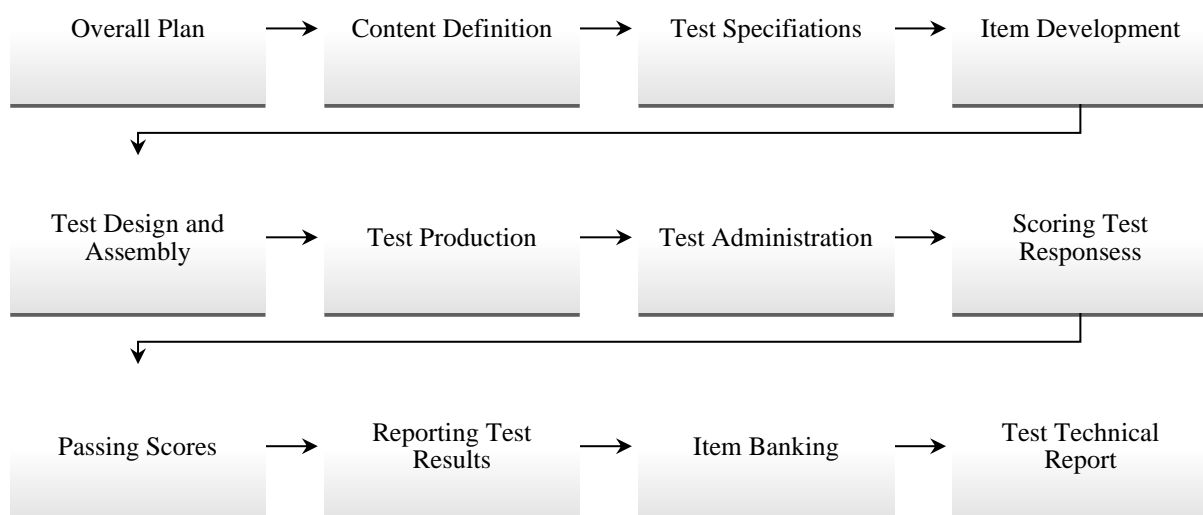
2.3.1. *Overall plan*

According to Downing (2006), the most crucial step of test development is to clearly state the purpose of the test in question. The purpose of the current test is to assess algorithmic thinking skills. The test items were planned in multiple-choice format to be applied in pen and paper style.

2.3.2. *Content definition*

In this step, the content of the test was described. The scope of the content was determined by including the basic algorithm structures and elements such as conditions, loops, operators, constants, and variables. These concepts were integrated within the framework of traditional games and game rules later in the procedure. A game pool was created, and question items were written. The games that make up the scope of the test are presented in Table 4.

Table 4. *Games covered in the test.*

Game Names			
Jump-rope	Hide-and-Seek	Hopscotch	Know and Keep Your Name
Relay	Day-and-Night	Musical Chairs	Puss-in-the-corner
Pounding Nails	Dodgeball	Handkerchief Grabber	Stop!
Word Derivation	Five Stones	Boom!	41 Sticks
Shopkeeper	Tawing	Captive	Grasshopper

2.3.3. *Test specifications*

In this step, the preferred test format, the number of items, the scoring rules, and the time limit are determined. It is recommended that the multiple-choice test type be preferred in cases of assessment at the lower levels of the cognitive field (Güler, 2015). This test type has two major advantages. Easy and objective scoring supports reliability, and the opportunity to measure a large number of outcomes ensures higher validity. It is also convenient as it can be applied to large audiences and it allows for rapid and precise assessment (Başol, 2019). Due to these strengths, a multiple-choice test format was selected for the tool in the study. The assessment tool was first drafted with a total of 27 items, 2 of which were sample items. In order to estimate the test duration, a pilot application was conducted on a group with similar characteristics as the sample group.

2.3.4. *Item development*

The actions undertaken at this stage are shown in Figure 2. Downing (2006) suggests writing items in the specified number and type and then having them reviewed by field experts. The experts' feedback is used to revise the items. In this study, the game was picked from the pool of games, and items were developed by taking a snapshot of the game played by certain characters within the framework of the game rules. The initial items consisted of three options "yes", "no" and "maybe". The items were reviewed by three experts. It was suggested by the experts to write options more relevant to the items. Thus, the items were revised accordingly. Worksheets were drawn up with the items and they were applied to seven students at the sixth-grade level for 7 weeks. It was checked whether the question items put the students through the steps in Vascencolos's model (2007). The worksheets were collected from the students, then they were reviewed, and matrices were created by two experts independently. The agreement rate between the experts was found 0.90 with the reliability formula of Miles and Huberman (1994). According to the reliability formula of Miles & Huberman (1994) used in the analysis of qualitative data, the fact that the result of Reliability = Consensus / (Consensus + Disagreement) above 0.80 indicates that the reliability level is high. The findings which were reached through consensus are presented in Table 5. It was seen that the students were able to

determine the basic algorithm structures and elements in the games and transfer them to the line flow charts, and as a result, the question items were proved to be compatible with the model of Vascencolos (2007).

Table 5. *Results of test item check against Vascencolos's (2007) model.*

	Participant						
Theme	P1	P2	P3	P4	P5	P6	P7
Read and Comprehend the Problem Statement	7	7	7	7	7	7	7
Select Theoretical Concepts That May Be Applied	7	6	7	7	7	6	6
Qualitative Description of the Problem	7	6	7	7	6	5	6
Formulization of a Solution Strategy	6	6	7	6	7	6	7
Test and Description of the Solution	7	7	7	7	7	7	7

At the end of this stage, a discussion was held with the participant group regarding the items and the number of options, and it was decided to increase the number of options to 4.

Figure 2. *Diagram of item development process.*
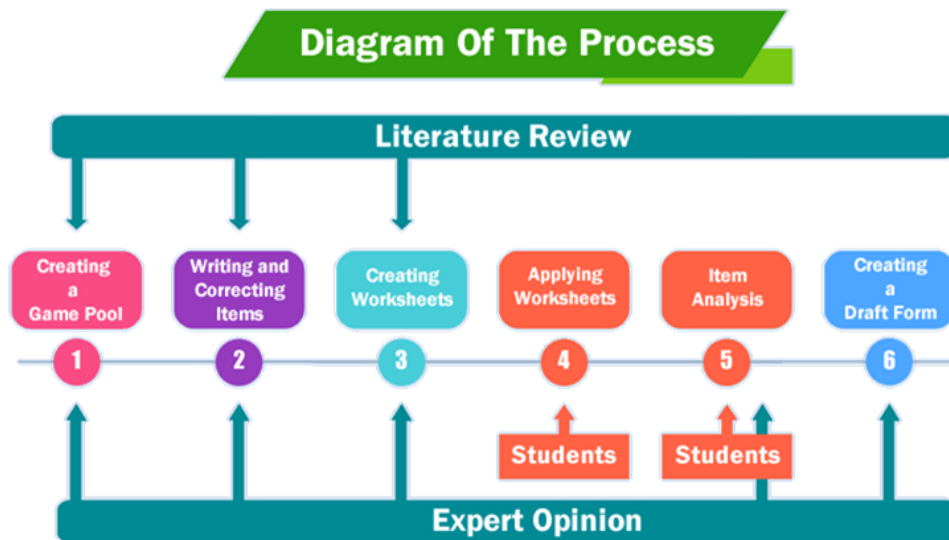
2.3.5. *Test design and assembly*

According to Downing (2006), the items are included in the test form in a way that minimizes the cognitive load on the respondents. The practices and the data obtained up to this stage were discussed with two experts. As a result, the draft form was obtained. The suitability of the items in the draft form and the algorithmic structures and elements embedded in the items were reviewed by three experts. The test was finalized accordingly. A sample item is shown below. Once the draft form was ready, it was given to 14 students in the sixth grade in order to appraise the time needed to answer the test items and to make a preliminary assessment of the items before starting the pilot application. Based on this trial, it was decided to fix the test duration to 40 minutes.

Sample Item: Five players stand around a circle. Ali starts counting by saying "One!" and counting continues clockwise one by one. The players corresponding to 5 and its multiples have to shout "Boom!" instead of "Five" or so on. Those who fail to do so are eliminated and the counting continues from where it left off. When it is Ali's turn for the third time, one person has been eliminated from the game.

In this case, which of the following is true?

A) Ali has said "Boom!" once.
B) Ali has said "Boom!" twice.
C) Ali has said "Boom!" three times.
D) Ali has never said "Boom!".

2.3.6. *Test production*

At this stage, the algorithmic thinking skill assessment test was published. The printout was checked carefully for typos.

2.3.7. *Test administration*

To make sure that the physical conditions were equal for all students during the application of the test, the students were given the test in their classrooms, and they were supervised by their teachers. Application permission from the Directorate of National Education, ethics committee report from Trabzon University and informed consent form from student parents were obtained for students to participate in the process. The advisors who supported the test in different classes during the evaluation process of the test were asked to inform the students that the exam would be completed within 1 class hour (40 minutes), that no course scores of the students would change with the test results, and that the purpose of the test was to measure the algorithmic thinking skills of the students. Thus, it is aimed to help students easily transfer their existing knowledge to paper without worrying emotionally during the evaluation phase. Similar information was given to the students during the application of the test, in which the researcher participated as an evaluator.

2.3.8. *Scoring test responses*

Since the 24 6th-grade students who participated in the development phase of the algorithmic thinking skill test items and the 256 6th-grade students who participated in the evaluation phase were in different schools in different provinces, there was no interaction between them. During the item development phase and evaluation process of the test, all students completed the evaluation simultaneously within one class hour (40 minutes). However, the only difference is that the evaluation for the item development phase of the test is made at a time before the test evaluation process. The students' responses in the test were coded as 0 or 1 in an MS Excel sheet by the researcher depending on the meaning of the responses. Then, the item difficulty indexes and item discrimination indexes were calculated, and a transaction was performed to determine the items to be included in the final test. Additionally, the KR-20 reliability calculation method was used to compute the reliability of the test.

2.3.9. *Passing scores*

In Downing's model (2006), the passing score of a test is calculated according to the absolute or relative criteria rules. In this study, the opinions of three field experts were asked and it was decided to determine the score value of each item separately in calculating the passing score of the test. It was advised by the experts to carry out two actions in determining the scores of the items. The first thing to do is to classify the items by difficulty level based on the item difficulty index value. In the second step, points are appointed to the basic algorithm structures and elements in the item by judging the importance of those elements and the structures for the solution of the problem. According to Başol (2019), item difficulty index values between 0.85 and 1.00 refer to very easy items, those between 0.61 and 0.84 are considered easy, items

ranging from 0.40 to 0.60 are considered medium difficulty, difficulty index between 0.16 and 0.39 refer to difficult items, and the values between 0.00 and 0.15 indicate very difficult items. According to the item difficulty index calculation, 16 of the items were found at medium difficulty and six items were easy. The difficulty levels of the items and their classification according to the elements they cover are shown in Table 6 (After the item statistics were made, three test items were removed and item numbers were recoded between 1-22.).

Table 6. *Test items by difficulty level and contained elements.*

Item Difficulty Level	Algorithm Structures and Elements Contained in the Item	No of Items	Item No
Medium Difficulty	Condition, Loop, Constant, Variable, Operator	12	4,7,8,9,10,14,15,18,19,20,21,22
	Condition, Loop, Variable, Operator	2	1,5
	Condition, Constant, Variable, Operator	2	2,17
Easy	Condition, Loop, Constant, Variable, Operator	1	12
	Condition, Constant, Variable, Operator	2	11,16
	Condition, Loop, Operator	1	13
	Condition, Variable, Operator	2	3,6

Definitions: Condition: Probability situation, Loop: Recurring situation, Constant: Unchanged in value, Variable: Changed value, Operator: Mathematical operations.

The algorithm structures and elements contained in each item were evaluated relative to each other. According to the experts in the study, condition and loop were of equal importance and more important than the other elements of the algorithm. The remaining elements, which are constant, variable and operator, were attached to the same level of importance. The scores of the items were calculated by taking into consideration the scores given by the experts for the basic structures and elements of the algorithm.

The scores for the basic algorithm structures and elements embedded in the items of medium difficulty are given in Table 7. The scores of the items of medium difficulty varied between 3.5 and 5.5 points depending on the basic structures and elements of the algorithm contained in the respective items.

Table 7. *Basic algorithm structures and element scores in items of medium difficulty.*

Basic Algorithm Structure and Element	Score
Condition	2
Loop	2
Constant	0.5
Variable	0.5
Operator	0.5

The scores for the basic structures and elements of the algorithm covered in easy items are presented in Table 8. It can be seen that the scores of the easy-level items range from 2.75 to 4 points, depending on the structures and elements contained in the respective items.

Table 8. *Basic algorithm structure and element scores in easy items.*

Basic Algorithm Structure and Element	Score
Condition	1.25
Loop	1.25
Constant	0.5
Variable	0.5
Operator	0.5

A student who answers all the test items correctly gets 100 points. Table 9 gives details about the scores of the test items.

Table 9. *Scores of test items.*

Items of Medium Difficulty		
No of Items	Item No	Item Score
12	4,7,8,9,10,14,15,18,19,20,21,22	5.5
2	1,5	5
2	2,17	3.5
Easy Items		
1	12	4
1	13	3
2	11,16	2.75
2	3,6	2.25

The Angoff method was preferred to determine the cut-off score of the test. It is the most widely used method in determining the cut-off score for tests (Demir & Köse, 2014). In the Angoff method, experts analyze the test items one by one and estimate the correct answer rate for each item for 100 students. Then, the average of the experts' estimates for each item is calculated to set the minimum passing score for the items separately. Finally, the passing score of the test is determined by taking the average of the minimum passing scores of all items. In this study, the estimates were made by three experts, and they are shown in Table 10 below. As instructed in the method above, the minimum passing scores for all of the items were determined by taking the average of the estimates reported by the experts for the items in the first place. Based on the average of the minimum passing scores, the passing score of the algorithmic thinking skill assessment test was set as 67.23.

Table 10. *Angoff method results for the algorithmic thinking skill test.*

Item No	Expert 1	Expert 2	Expert 3	Minimum Passing Score
Item 1	50	60	70	60
Item 2	60	40	50	50
Item 3	80	75	80	78.3
Item 4	75	65	70	70
Item 5	55	65	60	60
Item 6	72	64	80	72
Item 7	76	70	60	68.6
Item 8	55	60	75	63.3
Item 9	50	55	65	56.6
Item 10	80	75	75	76.6
Item 11	60	70	75	68.3
Item 12	85	80	85	83.3
Item 13	70	80	85	77.5
Item 14	40	55	65	78.3
Item 15	45	50	50	48.3
Item 16	60	60	65	61.6
Item 17	65	50	70	61.6
Item 18	55	55	65	58.3
Item 19	70	55	65	63.3
Item 20	75	70	80	78.3
Item 21	70	65	75	70
Item 22	70	80	75	75
Final Test Cut-off Score				67.23

2.3.10. *Reporting test results*

Downing (2006) states that students who take a test have the right to receive a report on their test performance. It is considered important to give a feedback report on student mistakes in clear and understandable language. For this reason, a report was created after assessing the student responses.

2.3.11. *Item banking*

According to Downing (2006), it is essential to safely store the items that are regarded as effective in evaluations in case they are needed for developing a new test form or developing a different version of the test. Therefore, the test items were saved in a safe platform.

2.3.12. *Test technical report*

Downing (2006) points out that all data regarding test development activities must be reported in complete. Again, the entire development process along with the findings concerning test reliability and validity are elaborated here.

3. FINDINGS

3.1. Validity and Reliability Studies

Test development becomes final upon the collection of data for validity and reliability. Validity means the determination of the extent to which the assessment tool measures the intended construct without confusing it with other features (Büyüköztürk et al., 2020). Reliability is defined as the quality of the assessment tool being free from random errors (Baykul et al., 2003; Güler, 2015). In this study, the results of the validity and reliability analyses of the test are presented under the following headings.

3.1.1. *Content validity*

Content validity relates to the extent the test items are competent in measuring the behaviors intended to be measured. The number and quality of test items are important to create a test

with high content validity. For content validity, expert opinion is frequently sought regarding the suitability and ability of the item to measure the intent situation. A table of specifications is created in light of the expert feedback (Büyüköztürk et al., 2020). In this study, 4 Information Technologies experts examined the 25 items of the test and produced the table. The specification table revealed that among the basic algorithm concepts, the condition was available in 24 items, the loop was in 16 items, the constant in 19 items, the variable in 25 items, and the operator was available in 24 items.

3.1.2. *Construct validity*

Construct validity concerns the adequacy of the test scores in measuring the construct targeted to be measured by the test (Büyüköztürk et al., 2020). To check construct validity, the items were analyzed according to the lower-upper group differences. The upper and lower groups were identified by applying the "27% rule" offered by Kelley (1939). In this scope, first of all, the total scores obtained from the test were calculated and the students were ranked from the highest to the lowest. The 27% extreme group that got the highest scores (68 people with the highest scores) was placed in the upper group, while the other 27% extreme group that got the lowest scores (68 people with the lowest scores) was named as the lower group. Finally, an independent sample t-test was applied on the difference between the upper and lower groups of the items. A significant difference between the groups shows that the items have enough discriminatory capacity to tell the proficient from the nonproficient students in terms of algorithmic thinking (Yıldırım & Şimşek, 2006). The data obtained from the independent sample t-test performed on the final test are given in Table 11. The results revealed that the items were significant ($p<0.01$) in distinguishing the students in the lower group from those in the upper group.

Table 11. *Independent sample t-test results based on lower-upper group difference.*

Item No	27% Lower-Upper Group Difference p Value
Item 1	0.000**
Item 3	0.000**
Item 4	0.000**
Item 5	0.000**
Item 6	0.000**
Item 7	0.000**
Item 8	0.000**
Item 9	0.000**
Item 11	0.000**
Item 12	0.000**
Item 13	0.000**
Item 14	0.000**
Item 15	0.000**
Item 16	0.000**
Item 17	0.000**
Item 19	0.000**
Item 20	0.000**
Item 21	0.000**
Item 22	0.000**
Item 23	0.000**
Item 24	0.000**
Item 25	0.000**

**p<0.01

3.2. Reliability Analysis

3.2.1. *Item discrimination index*

Item discrimination index expresses the capacity of each test item to distinguish a high performer from a low performer. In other words, it is answering an item correctly by a high-achieving student whereas being answered incorrectly by a low-achieving student. The item discrimination index is calculated by subtracting the number of respondents with correct answers in the upper group from the number of those in the lower group and then dividing the result by half of the whole group. Item discrimination index value varies between -1 and 1 (Bayrakçeken, 2015).

Item discrimination index value (rjx);

rjx>=0.40 signifies a very good item,
0.30<=rjx>=0.39 signifies a good item that can be kept in scale without amendment,
0.20<=rjx>=0.29 signifies an item that needs correction and improvement,
0.19<rjx signifies an item that ought to be omitted (Büyüköztürk et al., 2020). The item discrimination index results obtained for the test in this study are given in Table 12. It was found that Items No 2, 10 and 18 were not good items so they were removed from the scale.

3.2.2. *Item difficulty index*

Item difficulty index indicates the correct answer rate for each item in an assessment tool. It is calculated by summing up the respondents in both the upper and lower group providing a correct answer for a given item and then finding the ratio of this sum to the whole group. Item difficulty index can take a value between 0 and 1. A value close to 0 marks a difficult item, while values close to 1 signal easy items (Bayrakçeken, 2015). The item difficulty index results for the test in this study are given in Table 12. The average item difficulty index was found to be 0.52, which implies that the items in the assessment tool were of medium difficulty.

Table 12. *Item difficulty index and item discrimination index results for the test.*

Item No	Item Difficulty Index	Item Discrimination Index
Item 1	0.59	0.42
Item 2	**0.13**	**0.10**
Item 3	0.40	0.42
Item 4	0.69	0.55
Item 5	0.44	0.33
Item 6	0.52	0.44
Item 7	0.75	0.30
Item 8	0.45	0.32
Item 9	0.52	0.55
Item 10	**0.16**	**0**
Item 11	0.40	0.39
Item 12	0.50	0.67
Item 13	0.63	0.51
Item 14	0.73	0.47
Item 15	0.78	0.36
Item 16	0.40	0.54
Item 17	0.40	0.60
Item 18	**0.22**	**0.17**
Item 19	0.61	0.64
Item 20	0.46	0.51
Item 21	0.41	0.50
Item 22	0.47	0.48
Item 23	0.51	0.55
Item 24	0.45	0.58
Item 25	0.40	0.51
Item Difficulty Index Mean: 0.52, Discriminatory Index Mean: 0.48		

3.2.3. *KR-20 reliability analysis*

The KR-20 formula is applicable in cases where the responses to test items are scored as 0 (wrong) or 1 (correct) (Büyüköztürk et al., 2020). As the KR-20 value approaches 1, it is assumed that the internal consistency increases and the test is a homogeneous tool that measures similar features, corresponding to higher levels of reliability. By contrast, the test is regarded to have low reliability as the KR-20 value approaches 0 (Çetin, 2019). The KR-20 result for the current test is shown in Table 13. As can be understood from the table, the reliability of the test was high (KR-20= 0.836> 0.70).

Table 13. *KR-20 reliability analysis result.*

Test	KR-20	No of Items
Algorithmic Thinking Skill Assessment Test	0.836	22

4. DISCUSSION and CONCLUSION

In this study, an algorithmic thinking skill assessment test was developed through contextualization with traditional games. The test ultimately consisted of 20 conventional children's games for answering 22 relevant items. Either the rules of the games were explained or snapshot situations were taken from the games, and scenarios were invented accordingly. Problem situations were given along with correct and incorrect options. Consequently, a measurement tool was designed whereby students' algorithmic thinking skills were measured in the context of common traditional games. Previous studies assessing algorithmic thinking skills through games were also investigated. To begin with, in Gürbüz et al. (2012), the students were provided with the knowns like the sun, temperature, humidity and wind and they were asked to prepare a weather forecast in reference to the given facts and figures. In the study of Zhao and Shute (2019), there was a character in the game and the students were supposed to bring it to the target point by giving the right commands. Similarly, in the study by Czakoova (2020), a three-level game was presented and the students were expected to animate the character and thus collect points by giving the right commands during the first two levels. In the third level, the character must reach the target most effectively and shortly. Again, in Kazimoglu's (2020) study, there was a robot and the students were instructed to make it to the target in the most effective way by means of giving commands. Another example was a study by Chen and Chi (2020). Two groups of students acted as pirate gangs trying to seize the treasure and they had to develop strategies to achieve their goal.

With the collaboration of the relevant field experts, it was understood that the majority of the test items included condition, loop, constant, and variable and operator structures of the algorithm. Some of the items contained three of the algorithm structures and elements and some others contained four of them. Moreover, the study data demonstrated that "condition" became the most frequently used algorithm element evident in 22 items, but "loop" was the least seen element with 16 appearances. Thanks to this property, this new assessment test of algorithmic thinking skills seems to be an outstanding one in the literature. In contrast to the current study, the previous studies generated algorithmic thinking skill assessment tools that address only one or two of the basic algorithm structures and elements at one time (Hsu & Wang, 2018; Tsukamoto et al., 2017). For instance, Tsukamoto et al. (2017) developed a measurement instrument for primary school students' algorithmic thinking skills. It was built around sequential operations, conditional branches, and iterative operations. There were three items in the tool, each intending to measure one specific concept. Likewise, Hsu and Wang (2018) introduced an achievement test to assess algorithmic thinking skills, and the item added for debugging targeted conditionals as an algorithm structure. Another item assigned a task that entails using both condition and loop among basic algorithm structures and elements.

The resulting test is a valid and highly reliable instrument contextualized in the most common local children's games. According to Yılmaz (2020), when all the steps in a traditional game process are sequentially used and associated with a concept chosen from any discipline, students will have unconsciously learned the target concept while completing the game task in order to achieve success. Hence, it seems beneficial to integrate traditional games, which have survived as cultural heritage to the present day, into the education process.

In the study, content validation of the test was performed by taking expert opinion. A team of eight experts provided support for this particular job. Of the experts, seven were specialists of computer and instructional technology education and they were responsible for the content and applicability of the items. The other expert was a specialist in Turkish Language and Literature for checking the spelling and grammar rules in the texts. This step was in congruence with the literature as there were other studies seeking expert opinion to ensure the content validity of the test (Kocagül-Sağlam & Ünal-Çoban, 2018). As for construct validity, several methods were found in the literature, but one of the most extensively used ones was to perform an independent sample t-test on the difference between the upper and lower groups, as conducted in the present case (Özden & Yenice, 2021). Thus, both content and construct validation methods used in this study are in agreement with the literature.

When it comes to determining the difficulty indexes of the items in the algorithmic thinking skill assessment test, the values were noted between 0.40 and 0.78, resulting in the average difficulty index equal to 0.52 for the entire test. These figures reveal that the algorithmic thinking skill assessment test developed here is difficult at the intermediate level. Furthermore, the discrimination indexes of the individual items varied between 0.30 and 0.67, and the average discrimination index value of the test was found to be 0.48. In general, it is desirable to have a measurement tool with high validity and reliability. The meeting of these criteria is checked by looking at test results, particularly at the item discrimination index and item difficulty index values of each item (Güler, 2015). Although the item difficulty index values are not equal to 0.50, it is favourable to have the average test difficulty close to 0.50 for a successful test development process (Bayrakçeken, 2015). Discrimination index values between 0.30 and 0.40 classify good items, values greater than 0.40 classify very good items, and values lower than 0.20 imply that the items are too weak to remain on the scale (Özçelik, 2013). These threshold values support the quality of the test in that it consists of only very good and good items and it is a highly discriminatory instrument as a whole.

Lastly, the KR-20 coefficient was 0,83 which is far greater than the acceptable lower limit of 0.70. It is thus obvious that the reliability of the algorithmic thinking skills assessment test is at a satisfactory level (Büyüköztürk et al., 2020). The Kuder-Richardson (KR-20) reliability determination method is used to examine the internal consistency between test scores, in which responses to test items are computed as 0 (false) and 1 point (correct). This study is congruent with the related literature since the KR-20 reliability analysis was used for test reliability in most studies (Karatay & Doğan, 2016; Özden & Yenice, 2021; Şardağ & Kocakülah, 2016). Like many other aspects, the reliability calculation method used in this study is also similar to the previous studies in the literature.

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**: Trabzon University, 13.08.2021, E-81614018-000-704.

Contribution of Authors

The authors contributed equally to all the stages of the study.

Orcid

Emre Zengin https://orcid.org/0000-0002-6643-7550
Yasemin Karal https://orcid.org/0000-0003-4744-4541

REFERENCES

Altakrouri, B., & Schrader, A. (2012, September). Towards dynamic natural interaction ensembles. In *the 26th BCS Conference on Human Computer Interaction 26* (pp. 1-10). http://dx.doi.org/10.14236/ewic/HCI2012.0

Anderson, M.L. (2003). Embodied cognition: A field guide. *Artificial intelligence*, *149*(1), 91-130. https://doi.org/10.1016/S0004-3702(03)00054-7

Apostolellis, P., Stewart, M., Frisina, C., & Kafura, D. (2014). RaBit EscAPE: A board game for computational thinking. In *Proceedings of the 2014 conference on Interaction design and children* (pp. 349-352). http://dx.doi.org/10.1145/2593968.2610489

Ayala, N.A.R., Mendívil, E.G., Salinas, P., & Rios, H. (2013). Kinesthetic learning applied to mathematics using kinect. *Procedia Computer Science*, *25*, 131-135. https://doi.org/10.1016/j.procs.2013.11.016

Aytekin, A., Çakır, F.S., Yücel, Y.B., & Kulaözü, İ. (2018). The place and importance of algorithms in our lives. *Eurasian Journal of Social and Economic Studies*, *5*(7), 143-150. https://dergipark.org.tr/tr/pub/asead/issue/41013/495619

Başol, G. (2019). Measurement and evaluation in education. *Pegem Citation Index*, 001-307.

Baykul, Y., Gelbal, S., & Kelecioğlu, H. (2003). *Measurement and evaluation in education for Anatolian teacher high schools*. National Education Printing House.

Bayrakçeken, S. (2015). Test development. E. Karip (Ed.)*, In Measurement and Evaluation* (s. 292-322). Pegem Academy.

Bellocchi, A., King, D.T., & Ritchie, S.M. (2016). Context-based assessment: Creating opportunities for resonance between classroom fields and societal fields. *International Journal of Science Education*, *38*(8), 1304-1342. https://doi.org/10.1080/09500693.2016.1189107

Borkulo, S., Chytas, C., Drijvers, P., Barendsen, E., & Tolboom, J. (2021). Computational thinking in the mathematics classroom: Fostering algorithmic thinking and generalization skills using dynamic mathematics software. In *The 16th Workshop in Primary and Secondary Computing Education* (pp. 1-9). https://doi.org/10.1145/3481312.3481319.

Brown, W. (2015). *Introduction to algorithmic thinking.* https://raptor.martincarlisle.com/Introduction%20to%20Algorithmic%20Thinking.doc

Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö.E., Karadeniz, Ş., & Demirel, F. (2020). *Scientific research methods*. Pegem Publications.

Chen, K.Z., & Chi, H.H. (2022). Novice young board-game players' experience about computational thinking. *Interactive Learning Environments*, *30*(8), 1375-1387. https://doi.org/10.1080/10494820.2020.1722712

Chuechote, S., Nokkaew, A., Phongsasithorn, A., & Laosinchai, P. (2020). A neo-piagetian analysis of algorithmic thinking development through the "sorted" digital game. *Contemporary Educational Technology*, *12*(1), 1-15. http://dx.doi.org/10.30935/cet.685959

Czakóová, K. (2020). *Developing algorithmic thinking by educational computer games*. Paper presented at the Conference eLearning and Software for Educationtional, Romania. http://dx.doi.org/10.12753/2066-026X-20-003

Czakóová, K., & Udvaros, J. (2021). Applications and games for the development of algorithmic thinking in favor of experiential learning. In *EDULEARN21 Proceedings* (pp. 6873-6879). IATED. https://doi.org/10.21125/edulearn.2021.1389

Çetin, B. (Ed.) (2019). *Measurement and evaluation in education*. Anı Publishing.

Debabi, W., & Bensebaa, T. (2016). Using serious game to enhance learning and teaching algorithmic. *Journal of e-Learning and Knowledge Society*, *12*(2). http://dx.doi.org/10.20368/1971-8829/1125

Demir, O., & Köse, İ.A. (2014). Comparison of cutoff scores determined by Angoff, Nedelsky and Ebel standard setting methods. *Journal of Mersin University Faculty of Education*, *10*(2), 14-27. https://dergipark.org.tr/en/pub/mersinefd/issue/17394/181823?publisher=mersin

Doleck, T., Bazelais, P., Lemay, D.J., Saxena, A., & Basnet, R.B. (2017). Algorithmic thinking, cooperativity, creativity, critical thinking, and problem solving: Exploring the relationship between computational thinking skills and academic performance. *Journal of Computers in Education*, *4*(4), 355-369. https://doi.org/10.1007/s40692-017-0090-9

Downing, S.M. (2006). Twelve steps for effective test development. *Handbook of test development*, *3*, 25. https://doi.org/10.4324/9780203874776.ch1

Elshahawy, M., Aboelnaga, K., & Sharaf, N. (2020). Codaroutine: A serious game for introducing sequential programming concepts to children with autism. *In 2020 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1862-1867). IEEE. http://dx.doi.org/10.1109/EDUCON45650.2020.9125196

Erümit, K.A., Karal, H., Şahin, G., Aksoy, D.A., Aksoy, A., & Benzer, A.İ. (2018). A model proposal for teaching programming: programming in seven steps. *Education and Science*, *44*(197), 1-29. http://dx.doi.org/10.15390/EB.2018.7678

Evripidou, S., Amanatiadis, A., Christodoulou, K., & Chatzichristofis, S.A. (2021). Introducing algorithmic thinking and sequencing using tangible robots. *IEEE Transactions on Learning Technologies*, *14*(1), 93-105. https://doi.org/10.1109/TLT.2021.3058060

Fensham, P.J., & Rennie, L.J. (2013). Towards an authentically assessed science curriculum. In *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 69-100). Springer.

Figueiredo, M., Amante, S., Gomes, H.M.D.S.V., Gomes, M.A., Rego, B., Alves, V., & Duarte, R.P. (2021). Algorithmic thinking in early childhood education: Opportunities and supports in the Portuguese context. In *EDULEARN21 Proceedings* (pp. 9339-9348). IATED. https://doi.org/10.22521/edupij.2022.112.3

Futschek, G., & Moschitz, J. (2010). Developing algorithmic thinking by inventing and playing algorithms. *Proceedings of the 2010 constructionist approaches to creative learning, thinking and education: Lessons for the 21st century (constructionism 2010)*, 1-10.

Gall, M.D., Barg, W.R., & Gall, j.P. (1996). Educational research: an introduction (6th ed.). Longman Publishing.

Geisinger, K.F. (2016). 21st century skills: What are they and how do we assess them?. *Applied Measurement in Education*, *29*(4), 245-249. https://doi.org/10.1080/08957347.2016.1209207

Güler, N. (2015). *Measurement and evaluation in education* (7th ed.). Pegem Academy Publishing.

Gürbüz, H., Evlioğlu, B., Erol, Ç.S., Gülseçen, H., & Gülseçen, S. (2017). "What's the weather like today?": A computer game to develop algorithmic thinking and problem solving skills of primary school pupils. *Education and Information Technologies*, *22*(3), 1133-1147. https://doi.org/10.1007/s10639-016-9478-9

Hsu, C.C., & Wang, T.I. (2018). Applying game mechanics and student-generated questions to an online puzzle-based game learning system to promote algorithmic thinking skills. *Computers & Education*, *121*, 73-88. http://dx.doi.org/10.1016/j.compedu.2018.02.002

Hu, C. (2011). Computational thinking: What it might mean and what we might do about it. In Proceedings of the 16th annual joint conference on innovation and technology in computer science education (pp. 223-227). ACM. http://dx.doi.org/10.1145/1999747.1999811

Huizinga, J. (1955). Homo ludens: A study of the play element in culture. The Beacon Press. https://doi.org/10.2307/2087716

Iacolina, S.A., Lai, A., Soro, A., & Scateni, R. (2010). Natural interaction and computer graphics applications. In *Eurographics italian chapter conference* (pp. 141–146).

Jančec, L., & Vujičić, L. (2021). Project "Algorithmic Thinking Skills through Play-Based Learning for Future's Code Literates". In *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 641-644). IEEE. https://doi.org/10.23919/MIPRO52101.2021.9597078

Kanaki, K., & Kalogiannakis, M. (2022). Assessing algorithmic thinking skills in relation to gender in early childhood. *Educational Process: International Journal*, *11*, 44-59. https://dx.doi.org/10.22521/edupij.2022.112.3

Karakasis, C., & Xinogalos, S. (2020). BlocklyScript: Design and pilot evaluation of an RPG platform game for cultivating computational thinking skills to young students. *Informatics in Education*, *19*(4), 641-668. https://doi.org/10.15388/infedu.2020.28

Karatay, R., & Doğan, F. (2016). Development of science process skill scale of 7th grade secondary school students. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi, 27*, 1-8. http://dx.doi.org/10.14582/DUZGEF.548

Kátai, Z. (2015). The challenge of promoting algorithmic thinking of both sciences-and humanities-oriented learners. *Journal of Computer Assisted Learning*, *31*(4), 287-299. https://doi.org/10.1111/jcal.12070

Kazimoglu, C. (2020). Enhancing confidence in using computational thinking skills via playing a serious game: A case study to increase motivation in learning computer programming. *IEEE Access*, *8*, 221831-221851. http://dx.doi.org/10.1109/ACCESS.2020.3043278

Keane, T. (2012). Leading with technology: 21st century skills= 3Rs+ 4Cs. *Australian Educational Leader*, *34*(2), 44. https://search.informit.org/doi/10.3316/informit.895597893475453

Khanlari, A. (2013). Effects of robotics on 21st century skills. *European Scientific Journal*, *9*(27). https://doi.org/10.19044/esj.2013.v9n27p%25p

Kim, B., Park, H., & Baek, Y. (2009). Not just fun, but serious strategies: Using meta-cognitive strategies in game-based learning. *Computers & Education*, *52*(4), 800-810. https://doi.org/10.1016/j.compedu.2008.12.004

Kiss, G., & Arki, Z. (2017). The influence of game-based programming education on the algorithmic thinking. *Procedia-Social and Behavioral Sciences*, *237*, 613-617. http://dx.doi.org/10.1016/j.sbspro.2017.02.020

Kocagül-Sağlam, M. & Ünal-Çoban, G. (2019). Developing the reasoning skills test for science teachers and teacher candidates. *Elementary Education Online*, *17*(3). https://doi.org/10.17051/ilkonline.2018.466374

Kosmas, P., Ioannou, A., & Retalis, S. (2018). Moving bodies to moving minds: A study of the use of motion-based games in special education. *TechTrends*, *62*, 594-601. https://doi.org/10.1007/s11528-018-0294-5

Kosmas, P., & Zaphiris, P. (2023). Improving students' learning performance through Technology-Enhanced Embodied Learning: A four-year investigation in classrooms. *Education and Information Technologies*, 1-24. https://doi.org/10.1007/s10639-022-11466-x

La Belle, T.J., Moll, L.C., & Weisner, T.S. (1979). Context-based educational evaluation: A participant research strategy. *Educational Evaluation and Policy Analysis*, *1*(3), 85-94. https://doi.org/10.2307/1164160

Lee, T.Y., Mauriello, M.L., Ahn, J., & Bederson, B.B. (2014). CTArcade: Computational thinking with games in school age children. *International Journal of Child-Computer Interaction*, *2*(1), 26-33. http://dx.doi.org/10.1016/j.ijcci.2014.06.003

Lertlapnon, T., Lueangrungudom, N., & Vittayakorn, S. (2022). Protobot: An Educational Game for Algorithmic Thinking. In *2022 14th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 79-84). IEEE. https://doi.org/10.1109/ICITEE56407.2022.9954081

Li, J., Lin, Y., Sun, M., & Shadiev, R. (2020). Socially shared regulation of learning in game-based collaborative learning environments promotes algorithmic thinking, learning participation and positive learning attitudes. *Interactive Learning Environments*, 1-12. https://doi.org/10.1080/10494820.2020.1857783

Lin, S.Y., Chien, S.Y., Hsiao, C.L., Hsia, C.H., & Chao, K.M. (2020). Enhancing computational thinking capability of preschool children by game-based smart toys. *Electronic Commerce Research and Applications*, 44, 101011. https://doi.org/10.1016/j.elerap.2020.101011

McFarlane, A., Sparrowhawk, A., & Heald, Y. (2002). *Report on the educational use of games*. TEEM (Teachers evaluating educational multimedia). Cambridge.

Mezak, J., & Papak, P.P. (2018). Learning scenarios and encouraging algorithmic thinking. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 0760-0765). IEEE. https://doi.org/10.23919/MIPRO.2018.8400141

Mumcu, H.Y., & Yıldız, S. (2018). The investigation of algorithmic thinking skills of 5th and 6th graders according to different variables. *MATDER Journal of Mathematics Education*, *3*(2), 18-26. https://dergipark.org.tr/en/pub/med/issue/41621/453450

Özçelik, D.A. (2013). *Assessment and Evaluation in Schools Teacher's Handbook*. Pegem Academy.

Özden, B., & Yenice, N. (2021). Developing a scientific inquiry skills test for secondary school 7th and 8th grade students. *Journal of Mersin University Faculty of Education*, *17*(1), 112-131. https://doi.org/10.17860/mersinefd.726360

Paloma, G.F. (Ed.). (2017). *Embodied Cognition. Theories and applications in education science*. Nova Science Publishers. https://hdl.handle.net/11386/4712878

Paspallis, N., Kasenides, N., & Piki, A. (2022). A Software Architecture for Developing Distributed Games that Teach Coding and Algorithmic Thinking. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 101-110). IEEE. https://doi.org/10.1109/COMPSAC54236.2022.00023

Pivec, M., & Kearney, P. (2007). Games for learning and learning from games. *Informatica*, *31*(4).

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE)*, *1*(1), 21-21.

Sarı, U., Pektaş, H.M., Şen, Ö.F., & Çelik, H. (2022). Algorithmic thinking development through physical computing activities with Arduino in STEM education. *Education and Information Technologies*, 1-21. https://doi.org/10.1007/s10639-022-10893-0

Scharf, F., Winkler, T., & Herczeg, M. (2008). Tangicons: Algorithmic reasoning in a collaborative game for children in kindergarten and first class. Paper presented at the 7th International Conference on Interaction Design and Children, USA. http://dx.doi.org/10.1145/1463689.1463762

Shang, J., Ma, S., Hu, R., Pei, L., & Zhang, L. (2019). *Game-based learning in future school. In shaping future schools with digital technology*. Springer. http://dx.doi.org/10.1007/978-981-13-9439-3_8

Sungkaew, K., Lungban, P., & Lamhya, S. (2022). Game development software engineering: digital educational game promoting algorithmic thinking. *International Journal of Electrical and Computer Engineering (IJECE)*, *12*(5), 5393-5404. http://dx.doi.org/10.11591/ijece.v12i5.pp5393-5404

Şardağ, M., & Kocakülah, A. (2016). Developing a science process skills test for eighth grade students. *Journal of Sakarya University Faculty of Education*, *31,* 1-32. https://dergipark.org.tr/tr/pub/sakaefd/issue/24690/261073

Taasoobshirazi, G., & Carr, M. (2008). A review and critique of context-based physics instruction and assessment. *Educational Research Review*, *3*(2), 155-167. https://doi.org/10.1016/j.edurev.2008.01.002

Thomas, J.O., Odemwingie, O.C., Saunders, Q., & Watlerd, M. (2015). Understanding the difficulties African-American middle school girls face while enacting computational algorithmic thinking in the context of game design.

Thomas, J.O., Rankin, Y., Minor, R., & Sun, L. (2017). Exploring the difficulties African-American middle school girls face enacting computational algorithmic thinking over three years while designing games for social change. *Computer Supported Cooperative Work (CSCW)*, *26*(4), 389-421. https://doi.org/10.1007/s10606-017-9292-y

Tsukamoto, H., Oomori, Y., Nagumo, H., Takemura, Y., Monden, A., & Matsumoto, K.I. (2017). Evaluating algorithmic thinking ability of primary schoolchildren who learn computer programming. In *2017 IEEE Frontiers in Education Conference (FIE)* (pp. 1-8). IEEE. http://dx.doi.org/10.1109/FIE.2017.8190609

Turchi, T., Fogli, D., & Malizia, A. (2019). Fostering computational thinking through collaborative game-based learning. *Multimedia Tools and Applications*, *78*(10), 13649-13673. https://doi.org/10.1007/s11042-019-7229-9

Turnuklu, A. (2000). A qualitative research technique that can be used effectively in educational research: Interview. Educational management in theory and practice, *24*(24), 543-559.

Vasconcelos, J. (2007). *Basic Strategy for Algorithmic Problem Solving*.   http://www.cs.jhu.edu/~jorgev/cs106/ProblemSolving.html

Wangenheim, C., Medeiros, G., Missfeldt Filho, R., Petri, G., da Cruz Pinheiro, F., Ferreira, M.N., & Hauck, J.C. (2019). SplashCode--A Board Game for Learning an Understanding of Algorithms in Middle School. *Informatics in Education*, *18*(2), 259-280. http://dx.doi.org/10.31235/osf.io/2qbnp

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*, 625-636. https://doi.org/10.3758/BF03196322

Yadav, A., Gretter, S., Good, J., McLean, T. (2017). *Computational Thinking in Teacher Education*. In: Rich, P., Hodges, C. (eds) Emerging Research, Practice, and Policy on Computational Thinking. Educational Communications and Technology: Issues and Innovations. Springer, Cham. https://doi.org/10.1007/978-3-319-52691-1_13

Yıldırım, A., & Şimşek, H. (2006). *Qualitative research methods in the social sciences*. Seçkin Publishing.

Yılmaz, E.A. (2020). *By the power of games: an introduction to the science of gamification*. Epsilon Publishing House. https://dergipark.org.tr/tr/pub/tk/issue/56680/784921

Zsakó, L., & Szlávi, P. (2012). ICT Competences: Algorithmic Thinking. *Acta Didactica Napocensia*, *5*(2), 49-58.

Anxiety scale for parents of gifted children (ASPGC): Validity and reliability study

Kübra Yiğit Tekel [1,*], **Melike Şule Yıldız** [1], **Duygu Mutlu Bayraktar** [2], **Marilena Zinovia Leana Taşcılar** [3]

[1]Istanbul University-Cerrahpasa, Institute of Graduate Studies, Program in Education of Gifted Education, Türkiye
[2]Istanbul University-Cerrahpasa, Department of Computer Education and Instructional Technologies, Department of Computer and Education Technologies Education, Türkiye
[3]Istanbul Medeniyet University, Faculty of Education Sciences, Department of Special Education, Türkiye

Abstract: This study aimed to develop a valid and reliable measurement tool to determine the anxiety levels of parents with gifted children. The participants were selected using a convenient sampling method. The sample of this study consisted of 550 parents of gifted middle school students attending 12 Science and Art Centers throughout Turkey in the 2022-2023 academic year. This study utilized the "Anxiety Scale for Parents of Gifted Children," developed by the researchers in this study, which consisted of 30 items, to obtain quantitative data. A Personal Information Form was used for demographic information. The scale's validity was assessed through face, content, and construct validity, and its reliability was examined using Cronbach's alpha coefficient, item discrimination, and test-retest analyses. Expert opinion and a pilot study were conducted for face and content validity, whereas EFA and CFA were performed for construct validity. The data of 300 randomly selected participants (229 females and 71 males) were used for EFA, while the data of the remaining 250 participants (189 females and 61 males) were used for CFA. As a result of the EFA, eleven items were excluded, and CFA was conducted. The analysis yielded the following fit indices for the scale: RMSEA= .054, GFI= .912, CFI= .937, and IFI= .939. The Cronbach's alpha reliability coefficient was .890. Based on the analyses, a valid and reliable scale comprising 19 items and five factors was developed to assess the anxiety levels of parents with gifted children.

1. INTRODUCTION

Family is a critical environment in which individuals learn by observing their emotional reactions. The provision of a suitable emotional environment plays a crucial role in shaping one's perspective on life, achievement, and talent development. Family environments that provide emotional support are ideal for gifted children as well (Ataman, 2014; Bildiren, 2011). Parenting, in itself, can be a demanding and complex role, and being a parent of a gifted child can be both rewarding and exhausting (Köksal, 2020; Moon, 2004). The notion that parenting gifted children is easier and that families in such situations are luckier is a misconception (Avcı & Demirok, 2022). According to O̅urlu and Yaman's (2013) research, the number of parents

---

who perceive having a gifted child as an honorable experience is comparable to those who find it challenging and tiring.

Gifted children have different developmental processes and characteristics than their peers. Consequently, parents may encounter certain difficulties that increase parental stress (Deater-Deckard, 1998). Parents of gifted children experience higher levels of stress than parents of typically developing children (Bishop, 2012).

The parents of gifted children possess distinct attributes, requirements, and concerns. According to Moon and Hall (1998), family members who impose their lifestyle on their exceptionally gifted children in order to enhance their potential may create a stressor. This circumstance may even be described as a "burden" for the parent, as noted by Guthrie (2019). On the other hand, the challenges brought by these distinct needs are often misunderstood or misinterpreted by the surrounding environment, which may lead to parental anxiety (Porter, 2005, p. 208; Sutherland, 2008). Families with gifted children may encounter challenges such as unrealistic expectations of the environment (Aydın & Bu a, 2020), power struggles, sibling conflicts (Ben-Artzey, 2020), the risk of underachievement (Moon & Hall, 1998), parental feelings of inadequacy (Öztabak, 2018), being labeled as "gifted," the responsibility of balancing developmental areas, choosing a school that will develop your extraordinary gifted (Johnsen, 2017) and inadequate environmental support (Bishop, 2012).

Parents of gifted children are also concerned about finding an educational institution that provides appropriate educational conditions and programming (Bildiren, 2011; Moon & Hall, 1998; Sutherland, 2008). While parents have concerns about their gifted children's future, the thought of not being able to fully support their development also causes anxiety (Çamdeviren, 2014). Factors such as feeling inadequate to meet their child's needs, difficulties in studying, school selection, problems with teachers, relatives, societal expectations, boredom at school, and the process of starting primary school contribute to parental stress (Bildiren, 2011). Often, there is a sense of guilt about being the guardian of their child's talent and, at the same time, a fear of inadvertently suppressing it (McMann & Oliver, 1988). There are concerns about enabling children to demonstrate their potential (Clark 2013).

In addition to these concerns, another area of anxiety related to the social-emotional development of gifted children has emerged, involving issues such as making friends, socializing, and communication (Shichtman, 1999; Sutherland, 2008). Due to the asynchronous development of gifted individuals, they often struggle to communicate with their peers, and the problems they experience in their friendships also impact their relationships with their parents (Köksal, 2020). Parents experience the difficulty of not being able to help their gifted children cope with their challenging emotions (Renati et al., 2017).

Keirouz (1990) identified six domains that encompass the concerns of parents with gifted children. These domains are as follows: concerns related to parental roles, concerns related to sibling relationships, concerns related to the achievement of gifted children's talents, concerns related to parent-school issues, and concerns related to the social-emotional and cognitive development of gifted children.

The activation of gifted children's capacities, their ability to enjoy life, and becoming happy and healthy individuals is associated with the attitudes of parents. Therefore, similar to all other children, parents' contribution is important in raising gifted children (Karaku , 2010). All these challenges require parents of gifted children to possess various skills and strategies, and the resulting concerns and stress are reflected as anxiety for parents (Ero lu-Garip et al., 2022).

Several scales have been developed to assess parental anxiety. These include the State-Trait Anxiety Inventory (Spielberger et al., 1970), Parental Stress Scale (Berry & Jones, 1995), Generalized Anxiety Disorder Scale (GAD-7) (Spitzer et al., 2006), and Perceived Stress Scale (Cohen et al., 1994). The State-Trait Anxiety Inventory, developed by Spielberger et al. (1970),

was adapted to the Turkish culture by Öner and LeCompte (1998). The inventory considers anxiety in terms of two factors: state anxiety and trait anxiety. Anxiety was measured using 40 items. This scale has been used in many studies conducted in Turkey on adult and parental anxiety (Alisinano lu & Uluta 2003).

There are two measurement tools developed to measure parents' stress in Turkey. The first of these is the Mother and Father Stress scale developed by Özmen and Özmen (2012). The scale aims to measure the difficulties experienced by parents in their relationships with their children in daily life. It consists of 16 items and a single dimension. The second study is the Parenting Stress scale by Aydo an and Özbay (2017). The scale addresses the stress associated with being a parent in a single dimension. Considering the existing scales, it seems that there are a limited number of measurement tools to determine parenting stress. Additionally, existing scales do not include the concerns of parents of gifted children, who develop differently from their peers.

Considering the scales used in national and international studies on the anxiety levels of parents of gifted children, it is seen that scales developed for parents of typically developing children or those with special needs are used (e.g., Bishop, 2012; Eren et al. 2018; Francis, 2014). It seems that a scale measuring the parental concerns mentioned in the literature is needed. It is thought that the current study will fill this gap in the field.

According to five-factor personality theory, emotional balance and harmony constitute an important part of personality traits (Costa & McCrae, 2012). Parents' emotional balance and calmness contribute to a more positive development of children. However, parents' anxious attitudes may reduce life satisfaction (Dost et al., 2019). Parents' emotional inconsistencies may negatively affect young people's social competence (Bilgin, 2017). It is known that parental anxiety has a significant impact on anxiety and perfectionism of gifted children (Yazıcı, 2019). Therefore, developing a measurement tool to determine the concerns of parents of gifted children is important not only for the child's development but also for providing accurate guidance and consultancy support to parents and students. Creating a support group where parents of gifted children can face similar difficulties (Köksal, 2020), it is necessary to measure parents' concerns in order to develop support programs for these difficulties (Yıldız & Altay, 2021). By identifying parents' concerns, this measurement tool can help educators and mental health professionals to support these families more effectively. Furthermore, knowing the difficulties experienced by the family helps bridge the understanding gap between the school and family (Keirouz, 1990).

The objective of this study was to create a reliable and valid measurement tool for assessing the anxiety levels of parents of gifted children, utilizing the existing body of literature as a reference. To achieve this aim, efforts have been made to address the following research questions:

⎰ Is the scale developed to assess the parents' anxieties with gifted children valid?
⎰ Is the scale developed to assess the parents' anxieties with gifted children reliable?
⎰ What are the dimensions of a reliable and valid anxiety scale for parents with gifted children based on a solid theoretical foundation and empirical evidence?

2. METHOD

2.1. Sample

The study population consisted of parents of gifted children residing in various cities. The study group, representing these parents, was formed using a convenience sampling method by reaching out to the parents of Science and Art Centers (SACs) in various cities who volunteered to participate in this research. This sampling method was preferred because it facilitated data collection (Frankel & Wallen, 2008). Yoo and Moon (2006) emphasize that gifted individuals have different counseling needs depending on their age level. Since it is thought that the

concerns of parents of gifted children vary according to their children's developmental period, 550 parents of gifted children at the secondary school level voluntarily participated in the study. While the majority of the participants (76%) were mothers, the remaining 24% were fathers. Parents from Istanbul, Konya, Bursa, A rı, Kütahya, anlıurfa, Artvin, Malatya, Gaziantep, Adana, Denizli, and Manisa contributed to this study. Demographic characteristics of the participating parents and their demographic information are presented in Table 1.

Table 1. *Descriptive analysis of the demographic information of the participants.*

Variable	Options	*f*	%
Age	Between the ages of 31 and 40	253	46
	Between the ages of 41 and 50	278	50.5
	51 years and older	19	3.5
Grade level of the child	5th Grade	205	37.3
	6th Grade	162	29.5
	7th Grade	141	25.6
	8th Grade	42	7.6

As seen in Table 1, when the ages of the parents participating in the study were examined, more than half (50.5%) were between the ages of 41-50. Of the children, 37.3 were in the 5th grade, 29.5 percent in the 6th grade, 25.6 percent in the 7th grade and 7.6 percent in the 8th grade. When the educational level of the participating parents was examined, almost half (47.82%) had a licence degree and 22% had completed high school.

The data of 300 randomly selected participants (229 females and 71 males) were utilized for Exploratory Factor Analysis (EFA). The remaining 250 participants' data (189 females and 61 males) were used for Confirmatory Factor Analysis (CFA). Ho (2006) stated that the sample size for factor analysis should exceed five times the number of items. According to these standards, it can be seen that the sample of 550 individuals, which was used for this 30-item questionnaire in the current study, is sufficient.

2.2. The Development Process of the Data Collection Tool

The process of developing the anxiety scale for parents of gifted children consists of the following stages: identifying the problem situation, conducting a literature review related to the problem, item generation, obtaining expert opinions, conducting a pilot study, performing validity and reliability analysis, and presenting the final version of the scale.

For the literature review, previous scale studies on parental anxiety (Akkök, 1989; Berry & Jones, 1995; Kaner, 2001; Özmen & Özmen, 2012; Spielberger, 1970) and research on the anxieties of parents with gifted children were examined (Bildiren, 2011; Bishop, 2012; Clark, 2013; Çamdeviren, 2014; Porter, 2005; Sutherland, 2008;). In line with the studies examined in the literature, it has been observed that the concerns of parents with "normal" children and parents with specially talented children differ. This difference indicates the need for a scale (Bishop, 2012). The literature review revealed that parents might experience anxieties related to their gifted child's academic development, emotional development, social development, challenges in interaction, high expectations, and inadequacy of educational opportunities.

A 43-item questionnaire was developed to evaluate the parents' concerns with gifted children. The first section of the form included questions about parents' demographic information, while the second section consisted of items related to concerns grouped into five dimensions based on existing literature: concerns about academic development, concerns about emotional development, concerns about social development and interaction challenges, concerns arising from high expectations, and concerns related to educational opportunities and resources.

Special attention was attached to ensuring the clarity of the scale items and avoiding multiple judgments in their construction.

A total of 43 items were created and the opinions of five experts were sought, two from the field of Measurement and Evaluation for the scope and face validity of the items and three from the fields of Special Education and Psychological Counseling and Guidance. The draft form of the scale was sent to the experts as the "Expert Evaluation Form." In the evaluation form, experts expressed their opinions regarding each dimension and item as appropriate/not appropriate and provided item suggestions and correction recommendations in the case of not being appropriate. The content validity ratio (CVR) of the scale was calculated using Lawshe's (1975) technique. The content validity ratios (CVR) of each scale item presented to the expert opinion were calculated. The content validity index (CVI) of the scale was 0.92. This value meets the requirement proposed by Lynn (1968) to be greater than 0.83 and is considered sufficient. Based on the opinions of the experts and the content validity ratios, 12 items with similar content and limited ability to measure the specified characteristic were removed from the scale. Items that were not suitable for dimensions were placed under the relevant dimensions. Items with semantic errors and those measuring multiple skills were corrected, resulting in a trial form of the scale consisting of 31 items.

The pilot implementation phase was conducted with ten different parents using the trial form of the scale through interviews to assess the comprehensibility of the items. Based on the feedback received during the pilot implementation phase and the consensus among the researchers, one item was removed, resulting in the final version of the scale consisting of 30 items. During the pilot application phase, the participants read the scale items using the think-aloud technique. However, a single item is not sufficiently understood. The items were removed from the scale based on the consensus of researchers. The scale form was the final form of 30 items.

In the final stage, the scale was administered to parents of gifted children who received education at the middle school levels. The scale presented to parents is a 5-point Likert scale, with the addition of the option "I have no opinion." The parents' degree of agreement with the items in the scale is classified as follows: "I have no opinion" (0), "Strongly Disagree" (1), "Disagree" (2), "Neutral" (3), "Agree" (4), and "Strongly Agree" (5). Since there were no reverse-scored items in the scale, the obtained scores were summed to determine the level of parental anxiety. Since there are no reverse items on the scale, the parental anxiety level was determined by adding the scores obtained. The score range to be taken from the scale is between 0-95. High scores indicate high parental anxiety (Aydo an & Özbay, 2017; Çekiç et al., 2015).

2.3. Data Analysis

During the data analysis process, the SPSS 26.00 package program was used for EFA and the AMOS 24.00 package program for CFA. To test the construct validity and reliability of the scale, a sufficient number of data points is required (Akbulut, 2010). Comrey and Lee (1992) have stated that a sample size of 300 is good, while a sample size of 500 is very good. Ho (2006) emphasizes that the sample size should be at least five times the number of items in the scale. The scale was filled out by 550 participants in this study. In light of this information, it can be said that the sample size is appropriate for the validity and reliability study of the scale.

To assess the construct validity of the Anxiety Scale for Parents of Gifted Children (ASPGC), both EFA and CFA were conducted. The adequacy of the data for EFA was assessed using the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity. EFA examines the relationships between different items by grouping them into factors (Salkind, 2015).

To test the obtained structure using EFA, CFA was conducted. In CFA, model fit indices, such as $^2/df$, RMSEA, AGFI, CFI, GFI, NFI, IFI, and TLI, were considered to examine the model. The reliability analysis of the scale was assessed using Cronbach's alpha coefficient, item

discrimination index, and test-retest reliability. The discriminant reliability analysis of the scale was conducted using an independent samples t-test.

3. FINDINGS

In this section, the statistical procedures conducted within the scope of this study are presented, along with the findings obtained accordingly, following the order of the research questions. In this context, EFA, CFA, item-total correlations, comparisons of sub-upper groups, and test-retest analyses were performed. The first research question of this study was formulated as "Is the scale developed to assess the anxieties of parents with gifted children valid?" To determine whether this scale, which describes parental anxiety, meets the criteria of validity and reliability, separate sections entitled "Findings regarding the Validity of the Scale" and "Findings regarding the Reliability of the Scale" were presented.

3.1. Findings Regarding the Validity

In light of the first research question, both exploratory and confirmatory factor analyses were conducted to test the construct validity of the scale. First, the findings related to EFA are presented. Subsequently, the findings regarding CFA were reported.

3.1.1. *Findings of exploratory factor analysis*

To conduct the EFA, the suitability of the data was tested. Frankel & Wallen (2008) stated that the data could be appropriate for factor analysis if the Kaiser-Meyer-Olkin (KMO) value is greater than .60 and if Bartlett's test of sphericity yields a significant result ($p=0.00<.05$). According to the analysis conducted, the KMO value was.918, and Bartlett's test of sphericity yielded a significant result 4335.166 ($p=.000$). Hutcheson and Sofroniou (2006) state that KMO values between .8 and .9 indicate great suitability and values above 0.90 indicate the best fit. These results indicated that the data were suitable for the analysis.

Exploratory factor analysis is used to preserve the total variance contained in the measured variables and turn it into a component containing fewer variables (Park et al., 2002). It is recommended that these values should be higher than .40 (Field, 2009). The common variance values of the items in the scale vary between .41 and .67. Accordingly, the common variance values of the items were deemed to be appropriate for the scale.

Figure 1. *Line graph of eigenvalues of scale items.*
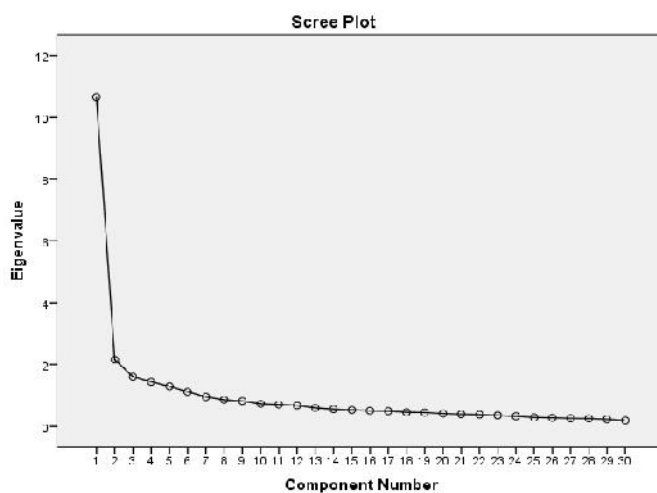
As shown in Figure 1, there are five breakpoints, indicating that the number of factors is limited to five. The eigenvalues and variance percentages of the five-factor scale are presented in Table 2.

Table 2. *Eigenvalues and variance percentages of items in the scale.*

Factor	Eigenvalue	% of variance	% of total variance
Factor 1	6.801	35.797	35.797
Factor 2	1.573	8.277	44.074
Factor 3	1.432	7.539	51.613
Factor 4	1.216	6.401	58.014
Factor 5	1.005	5.292	63.306

The 1st factor contributed 35.79% (eigenvalue= 6.801) of the total variance, the 2nd factor contributed 8.27% (eigenvalue=1.573), the 3rd factor contributed 7.53% (eigenvalue=1.432), the 4th factor contributed 6.401% (eigenvalue=1.226), and the 5th factor contributed 5.29% (eigenvalue=1.005) of the total variance. The value of the 1st factor was approximately five times greater than that of the 2nd factor, and this significant decrease was observed, as shown in Figure 1.

As shown in Figure 1, it can be observed that the line graph continues horizontally after the 5th factor, and there were no significant decreases between the factors. The contribution of these factors to the total variance decreased. The scale was constructed with a five-factor structure, accounting for 63.30% of the scale. Akbulut (2010) states that explained variance should be higher than unexplained variance. Therefore, it can be said that the variance explained by the scale is appropriate.

In exploratory factor analysis, the oblique rotation (direct oblimin) method was used to better explain the structure of the factor. This method is preferred when factors are interrelated (Tabachnick & Fidell, 2007).

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	5
Anxiety related to emotional development	I 22	.756				
	I 15	.724				
	I 12	.671				
	I 19	.554				
Anxiety related to difficulties in social development and interaction	I 18		.760			
	I 17		.747			
	I 24		.698			
	I 21		.566			
Anxiety related to educational opportunities and possibilities	I 26			-.862		
	I 25			-.844		
	I 30			-.677		
	I 9			**-.494**	.331	
Anxiety related to academic development	I 1				.814	
	I 2				.657	
	I 3				.645	
	I 5				.629	
Anxiety related to high expectations	I 11					-.841
	I 10					-.714
	I 13	.320				**-.638**
Eigenvalues		6.801	1.573	1.432	1.216	1.005
Explained variance		%35.797	%8.277	%7.539	%6.401	%5.292
Total variance		63.306				

Table 3 presents the loadings of items falling under these factors as a result of oblique rotation. The pattern matrix reflected the strength of each item within a given factor. According to Field (2009), the loadings of items in the pattern matrix should be higher than .40. In light of these findings, the factor loadings in the pattern matrix appear to be appropriate. The total explained variance should be 50% or higher (Thompson, 2004). According to the findings in Table 3, it can be said that the factor loadings in the pattern matrix and the explained variance are appropriate. The developed parental anxiety scale was grouped under five factors with eigenvalues greater than 1, and explained 63.30% of the total variance.

The first factor consisted of four items (12, 15, 19, and 22), and the loading values within the factor ranged from .554 to .756. This factor was labeled as "anxiety related to emotional development." The second factor, composed of four items (17, 18, 21, and 24), exhibits loading values ranging from .566 to .760, indicating "anxiety related to difficulties in social development and interaction." The third factor comprised four items (9, 25, 26, and 30), with loading values ranging from .494 to .862. These items pertain to "anxiety related to opportunities and possibilities in education." The fourth factor, consisting of four items (1, 2, 3, and 5), displayed loading values ranging from .629 to .814. It was labeled as "anxiety related to academic development." The fifth factor, composed of 3 items (10, 11, and 13), has loading values ranging from .638 to .841, representing "anxiety related to high expectations." (Table 4).

Table 4. *Sample items of the scale.*

Parental concerns scale expressions for having a gifted child	Rotated factor loadings
Anxiety related to emotional development	
19 It concerns me when my child experiences intense emotions.	.554
Anxiety related to difficulties in social development and interaction	
21 It concerns me if my child does not prefer to cooperate with their peers when necessary.	.566
Anxiety related to educational opportunities and possibilities	
30 I worry that my child's potential was not notice during the education process.	-.677
Anxiety related to academic development	
1 It concerns me if my child does not make the necessary effort thinking that she/he will be successful without studying.	.814
Anxiety related to high expectations	
13 It concerns me that teachers set high expectations for my child.	-.638

In summary, EFA was performed with an oblique rotation method on 30 items. Items with eigenvalues greater than one were grouped under five factors, which accounted for 63.30% of the scale's variance. When determining which items should be removed from the scale, a minimum factor loading of .30 was considered. Items were not allowed to have significant loadings on multiple factors (if an item loaded on two factors, the difference between the loadings should be at least .100). Furthermore, attention was paid to factors consisting of three or more items. Consequently, items 6th, 29th, 7th, 8th and 23rd items were removed due to overlap; the 16th, 14th and 27th items had factor loadings below .40, and items 4th, 20th and 28th items were removed as their content did not match the respective factor. After conducting another EFA with 19 items, a five-factor structure was obtained, explaining 63.30% of the scale's variance. The factors were named based on information from the literature (Bildiren, 2011; Bishop, 2012; Clark, 2013; Eren et. al, 2018; Eroǧlu-Garip et al., 2022; Keirouz, 1990; Öztabak, 2018; Renati et al., 2016).

3.1.2. *Confirmatory factor analysis findings*

The anxiety scale, previously identified as having five dimensions in the EFA, was tested using CFA with a different sample of 250 participants. Based on the analysis results, the goodness-of-fit index was examined and covariance was depicted between Items 1 and 5 and Items 26 and 25 in the model presented in Figure 2.

Figure 2. *The model for Confirmatory Factor Analysis of the scale items.*
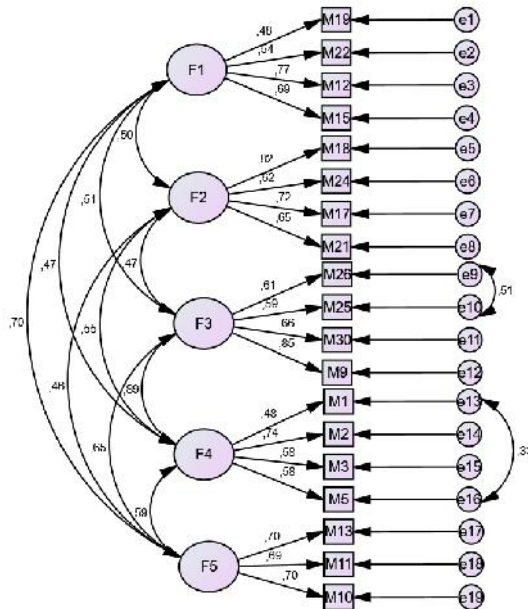
Table 5. *The Fit Values obtained in CFA.*

N	χ^2	df	χ^2/df	RMSEA	AGFI	AIC	NFI	GFI	CFI	IFI	TLI
250	242.705	140	1.734	.054	.880	342.705	.866	.912	.937	.939	.923

A structural equation model was constructed using the AMOS 24.00 software. When examining the goodness-of-fit index of the five-factor model, it can be observed that χ^2/df = 1.734; RMSEA=0.054; AGFI= 0.880; GFI= 0.912; NFI= 0.866; CFI= 0.937; IFI= 0.939; TLI= 0.923 (Table 5). In the literature, a χ^2/df value of 2.5 and below is considered a perfect fit, while GFI, NFI, CFI, IFI, and TLI values are greater than .90, AGFI value is greater than .85, and RMSEA value below .08 is considered a good fit (Brown, 2006; Byrne, 2016; Hooper et al., 2008; Hu & Bentler, 1999; Kline, 2010; Schermelleh-Engel et al., 2003;).

According to the goodness-of-fit values in Table 5, it can be stated that the NFI values are slightly below the acceptable threshold, while RMSEA, AGFI, GFI, CFI, IFI, and TLI fit indices indicate an acceptable fit. Additionally, the observed fit values for χ^2/df indicate an excellent fit. It can be concluded that the model created using the obtained data exhibits a good fit.

3.2. Findings Regarding the Reliability of the Scale

In light of the second research question of this study, the reliability of the scale's items and factors was analyzed using Cronbach's alpha coefficient, test-retest method, item-total score correlation, and item discriminability method by comparing the 27% sub-upper groups. In terms of the reliability of the scale, Cronbach's alpha coefficient value should be equal to or greater than .70 (Frankel & Wallen, 2008).

Table 6. *Reliability coefficients of the Factors.*

Factors	Cronbach's Alpha ( )	Number of Items
Anxiety related to emotional development	.734	4
Anxiety related to difficulties in social development and interaction	.807	4
Anxiety related to educational opportunities and possibilities	.811	4
Anxiety related to academic development	.719	4
Anxiety related to high expectations	.751	3
Total	.890	19

Based on the analysis results, the reliability coefficients for the factors were calculated as follows: 1st Factor (anxiety related to emotional development) .734; 2nd Factor (anxiety related to difficulties in social development and interaction) .807; 3rd Factor (anxiety related to educational opportunities and possibilities) .811; 4th Factor (anxiety related to academic development) .719; 5th Factor (anxiety related to high expectations) .751. The Cronbach's alpha coefficient for the overall scale was .890 (Table 6). The reliability coefficient of the scale and its factors being above .70 indicates that it is a reliable measurement tool.

Table 7. *Results of item analysis.*

Factors	Item No	Item-Total Correlation	*t* upper %27- lower%27
Anxiety related to emotional development	19	.443	8.99*
	22	433	10.97*
	12	.566	14.03*
	15	.490	10.94*
Anxiety related to difficulties in social development and interaction	18	.570	12.33*
	24	.511	10.48*
	17	.513	11.53*
	21	.519	11.01*
Anxiety related to educational opportunities and possibilities	26	.537	10.70*
	25	.557	10.99*
	30	.560	10.78*
	9	.597	13.17*
Anxiety related to academic development	1	.470	10.26*
	2	.538	11.68*
	3	.434	10.53*
	5	.482	10.06*
Anxiety related to high expectations	13	.539	12.43*
	11	.496	12.82*
	10	.542	12.12*

*$p< .001$

As shown in Table 7, it can be observed that the total correlation values of the items in the scale range from .433 to .597, and the *t*-values were significant ($p<.001$). A total item correlation value of .30 or higher indicates a high internal consistency of the scale items (Frankel & Wallen,

2008). This suggests that the scale items sample the same behaviors, and that the measurement tool differentiates the anxieties of parents with gifted children.

3.2.1. *Item discrimination analysis*

An analysis was conducted using an independent sample t-test to examine the item discrimination property of the scale. A comparison was made between the anxiety scores of the 27% sub-upper groups (Table 8).

Table 8. *Item discrimination analysis.*

Measure	N	x	Std deviation	df	t	p
1 (upper %27)	149	81.04	5.40	296	31.18	0.00
2 (lower %27)	149	49.75	8.57			

The analysis results indicated a statistically significant difference between the 27% sub-upper groups (t= 31.18, p= 0.00, Table 8). The developed scale is capable of distinguishing the upper and lower groups from each other significantly. It can be said that the anxiety scale for parents of gifted children has high distinctiveness power.

3.2.2 *Test-retest method*

The Parental Anxiety Scale was re-administered to the same parents (n= 50) within a specific time interval. In the second application 50 participants responded. Bonett and Wright (2014) suggest that 30 samples may be sufficient to measure reliability if there is a strong relationship between the scale items. The test-retest reliability of the scale was calculated using the Pearson correlation coefficient. The correlation coefficients were calculated as .724 for Factor 1, .685 for Factor 2, .590 for Factor 3, .677 for Factor 4, .571 for Factor 5, and .775 for the overall scale (p<.01). In the literature, it has been indicated that the correlation coefficient should be at least .70 for the scale to demonstrate stability (Özdamar, 2004, as cited in Akbulut, 2010). Thus, according to the test-retest method, the Parental Anxiety Scale has exhibited the characteristics of a reliable measurement tool.

In line with the conducted analysis studies, it has been concluded that the anxiety scale for parents of gifted children is a valid and reliable measurement tool. The subdimensions of the scale, as indicated in Figure 3, align with the third research question of this study.

Figure 3. *Subdimensions of anxiety scale for parents of gifted child.*
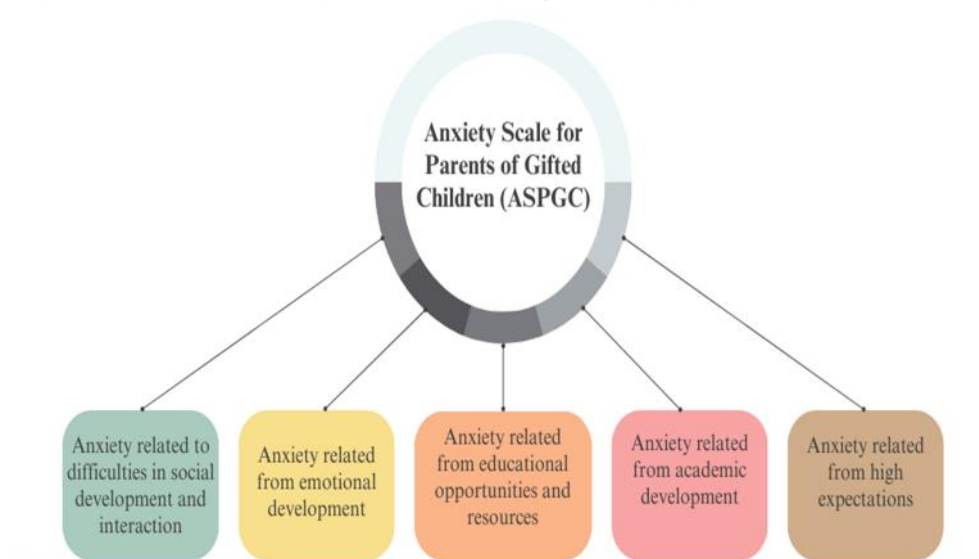
4. DISCUSSION and CONCLUSION

This study aimed to create a valid and reliable measurement tool to measure the anxiety levels of parents of gifted children. A validity and reliability study was conducted for the developed scale with the participation of 550 parents of gifted children. The 43 items in the draft of the scale were evaluated in light of the opinions of five experts, and on this basis, 12 items were removed from the scale. As a result of this process, the content validity index of the scale was found to be .92, which indicates high content validity (Lynn, 1968) and that the scale items cover issues related to the anxiety of parents with gifted children. A pilot study was conducted with 10 parents on the 31-item trial form, and based on feedback, another item that was unclear and vague was removed.

To test the construct validity of the scale, both EFA and CFA were conducted. The KMO value was .918, indicating that the sample size was adequate for EFA. Additionally, Bartlett's test of sphericity yielded a significant result, further confirming the suitability of the sample for factor analysis. Based on the oblique rotation analysis, a five-factor structure emerged. Care was taken to ensure that each factor consisted of at least two items. These factors were labeled as anxiety related to difficulties in social development and interaction, anxiety related to emotional development, anxiety related to educational opportunities and resources, anxiety related to academic development, and anxiety related to high expectations.

The EFA results indicated that the five factors accounted for 63.30% of the scale's variance. Consequently, items 6th, 29th, 7th, 8th and 23rd items were removed due to overlap; the 16th, 14th and 27th items had factor loadings below .40, and items 4th, 20th and 28th items were removed as their content did not match the respective factor.

To confirm the five-factor model, CFA was conducted. The fit indices of the five-factor model were as follows: χ^2/df =1.734; RMSEA = 0.054; AGFI = 0.880; GFI = 0.912; NFI = 0.866; CFI = 0.937; IFI = 0.939; TLI = 0.923. These values suggest a good fit for the model. The reliability coefficient of the overall scale was calculated as .890, indicating satisfactory internal consistency. The results of the internal consistency analysis for each subscale were as follows: first factor= .734 second factor= .807, third factor= .811, fourth factor= .719, fifth factor= .751. These values reflect acceptable levels of internal consistency for each factor.

Through the conducted analyses, the final version of the scale consisted of 19 items and five dimensions. These dimensions align with the frequently expressed anxieties and challenges reported by parents of gifted children in the literature (Bildiren, 2011; Keirouz, 1990; Moon & Hall, 1998; Shichtman, 1999; Sutherland, 2008; Sutherland, 2008). In a qualitative study conducted by Renati et al. (2016), the researchers examined the anxiety factors experienced by 49 parents of gifted children. The study identified three main categories of parental concerns: individual-related problems in meeting the needs of gifted children, family related problems, and social difficulties (e.g., lack of peer and school support). According to Ihlamur (2017), parents struggle with the characteristics of gifted children. Öztabak (2018) identified the concerns experienced by parents of gifted children in various dimensions, including family dynamics, sibling relationships, environmental factors, and developmental characteristics of gifted children. Families may require additional time and resources to meet the needs of gifted children. Additionally, parents may face challenges related to gifted children (Alberta Learning, 2004). In a study addressing the difficulties of parents with gifted children, Karakuş (2010) found that parents expressed challenges related to teachers with high expectations and lack of knowledge about gifted education; difficulties in social interactions of gifted children with peers, environment, and siblings; reluctance to do homework; failure to develop study habits; and parents' inability to meet the needs of children regarding their areas of interest. Akarsu (2004) described the challenges experienced by parents of gifted children in four categories: dealing with the family's interaction with the environment and school, coping with their gifted child, and managing differences within the family. In a study by Dalgıç (2017) investigating

the challenges faced by parents of gifted children aged 3-6, the findings showed that parents primarily encountered problems related to the social-emotional characteristics of gifted individuals and inadequacies in formal educational institutions. Studies have shown that the concerns of gifted children's parents are addressed in different dimensions or in a single dimension. These studies showed that the scale is consistent and comprehensive with its items and dimensions, as can be seen from the analysis results.

The difficulties experienced by parents regarding their gifted children can affect their anxiety levels; having anxious parents increases the anxiety of gifted children (Yazıcı, 2019). Considering the potential effects of parental anxiety on gifted children, the scale developed can test the effectiveness of family education studies for parents with high anxiety. Additionally, it may enable the preparation of comprehensive intervention programs that focus on parental anxiety. Accurately measuring parental concerns can also provide practical results in the psychological counseling process.

This study developed a scale to determine the anxiety levels of parents of gifted children studying at the secondary school level. The usability of the scale is based on the results of validity and reliability analyses, but it has some limitations. One of these limitations is that the participants who completed the scale had different socioeconomic levels and educational backgrounds. Another limitation is the difficulty in reaching parents who have gifted children.

It may be recommended to use different measurement tools to diversify the data sources in counseling and guidance activities for parents. In future studies, the current scale may be used to test the effectiveness of family education studies aimed at reducing anxiety experienced by parents of gifted children, developing competent parenting skills, and increasing family relationships.

Acknowledgments

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 University-Cerrahpasa Social and Humanities Research Ethics Committee ethical approval number 2023/4 dated 10.01.2023.

Contribution of Authors

Kübra Yiğit Tekel: Conception, Desing, Funding, Materials, Data Collection and Processing, Analysis and Interpretation, Literature Review, Writing. **Melike Şule Yıldız:** Conception, Desing, Funding, Materials, Data Collection and Processing, Analysis and Interpretation, Literature Review. **Duygu Mutlu Bayraktar:** Desing, Supervision, Materials, Data Collection and Processing, Analysis and Interpretation. **Marilena Zinovia Leana Taşcılar:** Conception, Supervision, Literature Review, Critical Review.

Orcid

Kübra Yi it Tekel https://orcid.org/0000-0002-5456-3568
Melike ule Yıldız https://orcid.org/0000-0002-7901-1081
Duygu Mutlu Bayraktar https://orcid.org/0000-0002-2276-3768
Marilena Zinovia Leana Ta cılar https://orcid.org/0000-0002-1271-0371

REFERENCES

Akarsu, F. (2004). *Üstün yetenekli çocukların ailelerinin sorunları [Problems of families of gifted children].* In M.R. irin, A. Kulaksızo lu, & A.E. Bilgili (Eds.), I. Türkiye Üstün Yetenekli Çocuklar Kongresi Seçilmi Makaleler Kitabı [1st Turkish Gifted Children Congress Selected Articles Book]. 447-460. Children's Foundation.

Akbulut, Y. (2010). *Sosyal bilimlerde SPSS uygulamaları [SPSS applications in social sciences].* Ideal Kültür.

Alisinano lu, F., & Uluta , . (2003). Çocukların kaygı düzeyleri ile annelerinin kaygı düzeyleri arasındaki ili kinin incelenmesi [A study on the relationship between children's anxiety levels and their mother's anxiety levels]. *Education and Science, 28*(128), 65-71.

Alberta Learning (2004). *A handbook for parents of children who are gifted and talented.* Alberta Learning.

Avcı, G., & Sakallı-Demirok, M. (2022). Özel yetenekli olmak ve özel yetenekli çocu a sahip olmak: Özel yetenekli birey ve ebeveyn görü leri [Being talented and having a special talented child: Opinions of special talented individuals and parents]. *Journal of Interdisciplinary Educational Research*, *6*(11), 39-54.

Aydın, Y.E., & Bu a, A. (2020). Özel yetenekli çocukların ebeveynlerinin akılcı olmayan inançları ve ebeveyn yetkinliklerinin incelenmesi [Investigation of the parents' irrational beliefs and parental self-efficacy of parents of gifted children]. *OPUS International Journal of Society Researches*, *15*(22), 1373-1407. https://doi.org/10.2646 6/opus.640449

Aydo an, D., & Özbay, Y. (2017). Ebeveynlik stres ölçe i geçerlik ve güvenirlik çalı ması [The parenting stress scale: The validity and the reliability study]. *International Journal of Early Childhood Education Studies, 2*(2), 24-38. http://ijeces.hku.edu.tr/tr/download/ article-file/359422

Berry, J.O., & Jones, W.H. (1995). Parental stress scale. Initial psychometric evidence. *Journal of Social and Personal Relationships*, *12*(3), 463-472. https://doi.org/10.1177/02654075 95123009

Bildiren, A. (2011). *Üstün yetenekli çocuklar aileler ve ö retmenler için bir kılavuz [A guide for gifted children, families and teachers].* Dogan.

Bilgin, M. (2017). Ergenin be faktör ki ilik özellikleri ile sosyal yetkinlik, bili sel hatalar ve anne baba be faktör ki ilik özellikleri ili kisi [The relationship between five personality traits of adolescent and social self-efficacy, cognitive errors and five personality traits of parents]. *Gaziantep University Journal of Social Sciences, 16*(1), 230-250. http://doi. org/10.21547/jss.284859

Bishop, B. (2012) *Parenting stress in mothers of gifted and typical children* [Master's thesis, University of Alberta]. Educational & Research Archieve.

Bonett, D.G., & Wright, T.A. (2014). Cronbach's alpha reliability: Interval estimation, hypothesistesting, and sample size planning. *Journal of Organizational Behavior*, 36(1), 3-15. https://doi.org/10.1002/job.1960

Brown, T.A. (2006). *Confirmatory factor analysis: For applied research*. Guilford Press.

Bryne, M.B. (2016). *Structural equation modeling with AMOS. Basic concepts, applications and programming.* (3th ed.). Routledge

Clark, B. (2013). *Üstün zekalı olarak büyümek evde ve okulda çocukların potansiyellerini geli tirmek [Growing up gifted: Developing the potential of children at school and at home].* (Ü. O urlu, & F. Kaya, Trans.). Nobel.

Cohen, S., Kamarck, T., & Mermelstein, R. (1994). Perceived stress scale. *Measuring Stress: A Guide for Health and Social Scientists*, *10*(2), 1-2.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education* (5th ed.). Routledge.

Comrey, A.L., & Lee, H.B. (1992). *A first course in factor analysis* (2nd ed.). Erlbaum.

Costa, P.T., & McCrae, R.R. (2012). The five-factor model, five-factor theory, and interpersonal psychology. *Handbook of Interpersonal Psychology*, 91-104. https://doi.org/10.1002/9781118001868.ch6

Çamdeviren, . (2014). *Bilim ve sanat merkezine (B LSEM) devam eden üstün yetenekli çocukların anne babalarının kar ıla tıkları güçlükler: Sakarya ili örne i. [The difficulties that parents of the gifted children attending Science and Arts Center encounter]* (Publication no. 353104) [Master's thesis, Sakarya University]. Council of Higher Educational Thesis Center.

Çekiç, A., Akba , T., & Hamamcı, Z. (2015). Anne Baba Stres Ölçe i'nin Türkçe'ye uyarlaması: geçerlik ve güvenirlik çalı ması [Adaptation of Parenting Stress Index into Turkish: The validity and the reliability study]. *Gaziantep University Journal of Social Sciences, 14*(3), 647-667. https://doi.org/10.21547/jss.256755

Dalgıç, N. . (2017). *Üstün yetenekli veya zekâlı çocu a sahip anne-babaların kar ıla tıkları sorunların ve beklentilerinin belirlenmesi [Determining the problems and expectations faced by parents of gifted or intelligent children]* [Master's thesis, Eastern Mediterranean University]. Eastern Mediterranean University Libraries.

Deater-Deckard, K. (1998). Parenting stress and child adjustment: Some old hypotheses and new questions. *Clinical Psychology: Science and Practice, 5*(3), 314-332. https://doi.org/10.1111/j.1468-2850.1998.tb00152.x

Deniz, M.E., Dilmaç, B., & Arıcak, O.T. (2009). Engelli çocu a sahip olan ebeveynlerin durumluk-sürekli kaygı ve ya am doyumlarının incelenmesi [An analysis of life satisfaction and state-trait anxiety of the parents with handicapped children]. *International Journal of Human Sciences*, *6*(1), 953-968. https://www.j-humansciences.com/ojs/index.php/IJHS/article/view/727/389

Dettman, D.F., & Colangelo, N. (1980). A functional model for counseling parents of gifted children. *Gifted Child Quarterly*, *24*(4), 158-161. https://doi.org/10.1177/0016986280024004

Dost, M.T., Aytaç, M., & Uysal, N.K. (2019). Ebeveynlik biçiminin ki ilik özellikleri, benlik saygisi ve ya am doyumunu yordama gücü [Predictive power of parenting styles on personality traits, selfesteem, and life satisfation]. *Electronic Journal of Social Sciences*, *18*(71), 1146-1165. https://doi.org/10.17755/esosder.472964

Eren, F., Ömerelli-Çete, A., Avcıl, S., & Baykara, B. (2018). Üstün yetenekli çocuklarda ve ailelerinde duygusal ve davranı sal özellikler [Emotional and behavioral characteristics of gifted children and their families]. *Archives of Neuropsychiatry, 55*(2), 105-112. https://doi.org/10.5152/npa.2017.12731

Ero lu-Garip, S., Karada , F., & Yıldız-Demirta , V. (2022). Özel yetenekli çocukların ebeveyni olmak ve erken çocuklukta özel yetenekli olmak: Ebeveyn görü leri [Parenting of gifted children and being gifted in early childhood: Parental views]. *Trakya University Journal of Social Science*, *24*(1), 197-226. https://doi.org/10.26468/trakyasobed.885478

Field, A. (2009). *Discovering statistics using SPSS*. SAGE.

Francis, S.E. (2014). The role of parental anxiety sensitivity in parent reports of child anxiety in treatment seeking families. *Clinical Child Psychology and Psychiatry*, *19*(1), 111-124. https://doi.org/10.1177/1359104512470055

Frankel, J.R., & Wallen, N.E. (2008*). How to design and evaluate research in education.* McGraw Hill Higher Education.

Guthrie, K.H. (2019). "Nothing is ever easy": Parent perceptions of intensity in their gifted adolescent children. *The Qualitative Report, 24*(8), 2080-2101. https://doi.org/10.46743/2160-3715/2019.3598

Hair, J.F.J., Black, W.C., Babin, B.J., & Anderson, R.E. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning.

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

Ho, R. (2006). *Handbook of univariate and multivariate data analysis and interpretation with SPSS*. Chapman and Hall/CRC.

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.

Ihlamur, . (2017). *Üstün yetenekli çocuklara sahip ailelerin danı manlık ihtiyaçlarının saptanması [Identification of the requirements for counseling for the families with gifted and talented children]* (Publication no. 472165) [Master thesis, stanbul Ticaret University]. Council of Higher Educational Thesis Center.

Johnsen, S.K. (2017). School choice and gifted education. *Gifted Child Today, 40*(2), 69-69. https://doi.org/10.1177/1076217517691194

Kaner, S. (2001). Aile stresini de erlendirme ölçe i (The questionnaire on resource and stress-F) faktör yapısı, güvenirlik ve geçerlik çalı maları [Factor structure, reliability and validity studies of the Family Stress Assessment Scale]. *Ankara University Journal of Faculty of Educational Sciences, 34*(1), 1-10. https://doi.org/10.1501/Egifak_00000000 39

Karaku , F. (2010). Üstün yetenekli çocukların anne babalarının kar ıla tıkları güçlükler [Difficulties that families of gifted students face]. *Mersin University Journal of The Faculty of Education*, *6*(1), 127-144. https://doi.org/10.17860/efd.89448

Keirouz, K.S. (1990). Concerns of parents of gifted children: A research review. *Gifted Child Quarterly*, *34*(2), 56-63. https://doi.org/10.1177/001698629003400202

Kline, R.B. (2010). *Principles and practice of structural equation modeling*. Guilford Press

Koca, A. (2017). *Engelli çocu a sahip olan ve engelli çocu a sahip olmayan annelerin ölüm kaygısı ve durumluk sürekli kaygı düzeyleri arasındaki farklıla manın incelenmesi [Investigation of the differences between death anxiety and state-trait anxiety levels of mothers with a disabled child and without a disabled child]* (Publication no. 486225) [Master thesis, Hasan Kalyoncu University]. Council of Higher Educational Thesis Center.

Köksal, A. (2020). Özel yetenekli bir çocu un ebeveyni olmak [Being a parent of a gifted child]. *Çocuk ve Medeniyet*, *5*(10), 495-502. https://doi.org/10.47646/CMD.2020.220

Lawshe, C.H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563-575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Lynn, M.R. (1986). Determination and quantification of content validity. *Nursing Research*, *35*(6), 382-386. https://doi.org/10.1097/00006199-198611000-00017

McMann, N., & Oliver, R. (1988). Problems in families with gifted children: Implications for counselors. *Journal of Counseling and Development*, *66*(6), 275-278. https://doi.org/10.1002/j.1556-6676.1988.tb00867.x

Moon, S.M., & Hall, A.S. (1998). Family therapy with intellectually and creatively gifted children. *Journal of Marital and Family Therapy*, *24*(1), 59-80. https://doi.org/10.1111/j.1752-0606.1998.tb01063.x

Moon, S.M. (Ed.). (2004). *Social/emotional issues, underachievement, and counseling of gifted and talented students.* Corwin Press.

O urlu U., & Yaman Y. (2013). Ebeveynlerin üstün zekalı ve yetenekli çocuklarıyla ilgili rehberlik ihtiyaçları [Guidance needs of gifted and talented children's parents]. *Turkish Journal of Giftedness and Education*, *3*(2), 81-94.

Öner, N., & Le Compte, A. (1983). *Durumluk-Sürekli Kaygı Envanteri El Kitabı [Trait-trait anxiety inventory manual].* Bo aziçi University Press.

Özmen, S.K., & Özmen, A. (2012). Anne Baba Stres Ölçe i'nin geli tirilmesi [Parent Stress Scale development]. *Milli E itim Dergisi, 42*(196), 20-35.

Öztabak, M.Ü. (2018). Özel/üstün yetenekli çocukların ailelerine yönelik rehberlik ve psikolojik danı manlık [Guidance and psychological counseling for gifted/talented children's parent]. *Journal of Educational and Humanities*: *Theory and Practice*, *9*(17), 79-106.

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

Renati, R., Bonfiglio, N.S., & Pfeiffer, S. (2016). Challenges raising a gifted child: Stress and resilience factors within the family. *Gifted Education International, 33*(2), 145-162. https://doi.org/10.1177/0261429416650948

Salkind, N. (2015). *statistikten nefret edenler için istatistik* [Statistics for people who (think they) hate statistics]. (A. Çuhadaro lu, Trans.). Pegem.

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.

Shichtman, D. (1999). *Concerns and problems of parents of academically successful gifted children* (Publication No: 9939553) [Doctoral dissertation, Columbia University]. ProQuest Dissertation & Theses Global.

Spielberger, C.D., Gorsuch, R.C., & Luschene. R.E. (1970). *Manual for the State-Trait Anxiety Inventory.* Consulting Psychologists Press.

Spitzer, R.L., Kroenke, K., Williams, J.B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, *166*(10), 1092-1097. https://doi.org/10.1001/archinte.166.10.1092

Sutherland, M. (2008). *Developing the gifted and talented young learner.* Sage Publication.

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

Webb, J.T. (1993). Nurturing social-emotional development of gifted children. In K.A. Heller, F.J. Mönks, & A.H. Passow (Eds.), *International handbook of research and development of giftedness and talent* (pp. 525–538). Pergamon Press.

Yazıcı, E. (2019). *Level and sources of anxiety affecting gifted students who attend science and art centers* (Publication no. 588386) [Master's thesis, Bilkent University]. Council of Higher Educational Thesis Center.

Yıldız, S., & Altay, N. (2021). The parenting attitudes and effects on their gifted children: A literature review. *Journal for the Education of Gifted Young Scientists, 9*(2), 123-132. http://dx.doi.org/10.17478/jegys.864037

Yoo, J.E., & Moon, S.M. (2006). Counseling needs of gifted students: An analysis of intake forms at a university-based counseling center. *Gifted Child Quarterly*, *50*(1), 52-61. https://doi.org/10.1177/001698620605000106

Research Article

Development of a knowledge test about scientists for secondary school students

Pelin Yıldırım[1*], **Gonca Keçeci**[2], **Fikriye Kırbağ Zengin**[2]

[1]Fırat University, Institute of Education Sciences, Elazığ, Türkiye
[2]Fırat University, Faculty of Education, Department of Mathematics and Science Education, Elazığ, Türkiye

Abstract: This study aims to develop a knowledge test to assess secondary school students' awareness of scientists. After determining the topics that the questions in the test would measure, 50 multiple-choice questions were prepared and presented to six experts in the field. Following the experts' feedback, the number of questions in the test was reduced to 44 due to issues related to question clarity, alignment with the curriculum, and cognitive appropriateness. The 44-question test underwent item analysis. The prepared test was subjected to a preliminary pilot application with 222 secondary school students. As a result of the preliminary pilot application, two questions with discrimination indices of 0.19 or lower were removed, and the final version of the test was arranged to consist of 42 questions for the pilot application. The pilot application was conducted with 211 secondary school students. In this application, secondary school students were administered the Knowledge Test of Scientists consisting of 42 questions, and item analysis was performed again. Two questions with discrimination indices of 0.19 or lower were removed from the pilot application. As a result of the pilot application, the Knowledge Test of Scientists had a KR-20 reliability coefficient of 0.94, and the Spearman Brown two-half test correlation coefficient was calculated as 0.92. The final version of the scale was arranged to consist of 40 items. These results demonstrate that the test is highly reliable, meaning that the test results accurately reflect the measured attribute and are replicable.

1. INTRODUCTION

Education is a process aimed at changing individuals' behaviors in the desired direction (Ertürk, 1972). This process encompasses various goals such as acquiring new knowledge, correcting erroneous behaviors, or altering undesirable behaviors. To achieve the intended changes in education, careful planning and control mechanisms are required (Turgut & Baykul, 2012).

The achievements and challenges of the education system play a significant role in improving this system. To overcome the issues encountered in education and to make future educational activities more effective, it is important to assess educational processes and analyze their outcomes (Ülger, 2021). Therefore, the assessment process in education holds great importance. Assessment is a process of gathering information, analyzing it, and making decisions based on criteria. These decisions are used to enhance educational processes and improve student achievement (Turgut & Baykul, 2012).

Education plays a crucial role in contributing significantly to the development of society, and decisions made in this context typically revolve around student success. However, making accurate and fair decisions becomes challenging without having a solid foundation in student achievement. Assessing student success forms the basis of educational decisions, and the reliability and validity of these assessments are of critical importance (Sarıbaş & Babadağ, 2015; Turgut & Baykul, 2012). Alongside these assessments, evaluations based on the appropriateness of criteria are also essential. Unfortunately, there is often a lack of awareness and understanding about measurement and evaluation in our country. This deficiency can lead to complications in measurement results, criteria, and evaluation outcomes, resulting in incorrect decisions. Decisions may be based on unreliable measurements, valid criteria may be chosen or misinterpreted, and in some cases, decisions may be made without utilizing any evaluation criteria at all. Such issues jeopardize the reliability and effectiveness of educational decisions (Turgut & Baykul, 2012). Therefore, more effort should be invested in assessing student success and allowing these assessments to shape educational decisions. Additionally, the focus should be on reliable and valid tools for measuring student achievement. In this way, educational decisions can become fairer and more evidence based. Overcoming these problems can contribute to the increased effectiveness of our education system. Multiple-choice tests, which are among the traditional measurement and evaluation techniques, can still be considered to play a significant role in determining students' learning levels today. This measurement method offers advantages such as minimizing subjective effects during scoring, providing greater content validity compared to other measurement methods, and making measurement and evaluation easier, especially in large groups. Therefore, it is considered a preferred measurement and evaluation technique by teachers (Çardak & Selvi, 2018; Doğan, 2009).

It is possible to observe that multiple-choice tests are frequently used in scientific studies related to education. In such studies, it is essential to meticulously follow the standard test development steps when creating achievement tests. These types of studies can contribute significantly to educational sciences by providing important information about groups' learning levels and achievements (Çardak & Selvi, 2018).

One of the fundamental goals of today's education system is to increase students' awareness of historical and cultural values (Meydan & Akkuş, 2014). However, the level of knowledge among students regarding scientists and their contributions can often be insufficient (Bıçak & Bilir, 2023). Educational programs often do not emphasize these important figures enough, leaving students with a limited understanding of the lives and scientific contributions of scientists. This limitation in knowledge can limit the awareness of the younger generation towards historical and cultural heritage, potentially causing them to miss out on valuable sources of inspiration (Arabacı & Dönel Akgül, 2022; Özdemir, 2022; Yıldırım & Keçeci, 2022). In addition, a detailed literature review has been conducted to determine the general perception and thoughts about scientists in the field of education. Within the scope of these studies, various methods such as open-ended questions, interviews, and drawings have been observed (Ağgül Yalçın, 2012; Bıçak & Bilir, 2023; Bilir et al., 2020; Duran & Bayar, 2019; Durukan, 2017; Karaçam et al., 2014; Özdemir, 2017; Özdeş & Aslan, 2019). As a result of the comprehensive literature review, it has been identified that there is only one achievement test related to scientists (Karasu, 2019). The achievement test developed by Karasu (2019) consists of 20 questions and includes questions about both Turkish and foreign scientists, focusing entirely on their inventions. On the other hand, the test developed within the scope of this study has a broader perspective. This 40-question test contains only questions about Turkish scientists, covering fundamental information about scientists to provide a comprehensive perspective. Additionally, the questions focus on the general contributions of scientists, providing a different emphasis from the previous study. This study particularly focuses on Turkish scientists and differentiates itself from previous studies by increasing the number of

questions. In this way, it aims to more detailedly assess students' knowledge about Turkish scientists and provide a more comprehensive measurement tool compared to previous studies.

At this point, it becomes evident that there is a lack of knowledge among secondary school students regarding scientists. The question of whether there is a deficiency in students' knowledge about who scientists are, what they do, and their contributions to society forms the basis of the problem. Simultaneously, the absence of an appropriate measurement tool to assess students' awareness of scientists poses a problem. In this context, the main problem addressed by this study is to tackle the lack of knowledge among secondary school students regarding scientists within an educational system that aims to increase awareness of historical and cultural values. This problem highlights the need to address the deficiency in knowledge about scientists and the absence of a suitable measurement tool to assess awareness. The study aims to develop a reliable and valid test to determine students' awareness of scientists. In line with this purpose, the fundamental study question of the study is, "How can the reliability of an achievement test developed to assess secondary school students' awareness of scientists be evaluated, and to what extent can the results be replicated?"

This test can serve the purpose of evaluating students' levels of knowledge about scientists, identifying deficiencies in science education, and providing support to students in these areas. This situation brings the potential to enhance science instruction. Furthermore, students' interest in scientists and science can be influenced by their own awareness. The developed test has the potential to increase students' interest in science by helping them better understand the importance and contributions of scientists. In addition to all these aspects, this study can contribute to the field's knowledge by introducing a new study and assessment tool to science education and assessment literature.

2. METHOD

This study focused on the development of an achievement test aimed at assessing secondary school students' awareness of scientists. This measurement tool is a type of achievement test used to evaluate secondary school students' knowledge about scientists. The test was designed to measure students' knowledge levels on topics such as who scientists are, the type of work they engage in, and their contributions to society.

The development process of the test followed the test development steps outlined by Crocker and Algina (1986). Initially, the objectives of the test were determined. Subsequently, a pool of 50 test items was created during this phase, and expert opinions were sought to gather necessary feedback. Based on the evaluations provided by experts, necessary corrections were made, and six items were excluded, reducing the number of questions to 44. In the process of developing the achievement test, a two-stage implementation was carried out, including the preliminary pilot and subsequent pilot. The preliminary pilot application and the subsequent pilot application were conducted to evaluate the effectiveness, discriminability, difficulty, and overall effectiveness of the test items. Based on the results of the preliminary pilot application, two items with discrimination indices below 0.19 (items 2 and 8) were removed from the test, some items were revised, and the number of items was reduced from 44 to 42 (see Table 4). Subsequently, the test consisting of 42 items underwent a pilot application. The pilot application was conducted to assess the impact of the corrections and to observe the overall performance of the test. This stage provided additional information on the reliability and validity of the test, examining the consistency of items among students and the accurate measurement of the intended topics. According to the results of the pilot application, two items with discrimination indices below 0.19 (items 1 and 40) were removed, reducing the number of items from 42 to 40 (see Table 6). Finally, the final test was composed of 40 distinct items, and test statistics were determined based on the results of the pilot application. A set of guidelines was created for evaluating test scores.

2.1. Study Group

The sample for the test consisted of a total of 433 students (222 for the preliminary pilot and 221 for the pilot application) who were enrolled in three different secondary schools in the city center of Elazig during the 2021-2022 academic year. These students were from 5th, 6th, 7th, and 8th grades. Yıldırım and Şimşek (2008) suggested maximizing diversity in the sample selection by conducting the study in schools with different levels of achievement and socio-economic statuses. Therefore, special attention was paid to schools with varying levels of achievement and socio-economic status in this study. Additionally, the selection of schools was influenced by their central location and easy accessibility.

As highlighted by Nunnally (1978), the sample size for testing exam items in a small student group should be at least five times the number of items in the test. The objective at this stage is to assess student responses and observe elements such as exam duration, test difficulty level, and confusion related to the items. In other words, the sample size is determined to evaluate the initial applications of the items and comprehend the overall structure of the test. This pertains to an initial pilot study aiming to assess the success of the test by evaluating students' reactions to the exam items. In the preliminary pilot application, findings obtained using Crocker and Algina's (1986) test development stages demonstrate the successful steps taken to ensure the reliability and validity of the test results. Consequently, the test has become a reliable measurement tool suitable for its intended purposes. The distribution of students by grade levels who participated in the preliminary pilot and pilot applications is shown in Table 1.

Table 1. *Distribution of students by grade levels.*

Type of application	Grade level	*f*
Preliminary pilot	5th grade	27
	6th grade	56
	7th grade	102
	8th grade	37
Pilot	5th grade	40
	6th grade	55
	7th grade	70
	8th grade	46

Table 1 shows the number of students from different grade levels who participated in various stages of the study. In the preliminary pilot application, 222 students participated. The majority of participants were from the 7th grade (102 students), followed by the 6th grade (56 students), 8th grade (37 students), and 5th grade (27 students). In the pilot application, 211 students participated. Again, the majority of participants were from the 7th grade (70 students), followed by the 6th grade (55 students), 5th grade (40 students), and 8th grade (46 students).

During the course of the study, necessary permissions were obtained from the Elazig Provincial Directorate of National Education. These permissions ensured that the study was conducted in compliance with legal and administrative requirements. Data confidentiality was rigorously maintained, and various measures were taken for this purpose. Participant selection was based on a voluntary basis, and Informed Consent Forms were signed by the parents of students who wished to participate in the study. This step ensured that participation was entirely voluntary and informed. During the data collection process, the personal information of participants was kept confidential and used only for the purpose of the study. This is a critical step in protecting privacy. Secure databases and encryption methods were used for data storage and analysis, ensuring protection against unauthorized access to the data. In addition, access to the obtained data was restricted to the authors of the study and authorized personnel only. This ensures that the data is used under the supervision of authorized individuals only. In conclusion, full

compliance with privacy policies ensured the highest level of data privacy and security. These steps ensured that the study was conducted in accordance with ethical standards and legal requirements.

2.2. Test Development Process

In accordance with the purpose of the study, the test development process was carried out in steps, and the test development steps proposed by Crocker and Algina (1986) were carefully followed. The main stages of creating the test can be listed as follows:

Goal Setting: The main objective of the study is to develop a valid and reliable knowledge test for scientists.

Defining Behaviors: The behaviors or topics that the test aims to measure were determined.

Development of Table of Specification: A specification table was created, indicating which behaviors would be included in the test and how each behavior would be measured.

Development of Item Pool: Based on the specification table, a pool of multiple-choice questions was prepared to measure each behavior.

Examination and Arrangement of Items: Prepared questions were examined and arranged by experts. The clarity of questions, the quality of distractors, and alignment with the curriculum were evaluated.

Creating a Trial Form: A trial form consisting of reviewed and edited questions was prepared.

Pilot Test: The prepared trial form was administered to a selected group of students for a pilot test. This stage aimed to understand how the test worked under field conditions and to identify possible problems.

Statistical Analysis of Items: The data obtained during the pilot test were used for the statistical analysis of the items. Correct answers were scored as 1 point, incorrect answers, questions with multiple markings, and unanswered questions were scored as 0 points. Item analysis was conducted to calculate item statistics, and items that did not meet the criteria or performed poorly were identified and removed from the test. This analysis was conducted to assess how the items performed in terms of the validity and reliability of the test.

Determining Statistical Properties: Statistical properties of the items such as difficulty levels and discrimination indices were evaluated. This provides information about the reliability and validity of the test.

Calculation of Test Scores: Based on the obtained data, participants' test scores were calculated and the results were analyzed.

In summary, after determining the topics to be measured in the test (basic information about scientists and their contributions), 50 multiple-choice questions were prepared. These questions were presented to six experts in the field, including four academic staff members (two professors, one associate professor, one Ph.D.) and two science teachers. Test item evaluations, focusing on the diverse academic titles and experience levels among experts, aimed to provide a more comprehensive and balanced perspective on the study. These evaluations aimed to develop questions that accurately measure students' knowledge, ensuring grammatical correctness and clarity in meaning. The expert panel, consisting of a professor, associate professor, and Ph.D. with extensive knowledge and experience in the field, allowed for a thorough assessment of the scientific content of the test. Additionally, an academic staff member from a state university specializing in Turkish education assessed the grammatical and semantic accuracy of the test questions. By focusing on grammatical rules and clarity, this expert contributed to the test's effectiveness in terms of language. The two science teachers in the expert panel provided an evaluation from a student perspective, being directly involved with students and having mastery over the curriculum. The teachers' suggestions aimed at revising

test questions to make them suitable, understandable, and aligned with learning objectives for students. The revisions proposed by experts generally aimed to improve the clarity of questions, correct grammatical errors, and ensure students could answer correctly. Furthermore, experts provided suggestions for reviewing cognitive levels and difficulty levels to enhance the test's relevance to its purpose and measurement of student achievements. In this way, experts combined different perspectives to offer a comprehensive assessment to enhance the overall quality of the test. As a result of expert opinions, the number of questions in the test was reduced from 50 to 44. The test, reduced to 44 questions, underwent an item analysis process. This analysis was employed to ensure the structural validity of the test and to enhance the content, clarity, and difficulty levels of the questions (Turgut, 1992).

The prepared test was administered to a total of 222 secondary school students, with 107 being girls and 115 boys (27 students at the 5th-grade level, 56 students at the 6th-grade level, 102 students at the 7th-grade level, and 37 students at the 8th-grade level) in the preliminary pilot application. The test scores of this group of students were arranged in order of achievement, and sixty students, equivalent to 27% of the total number of students, were selected to form the lower and upper groups. To assess the structural validity of the test items, item analysis was conducted, and the discrimination indices of the test items were examined. According to the evaluation conducted by Crocker and Algina (1986), items with a discrimination index greater than 0.40 were considered to be excellent, while items with an index value between 0.30 and 0.39 were considered good. On the other hand, items with an index value between 0.20 and 0.29 were identified as items that needed correction and improvement, and items with a discrimination index of 0.19 or lower were deemed to be removed from the test. As a result of the preliminary pilot application, two items with a discrimination index of 0.19 or lower were removed from the test, reducing the number of questions to 42. The test, consisting of 42 questions, was administered to 211 secondary school students (40 students at the 5th-grade level, 55 students at the 6th-grade level, 70 students at the 7th-grade level, and 46 students at the 8th-grade level) through the pilot application, and item analysis was conducted once again. As a result of the item analysis, two items with a discrimination index of 0.19 or lower were removed, reducing the number of questions to 40. The final version of the scale was arranged to include 40 items.

2.3. Data Analysis

The data collected were subjected to analysis through SPSS 22 software, and assessments of the validity and reliability of the test were carried out. Methods used to assess the reliability of a test can be categorized into two main categories: single-administration methods and two-administration methods. The various methods used to assess measurement reliability include test-retest, alternate form, split-half, interrater, Cronbach's alpha coefficient, Kuder-Richardson 20 (KR-20), and Hoyt's variance analysis. These methods are typically evaluated using the Pearson product-moment correlation coefficient (Crocker & Algina, 1986; Linn & Gronlund, 2000; Mehrens & Lehmann, 1991). Among the single-administration methods, which are based on the scores obtained by administering a test to a group only once, are KR-20 and 21, Cronbach's alpha, Hoyt's Variance Analysis, and the Test Split-Half methods (Karip, 2012). In this study, the reliability of the measurement results was calculated using split-half tests and KR-20 reliability types.

3. RESULTS

In this section, data related to the validity and reliability studies of the knowledge test developed for scientists are presented. These data have been collected to assess the measurement-related validity and reliability of the test. Validity studies aim to determine whether the test accurately reflects the concept or skill it intends to measure. Reliability studies, on the other hand, assess whether the test provides consistent and reliable results. The data analysis methods and results used in this study provide valuable information about the scientific validity and reliability of

the test. Therefore, these data will assist in understanding how reliable and valid the use and interpretation of the test scores are.

Within the scope of the study, eight science textbooks used in secondary schools affiliated with the Ministry of National Education (MoNE) during the 2020-2021 academic year were examined. The distribution of Turkish/Turkish-origin scientists mentioned in the textbooks according to grade levels is presented in Table 2.

Table 2. *Distribution of Turkish/Turkish-origin scientists mentioned in secondary school science textbooks by grade levels.*

Grade level	The mentioned Turkish/Turkish-origin scientists	*f*
5th grade (MEB Publications and SDR Vertical Publishing)	Ali Kuscu, Fatih Sultan Mehmet, Hezarfen Ahmet Celebi, Ibnu'l-Heysem, Vecihi Hurkus	5
6th grade (Sevgi Publications, MEB Publications-1, MEB Publications-2)	Ali Kuscu, Aziz Sancar, Canan Dagdeviren, Associate Professor Dr. Ozgur Sahin, El-Memun, Ibnu'l-Heysem, Ibn-i Sina, Mimar Sinan, Ulug Bey	9
7th grade (Tutku Publishing, MEB Publications)	Ali Kuscu, Aziz Sancar, Ibnu'l-Heysem, Ulug Bey	4
8th grade (SDR Vertical Publishing)		-

Table 2 reflects which scientists are presented to students in secondary school Science textbooks and in which grade levels. According to this table, there are the names of five scientists in the 5th-grade textbooks. The 6th-grade textbooks feature nine different scientists, making it the grade level with the highest number of scientist mentions compared to other grade levels. In 7th-grade textbooks, the names of four scientists are mentioned, while in 8th-grade textbooks, there is no mention of a Turkish/Turkish-origin scientist.

The developed knowledge test is focused exclusively on Turkish-Islamic scholars. In this context, there are no items related to contemporary influential scientists, such as Aziz Sancar and Canan Dagdeviren. This choice reflects the test's emphasis on a more historical and cultural perspective. The selection of scientists based on specific criteria in organizing the test aims to enable students to acquire knowledge at levels appropriate to their grade levels. While determining the number of items represented by each scientist, their impact on the history of science and the general knowledge level of students have been taken into account. Certain scientists are represented by more items due to their historical significance and contributions. For instance, prominent figures like Ebu Bekir er-Razi, Ibnu'l-Heysem, and el-Cezeri are represented by more items because of their extensive influence in various fields. This approach provides students with an opportunity to gain in-depth knowledge about the work of specific scientists and explore different aspects of their contributions. Furthermore, to enhance diversity in the test, a fair distribution among different scientific fields has been ensured. An attempt has been made to establish a balance among scientists focusing on various fields such as physics, mathematics, and astronomy. This diversity offers students the chance to explore different disciplines of science and enhance their overall knowledge levels. In this context, the distribution of items in the knowledge test according to scientists in the preliminary pilot application is presented in Table 3.

Table 3 displays which scientists the items in the knowledge test are based on and numerically represents the representation of these scientists in the test content. In total, there are items related to 16 different scientists in the knowledge test. Ebu Bekir er-Razi is the most represented scientist in the table with six different items. Following him are el-Cezeri and Abdurrahman el-Hazini, both represented with five different items.

Table 3. *Distribution of items in the knowledge test according to scientists in the preliminary pilot application.*

Scientists	Items
Ebu Bekir er-Razi	8-10-21-25-37-43
el-Cezeri	1-3-24-36-42
Abdurrahman el-Hazini	4-17-38-41
Ibnu'l-Heysem	5-11-29-44
Cabir ibn Hayyan	9-20-34
Farabi	12-19-28
Fergani	16-35-40
Ibn Sina	15-18-23
Ibnu'n-Nefis	6-13-32
Ulug Bey	22-26-33
Ali Kuscu	7-27
Aksemseddin	30
Biruni	14
Hezarfen Ahmet Celebi	31
John Dalton	2
Kindi	39

As a result of the item analysis conducted, the difficulty levels (p) and discrimination indices (d) of the items in the knowledge test within the scope of the preliminary pilot application are presented in Table 4.

Table 4. *Difficulty levels (p) and discrimination indices (d) of the items in the preliminary pilot application knowledge test.*

Questions	Groups	A	B	C	D	Blank	Filled	p	d
1	Upper	56	-	2	2	0	60	0.61	0.63
	Lower	18	14	17	11	0	60		
2	Upper	-	-	60	-	0	60	0.92	0.15
	Lower	3	4	51	2	0	60		
3	Upper	1	11	33	15	0	60	0.36	0.36
	Lower	6	17	11	26	0	60		
4	Upper	1	13	5	41	0	60	0.43	0.50
	Lower	9	22	18	11	0	60		
5	Upper	-	49	10	1	0	60	0.52	0.58
	Lower	5	14	22	19	0	60		
6	Upper	43	3	12	2	0	60	0.45	0.51
	Lower	12	16	27	5	0	60		
7	Upper	-	3	57	-	0	60	0.74	0.41
	Lower	8	8	32	12	0	60		
8	Upper	25	16	3	16	0	60	0.25	0.01
	Lower	14	18	13	15	0	60		
9	Upper	6	46	6	2	0	60	0.52	0.48
	Lower	26	17	13	4	0	60		
10	Upper	8	34	16	2	0	60	0.40	0.33
	Lower	16	14	23	7	0	60		
11	Upper	1	43	9	7	0	60	0.47	0.48
	Lower	8	14	21	17	0	60		

12	Upper	1	7	36	16	0	60	0.41	0.36
	Lower	9	20	14	17	0	60		
13	Upper	4	1	51	4	0	60	0.61	0.46
	Lower	9	12	23	16	0	60		
14	Upper	16	37	5	2	0	60	0.40	0.43
	Lower	29	11	15	5	0	60		
15	Upper	3	7	10	40	0	60	0.40	0.51
	Lower	17	13	21	9	0	60		
16	Upper	3	41	5	11	0	60	0.50	0.35
	Lower	13	20	14	13	0	60		
17	Upper	39	5	11	5	0	60	0.49	0.31
	Lower	20	10	10	20	0	60		
18	Upper	3	43	7	7	0	60	0.43	0.56
	Lower	21	9	19	11	0	60		
19	Upper	8	47	3	2	0	60	0.52	0.51
	Lower	10	16	14	20	0	60		
20	Upper	10	4	42	4	0	60	0.48	0.43
	Lower	18	13	16	13	0	60		
21	Upper	40	11	9	-	0	60	0.46	0.40
	Lower	16	13	13	18	0	60		
22	Upper	50	6	3	1	0	60	0.50	0.65
	Lower	11	20	16	13	0	60		
23	Upper	4	55	-	1	0	60	0.62	0.58
	Lower	17	20	15	8	0	60		
24	Upper	4	45	7	4	0	60	0.46	0.56
	Lower	19	11	13	17	0	60		
25	Upper	5	42	10	3	0	60	0.47	0.45
	Lower	9	15	17	19	0	60		
26	Upper	4	37	15	4	0	60	0.47	0.28
	Lower	8	20	25	7	0	60		
27	Upper	6	39	4	11	0	60	0.42	0.45
	Lower	14	12	22	12	0	60		
28	Upper	1	43	11	5	0	60	0.49	0.45
	Lower	17	16	25	2	0	60		
29	Upper	2	3	7	48	0	60	0.54	0.51
	Lower	6	20	17	17	0	60		
30	Upper	54	3	3	-	0	60	0.52	0.75
	Lower	9	18	14	19	0	60		
31	Upper	-	55	2	3	0	60	0.65	0.53
	Lower	9	23	20	8	0	60		
32	Upper	39	5	11	5	0	60	0.40	0.48
	Lower	10	18	22	10	0	60		
33	Upper	4	5	49	2	0	60	0.53	0.56
	Lower	16	15	15	14	0	60		
34	Upper	-	12	45	3	0	60	0.45	0.58
	Lower	14	26	10	10	0	60		
35	Upper	6	10	40	4	0	60	0.42	0.48
	Lower	17	19	11	13	0	60		
36	Upper	7	23	24	6	0	60	0.30	0.20

								p	d
	Lower	14	12	12	22	0	60		
37	Upper	14	3	40	3	0	60	0.45	0.41
	Lower	15	20	15	10	0	60		
38	Upper	7	14	37	2	0	60	0.46	0.30
	Lower	10	20	19	11	0	60		
39	Upper	5	8	30	17	0	60	0.38	0.23
	Lower	17	15	16	12	0	60		
40	Upper	2	24	27	7	0	60	0.30	0.28
	Lower	16	20	19	14	0	60		
41	Upper	6	6	41	7	0	60	0.40	0.55
	Lower	17	18	8	17	0	60		
42	Upper	7	5	13	35	0	60	0.37	0.41
	Lower	16	12	22	10	0	60		
43	Upper	3	2	25	3	0	60	0.48	0.60
	Lower	11	18	20	11	0	60		
44	Upper	21	3	4	12	0	60	0.25	0.20
	Lower	9	17	17	17	0	60		

p: difficulty
d: discrimination

When examining Table 4, it can be seen that the difficulty indices of the test items range from 0.25 to 0.92. Items within this range indicate different levels of difficulty. Additionally, the discrimination indices of the items also vary between 0.01 and 0.75. These values indicate how well the items differentiate between different ability groups. A total of 31 items with a discrimination index of 0.40 or higher were used in the test, and it can be said that these items better measure the differences between different ability groups. Similarly, six items with a discrimination index between 0.30 and 0.39 were also included in the test, indicating an acceptable level of discriminative ability. Five items with a discrimination index between 0.20 and 0.29 (items 26, 36, 39, 40, and 44) were included in the test after necessary adjustments were made. These adjustments were found to enhance the discriminative power of the items. However, two items with a discrimination index of 0.19 or lower, related to John Dalton (a general question) and Ebu Bekir er-Razi, were removed from the test. As a result of these steps, following the preliminary pilot application, the reliability coefficient of the test was calculated as 0.87 for KR-20 and as 0.78 for the Spearman-Brown two-half test correlation value. This high reliability value indicates that the test's internal consistency has been achieved. Consequently, the final version of the scale has been organized to consist of 42 items for the pilot application. The distribution of items in the knowledge test according to scientists in the pilot application is presented in Table 5.

Table 5 indicates the basis of the knowledge test used in the pilot application and numerically represents the representation of various scientists in the test content. In total, there are items related to 15 different scientists in the knowledge test. Ebu Bekir er-Razi and el-Cezeri are the most represented scientists in the table with five different items each. Following them are Abdurrahman el-Hazini and Ibnu'l-Heysem, both represented with four different items.

The Science Knowledge Test consisting of 42 questions was administered to secondary school students through a pilot application, and item analysis was conducted again. The results of the item analysis, including item difficulty levels, discrimination indices, item-total correlations, and t-values, are presented in Table 6.

Table 5. *Distribution of items in the knowledge test according to scientists in the pilot application.*

Scientists	Items
Ebu Bekir er-Razi	10-21-25-37-43
el-Cezeri	1-3-24-36-42
Abdurrahman el-Hazini	4-17-38-41
Ibnu'l-Heysem	5-11-29-44
Cabir ibn Hayyan	9-20-34
Farabi	12-19-28
Fergani	16-35-40
Ibn Sina	15-18-23
Ibnu'n-Nefis	6-13-32
Ulug Bey	22-26-33
Ali Kuscu	7-27
Aksemseddin	30
Biruni	14
Hezarfen Ahmet Celebi	31
Kindi	39

Table 6. *Difficulty levels (p), discrimination indices (d), item-total correlations, and t-values of the items in the knowledge test in the pilot application.*

Questions	Groups	A	B	C	D	Blank	Filled	p	d	r	t^2
1	Upper	54*	-	1	2	0	57	0.99	0.08	0.16	1.59
	Lower	49	2	3	3	0	57				
2	Upper	11	8	32*	6	0	57	0.37	0.36	0.43	4.34***
	Lower	23	5	11	8	0	57				
3	Upper	4	9	7	37*	0	57	0.40	0.49	0.47	6.12***
	Lower	13	14	21	9	0	57				
4	Upper	3	45*	2	7	0	57	0.56	0.45	0.44	5.47***
	Lower	6	19	14	18	0	57				
5	Upper	40*	8	3	6	0	57	0.45	0.49	0.47	5.99***
	Lower	12	28	7	10	0	57				
6	Upper	-	-	54*	3	0	57	0.75	0.38	0.39	5.30***
	Lower	3	9	32	13	0	57				
7	Upper	4	42*	3	8	0	57	0.52	0.42	0.36	4.92***
	Lower	19	18	11	9	0	57				
8	Upper	12	31*	4	10	0	57	0.39	0.29	0.52	3.39***
	Lower	7	14	24	12	0	57				
9	Upper	3	42*	11	1	0	57	0.46	0.54	0.52	6.88***
	Lower	12	11	25	9	0	57				
10	Upper	17	2	34*	4	0	57	0.44	0.29	0.31	3.32***
	Lower	27	15	17	8	0	57				
11	Upper	3	5	46*	3	0	57	0.54	0.52	0.51	6.58***
	Lower	3	23	16	15	0	57				
12	Upper	10	34*	3	10	0	57	0.37	0.43	0.50	5.36***
	Lower	27	9	16	5	0	57				
13	Upper	6	5	10	36*	0	57	0.36	0.52	0.55	6.88***
	Lower	14	18	19	6	0	57				
14	Upper	8	39*	7	3	0	57	0.47	0.42	0.39	4.92***
	Lower	20	15	19	3	0	57				
15	Upper	38*	6	4	9	0	57	0.47	0.38	0.40	4.43***
	Lower	16	13	18	10	0	57				
16	Upper	4	42*	7	2	0	57	0.45	0.56	0.58	7.22***
	Lower	20	10	12	15	0	57				

								p	d	r	t
17	Upper	2	45*	6	5	0	57	0.52	0.52	0.50	6.56***
	Lower	7	15	22	13	0	57				
18	Upper	6	4	41*	6	0	57	0.50	0.42	0.36	4.91***
	Lower	15	22	17	3	0	57				
19	Upper	38*	5	10	4	0	57	0.45	0.42	0.36	4.93***
	Lower	14	21	17	5	0	57				
20	Upper	46*	3	5	3	0	57	0.49	0.63	0.63	8.62***
	Lower	10	25	17	5	0	57				
21	Upper	3	52*	1	1	0	57	0.63	0.56	0.54	7.57***
	Lower	9	20	14	14	0	57				
22	Upper	5	42*	2	8	0	57	0.46	0.54	0.57	6.88***
	Lower	12	11	19	15	0	57				
23	Upper	9	41*	4	3	0	57	0.47	0.49	0.44	5.72***
	Lower	21	13	11	12	0	57				
24	Upper	10	34*	6	7	0	57	0.44	0.29	0.36	3.20***
	Lower	5	17	21	14	0	57				
25	Upper	4	35*	10	8	0	57	0.40	0.42	0.50	5.02***
	Lower	17	11	19	10	0	57				
26	Upper	10	38*	5	4	0	57	0.47	0.38	0.40	4.43***
	Lower	16	16	14	11	0	57				
27	Upper	8	3	1	45*	0	57	0.56	0.45	0.46	5.47***
	Lower	5	10	23	19	0	57				
28	Upper	50*	4	3	-	0	57	0.54	0.66	0.63	9.53***
	Lower	12	24	15	6	0	57				
29	Upper	1	52*	3	1	0	57	0.65	0.50	0.56	6.72***
	Lower	7	23	18	9	0	57				
30	Upper	37*	7	8	5	0	57	0.42	0.45	0.46	5.51***
	Lower	11	21	14	11	0	57				
31	Upper	3	2	48*	4	0	57	0.53	0.61	0.51	8.26***
	Lower	17	15	13	12	0	57				
32	Upper	8	4	41*	4	0	57	0.46	0.50	0.51	6.27***
	Lower	14	25	12	6	0	57				
33	Upper	12	5	37*	3	0	57	0.42	0.43	0.21	5.22***
	Lower	17	13	12	15	0	57				
34	Upper	10	13	24*	10	0	57	0.28	0.28	0.28	3.48***
	Lower	17	18	8	15	0	57				
35	Upper	6	5	39*	7	0	57	0.44	0.47	0.45	5.73***
	Lower	10	19	12	16	0	57				
36	Upper	4	6	34*	13	0	57	0.42	0.33	0.34	3.78***
	Lower	10	21	15	11	0	57				
37	Upper	10	7	29*	11	0	57	0.40	0.21	0.22	2.32**
	Lower	6	13	17	21	0	57				
38	Upper	12	13	25*	7	0	57	0.28	0.29	0.30	3.68***
	Lower	14	17	8	18	0	57				
39	Upper	6	8	38*	5	0	57	0.38	0.56	0.56	7.46***
	Lower	17	23	6	11	0	57				
40	Upper	6	5	6	40*	0	57	0.64	0.10	0.09	1.17
	Lower	12	9	12	34	0	57				
41	Upper	4	7	4	42*	0	57	0.44	0.57	0.50	7.57***
	Lower	6	27	15	9	0	57				
42	Upper	36*	6	7	8	0	57	0.38	0.63	0.59	6.18***
	Lower	8	21	17	11	0	57				

p: difficulty, d: discrimination, r: item-total correlation, t: value between lower 27% - upper 27%
^{1}n=211 ^{2}n$_1$=n$_2$=57   *correct answer   **p<.05   ***p<.001

Based on the results presented in Table 6, it is observed that there are 29 items with a discrimination index of 0.40 or higher, five items with a discrimination index ranging from 0.30 to 0.39, and six items with a discrimination index between 0.20 and 0.29. Additionally, there are two items with a discrimination index of 0.19 or lower related to el-Cezeri and Fergani, which have been removed from the test. After removing these two items, the item-total correlations for all remaining items in the test range from 0.21 to 0.63. A t-value of $p < .001$ is significant for 39 items, and a t-value of $p < .05$ is significant for one item.

Based on the pilot study results, the KR-20 reliability coefficient for the Scientist Knowledge Test is 0.94, and the Spearman-Brown two-half test correlation value is 0.92. In light of these findings, the final version of the scale has been arranged to include 40 items. The Scientists Knowledge Test is included in the Appendix.

The final version of "The Science Scientists Knowledge Test" consists of items specific to a total of 15 different scientists. The distribution of the items related to the scientists in the final version of the test is presented in detail in Table 7.

Table 7. *Relationship of scientists with items in the Science Scientists Knowledge Test.*

Scientists	Related items
Ebu Bekir er-Razi	10-21-25-37-43
Abdurrahman el-Hazini	4-17-38-41
el-Cezeri	3-24-36-42
Ibnu'l-Heysem	5-11-29-44
Cabir ibn Hayyan	9-20-34
Farabi	12-19-28
Ibn Sina	15-18-23
Ibnu'n-Nefis	6-13-32
Ulug Bey	22-26-33
Ali Kuscu	7-27
Fergani	16-35
Aksemseddin	30
Biruni	14
Hezarfen Ahmet Celebi	31
Kindi	39

When Table 7 is examined, it is observed that Ebu Bekir er-Razi, Ibnu'l-Heysem, and el-Cezeri are the focal points of more items in the test. In the developed knowledge test, some scientists like Aksemseddin, Hezarfen Ahmet Celebi, Biruni are represented by only one item.

4. DISCUSSION and CONCLUSION

This study aims to develop a knowledge test that meets the validity and reliability requirements to assess secondary school students' awareness of scientists. In line with this objective, the stages of test development have been followed to create a valid and reliable knowledge test. Upon reviewing the existing literature, it was found that there is only one achievement test available for assessing the awareness of scientists. However, similar test development procedures have been observed to be used in scales designed for science courses (Akbulut & Çepni, 2013; Keçeci et al., 2019; Yazıcı et al., 2022).

The results of the item analysis conducted within the scope of the study indicate that the developed test ensures structural validity. While the majority of items in the test have item discrimination index values ranging from 0.21 to 0.66, most of the items in the developed knowledge test have item discrimination index values of 0.40 and above. Additionally, the mean item discrimination index value for the test is 0.45. These findings demonstrate that the

developed test has been designed to be valid and aligned with its intended measurement domain. Structural validity is a critical indicator of measurement quality that assesses the internal consistency of a measurement instrument and the harmony among its items. The high item discrimination index values for the test items signify that these items effectively represent the measured construct and are in line with the test's intended purpose (Tekin, 2010). Therefore, establishing the structural validity of the test aids in confirming the reliability and validity of the test results. Study suggests that tests designed for science courses generally have mean item discrimination index values of 0.40 and above (Açıkgöz & Karslı, 2015; Demir et al., 2016). These findings indicate that the developed knowledge test is consistent with the mean item discrimination index value. Furthermore, all items in the test exhibit item-total correlations ranging from 0.21 to 0.63. There is one item with a significant T-value at the 0.001 level and another at the 0.005 level. Variations in item-total correlations provide further insights into the extent to which items contribute to the overall internal consistency and the measurement objective. High correlation values indicate that the item aligns with the test and enhances the overall reliability of the measurement, whereas low correlation values may raise questions about the consistency between the item and the measurement. These results collectively demonstrate that the test is a reliable and valid measurement instrument in its entirety.

The difficulty levels of the test were assessed using item difficulty index (p) values. These values range from 0.28 to 0.75, indicating that some items are challenging for students while others are relatively easy. The most difficult items are item 33 and item 37, whereas the easiest item is item 5. It can be expressed that the test is generally easy for students, and the questions are appropriate for their comprehension levels, with an average item difficulty index of 0.45. These index values are crucial for evaluating the difficulty levels of the items (Kan, 2011). These results are consistent with similar studies, which have shown that tests developed for science education generally have a moderate to high difficulty level. Additionally, the item difficulty index value of this knowledge test is similar to that of other tests designed for science education, indicating its comparability with these tests (Aymen Peker & Taş, 2019; Bolat & Karamustafaoğlu, 2019).

The reliability of the developed Science Scientists Knowledge Test was evaluated using the KR-20 coefficient, which was calculated as 0.94. This value falls within the range of $0.60 \leq \alpha < 0.90$, indicating that the test scores are highly reliable (Can, 2014). These results provide strong support for the high reliability of the students' test scores. Additionally, in line with the reliability levels of similar tests in the literature, the reliability of the test has been robustly demonstrated (Açıkgöz & Karslı; Saraç, 2018). However, despite the achieved high reliability, there are certain limitations regarding the generalizability of the test. Specifically, the focus of the test on Turkish-Islamic scholars may restrict the potential student groups to whom the test can be applied within a certain cultural and religious context. This situation could limit the overall validity of the test and may impact its applicability to student groups from different cultures or disciplines.

In conclusion, this study makes a significant contribution to the development of a reliable test aimed at assessing middle school students' awareness of scientists in the field of science education. The study can guide science education practitioners and researchers in providing an effective tool for measuring students' perceptions of science. Furthermore, future studies may assess the applicability of the test in a broader student population and contribute further to understanding learning efforts in science education. The analysis of items and the assessment of the test's reliability provide a substantial contribution to test development methodology. This study demonstrates that the steps in the knowledge test development process were followed accurately, and the test is suitable for its intended measurement purpose. Consequently, it can guide other researchers and educational professionals in test development. These aspects

indicate that the study enriches the knowledge in the field of science education and has the potential to offer an effective tool for measuring students' awareness of scientists.

5. RECOMMENDATIONS

Based on the results of this study, the following recommendations can be made:

- This study involved the development of a knowledge test to assess secondary school students' awareness of scientists. It is recommended that similar knowledge tests be developed for a wider range of topics and fields. This would allow for the evaluation of scientific awareness among students at different educational levels and across different disciplines.

- The Science Scientists Knowledge Test that has been developed has proven to be a powerful tool for measuring students' awareness of scientists. This test may encourage students to show more interest in science-related career opportunities and scientific study. Therefore, it is recommended that this test be widely used in schools and educational institutions.

- This study focused on assessing awareness of scientists. Future study could explore other aspects of students in the field of science education, leading to a more comprehensive understanding. This could provide more opportunities to enhance science education.

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**: Firat University Social Sciences and Humanities Research Ethics Committee, 23/11/2020-425157.

Contribution of Authors

Pelin Yıldırım: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing-original draft preparation, Writing – review & editing. **Gonca Keçeci**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing-original draft preparation. **Fikriye Kırbağ Zengin**: Conceptualization, Project administration, Resources, Validation, Visualization.

Orcid

Pelin Yıldırım https://orcid.org/0000-0003-4425-2472
Gonca Keçeci https://orcid.org/0000-0002-2582-3850
Fikriye Kırbağ Zengin https://orcid.org/0000-0002-0547-8746

REFERENCES

Açıkgöz, M., & Karslı, F. (2015). Alternatif ölçme-değerlendirme yaklaşımları kullanılarak iş ve enerji konusunda geliştirilen başarı testinin geçerlilik ve güvenirlik analizi [Validity and reliability analysis of achievement test which developed for work and energy, using alternative measure-evaluation techniques]. *Amasya Education Journal*, *4*(1), 1-25. Erişim adresi: https://dergipark.org.tr/en/download/article-file/19638

Ağgül Yalçın, F. (2012). Öğretmen adaylarının bilim insanı imajlarının bazı değişkenlere göre incelenmesi [Investigation of prospective teachers' image of scientist with respect to some variables]. *Elementary Education Online*, *11*(3), 611-628. https://dergipark.org.tr/tr/download/article-file/90555

Akbulut, H.İ., & Çepni, S. (2013). Bir üniteye yönelik başarı testi nasıl geliştirilir?: İlköğretim 7. sınıf kuvvet ve hareket ünitesine yönelik bir çalışma [How to develop an achievement

test for a unit?: A study for grade 7 force and motion unit]. *Amasya Education Journal*, *2*(1), 18-44. https://dergipark.org.tr/en/download/article-file/19600

Arabacı, S., & Dönel Akgül, G. (2022). Fen Bilimleri ders kitaplarına bir bakış: Türk İslam bilim insanları [A overview of science textbooks: Turkish Islamic scientists]. *International Journal of Social Sciences Academy*, *4*(9), 386-404. https://doi.org/10.479 94/usbad.1068516

Aymen Peker, E., & Taş, E. (2019). Üç aşamalı kavram tanı testi geliştirme ve geçerlik çalışması: 5. sınıf canlılar dünyasını gezelim tanıyalım ünitesi [Three tier concept diagnostic test development and validity study: Let's travel and learn about the living world unit of fifth grade]. *Journal of Computer and Education Research*, *7*(14), 515–539. https://doi.org/10.18009/jcer.602223

Bıçak, C.Y., & Bilir, V. (2023). Fen Bilimleri ders kitaplarında yer alan bilim insanları ile ilgili bilgilerin öğretmen gözünden değerlendirilmesi [Evaluation of the information regarding scientists in the science textbooks from the teacher's perspective]. *Mehmet Akif Ersoy University Journal of Education Faculty*, *65*, 487-511. https://doi.org/10.21764/maeuef d.1107010

Bilir, V., Tatlı, A., Yıldız, C., Emiroğlu, B.B., Ertuğrul, D., & Sakmen, G. (2020). Argümantasyon tabanlı öğrenme yaklaşımında kullanılan argümantasyon tekniklerinin ortaokul sekizinci sınıf öğrencilerinin bilim insanı imajları üzerine etkisi [The effect of argumentation techniques used in argumentation based learning approach on 8th grade secondary school students' image of scientist]. *Journal of Gazi University Gazi Education Faculty*, *40*(2), 481-510. https://dergipark.org.tr/en/download/article-file/722283

Bolat, A., & Karamustafaoğlu, S. (2019). "Vücudumuzdaki sistemler" ünitesi başarı testi geliştirme: Geçerlik ve güvenirlik [Development achievement of test on the unit "systems in our body": Validity and reability]. *Gazi Journal of Education Sciences, 5*(2), 131-159. https://doi.org/10.30855/gjes.2019.05.02.008

Can, A. (2014). *SPSS ile bilimsel araştırma sürecinde nicel veri analizi [Quantitative data analysis in the scientific research process with SPSS]* (2nd ed.). Pegem Akademi Yayıncılık.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston.

Çardak, Ç.S., & Selvi, K. (2018). Öğretim ilke ve yöntemleri dersi için bir başarı testi geliştirilmesi [Developing an achievement test for the teaching principles and methods course]. *Mediterranean Educational Research Journal*, *12*(26), 379-406. https://doi.org/ 10.29329/mjer.2018.172.19

Demir, N., Kızılay, E., & Bektaş, O. (2016). 7. sınıf çözeltiler konusunda başarı testi geliştirme: Geçerlik ve güvenirlik çalışması [Development of an achievement test about solutions for 7th graders: A validity and reliability study]. *Necatibey Faculty of Education Electronic Journal of Science and Mathematics Education*, *10*(1). https://doi.org/10.17522/nefefme d.52947

Doğan, N. (2009). *Çoktan seçmeli testler [Multiple choice tests]*. H. Atılgan (Eds.), Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education] içinde (3rd ed.) (pp. 223-268). Anı Yayıncılık.

Duran, E., & Bayar, A. (2019). İlkokul öğrencilerinin bilim ve bilim insanına ilişkin algıları [The perceptions of primary school students on science and scientist]. *Social Sciences Research Journal*, *8*(4), 14-29. https://dergipark.org.tr/en/download/article-file/870418

Durukan, Ü. G. (2017). Fen Bilimleri ders kitaplarında yer alan bilim insanları [Scientists in science textbooks]. *Karadeniz Sosyal Bilimler Dergisi*, *9(Kadın özel sayısı)*, 321-344. https://search.trdizin.gov.tr/tr/yayin/detay/265362/

Ertürk, S. (1972). *Eğitimde "program" geliştirme [Developing a "program" in education]*. Hacettepe Üniversitesi Basımevi.

Karaçam, S., Aydın, F., & Digilli, A. (2014). Fen ders kitaplarında sunulan bilim insanlarının basmakalıp bilim insanı imajı açısından değerlendirilmesi [Evaluation of scientists represented in science textbooks in terms of stereotype scientist image]. *Ondokuz Mayis University Journal of Faculty of Education*, *33*(2), 606-627. https://doi.org/10.7822/om uefd.33.2.19

Karasu, B. (2019). *Ortaokul Fen Bilimleri derslerinde işlenen bilim insanlarına yönelik formüllü tişört yönteminin akademik başarıya etkisi [The effect on academic success of formulated t-shirt method for teaching scientists' studies in middle school science course]* [Master's thesis]. Ağrı İbrahim Çeçen University.

Karip, E. (Ed.). (2012). *Ölçme ve değerlendirme [Measurement and evaluation]* (5th ed.). Pegem Akademi.

Keçeci, G., Yıldırım, P., & Kırbağ Zengin, F. (2019). Sistemler akademik başarı testi: Geçerlik ve güvenirlik çalışması [Systems academic achievement test: Validity and reliability study]. *National Journal of Education Academy*, *3*(1), 96-114. https://doi.org/10.32960/ uead.514188

Linn, R.L., & Gronlund, N.E. (2000). *Measurement and assessment in teaching* (8th ed.). Merrill/Prentice-Hall.

Mehrens, W.A., & Lehmann I.J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). Harcourt Brace College.

Meydan, A., & Akkuş, A. (2014). Sosyal bilgiler öğretiminde müze gezilerinin tarihi ve kültürel değerlerin kazandırılmasındaki önemi [The importance of getting students adopt the historical and cultural values through museum visits in social studies teaching]. *Marmara Coğrafya Dergisi*, *29*, 402-422. https://doi.org/10.14781/mcd.92390

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

Özdemir, E. (2017). *Ortaokul öğrencilerinin bilim insanı imajı hakkındaki görüşleri [Image of secondary school students' views of scientist]* [Unpublished master's thesis]. Necmettin Erbakan University.

Özdemir, E. (2022). Ortaokul Fen Bilimleri ders kitaplarının bilim insanları bağlamında incelenmesi [Analysis of secondary school Science textbooks in the context of scientists]. *Journal of Research in Education and Society*, *9*(2), 403-429. https://doi.org/10.51725/e tad.1098860

Özdeş, S., & Aslan, O. (2019). Kız öğrencilerin bilim insanı cinsiyetine yönelik algılarını ve bilim insanı olma isteklerini etkileyen faktörler [The factors affecting the girl students' perceptions about the gender of scientist and their willingness to become a scientist]. *Ondokuz Mayis University Journal of Education Faculty*, *38*(1), 190-214. https://doi.org/10.7822/omuefd.501944

Saraç, H. (2018). Fen Bilimleri dersi 'maddenin değişimi' ünitesi ile ilgili başarı testi geliştirme: Geçerlik ve güvenirlik çalışması [Development of achievement test about science lesson 'change of matter' unit: Validity and reliability study]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *18*(1), 416-445. https://doi.org/10.17240/aibuefd.2018..-388815

Sarıbaş, S., & Babadağ, G. (2015). Temel eğitimin temel sorunları [Basic problems of primary education]. *Anatolian Journal of Educational Leadership and Instruction*, *3*(1), 18-34.

Tekin, H. (2010). *Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education]* (20th ed.). Yargı Yayınevi.

Turgut, M.F. (1992). *Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education]*. Saydam Matbaacılık.

Turgut, M.F. (1997). *Eğitimde ölçme ve değerlendirme metotları [Measurement and evaluation methods in education]*. Gül Yayınevi.

Turgut, M.F., & Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme [Measurement and evaluation in education]* (4th ed.). Pegem Akademi.

Ülger, K. (2021). Uzaktan eğitim modelinde karşilaşilan sorunlar-fırsatlar ve çözüm önerileri [Problems-opportunities and solution suggestions encountered in the distance education model]. *International Journal of Contemporary Educational Studies*, *7*(1), 393-412.

Yazıcı, F., Gül, Ş., & Sözbilir, M. (2022). "Bitki ve hayvanlarda üreme, büyüme ve gelişme" konusuna yönelik bir başarı testinin geliştirilmesi [The development of an achievement test towards "reproduction, growth and development in plants and animals" subject]. *Journal of Atatürk University Kazım Karabekir Education Faculty*, *46*, 11-19. https://doi.org/10.5152/AUJKKEF.2022.837043

Yıldırım, A., Şimşek, H. (2008). *Sosyal bilimlerde nitel araştırma yöntemleri [Qualitative research methods in the social sciences]*. Seçkin Yayıncılık.

Yıldırım, P., & Keçeci, G. (2022). Türk-İslam alimlerinin Fen Bilimleri ders kitaplarında ve öğretim programında yeri [The place of Turkish-Islamic scholars in Science textbooks and curriculum]. *Van Yüzüncü Yıl University Journal of Education*, *19*(2), 429-445. https://doi.org/10.33711/yyuefd.1064898

APPENDIX

Turkish version of Scientists Knowledge Test

BİLİM İNSANLARI BİLGİ TESTİ

Sevgili Öğrencim;

Bu çalışma ile fen alanında bilime yön vermiş Türk-İslam alimlerine ilişkin bilgi testinin geliştirilmesi, ortaokul öğrencilerinin bilgi düzeyinin belirlenmesi amaçlanmaktadır. Çalışmanın sonuçları yalnızca bilimsel amaçla kullanılacak olup, kişisel bilgileriniz kimseyle paylaşılmayacaktır.

Cinsiyetiniz   : ( ) Kız ( ) Erkek

Sınıf Düzeyiniz : ( ) 5. sınıf   ( ) 6. sınıf   ( ) 7. sınıf   ( ) 8. sınıf

1. Aşağıdakilerden hangisi filli su saatinde yer alan unsurlardan değildir?

A) Fil   B) Zümrüdüanka Kuşu C) Fıskiye   D) Hükümdar

2. Türkistan'da yetişen, yer çekimi ve terazilerle alakalı çalışmalar yapan fizik, astronomi ve matematik alimi aşağıdakilerden hangisidir?

A) Fatih Sultan Mehmed   B) Ali kuşçu

C) İbn Sina   D) Abdurrahman el-Hazini

3. Batı'da "*Alhazen*" ve "*Alhacen*" gibi isimlerle tanınan ve Kitabü'l-Menazir (Görüntüler Kitabı) isimli eseri olan Türk-İslam alimi aşağıdakilerden hangisidir?

A) Ali Kuşçu   B) İbnü'l-Heysem   C) Biruni   D) Piri Reis

4. Yaşadığı dönemde "*İkinci İbn Sînâ*" olarak anılan Türk-İslam alimi aşağıdakilerden hangisidir?

A) İbnü'n-Nefis   B) Uluğ Bey   C) İbnü'l-Heysem   D) el-Cezeri

5. "*Kuşçu*" veya "*Kuşi*" lakabıyla anılan, Uluğ Bey'den matematik ve astronomi dersleri alan Türk-İslam alimi aşağıdakilerden hangisidir?

A) İbnü'n-Nefis   B) Farabi   C) Ali Kuşçu   D) Mimar Sinan

6. Şaraptan, saf alkol elde eden ilk bilim insanı kimdir?

A) Biruni   B) Cabir İbn Hayyan   C) Evliya Çelebi   D) el-Cezeri

7. Kızıl ve kızamık hastalıklarının iki ayrı hastalık olduğunu ortaya koyan bilim insanı aşağıdakilerden hangisidir?

A) Harezmi   B) Ebu Bekir er-Razi   C) Abdurrahman el-Hazini   D) el-Cezeri

8. Aşağıdakilerden hangisi, İbnü'l-Heysem'in optikle ilgili en önemli eseridir?

A) Mizanü'l Hikme (Hikmetin Terazisi)

B) Kitabü'l-Menazir (Görüntüler Kitabı)

C) el-Kanun Fi't-Tıbb (Tıbbın Kanunu)

D) Risaletü'l-Fethiyye

9. Hava titreşimlerinden ibaret olan ses olayının ilk mantıklı açıklamasını hangi bilim insanı yapmıştır?

A) İbn Sina   B) Evliya Çelebi   C) Farabi   D) Uluğ Bey

10. Ciltte oluşan beyaz lekelerden söz eden ve bunun deride yer yer renk kaybı şeklinde ortaya çıkan bir hastalık olduğunu belirten Türk -İslam alimi kimdir?

A) Uluğ Bey   B) Ali Kuşçu   C) İbnü'n-NefisD) Biruni

11. "*Piknometre*" denilen bir aletle, cisimlerin özgül ağırlıklarını ilk defa ölçen Türk-İslam alimi aşağıdakilerden hangisidir?

A) Akşemseddin   B) Biruni   C) Piri Reis   D) Ali Kuşçu

12. Menenjit hastalığını ve türlerini ilk defa tespit eden bilim insanı aşağıdakilerden hangisidir?

A) el-Cezeri   B) Cabir İbn Hayyan

C) Musaoğulları (Benî Musa)   D) İbn Sina

13. Batı'da "*Alfraganus*" ismiyle çok meşhur olan Türk-İslam alimi aşağıdakilerden hangisidir?

A) İbnü'n-Nefis   B) Fergani   C) İbn Sina   D) Akşemseddin

14. Aşağıdakilerden hangisi, çeşitli maddelerin özgül ağırlıklarını tespit eden Türk-İslam alimlerinden biridir?

A) Abdurrahman el-Hazini   B) Piri Reis

C) İbnü'l-Heysem   D) Mimar Sinan

15. Batı'da "*Avicenna*" ismiyle şöhret yapan Türk-İslam alimi aşağıdakilerden hangisidir?

A) Fatih Sultan Mehmed   B) İbn Sina

C) Piri Reis   D) Mimar Sinan

16. Batı'da Alfarabius, Abunazar gibi isimlerle meşhur olan Türk-İslam alimi aşağıdakilerden hangisidir?

A) el-Cezeri   B) Farabi   C) İbn Sina   D) Mimar Sinan

17. Batı dünyasında "*Geber*" ismiyle tanınan, Doğu'da ve Batı'da kimyanın kurucusu kabul edilen Türk-İslam alimi aşağıdakilerden hangisidir?

A) el-Cezeri   B) Mimar Sinan   C) Cabir İbn Hayyan   D) Uluğ Bey

18. Hastanelere klinik sistemini ilk yerleştiren kişi aşağıdakilerden hangisidir?

A) Ebu Bekir er-Razi   B) Farabi

C) İbnü'l-Heysem   D) el-Cezeri

19. Aşağıdakilerden hangisi 15. yüzyılın en büyük gökbilimcilerinden biridir?

A) Uluğ Bey   B) Akşemseddin   C) Ebu Bekir er-Razi   D) Fatih Sultan Mehmed

20. Aşağıdakilerden hangisi Türk-İslam alimidir?

A) Galileo   B) Avicenna   C) Newton   D) Archimedes

21. el-Cezeri, en önemli bilimsel buluşlarını hangi alanda yapmıştır?

A) Matematik   B) Fizik   C) Kimya   D) Astronomi

22. Gırtlak sinirini keşfeden ve bu sinirin bazen sağ tarafta olduğunu da ifade eden Türk-İslam alimi aşağıdakilerden hangisidir?

A) Mimar Sinan   B) Ebu Bekir er-Razi

C) Abdurrahman el-Hazini   D) el-Cezeri

23. Uluğ Bey'in, kurduğu rasathanede gerçekleştirilen gözlemlerin sonuçlarının toplandığı dünyaca meşhur eserinin adı aşağıdakilerden hangisidir?

A) Mizanü'l Hikme (Hikmetin Terazisi)

B) Zic-i Uluğ Bey

C) Kitabü'l-Menazir (Görüntüler Kitabı)

D) el-Kanun Fi't-Tıbb (Tıbbın Kanunu)

24. Astronomiyle ilgili Risâle fi'l-Hey'e isimli Farsça eserine yeni ilaveler yaparak Risaletü'l-Fethiyye adıyla Arapça'ya çeviren ve Fatih Sultan Mehmed'e sunan Türk-İslam alimi kimdir?

A) el-Cezeri   B) Ali Kuşçu   C) Farabi   D) İbnü'n-Nefis

25. Müzikte sesleri notalarken ve sesleri bölümlerken logaritmayı icat eden Türk-İslam alimi aşağıdakilerden hangisidir?

A) İbn Sina   B) Farabi   C) Akşemseddin   D) el-Cezeri

26. Gelişmiş kamerayı tasarlayan Leonardo da Vinci'den çok daha önce karanlık odayı keşfeden Türk- İslam alimi aşağıdakilerden hangisidir?

A) Ali kuşçu   B) Farabi   C) İbn Sina   D) İbnü'l-Heysem

27. Tıp tarihinde ilk defa mikrop meselesini ortaya atan ve hastalıkların bu yolla bulaştığı fikrini öne süren Türk-İslam alimi kimdir?

A) Akşemseddin   B) Fatih Sultan Mehmed
C) Ali Kuşçu   D) Mimar Sinan

28. Yaptığı özel bir aletle İstanbul'da Galata Kulesi'nden havalanarak Boğaz'ı geçip Üsküdar'a indiği rivayet edilen Türk-İslam alimi aşağıdakilerden hangisidir?

A) el-Cezeri   B) Hezarfen Ahmed Çelebi   C) Mimar Sinan   D) Uluğ Bey

29. I. İdrar incelemesiyle şeker hastalığını teşhis ve tespit etmiştir.

II. Döneminde "*İkinci İbn Sînâ*" olarak anılmıştır.

III. Küçük kan dolaşımını keşfetmiştir.

IV. Filli su saati, en önemli buluşlarından biridir.

İbnü'n-Nefis ilgili, yukarıdaki bilgilerden hangileri yanlıştır?

A) I ve IV   B) Yalnız III
C) I, II ve III   D) Hepsi

30. Uluğ Bey'in 1421 yılında Semerkant'ta açtığı, dönemin en gelişmiş rasathanesi aşağıdakilerden hangisidir?

A) İstanbul Rasathanesi

B) Meraga Rasathanesi

C) Semerkant Rasathanesi

D) Hemedan Rasathanesi

31. Geliştirdiği kimyasal fırınlar sayesinde, bugün bildiğimiz kostik soda, sülfirik asit, arsenik asit ve nitrik asit gibi temel asitleri kimyasal yollar ile elde eden Türk-İslam alimi aşağıdakilerden hangisidir?

A) Uluğ Bey   B) Akşemseddin   C) Cabir İbn Hayyan   D) Farabi

32. Gezegenler gibi Güneş'in de kendi yörüngesinde Batı'dan Doğu'ya bir hareketi olduğunu ilk defa açıklayan Türk-İslam alimi aşağıdakilerden hangisidir?

A) Hezarfen Ahmed Çelebi   B) Farabi   C) Fergani   D) İbn Sina

33. I. En önemli bilimsel buluşlarını kimya alanında yapmıştır.

II. Doğu'da ve Batı'da kimyanın kurucusu kabul edilmektedir.

III. Filli su saati ve değişken şekilli fıskiye, Cezeri'nin en önemli buluşlarındandır.

IV. Fil, zümrüdüanka kuşu, seyis, hükümdar ve halı, el-Cezeri'nin filli su saatinde yer alan unsurlardandır.

el-Cezeri ile ilgili, yukarıdaki bilgilerden hangileri doğrudur?

A) I ve II   B) Yalnız III
C) III ve IV   D) Hepsi

34. Simyayı tıbbın hizmetine sunan ilk kişi kimdir?

A) İbnü'l-Heysem   B) Mimar Sinan   C) Ebu Bekir er-Razi   D) el-Cezeri

35. Aşağıdakilerden hangisi, Abdurrahman el-Hazini'nin en önemli eserlerinden biridir?

A) el-Kanun Fi't-Tıbb (Tıbbın Kanunu)

B) Risaletü'l-Fethiyye

C) Mizanü'l Hikme (Hikmetin Terazisi)

D) Kitabü'l-Menazir (Görüntüler Kitabı)

36. Gökyüzünün ve denizlerin mavi görünmelerinin nedenini ilk defa doğru bir şekilde açıklayan Türk-İslam alimi aşağıdakilerden hangisidir?

A) el-Cezeri   B) Abdurrahman el-Hazini

C) Kindi   D) Musaoğulları (Benî Musa)

37. Yapmış olduğu deneyler sonucunda bütün cisimlerin yerin merkezine doğru, bir kütle çekim kuvveti ile çekildiklerini gösteren ve yer çekimini keşfeden Türk-İslam alimi aşağıdakilerden hangisidir?

A) Cabir İbn Hayyan   B) İbn Sina   C) Abdurrahman el-Hazini   D) el- Cezeri

38. Aşağıdakilerden hangisi, el-Cezeri'nin en önemli bilimsel buluşlarından biridir?

A) Piknometre   B) Hikmet Terazisi   C) Usturlap   D) Filli su saati

39. Allerji üzerine ilk eser yazan alim aşağıdakilerden hangisidir?

A) Fatih Sultan Mehmed   B) Uluğ Bey   C) Piri Reis   D) Ebu Bekir er-Razi

40. Aşağıdakilerden hangisi önemli bir fizikçi ve optikçidir?

A) İbnü'l-Heysem   B) İbn Sina   C) Fatih Sultan Mehmed   D) Harezmi

Bilim İnsanları Bilgi Testi Cevap Anahtarı

1. C	2. D	3. B	4. A	5. C	6. B	7. B	8. B	9. C	10. C
11. B	12. D	13. B	14. A	15. B	16. B	17. C	18. A	19. A	20. B
21. B	22. B	23. B	24. B	25. B	26. D	27. A	28. B	29. A	30. C
31. C	32. C	33. C	34. C	35. C	36. C	37. C	38. D	39. D	40. A

Research Article

Current Effect of Mother-Child Memory Talk on Emotion Regulation, Self-Esteem, and Memory

Nilsu Borhan [iD][1*]

[1]Pamukkale University, Faculty of Humanities and Social Sciences, Department of Psychology, Denizli, Türkiye

Abstract: Children talking to their parents more frequently about past experiences tend to have higher emotion regulation skills and self-esteem in their future lives, which may lead to higher volume and richer emotional content in future memories. Previous research also indicated that self-esteem has a strong bond with emotion regulation skills. This study's aim is to measure the mediator roles of emotion regulation difficulty and self-esteem on the relationship between childhood maternal reminiscing frequency and the volume and emotional content of the current memory experienced with the mother. Additionally, mediator roles of self-esteem in the connection between past maternal reminiscing and emotion regulation difficulty, and emotion regulation difficulty in the relationship between self-esteem and the current memory variables (i.e., total words, total emotion and unique emotion words) are examined. Participants (N=124, the age range was 22-39) filled out Rosenberg Self-Esteem, Emotion Regulation Difficulty, and Family Reminiscence Scales and wrote down one negatively-charged recent memory about their mothers. Path analysis revealed significant positive associations between past maternal reminiscing and self-esteem, emotion regulation difficulty, and total and unique emotion words in recent memory, and negative association between self-esteem and emotion regulation difficulty, supporting half of the mediation hypotheses. Results supported the notion that the frequency of parent-child reminiscing conversations in childhood is a parameter of child development since it can show its prospective effect via improving self-esteem and emotional functioning. Gender differences were not evident for current memory variables but more research on this issue is needed to reach more precise conclusions.

1. INTRODUCTION

Mothers engage in conversations about shared or unshared past events with their children. It is well-established that such discussions are helpful for children in many ways related to cognitive, social, and emotional development (Laible, 2011). Some parental characteristics may improve or deteriorate child outcomes in the reminiscing context. To illustrate, when the mothers succeed in providing more elaborative, voluminous, and emotionally rich content during these discussions, the narrative skills, autobiographical memory, and self-concept development of the child are affected positively (Fivush, 2007; Fivush & Nelson, 2006; Peterson et al., 2007; Reese et al., 1993; Wang et al., 2010). Hence, the existing literature on maternal reminiscing has been based on the quality of the reminiscing conversations provided by the mothers rather than its frequency during childhood. In fact, some studies examined the

---

frequency of emotion-related words or total words used in a memory talk between mother-child dyads (e.g., Fivush, 2007; Fivush & Nelson, 2006) but they did not measure the overall frequency of memory talks in the daily lives of mother-child pairs. How frequently mothers offer an opportunity to discuss shared past events with their children may also affect child development substantially. In addition, the parent-child reminiscing literature did not present much information about written memories shared with the mother. Thus, this study focused on written memories of the young adults recently experienced with their mothers. Moreover, although the positive effects of mother-child reminiscing on the child's emotion management and self-esteem have been shown in separate studies (Goodvin & Romdall, 2013; Harris et al., 2017), the direct and indirect associations between maternal reminiscing, emotion regulation skills, and self-esteem have not been examined together extensively. What's more, the existing maternal reminiscing literature is mainly composed of studies examining child outcomes prospectively. To our knowledge, the number of studies dealing with parent-child reminiscing and related outcomes retrospectively with adults is scarce (Öner & Gülgöz, 2022). Hence, retrospective research demonstrating the underlying mechanisms of maternal reminiscing is needed to draw more precise conclusions. For all these reasons, the current study tried to answer the questions regarding the effect of the frequency of memory talks between mothers and children about shared past events on adults' recent self-esteem, emotion regulation skills, and volume and emotional content of written memories recently experienced with the mothers.

Former studies also suggest that other than parent characteristics, including the volume and emotional atmosphere of the reminiscing conversations, the child's gender may be an influential factor in child outcomes (Aznar & Tenenbaum, 2015; Buckner & Fivush, 2000; Fivush et al., 2003). In several studies, female children were found to have an advantage over boys regarding the total word usage and the intensity of emotional expression in the reminiscing context due to the differential socialization process imposed by their parents. However, some studies failed to indicate such gender differences, as well (e.g., Svane et al., 2022). Therefore, gender differences in the reminiscing context should be further investigated to clarify this issue. The present study tried to close the gap in the previous literature by delving into the gender effect on the written current memories experienced with mothers.

1.1. Theoretical Background

At first, parent-child reminiscing conversations or memory talks can be perceived as solely an associate of autobiographical memory development. Although its huge effect on autobiographical memory cannot be denied, child development can benefit from such conversations about past events in various ways. Under parental guidance, children try to learn how to combine the outside world with their inner worlds and differentiate themselves as separate individuals from others (Fivush, 2020). This can be attributed to the theory of mind, which refers to being able to make inferences about others' mental states, including thoughts, feelings, desires, etc. (Premack, & Woodruff, 1978). Parent-child conversations about shared past events are likely to include referents to both the child's and others' mental states. For instance, a mother may discuss their last conflict with her child about something at home. This mother would probably try to explain her perspective by mentioning her feelings and thoughts on this issue. Then, she will likely encourage her child to speak up about his unique and subjective perspective by addressing his feelings and thoughts. This may explain why parent-child memory talk may fasten the theory of mind development in children (Pavarini et al., 2013; De Rosnay & Hughes, 2006).

Since the theory of mind is about self-knowledge and emotional awareness, frequently having a past-themed conversation with the child by accepting, verbalizing, and valuing the child's perspective would motivate the child to rely on and confidently express his subjective experiences and feelings. In time, such children will become adults with high self-esteem who can regulate their emotions efficiently. Indeed, empirical evidence supported the relevancy of

parent-child past referenced talk with child's improved self-esteem and better emotion regulation skills in many previous studies (e.g., Bohanek et al., 2008; Ellis et al., 2014; Laible et al., 2013; Marshall & Reese, 2022).

Another theoretical rationale for parent-child memory talks and child development might be attachment theory (Bowlby, 1979) since such conversations might strengthen the attachment bond between the child and parent (Bost et al., 2006). If parents form a secure attachment with their children, they will be guaranteed to have resilient children (Darling Rasmussen et al., 2019). One of the resilience factors associated with attachment security would be emotional management. Securely attached children are capable of cultivating efficient emotion regulation strategies for themselves (Brumariu, 2015). Furthermore, according to several studies, attachment security may be regarded as one of the variables responsible for adulthood self-esteem (Pinquart, 2023). Thereby, parent-child past-referenced daily talks can show favorable child outcomes that are empowered self-esteem and emotion management via raising attachment security. Overall, the present study bases its expectations regarding the study hypotheses on the theory of mind and attachment theory. Hence, the study's results will enlighten the theoretical implications in this sense.

1.2. Effect of Mother – Child Reminiscing on Child's Self-esteem

Mothers are the primary figures in the lives of children who determine the child's sense of self or self-concept via their daily talks about shared or unshared past events (Bird & Reese, 2006; Fivush & Nelson, 2006; Song & Wang, 2020; Welch-Ross et al., 1999). Hence, it is highly likely that child's self-esteem will be affected by such conversations between the mother and the child. For instance, the study of Marshall and Reese (2022) indicated that being exposed to an elaborative maternal reminiscing style during early childhood was positively associated with self-esteem in emerging adulthood. Furthermore, the child's future self-esteem seems to depend on the emotional content of mother-child reminiscing (Bohanek et al., 2008). Conversing more emotionally with the child as a mother regarding past events might boost the bond between the mother and the child, which, in turn, is inclined to promote the child's self-esteem. In the longitudinal study conducted by Harris et al. (2017), how mother-child reminiscing may influence a child's future self-esteem was investigated with a sample of mothers and their 6-14 aged children. Mother-child pairs were given a reminiscing task in which they were supposed to reminisce about a past adverse event. For the reminiscing task, the researchers coded both the participants' emotion-related words and behaviors. Sixteen months later, it was evident that the self-esteem levels of the children whose mothers provided more detailed explanations for the emotions were higher due to the increases in their secure attachment levels. In a nutshell, it can be considered that mother-child reminiscing talk is crucial in shaping the child's self-development concerning self-esteem (Reese et al., 2007). Although the existing literature is insufficient to indicate the association between the frequency of the mother-child reminiscing conversations as compared to the content of these conversations, it can be expected that when the children engage in a higher rate of such conversations with their mothers in daily life, their self-esteem development still can be influenced in a favorable way. Therefore, one of the hypotheses of the present study was that those with more frequent childhood reminiscing conversations with their mothers are believed to reap more self-esteem in their adulthood.

1.3. Mother-child Reminiscing and Emotion Regulation Skills

Emotionally more elaborative maternal reminiscing style may determine not only the child's self-concept and, more specifically, self-esteem development but also how the child will learn to manage or regulate their emotions. More elaborate mothers create opportunities for the discussion of self and emotions by helping their children be aware of their own and others' mental states including feelings (Fivush, 2020). Hence, it is unsurprising to observe that more elaborate mothers would have children with higher self-esteem and better emotion regulation skills. For example, Goodvin and Romdall (2013) examined the effect of mother-child

reminiscing on the children's self-concept and emotion management. As expected, it was indicated that when the children were encouraged to talk more about the negative affect in the reminiscing task, they perceived themselves as having a less negative affect, which in turn was thought to be linked to better self-representations of the children regarding the management and regulation of negative affect. When mother-child conversations about past events are detailed in terms of emotional content, the emotion regulation skills of the child will likely be boosted (Ellis et al., 2014; Laible et al., 2013; Sales & Fivush, 2005). This is especially true for the memory talk for negatively charged memories since children can learn how to express themselves and their negative emotions via such conversations (Fivush et al., 2000). It may be because parents may refer to emotion-related words more when talking to their child about negatively charged past events compared to when a positively charged event is to be discussed (Lagattuta & Wellman, 2002). In light of the existing literature, we predicted that those being exposed to more frequent childhood reminiscing conversations with their mothers would have better emotion regulation skills in their adulthood.

1.4. The Direct Link between Past Maternal Reminiscing and Recent Memory with the Mother

When children are grown up with their mothers who are highly elaborate in daily reminiscing talks, they have a better memory for remembering and narrating the past events they experienced (Noel et al., 2019; Salmon & Reese, 2016; Wareham & Salmon, 2006; Waters et al., 2019; Valentino et al., 2014), and this effect can be prolonged until adulthood. Therefore, the characteristics of mother-child reminiscing are likely to be related to children's memory and narrative skills in their future lives. Moreover, referring emotions during memory talks is known to promote child's emotional development including emotional understanding (Laible et al., 2013; Leyva et al., 2021). Therefore, it would make sense to expect that if parents expose their children to reminiscing conversations about important past events frequently, their children will be more likely to use more words and emotion referents in their memories when they become adults.

1.5. Mediator Role of Emotion Regulation and Self-esteem on Past Maternal Reminiscing and Recent Written Memory Variables

Expressing positive and negative emotions rather than suppressing them generally indicates mental and physical health (Berry & Pennebaker, 1993). Although it may be perceived that negative emotional expression is not beneficial, it may not be the case; it has been known that suppression of negative affect is an essential characteristic of individuals with anxiety and mood disorders (Campbell-Sills et al., 2006; Fernandes et al., 2022). Furthermore, negative emotional expression is crucial for social life since it is a basis for initiating intimacy with others (Graham et al., 2008). Since maternal reminiscing was found to be associated with higher child well-being (Fivush et al., 2009), its higher frequency in childhood would likely end up with more emotional expression in adulthood, which is associated with better mental health. Therefore, it is probable that when the children had mothers who used to talk a lot about shared past events in their childhood, they are going to be more skilled in regulating their emotions and experience less emotion regulation difficulty, which in turn is likely to be associated with the richer volume and emotional content of the current memories shared with their mothers.

The direct link between self-esteem and emotional expression was evident in a bunch of studies. Low self-esteem was found to be an associate of emotional suppression. In one study, it was found that increased anxiety was associated with decreased self-esteem, which in turn was found to be related to the suppression of emotions as an emotional regulation strategy (Fernandes et al., 2022). Some of the previous research (Brown & Marshall, 2001; Lightsey et al., 2006) indicated negative associations between self-esteem and negative emotions, but such studies only measured individuals' emotions by using Positive and Negative Affect Schedule (PANAS) (Watson et al., 1988). The current literature has not provided any evidence for how

self-esteem might be related to emotional expression in written memories experienced with mothers. Although feeling negative emotions can be linked to low self-esteem, this may not be true when it comes to expressing negative emotions in a written memory. In fact, writing about negative experiences to express negative affect would be functional for mental health by enhancing emotional regulation. For instance, Ahmadi et al. (2010) found that when depressive individuals are encouraged to write down their negative experiences for emotional expression, their depressive symptoms decrease. The researchers explained this finding by claiming that writing about negative experiences might help reappraise these experiences and regulate negative emotions by having individuals face with challenges of life more efficiently and protecting their psychological well-being. Therefore, young adults with higher self-esteem are likely to write down their memories in a more detailed manner and to express their negative emotions more intensely when asked to write down their negatively charged memories with their mothers.

Former research also suggested that emotional regulation difficulty and self-esteem are negatively associated (Gomez et al., 2018; Mouatsou & Koutra, 2023; Surzykiewicz et al., 2022) since when the individuals have high self-esteem, they are likely to be good at regulating their emotions and have higher emotional intelligence (Rey et al., 2011). People with high self-esteem can express themselves without setting boundaries for their emotions and being overwhelmed by their negative emotions since they have the required emotion regulation skills. As a result, such confidence in themselves might result in experiencing negative affect differently than individuals with low self-esteem. Differential functioning of some cognitive processes for people with low and high self-esteem may be responsible here (Rimes et al., 2023). To illustrate, the study of Smith and Petty (1995) stated that individuals with high self-esteem tend to have different cognitions, such as recalling and thinking of less negative and more positive information, even when manipulated to deduce negative affect. Individuals with low self-esteem, on the other hand, had more negative thoughts when manipulated to induce negative affect. As explained before, numerous studies indicated that maternal reminiscing was related to child's self-development including self-esteem (e.g., Bohanek et al., 2008) Although past maternal reminiscing frequency can be directly linked to recent memory experienced with the mother, they can also be indirectly related to each other via the mediator role of improved self-esteem.

Final mediation hypothesis of the present study was about the indirect effect of self-esteem on recent memory experienced with the mother. When the connection between self-esteem and emotion regulation skills is considered (Gomez et al., 2018; Mouatsou & Koutra, 2023), it would be expected that emotion regulation difficulty would act as a mediator for the relationship between past reminiscing conversations between the mother-child dyad and variables of recent memory related to the mother.

1.6. Child's Gender and Its Effect on Mother-Child Reminiscing

When mothers engage in daily conversations with their children, they may consciously or unconsciously consider the child's gender and talk accordingly, which creates gender differences in mother-child reminiscing literature. When the child's gender is female, the parents may show a tendency to converse in a more voluminous and emotionally intense way as compared to when the child is male (Fivush et al., 2003; Kuebli & Fivush, 1992; Morawska, 2020; Pohárnok & Láng, 2021; Reese et al., 1996). As a result of this parental differences based on a child's gender, gender differences can be observed in child outcomes, too; girls tend to use more total words and emotion words in memory conversations (Aznar & Tenenbaum, 2015; Buckner & Fivush, 2000; Fivush et al., 2003; Reese et al., 1996). Yet, some studies did not claim gender differences in mother-child reminiscing. For instance, Svane et al. (2022) explored the gender differences in emotional memory conversations, and they did not find significant results neither for the parent's gender nor for the child's gender. However, the gender ideology

of the parents and the cultural context may have mattered in finding nonsignificant results since they conducted the study with a sample of Danish citizens of a mainly egalitarian culture. In a cultural context like Turkey, gender differences in mother-child reminiscing are probable to be observed since Turkey's cultural context is still under the influence of traditional gender ideology in many areas of life, including social, working, and within the family (Yüksel-Kaptanoğlu & Bernhardt, 2018). Considering the previous research, the expectation in the present study was that female participants would report more voluminous and emotionally intense written memories they recently experienced with their mothers.

In the meta-analytic review of Aznar and Tenenbaum (2020), 34 studies in the mother-child reminiscing literature were selected and examined regarding how frequent the emotional talk was between the mother-daughter and the mother-son dyads. They could not find significant results about such gender differences regarding the frequency of the mother-child emotion talk. Yet, although they considered the child's age a potential moderator variable, they did not consider or control the effect of emotion talk context. They included not only studies in which emotion talks in the context of reminiscing were investigated but also included studies in which emotion talks in different contexts such as free play or storytelling. Besides, they only looked at the frequency of such talks but not the content or level of elaboration. Overall, it can be concluded that gender differences in the mother-child reminiscing context are worth investigating with a Turkish sample.

1.7. Aims of the Study

This study tried to find out the connections between past maternal reminiscing frequency in childhood, current self-esteem, emotion regulation difficulty, and the emotional content and volume of negatively charged memory related to mother. To our knowledge, this study is one of the relatively rare studies in the literature in this sense. Considering the existing literature, it was hypothesized that individuals' past maternal reminiscing frequency and their recent self-esteem levels would be positively associated with the volume and emotional intensity of the written negative memories about their mothers. Moreover, past maternal reminiscing was expected to be negatively linked to emotion regulation difficulty, while it was thought to be positively related to self-esteem. Emotion regulation difficulty was predicted to be negatively associated with variables of current negative memory about the mother and self-esteem. Hence, serial and parallel mediation models, including self-esteem and emotion regulation difficulty as mediators, were proposed. Since age has been shown to be a correlate of self-esteem and emotion-related variables including emotional regulation and emotional expression in previous research (Bailey et al., 2020; Livingstone & Isaacowitz, 2021; Orth et al., 2018), its effect was decided to be controlled.

Other contributions of the present study were that the mediator roles of self-esteem and emotion regulation difficulty in the relationship between maternal reminiscing in childhood and the volume and emotional content of current negative memory experienced with the mothers were also examined with the simultaneous examination of serial and parallel mediation models. Therefore, it was proposed that emotional regulation difficulty would mediate the relationship between past maternal reminiscing frequency and current negative memory variables, and the relationship between self-esteem and current negative memory variables. It was also anticipated that self-esteem would mediate the association between past maternal reminiscing frequency and emotion regulation difficulty, and the association between past maternal reminiscing frequency and current negative memory variables. The proposed model of the study can be examined in Figure 1.

Figure 1. *The model of the current study.*
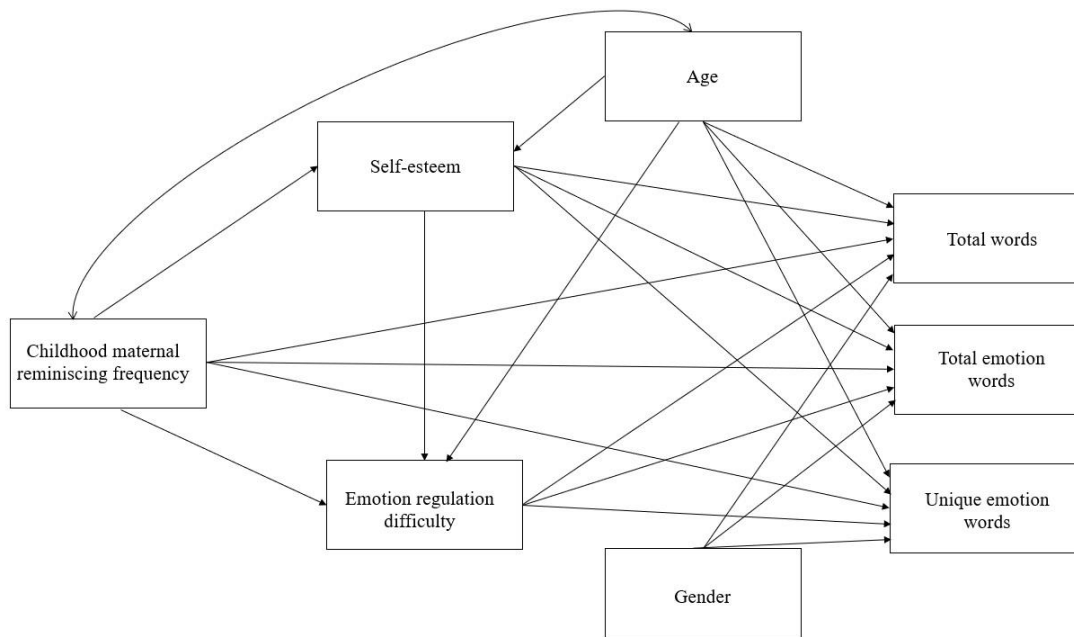
2. METHOD

2.1. Participants

The data for the current study was gathered from 124 participants, which consisted of 84 females (67.7%) and 40 males (32.3%). Using convenience technique in sampling, the questionnaires were shared on social media and with METU psychology students on SONA systems in which undergraduate psychology students get bonus points from the classes in return for their participation in the studies. Eligible participants had their mothers alive and were between the ages of 22 and 40 to target individuals in early adulthood. The age range was between 22 and 39 ($M = 24.88$, $SD = 3.90$). One hundred participants (80.6%) reported a bachelor's degree. Regarding parental information, the most frequent maternal education level was a high school degree [44 mothers (35.5%)]. More than half of the participants [69 participants (55.6%)] reported that they have a single sibling, and almost half of the participants said being the first child in the family [60 (48.4%)]. Moreover, only a few participants stated that their mothers are divorced or widowed [26 (21%)].

2.2. Measures

2.2.1. *Rosenberg self-Esteem scale (RSES) short form*

Rosenberg Self-Esteem Scale (RSES) is a continuous scale with ten items created by Rosenberg (1965) to assess individuals' self-worth. It has a single factor, and half of the items are reverse-coded. Responses are based on the 4-Likert type scale including options from "Strongly disagree" to "Strongly agree". An example item is "I take a positive attitude toward myself.". Psychometric characteristics of the Turkish version of RSES were evaluated by Çuhadaroğlu (1986) and found to be quite good. In our study, Cronbach's alpha value was found as .91.

2.2.2. *Family reminiscence scale (FARS)*

The Family Reminiscence Scale (FARS) was developed by Öner et al. (2020) to measure how frequently individuals used to engage in reminiscing talks with their parents about shared past events when they were a child. The FARS has ten five-point Likert items ranging from "never" to "very frequently". An example FARS item is "How often did your mother talk about any holiday experience that you had when you were a child?". It has a two-factor structure; one factor is "General Recurrent Events Reminiscence" (GRER), and the other one is "First-time Events Reminiscence" (FER). The Cronbach's alpha values were .86 and .75 for the GRER and

the FER, respectively, indicating good levels of internal consistency. In the present study, Cronbach's alpha value for the FER subscale was .81, while Cronbach's alpha value for the GRER subscale was .85.

2.2.3. *Emotion regulation difficulty scale – Short form (DERS-16)*

Emotion Regulation Difficulty Scale – Short Form (DERS-16) was a 5-point-Likert type scale (from 1 point referring to almost never to 5 point referring to almost always) with 16 items (Bjureberg et al., 2016). Higher scores mean more emotion regulation difficulty. Turkish version was created by Yiğit and Guzey Yiğit (2019). It has a five-factor structure, and the factors are "Clarity", "Goals", "Impulse", "Strategies", and "Non-acceptance". For total DERS-16, internal reliability was found as .92. For the factors Clarity, Goals, Impulse, Strategies, and Nonacceptance, internal reliabilities were .84, .84, .87, .87, and .78, respectively. In the present study, reliability coefficient was .94 for total DERS-16. For subscales, coefficients were .85, .83, .89, .90, .84 for Clarity, Goals, Impulse, Strategies, and Nonacceptance, respectively.

2.2.4. *Emotional coding of memories*

The outcome variable, the current negatively charged memory experienced with the mother, was operationally defined as the volume and emotional coding of these memories. Coding scheme of the memories included number of total words, number of total emotion and unique emotion words. The number of total words was counted to observe the volume of each memory reported by the participants. One component of the emotional content was the number of total emotion words used by the participants, including the repetitions of the exact emotion words. The second component of the emotional content was the number of unique emotion words used by the participants, which was calculated by excluding repeated emotion words so that how diverse the emotions used in the memories could be observed. To illustrate, if the participant uses an emotion word such as 'happiness' twice, it is counted as only one unique emotion word since the repetitions are excluded when calculating unique emotion words. However, for instance, if 'happiness' is present in the text twice, it is counted as two total emotion words since total emotion words count the repetitions.

2.3. Procedure

After receiving Middle East Technical University (METU) Ethics Committee's approval (0067-ODTUİAEK-2022), participants filled out an online survey, including an Informed Consent Form, Demographic Information Form, and other scales used. Data collection time was between January and June 2022. For the written memory task, they were instructed to write down one emotionally negative memory they experienced with their mothers in the recent past, detailed as much as possible. To prevent the priming effect, the memory task was at the beginning of the survey after informed consent and demographics forms. No specific time frame for the "recent past" expression was mentioned in the instruction of the memory task. This was because rather than the occurrence time of the event, it was believed that the emotional valence of the event would matter more for participants' remembering. In addition, if we limited the time of the event to the last week or month, the participants who did not experience any specific memory with their mothers in that particular period would not have participated in the study, which would mean some data loss. Hence, we tried to extend the participant pool as much as possible for such a qualitative study, in which it is hard to access participants. Two coders coded the data for the qualitative parts of the study, which included the memory task and its coding. The author, as an experienced coder in emotional coding of reminiscing conversations, trained the second coder, who was a graduate student in the related field of research.

2.4. Data Analysis

In order to investigate the inter-rater reliabilities of the three codes including the number of total words, total emotion, and unique emotion words, intraclass coefficients were calculated

by using IBM Statistical Package for Social Sciences (SPSS) 28. Structural equation modeling was utilized to test the hypotheses of the current study via AMOS 24.0 (Arbuckle, 2013). Specifically, the path analysis with observed variables was performed. Since age was observed to correlate significantly with outcome variables in the model, it was decided to add it as a control variable. Since the three outcome variables were correlated highly significantly, their error terms were covaried in the model. Bootstrapped standard errors were utilized to test indirect effects.

3. RESULTS

3.1. Inter-Rater Reliability of Memory Coding Schemes

To check inter-rater reliabilities for the coding schemes of written memories, a second coder took part in the coding process of the data. The first coder did the emotional coding of memories of the whole sample, and the second coder did the emotional coding of memories for of 20% of the sample. The interrater reliabilities for the number of total words, emotion words, and unique emotion words were found to be as 1.00, .981, and .957 respectively, indicating high levels of reliability (see Table 1).

Table 1. *Interrater reliabilities of the coding schemes.*

Coding Scheme	Intraclass Correlation Coefficient
Total word	1.00
Total emotion word	.981
Unique emotion word	.957

3.2. Descriptives, Intercorrelations, and Path Analysis

The descriptive statistics and intercorrelations among the study variables can be checked on Table 2 and Table 3. Path analysis pointed out that the proposed model exerted a good fit to the data (x^2 (*N* = 124, *df* = 4) = 6.01, *TLI* = .97, *CFI* = .99, *NFI* = .96, *RMSEA* = .06; *p* = .199). The paths from childhood maternal reminiscing frequency to self-esteem (β = .27, *p* = .001 95% CI [.12, .41] was significant. The paths from emotion regulation difficulty to total emotion words (β = .23, *p* = .036, 95% CI [.04, .40] and unique emotion words (β = .28, *p* = .008, 95% CI [.10, .45] were also significant whereas the path from emotion regulation difficulty to total number of words was not significant. The paths from childhood maternal reminiscing frequency to emotion regulation and from childhood maternal reminiscing frequency and from self-esteem to outcome variables (i.e., total words, total emotion words, & unique emotion words) were not significant, contradicting with the mediation hypotheses of the current study. Hence, two proposed parallel mediation models were not supported for the mediator roles of self-esteem and emotion regulation difficulty on the relationships between childhood maternal reminiscing frequency and outcome variables. However, as expected, self-esteem of the individuals was negatively associated with their emotion regulation difficulties (β = -.55, *p* < .001, 95% CI [-.68, -.37]. Additionally, there was a serial mediation for the link between childhood maternal reminiscing frequency and total emotion words and unique emotion words by self-esteem and emotion regulation difficulty. The indirect effect of self-esteem via emotion regulation difficulty on total emotion words (-.12, *p* = .011) and unique emotion words (-.15, *p* = .003) were significant, partially conforming to our expectation since the direction of the relationships between emotion regulation difficulty and outcome variables were the opposite of what was expected. Contrary to our expectations, the indirect effect of self-esteem via emotion regulation difficulty on total number of words was not significant. In line with the expectations, the indirect effect of childhood maternal reminiscing frequency on emotion regulation via self-esteem (-.15, *p* = .005) was significant, too. Unexpectedly, there were not any significant gender differences in current memory variables of total, total emotion, and unique emotion words (see Figure 2).

Table 2. *Descriptive statistics of the study variables (N = 124).*

Variables	Mean	SD	Min	Max
Age	24.88	3.90	22	39
Total words	55.75	45.51	2	234
Total emotion words	4.21	3.38	0	19
Unique emotion words	3.54	2.52	0	13
Self-esteem	28.65	6.42	13	40
Maternal reminiscing frequency	29.11	8.58	10	50
Emotion regulation difficulty	44.33	14.50	20	79

Table 3. *The intercorrelations and internal consistency reliabilities of the study variables (N = 124).*

Variables	1	2	3	4	5	6	7	8	9	10
1. Age	-	-.00	.18*	.31**	-.26**	-.19*	-.18*	.30**	-.17	-.30**
2. Gender		-	.12	.00	-.14	-.07	-.08	.14	-.10	-.04
3. Birth order			-	.60**	-.22*	-.08	-.11	-.00	-.21*	-.07
4. Siblings				-	-.21*	.00	-.03	-.00	-.21*	-.11
5. Total words					-	.73**	.71**	-.26**	-.01	.25**
6. Total emotion words						-	.97**	-.20*	.00	.28**
7. Unique emotion words							-	-.20*	.03	.31**
8. Self-esteem								(.92)	.21*	-.58**
9. Maternal reminiscing frequency									(.88)	-.07
10. Emotion regulation diffiulty										(.94)

Notes. Numbers on the Diagonal are Cronbach's Alpha coefficients. Gender was coded as "0" for females and "1" for males.
** Correlation is significant at the 0.01 level (2-tailed). * Correlation is significant at the 0.05 level (2-tailed).

Figure 2. *The standardized estimations of the proposed mediational model.*
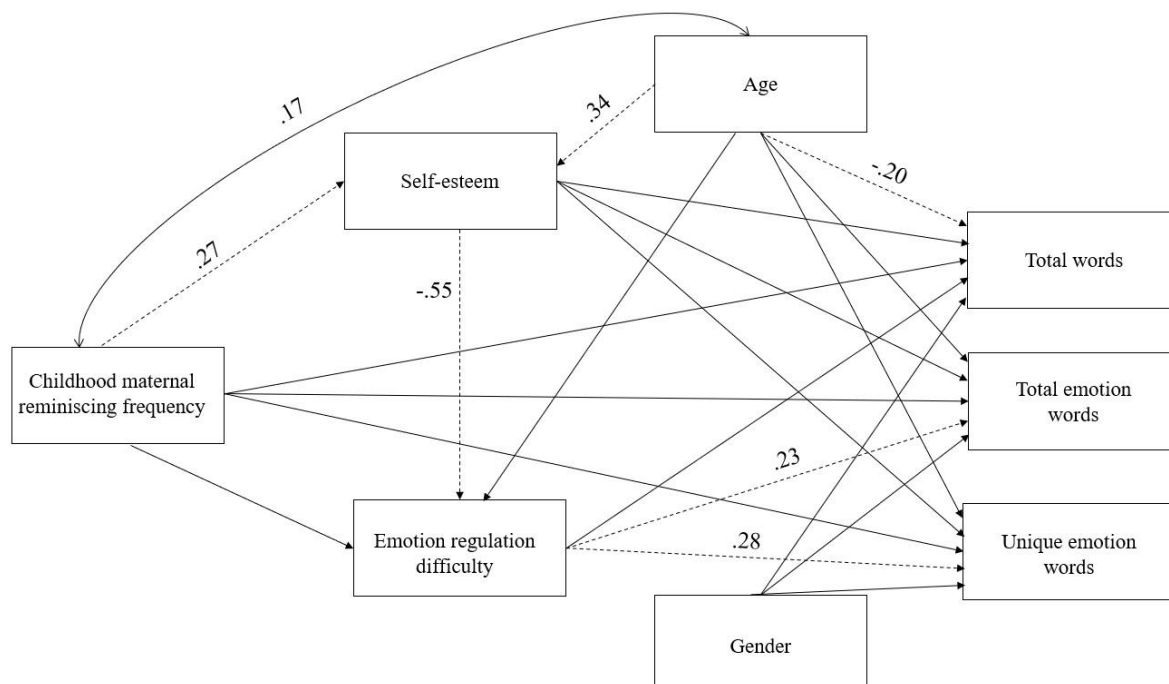
Note. Only significant paths indicated by dashed lines were reported.

4. DISCUSSION and CONCLUSION

In the current study, how the frequency of memory conversations with the mother in childhood may predict the recent negative memory shared with the mother via self-esteem and emotion regulation difficulty were examined. As a promoter of the previous literature, it was found that the frequency of reminiscing conversations in childhood predicted individuals' self-esteem levels positively (Fivush & Nelson, 2006; Harris et al., 2017; Marshall & Reese, 2022; Reese et al., 2007; Song & Wang, 2020). Surprisingly, the direct link between the frequency of shared past event talks with the mother did not predict the variables of individuals' current negative memory shared with the mother (i.e., number of total words, total emotion words, and unique emotion words). This may be stemmed from the fact that only the frequency of shared past talks with the mother was included rather than the volume and the emotional content of the past reminiscence. In the literature, studies finding the link between mother-child reminiscing and future child outcomes generally examined the content of those conversations rather than the frequency (Cook et al., 2023; Fivush & Nelson, 2006; Koh & Wang, 2021; Noel et al., 2019; Peterson et al., 2007; Wang et al., 2010). Moreover, the measurement differences in the two variables can also account for the nonsignificant association between them. While the zoom-in technique, the frequency count of words and emotion words, was utilized for current negative memory variables, the self-report measure in which the construct is given an estimated score was used to quantify the frequency of past maternal reminiscing (Grysman & Mansfield, 2020). Besides, although the Family Reminiscence Scale included items for both negatively and positively charged memories, the current memory shared with the mother had an emotionally negative charge. Future research can reexamine this relationship by using similar measurements for both variables.

The association between maternal past reminiscing frequency in childhood and the individuals' emotion regulation difficulty was insignificant, contradicting the anticipated results. However, self-esteem negatively predicted individuals' emotion regulation difficulties, which was congruent with the expectations. The indirect effect of past reminiscing on emotion regulation difficulty via self-esteem was also significant; the association between past reminiscing frequency and emotion regulation difficulty was fully mediated by self-esteem. The existing literature showed the link between being exposed to emotionally rich reminiscing style and having enhanced emotional knowledge and emotion regulation skills (Goodvin & Romdall, 2013; Laible et al., 2013; Leyva et al., 2021; Mitchell & Reese, 2022; Valentino et al., 2019). Previous research also indicated that when such reminiscing conversations between parents and children include a social aspect like resolving a conflict with others and necessary cognitive coping styles, children's positive self-concepts are likely to be encouraged (Goodvin & Romdall, 2013; Song & Wang, 2020). The current findings suggested that the social dimension of memory talks between parents and children may be a prerequisite for the emotional development of children. Only after parents try to teach their children how to express and control their emotions when they experience a conflict in their social relationships by discussing the adverse event in a detailed way children can have enhanced emotion management skills. One can assume that parents who engage in more frequent memory talks with their children about important life events are also attentive to reminiscing with their children about negatively charged memories from both emotional and social perspectives. Therefore, the significant indirect association between the frequency of past maternal reminiscence and emotion regulation difficulty via self-esteem might have been observed. The direct association can be evident only between the content of memory talks-rather than the frequency-and emotion regulation difficulty. In conclusion, this finding had a unique contribution to the literature in revealing that not only the emotional content but also the frequency of maternal reminiscing in childhood may be a determinant of a child's self-esteem and emotion regulation difficulty.

The emotion regulation difficulty of the individuals predicted the emotional content of the current negative memory shared with the mother. Findings demonstrated that the direction of the relationship was positive, which was in the reverse direction of what had been proposed previously. In other words, the findings implied that as the emotional regulation difficulty increases, number of total emotion words and unique emotion words in written negatively charged memory about the mother also increase. Although the expression of negative affect is thought to be functional in general as an indication of mental health (Berry & Pennebaker, 1993; Campbell-Sills et al., 2006; Chang et al., 2018; Fernandes et al., 2022; Gross & Cassidy, 2019; Roth et al., 2019) and expressive suppression was linked to psychopathology (McCullen et al., 2023; Young et al., 2019), when the frequency and diversity of negative emotions are too many, this would reflect dysfunctionality, which points out a curvilinear relationship between emotion regulation difficulty and emotional expression. As explained by Kennedy-Moore and Watson (2001), negative emotional expression may be like a two-sided medallion having pros and cons simultaneously; the use of negative emotional expression may signalize both the presence of nuisance in the individual and her effort to cope with the negative affect being experienced. Thus, although a child's and parents' mentioning diverse negative emotions and using as many words as possible during a conversation about a shared past event is beneficial for future child outcomes (Fivush & Salmon, 2023; Gross & Cassidy, 2019; Peterson et al., 2007; Reschke et al., 2023; Wang et al., 2010), this association may not still be valid when that child becomes an adult. There might be an ideal level for expressing negative affect in written narratives of adults; when it is too much or too little, it may lose functionality. These may explain why emotion regulation difficulty positively predicted the intensity of emotions in the current memory of individuals. Furthermore, emotional regulation difficulty was not a significant associate of total word usage. It was expected that emotion regulation difficulty was negatively related to number of total words used. This was because in childhood, poor language skills are associated with poor emotion management in many previous studies (Cohen & Mendez, 2009; Ren et al., 2016). Yet, we did not evaluate the participants' language skills in the present study; the number of words used may have nothing to do with adult language skills, although the number of words used by children may indicate their language skills substantially. Participants of the current study may have intentionally preferred to use fewer words for other reasons. Their language skills or proficiencies would have been assessed to see whether there is a connection between language and emotion management in adults.

Self-esteem did not directly predict the volume and emotional content of the current negative memory of the individuals they shared with their mothers, which again contradicted our predictions. The mediator role of self-esteem on the relationship between the frequency of past reminiscence with the mother and the volume and emotional content of the current negative memory shared with the mother was also not supported. However, there was a full mediation in which the indirect effects of self-esteem on total emotion words and unique emotion words were mediated by the emotion regulation difficulties of the individuals, which is in line with the previous literature. Many studies found a positive link between self-esteem and associates of emotion regulation skills such as emotional health, emotional competence, and emotional intelligence (Bibi et al., 2016; Coetzee et al., 2006; Gomez et al., 2018; Moksnes & Espnes, 2012; Mouatsou & Koutra, 2023; Surzykiewicz et al., 2022). All these findings underlined and supported the idea that parents invest in their children's future self-representations and emotional functioning by having frequent reminiscing conversations with them. In this sense, the theoretical implication of the current study would be the validation of the relationship between parental reminiscing, theory of mind, and attachment security. Frequently and regularly engaging in memory talks with the children by discussing children's and other people's internal states, including emotions, reinforces the theory of mind development, and the children can feel the bond between them and their parents and develop a secure attachment style. Thus, such children are observed to have higher self-confidence and potent emotional

management systems (Bohanek et al., 2008; Harris et al., 2017; Pavarini et al., 2013; De Rosnay & Hughes, 2006). In addition, these results suggest that when negative emotions are defined, validated, accepted, and allowed to be expressed under parental guidance during childhood, they can be processed and controlled effectively in adulthood, reducing the need to express and regulate such negative emotions. It seems that when the parents do not meet such needs in childhood, such deficiency still shows itself in their children's adulthood. Parents should know their daily interactions comprise a basis for their children's future quality of life in terms of mental and psychological well-being.

As the final purpose of the present study, gender differences in recent memory variables were tested, and no significant results were evident. One explanation for this finding would be about the sample characteristics in terms of gender ideology. During the interactions with their children, parents tend to exert behaviors mirroring their gender ideology (Gowda & Rodriguez, 2019; Halpern & Perry-Jenkins, 2016; Perales et al., 2021; Schroeder et al., 2019). For instance, it was found by Bulanda (2004) that fathers with more gender-egalitarian viewpoints devote more time with their children in terms of playing, engaging in leisure time activities, and having conversations as compared to fathers with more traditional views. Since the sample was composed of METU students who may have had familial backgrounds in which gender egalitarian values and beliefs might be more internalized, which might have prevented gender differences. Even if their familial backgrounds were more traditional regarding gender ideology, being a university student at METU, which provides a liberal context for the students, may have eliminated gender differences (Yüksek, 2022). As a limitation of the study, METU students may not represent Turkey's population.

A final important note about the nonsignificant gender effect would be the unequal group sizes; since the numbers of male and female participants were unequal, this discrepancy across the two groups may have affected the results. Therefore, this nonsignificant result should be taken cautiously. It was observed that male participants were less eager to participate in a study involving writing down a personal memory shared with their mothers. In fact, such a difference between males and females regarding shyness or self-disclosure can imply potential gender differences in the reminiscing context (Li et al., 2022). Hence, future studies should proceed to investigate gender differences in memory research further by including at least closer numbers of male and female participants. Since the data was composed of a qualitative part, which is demanding and time-consuming, the sample size was relatively low in the current study, which was another study limitation. Other limitations of this study included being a retrospective study based on the participants' self-reports, which means the participants' memory decreased accuracy for past reminiscence may have affected the results. Another limitation would be not having information about the emotional content and frequency of emotional content of past reminiscing talks with the mother in childhood. Therefore, future research may focus on developing a more specific scale measuring both the emotional content and the frequency of mother-child reminiscing. Alternatively, individuals can be asked to talk about or write down the childhood memories they shared with their mothers, and similar studies can be replicated to ensure the results. Similar studies can be designed to include a more representative sample to test potential gender differences in volume and emotional intensity of the current negative memory shared with the mother. Future research may also focus on the possible curvilinear relationship between emotion regulation skills and the expression of negative affect to examine whether there might be an ideal and healthy level of negative emotional expression associated with better emotion regulation skills.

This study was one of the few studies examining the direct and indirect relations between maternal reminiscing frequency in childhood, recent self-esteem and emotional regulation difficulty, and the content of recent written negative memory experienced with the mother. The findings underlined the worth of reminiscing conversations of the mother-child dyads for the

child's future self-esteem, emotion regulation skills, and emotional expression in recent memory. To some parents, simple daily conversations about past events might look trivial at first sight. Still, these results posit that the context of reminiscing conversations serves as a fundamental key for parents to raise a mentally and emotionally healthy individual. Therefore, social policies should target families and educators to train them on the importance of reminiscing conversations for child development. Parents should allocate their time to reminiscence with their children on a regular basis frequently. Apart from the frequency, the content of reminiscing conversations matters, too. Having an elaborate reminiscing style, including open-ended questions such as "wh questions," letting the child express his inner world with his emotions, and accepting and naming the child's positive and negative emotions should be parts of such conversations (Fivush et al., 2009; Marshall & Reese, 2022). Such content information should also be explained to the parents and educators. Preschool and primary school-aged children's families and educators in preschools and primary schools should compose the target populations.

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**: Middle East Technical University, Applied Ethics Research Center, 0067-ODTUİAEK-2022.

Orcid

Nilsu Borhan https://orcid.org/0000-0002-7964-6730

REFERENCES

Ahmadi, M., Abdollahi, M.H., Ramezani, V., & Heshmati, R. (2010). The impact of written emotional expression on depressive symptoms and working memory capacity in Iranian students with high depressive symptoms. *Procedia-Social and Behavioral Sciences*, *5*, 1610-1614. https://doi.org/10.1016/j.sbspro.2010.07.334

Arbuckle, J.L. (2013). Amos (Version 24.0), Computer Program, SPSS/IBM, Chicago.

Aznar, A., & Tenenbaum, H.R. (2015). Gender and age differences in parent–child emotion talk. *British Journal of Developmental Psychology*, *33*(1), 148-155. https://doi.org/10.1111/bjdp.12069

Aznar, A., & Tenenbaum, H.R. (2020). Gender comparisons in mother-child emotion talk: A meta-analysis. *Sex Roles*, *82*, 155-162. https://doi.org/10.1007/s11199-019-01042-y

Bailey, P.E., Brady, B., Ebner, N.C., & Ruffman, T. (2020). Effects of age on emotion regulation, emotional empathy, and prosocial behavior. *The Journals of Gerontology: Series B*, *75*(4), 802-810. https://doi.org/10.1093/geronb/gby084

Berry, D.S., & Pennebaker, J.W. (1993). Nonverbal and verbal emotional expression and health. *Psychotherapy and Psychosomatics*, *59*(1), 11-19. https://doi.org/10.1159/000288640

Bibi, S., Saqlain, S., & Mussawar, B. (2016). Relationship between emotional intelligence and self-esteem among Pakistani university students. *Journal of Psychology & Psychotherapy*, *6(4)*, 279. https://doi.org/1-6.10.4172/2161-0487.1000279

Bird, A., & Reese, E. (2006). Emotional reminiscing and the development of an autobiographical self. *Developmental Psychology*, *42*(4), 613. https://doi.org/10.1037/0012-1649.42.4.613

Bjureberg, J., Ljótsson, B., Tull, M.T., Hedman, E., Sahlin, H., Lundh, L.G., ... & Gratz, K.L. (2016). Development and validation of a brief version of the difficulties in emotion regulation scale: the DERS-16. *Journal of Psychopathology and Behavioral Assessment*, *38*, 284-296. https://doi.org/10.1007/s10862-015-9514-x

Bohanek, J.G., Marin, K.A., & Fivush, R. (2008). Family narratives, self, and gender in early adolescence. *The Journal of Early Adolescence, 28*(1), 153-176. https://doi.org/10.1177/0272431607308673

Bost, K.K., Shin, N., Mcbride, B.A., Brown, G.L., Vaughn, B.E., Coppola, G., ... & Korth, B. (2006). Maternal secure base scripts, children's attachment security, and mother–child narrative styles. *Attachment & Human Development*, *8*(3), 241-260. https://doi.org/10.1080/14616730600856131

Bowlby, J. (1979). The bowlby-ainsworth attachment theory. *Behavioral and Brain Sciences*, *2*(4), 637-638.

Brown, J.D., & Marshall, M.A. (2001). Self-Esteem and Emotion: Some Thoughts about Feelings. *Personality and Social Psychology Bulletin, 27*(5), 575-584. https://doi.org/10.1177/0146167201275006

Brumariu, L.E. (2015). Parent–child attachment and emotion regulation. *New Directions for Child and Adolescent Development*, *2015*(148), 31-45. https://doi.org/10.1002/cad.20098

Buckner, J.P., & Fivush IV, R. (2000). Gendered themes in family reminiscing. *Memory*, *8*(6), 401-412. https://doi.org/10.1080/09658210050156859

Bulanda, R.E. (2004). Paternal involvement with children: The influence of gender ideologies. *Journal of Marriage and Family*, *66*(1), 40-45. https://doi.org/10.1111/j.0022-2455.2004.00003.x

Campbell-Sills, L., Barlow, D.H., Brown, T.A., & Hofmann, S.G. (2006). Acceptability and suppression of negative emotion in anxiety and mood disorders. *Emotion, 6*(4), 587–595. https://doi.org/10.1037/1528-3542.6.4.587

Chang, V.T., Overall, N.C., Madden, H., & Low, R.S.T. (2018). Expressive suppression tendencies, projection bias in memory of negative emotions, and well-being. *Emotion, 18*(7), 925–941. https://doi.org/10.1037/emo0000405

Coetzee, M., Martins, N., Basson, J.S., & Muller, H. (2006). The relationship between personality preferences, self-esteem and emotional competence. *SA Journal of Industrial Psychology*, *32*(2), 64-73. https://doi.org/10.4102/sajip.v32i2.233

Cohen, J.S., & Mendez, J.L. (2009). Emotion regulation, language ability, and the stability of preschool children's peer play behavior. *Early Education and Development*, *20*(6), 1016-1037. https://doi.org/10.1037/12059-004

Cook, O.K., Coffman, J.L., & Ornstein, P.A. (2023). The development of children's autobiographical and deliberate memory through mother–child reminiscing. *Journal of Cognition and Development*, 1-16. https://doi.org/10.1080/15248372.2023.2225620

Çuhadaroğlu, F. (1986). *Adölesanlarda Benlik Saygısı* [Self-esteem in Adolescents] [P.h.D. – Doctor of Philosophy]. Hacettepe University.

Darling Rasmussen, P., Storebø, O.J., Løkkeholt, T., Voss, L.G., Shmueli-Goetz, Y., Bojesen, A.B., ... & Bilenberg, N. (2019). Attachment as a core feature of resilience: A systematic review and meta-analysis. *Psychological Reports*, *122*(4), 1259-1296. https://doi.org/10.1177/0033294118785557

De Rosnay, M., & Hughes, C. (2006). Conversation and theory of mind: Do children talk their way to socio-cognitive understanding?. *British Journal of Developmental Psychology*, *24*(1), 7-37. https://doi.org/10.1348/026151005X82901

Ellis, B.H., Alisic, E., Reiss, A., Dishion, T., & Fisher, P.A. (2014). Emotion regulation among preschoolers on a continuum of risk: The role of maternal emotion coaching. *Journal of Child and Family Studies*, *23*, 965-974. https://doi.org/10.1007/s10826-013-9752-z

Fernandes, B., Newton, J., & Essau, C.A. (2022). The mediating effects of self-esteem on anxiety and emotion regulation. *Psychological Reports*, *125*(2), 787-803. https://doi.org/10.1177/0033294121996991

Fivush, R. (2020). The emergence of autobiographical consciousness and the construction of an autobiographical self. In B., Sahin-Acar, & S., Gülgöz. (Eds). *Autobiographical memory development*. (1st ed., pp. 6-22). Routledge.

Fivush, R. (2007). Maternal reminiscing style and children's developing understanding of self and emotion. *Clinical Social Work Journal*, *35*, 37-46. https://doi.org/10.1007/s10615-006-0065-1

Fivush, R., Berlin, L., McDermott Sales, J., Mennuti-Washburn, J., & Cassidy, J. (2003). Functions of parent-child reminiscing about emotionally negative events. *Memory*, *11*(2), 179-192. https://doi.org/10.1080/741938209

Fivush, R., Brotman, M.A., Buckner, J.P., & Goodman, S.H. (2000). Gender differences in parent-child emotion narratives. *Sex Roles 42*, 233-253. https://doi.org/10.1023/A:1007091207068

Fivush, R., Marin, K., McWilliams, K., & Bohanek, J.G. (2009). Family reminiscing style: Parent gender and emotional focus in relation to child well-being. *Journal of Cognition and Development*, *10*(3), 210-235. https://doi.org/10.1080/15248370903155866

Fivush, R., & Nelson, K. (2006). Parent–child reminiscing locates the self in the past. *British Journal of Developmental Psychology*, *24*(1), 235-251. https://doi.org/10.1348/026151005X57747

Fivush, R., & Salmon, K. (2023). Maternal reminiscing as critical to emotion socialization. *Mental Health & Prevention*, *30*, 200281. https://doi.org/10.1016/j.mhp.2023.200281

Gomez, T., Quiñones-Camacho, L., & Davis, E. (2018). Building a sense of self: The link between emotion regulation and self-esteem in young adults. *UC Riverside Undergraduate Research Journal*, *12*(1). https://doi.org/10.5070/RJ5121039160

Goodvin, R., & Romdall, L. (2013). Associations of mother–child reminiscing about negative past events, coping, and self-concept in early childhood. *Infant and Child Development*, *22*(4), 383-400. http://dx.doi.org/10.1002/icd.1797

Gowda, A.S., & Rodriguez, C.M. (2019). Gender role ideology in mothers and fathers: Relation with parent-child aggression risk longitudinally. *Child Abuse & Neglect*, *96*, 104087. https://doi.org/10.1016/j.chiabu.2019.104087

Graham, S.M., Huang, J.Y., Clark, M.S., & Helgeson, V.S. (2008). The positives of negative emotions: Willingness to express negative emotions promotes relationships. *Personality and Social Psychology Bulletin*, *34*(3), 394-406. https://doi.org/10.1177/0146167207311281

Gross, J.T., & Cassidy, J. (2019). Expressive suppression of negative emotions in children and adolescents: Theory, data, and a guide for future research. *Developmental Psychology, 55*(9), 1938–1950. https://doi.org/10.1037/dev0000722

Grysman, A., & Mansfield, C.D. (2020). Narrative methods in autobiographical memory. In *Autobiographical Memory Development* (pp. 67-83). Routledge.

Halpern, H.P., & Perry-Jenkins, M. (2016). Parents' gender ideology and gendered behavior as predictors of children's gender-role attitudes: A longitudinal exploration. *Sex Roles*, *74*, 527-542. https://doi.org/10.1007/s11199-015-0539-0

Harris, M.A., Donnellan, M.B., Guo, J., McAdams, D.P., Garnier-Villarreal, M., & Trzesniewski, K.H. (2017). Parental co-construction of 5-to 13-year-olds' global self-esteem through reminiscing about past events. *Child Development*, *88*(6), 1810-1822. https://doi.org/10.1111/cdev.12944

Kennedy-Moore, E., & Watson, J.C. (2001). How and when does emotional expression help?. *Review of General Psychology*, *5*(3), 187-212. https://doi.org/10.1037//1089-2680.5.3.187

Koh, J.B.K., & Wang, Q. (2021). Mother–child reminiscing about emotionally negative events and children's long-term mental health. *Frontiers in Psychology*, *12*, 3877. https://doi.org/10.3389/fpsyg.2021.632799

Kuebli, J., & Fivush, R. (1992). Gender differences in parent-child conversations about past emotions. *Sex Roles*, *27*(11-12), 683-698. https://doi.org/10.1007/BF01954572

Lagattuta, K.H., & Wellman, H.M. (2002). Differences in early parent-child conversations about negative versus positive emotions: implications for the development of psychological understanding. *Developmental Psychology*, *38*(4), 564. https://doi.org/10.1037/0012-1649.38.4.564

Laible, D. (2011). Does it matter if preschool children and mothers discuss positive vs. negative events during reminiscing? Links with mother-reported attachment, family emotional climate, and socioemotional development. *Social Development*, *20*(2), 394-411. https://doi.org/10.1111/j.1467-9507.2010.00584.x

Laible, D., Panfile Murphy, T., & Augustine, M. (2013). Constructing emotional and relational understanding: The role of mother–child reminiscing about negatively valenced events. *Social Development*, *22*(2), 300-318. https://doi.org/10.1111/sode.12022

Leyva, D., Catalan Molina, D., Suárez, C., Tamis-LeMonda, C.S., & Yoshikawa, H. (2021). Mother-child reminiscing and first-graders' emotion competence in a low-income and ethnically diverse sample. *Journal of Cognition and Development*, *22*(4), 501-522. https://doi.org/10.1080/15248372.2021.1908293

Li, L., Chen, Y., & Liu, Z. (2022). Shyness and self-disclosure among college students: The mediating role of psychological security and its gender difference. *Current Psychology*, *41*(9), 6003-6013. https://doi.org/10.1007/s12144-020-01099-z

Lightsey, O.R., Jr., Burke, M., Ervin, A., Henderson, D., & Yee, C. (2006). Generalized self-efficacy, self-esteem, and negative affect. *Canadian Journal of Behavioural Science / Revue Canadienne des sciences du comportement, 38*(1), 72-80. https://doi.org/10.1037/h0087272

Livingstone, K.M., & Isaacowitz, D.M. (2021). Age and emotion regulation in daily life: Frequency, strategies, tactics, and effectiveness. *Emotion, 21*(1), 39-51. https://doi.org/10.1037/emo0000672

Marshall, S., & Reese, E. (2022). Growing Memories: Benefits of an early childhood maternal reminiscing intervention for emerging adults' turning point narratives and well-being. *Journal of Research in Personality*, *99*, 104262. https://doi.org/10.1016/j.jrp.2022.104262

McCullen, J.R., Counts, C.J., & John-Henderson, N.A. (2023). Childhood adversity and emotion regulation strategies as predictors of psychological stress and mental health in American Indian adults during the COVID-19 pandemic. *Emotion*, *23*(3), 805-813. https://doi.org/10.1037/emo0001106

Mitchell, C., & Reese, E. (2022). Growing Memories: Coaching mothers in elaborative reminiscing with toddlers benefits adolescents' turning-point narratives and wellbeing. *Journal of Personality*, *90*(6), 887-901. https://doi.org/10.1111/jopy.12703

Moksnes, U.K., & Espnes, G.A. (2012). Self-esteem and emotional health in adolescents– gender and age as potential moderators. *Scandinavian Journal of Psychology*, *53*(6), 483-489. https://doi.org/10.1111/sjop.12021

Morawska, A. (2020). The effects of gendered parenting on child development outcomes: A systematic review. *Clinical Child and Family Psychology Review*, *23*(4), 553-576. https://doi.org/10.1007/s10567-020-00321-5

Mouatsou, C., & Koutra, K. (2023). Emotion regulation in relation with resilience in emerging adults: The mediating role of self-esteem. *Current Psychology*, *42* 734-747. https://doi.org/10.1007/s12144-021-01427-x

Noel, M., Pavlova, M., Lund, T., Jordan, A., Chorney, J., Rasic, N., Brookes, J., Hoy, M., Yunker, W.K., & Graham, S. (2019). The role of narrative in the development of children's pain memories: influences of father–and mother–child reminiscing on children's recall of pain. *Pain*, *160*(8), 1866-1875. https://doi.org/10.1097/j.pain.0000000000001565

Orth, U., Erol, R.Y., & Luciano, E.C. (2018). Development of self-esteem from age 4 to 94 years: A meta-analysis of longitudinal studies. *Psychological Bulletin, 144*(10), 1045–1080. https://doi.org/10.1037/bul0000161

Öner, S., Ece, B., & Gülgöz, S. (2020). Family reminiscence scale: A measure of early communicative context. *Journal of Language and Linguistic Studies*, *16*(2), 849-863. https://doi.org/10.17263/jlls.759327

Öner, S., & Gülgöz, S. (2022). Adults' recollection of the earliest memories: early parental elaboration mediated the link between attachment and remembering. *Current Psychology*, 1-12. https://doi.org/10.1007/s12144-022-03811-7

Pavarini, G., de Hollanda Souza, D., & Hawk, C.K. (2013). Parental practices and theory of mind development. *Journal of Child and Family Studies*, *22*, 844-853. https://doi.org/10.1007/s10826-012-9643-8

Perales, F., Hoffmann, H., King, T., Vidal, S., & Baxter, J. (2021). Mothers, fathers and the intergenerational transmission of gender ideology. *Social Science Research*, *99*, 102597. https://doi.org/10.1016/j.ssresearch.2021.102597

Peterson, C., Sales, J.M., Rees, M., & Fivush, R. (2007). Parent–child talk and children's memory for stressful events. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *21*(8), 1057-1075. https://doi.org/10.1002/acp.1314

Pinquart, M. (2023). Associations of self-esteem with attachment to parents: A meta-analysis. *Psychological Reports*, *126*(5), 2101-2118. https://doi.org/10.1177/00332941221079732

Pohárnok, M., & Láng, A. (2021). Gender Differences in Mother-Child Conversations About Shame and Pride in a Hungarian Sample. *Europe's Journal of Psychology*, *17*(2), 58. https://doi.org/10.5964/ejop.2859

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?. *Behavioral and Brain Sciences*, *1*(4), 515-526.

Reese, E., Bird, A., & Tripp, G. (2007). Children's self-esteem and moral self: Links to parent-child conversations regarding emotion. *Social Development*, *16*(3), 460-478. https://doi.org/10.1111/j.1467-9507.2007.00393.x

Reese, E., Haden, C.A., & Fivush, R. (1993). Mother-child conversations about the past: Relationships of style and memory over time. *Cognitive Development*, *8*(4), 403-430. https://doi.org/10.1016/S0885-2014(05)80002-4

Reese, E., Haden, C.A., & Fivush, R. (1996). Mothers, fathers, daughters, sons: Gender differences in autobiographical reminiscing. *Research on Language and Social Interaction*, *29*(1), 27-56. https://doi.org/10.1207/s15327973rlsi2901_3

Ren, Y., Wyver, S., Xu Rattanasone, N., & Demuth, K. (2016). Social competence and language skills in Mandarin–English bilingual preschoolers: The moderation effect of emotion regulation. *Early Education and Development*, *27*(3), 303-317. https://doi.org/10.1080/10409289.2015.1066639

Reschke, P.J., Clifford, B.N., Brown, M., Siufanua, M., Graver, H., Cooper, A.M., ... & Coyne, S.M. (2023). Links between parent–child conversations about emotions and changes in children's emotion knowledge across early childhood. *Child Development*. https://doi.org/10.1111/cdev.13960

Rey, L., Extremera, N., & Pena, M. (2011). Perceived emotional intelligence, self-esteem and life satisfaction in adolescents. *Psychosocial Intervention*, *20*(2), 227-234. http://dx.doi.org/10.5093/in2011v20n2a10

Rimes, K., Smith, P., & Bridge, L. (2023). Low self-esteem: A refined cognitive behavioural model. *Behavioural and Cognitive Psychotherapy,* 1-16. https://doi.org/10.1017/S1352465823000048

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton University Press.

Roth, G., Vansteenkiste, M., & Ryan, R.M. (2019). Integrative emotion regulation: Process and development from a self-determination theory perspective. *Development and Psychopathology*, *31*(3), 945-956. https://doi.org/10.1017/S0954579419000403

Sales, J.M., & Fivush, R. (2005). Social and emotional functions of mother–child reminiscing about stressful events. *Social Cognition*, *23*(1), 70-90. https://doi.org/10.1521/soco.23.1.70.59196

Salmon, K., & Reese, E. (2016). The benefits of reminiscing with young children. *Current Directions in Psychological Science*, *25*(4), 233-238. https://doi.org/10.1177/0963721416655100

Schroeder, K.M., Bámaca-Colbert, M.Y., & Robins, R.W. (2019). Becoming more egalitarian: A longitudinal examination of Mexican-origin adolescents' gender role attitudes. *Developmental Psychology, 55*(11), 2311–2323. https://doi.org/10.1037/dev0000811

Smith, S.M., & Petty, R.E. (1995). Personality moderators of mood congruency effects on cognition: the role of self-esteem and negative mood regulation. *Journal of Personality and Social Psychology*, *68*(6), 1092. https://doi.org/10.1037/0022-3514.68.6.1092

Song, Q., & Wang, Q. (2020). Mother–child reminiscing about peer experiences in European American and Chinese immigrant families: The impact on children's relational self-concepts. *Asian American Journal of Psychology, 11*(1), 40-48. https://doi.org/10.1037/aap0000162

Surzykiewicz, J., Skalski, S.B., Sołbut, A., Rutkowski, S., & Konaszewski, K. (2022). Resilience and Regulation of Emotions in Adolescents: Serial Mediation Analysis through Self-Esteem and the Perceived Social Support. *International Journal of Environmental Research and Public Health*, *19*(13), 8007. https://doi.org/10.3390/ijerph19138007

Svane, R.P., Zaman, W., Merrill, N., Krøjgaard, P., & Fivush, R. (2022). Gender differences in emotional reminiscing in a Scandinavian sample. *Scandinavian Journal of Psychology*, *63*(3), 173-181. https://doi.org/10.1111/sjop.12802

Valentino, K., Cummings, E.M., Borkowski, J., Hibel, L.C., Lefever, J., & Lawson, M. (2019). Efficacy of a reminiscing and emotion training intervention on maltreating families with preschool-aged children. *Developmental Psychology, 55*(11), 2365-2378. https://doi.org/10.1037/dev0000792

Valentino, K., Nuttall, A.K., Comas, M., McDonnell, C.G., Piper, B., Thomas, T.E., & Fanuele, S. (2014). Mother–child reminiscing and autobiographical memory specificity among preschool-age children. *Developmental Psychology*, *50*(4), 1197. https://doi.org/10.1037/a0034912

Wang, Q., Doan, S.N., & Song, Q. (2010). Talking about internal states in mother–child reminiscing influences children's self-representations: A cross-cultural study. *Cognitive Development*, *25*(4), 380-393. https://doi.org/10.1016/j.cogdev.2010.08.007

Wareham, P., & Salmon, K. (2006). Mother–child reminiscing about everyday experiences: Implications for psychological interventions in the preschool years. *Clinical Psychology Review*, *26*(5), 535-554. https://doi.org/10.1016/j.cpr.2006.05.001

Waters, T.E., Camia, C., Facompré, C.R., & Fivush, R. (2019). A meta-analytic examination of maternal reminiscing style: Elaboration, gender, and children's cognitive development. *Psychological Bulletin*, *145*(11), 1082. https://doi.org/10.1037/bul0000211

Watson, D., Clark, L.A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

Welch-Ross, M.K., Fasig, L.G., & Farrar, M.J. (1999). Predictors of preschoolers' self-knowledge: Reference to emotion and mental states in mother-child conversation about past events. *Cognitive Development*, *14*(3), 401-422. https://doi.org/10.1016/S0885-2014(99)00012-X

Yiğit, İ., & Guzey Yiğit, M. (2019). Psychometric properties of Turkish version of difficulties in emotion regulation scale-brief form (DERS-16). *Current Psychology*, *38*, 1503-1511. https://doi.org/10.1007/s12144-017-9712-7

Young, K.S., Sandman, C.F., & Craske, M.G. (2019). Positive and negative emotion regulation in adolescence: links to anxiety and depression. *Brain Sciences*, *9*(4), 76. https://doi.org/10.3390/brainsci9040076

Yüksek, P. (2022). *University Youth and Politics in Turkey: The Case of Metu Engineering Students* [M.A. - Master of Arts]. Middle East Technical University.

Yüksel-Kaptanoğlu, İ., & Bernhardt, E. (2018). Education, religious practice and gender ideology in Turkey. Stockholm Research Reports in Demography, *16*(1), 3-31. https://doi.org/10.17045/sthlmuni.6281342.v1

APPENDICES

6.1. Demographic Information Form
Yaşınız: _____

Cinsiyetiniz: _____

Kardeşleriniz varsa cinsiyetleri neler ve siz kaçıncı çocuksunuz: _____

Annenizin ulaştığı en yüksek eğitim seviyesi nedir ? _____

Babanızın ulaştığı en yüksek eğitim seviyesi nedir ? _____

Ebeveynlerinizin medeni durumu nedir ? _____

6.2. Negatively Charged Memory About the Mother
1. Yakın zamanda **annenizle** yaşadığınız ve duygusal olarak **OLUMSUZ** bir anınızı düşünmenizi istiyoruz. Lütfen bu anıyı mümkün olan en ayrıntılı biçimde tarif ediniz._____

_____
_____
_____
_____
_____
_____
_____
_____
_____
_____

6.3. Family Reminiscence Scale
1. Anneniz birinci doğum gününüzü size ne sıklıkta anlatmıştır?

Hiç	Çok az	Biraz	Oldukça	Çok

2. Anneniz çocukken nasıl bir evde yaşadığınızı size ne sıklıkta anlatmıştır?

Hiç	Çok az	Biraz	Oldukça	Çok

3. Anneniz çocukken yaptığınız yaramazlıkları size ne sıklıkta anlatmıştır?

Hiç	Çok az	Biraz	Oldukça	Çok

4. Anneniz çocukken geçirdiğiniz hastalık veya yaralanmaları size ne sıklıkta anlatmıştır?

Hiç	Çok az	Biraz	Oldukça	Çok

5. Anneniz çocukken gittiğiniz bir tatili size ne sıklıkta anlatmıştır?

Hiç	Çok az	Biraz	Oldukça	Çok

6. Anneniz çocukluğunuzdaki oyuncaklarınızı size ne sıklıkta anlatmıştır?

Hiç	Çok az	Biraz	Oldukça	Çok

7. Anneniz çocukken sizi ağlatan bir olayı size ne sıklıkta anlatmıştır?

Hiç	Çok az	Biraz	Oldukça	Çok

8. Anneniz size çocukluğunuza ilişkin fotoğraf veya videolar göstererek çocukluğunuz üzerine ne sıklıkta konuşmuşlardır?

Hiç	Çok az	Biraz	Oldukça	Çok

9. Anneniz konuşmaya başladığınızı size ne sıklıkta anlatmıştır?

Hiç	Çok az	Biraz	Oldukça	Çok

10. Anneniz yürümeye başladığınızı size ne sıklıkta anlatmıştır?

Hiç	Çok az	Biraz	Oldukça	Çok

6.4. Rosenberg Self-Esteem Scale

Sizin için uygun olanı seçiniz.

1) Kendimi en az diğer insanlar kadar değerli buluyorum.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

2) Bazı olumlu özelliklerim olduğunu düşünüyorum.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

3) Genelde kendimi başarısız bir kişi olarak görme eğilimindeyim.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

4) Ben de diğer insanların birçoğunun yapabildiği kadar bir şeyler yapabilirim.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

5) Kendimde gurur duyacak fazla bir şey bulamıyorum.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

6) Kendime karsı olumlu bir tutum içindeyim.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

7) Genel olarak kendimden memnunum.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

8) Kendime karsı daha fazla saygı duyabilmeyi isterdim.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

9) Bazen kesinlikle kendimin bir işe yaramadığını düşünüyorum.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

10) Bazen kendimin hiç de yeterli bir insan olmadığını düşünüyorum.
a) Çok doğru b) Doğru c) Yanlış d) Çok yanlış

6.5. Emotion Regulation Difficulty Scale – Short Form

Duygu Düzenleme Güçlüğü Ölçeği-Kısa Form (DDGÖ-16)

Aşağıdaki ifadelerin size ne sıklıkla uyduğunu, her ifadenin yanında yer alan 5 dereceli ölçek üzerinden değerlendiriniz. Her bir ifadenin altındaki 5 noktalı ölçekten, size uygunluk yüzdesini de dikkate alarak, yalnızca bir tek rakamı yuvarlak içine alarak işaretleyiniz.

	Hemen hemen hiç (% 0-% 10)	Bazen (% 11-% 35)	Yaklaşık Yarı yarıya (% 36-% 65)	Çoğu zaman (% 66-% 90)	Hemen hemen her zaman (% 91-% 100)
1. Duygularıma bir anlam vermekte zorlanırım.					
2. Ne hissettiğim konusunda karmaşa yaşarım.					
3. Kendimi kötü hissettiğimde işlerimi bitirmekte zorlanırım.					
4. Kendimi kötü hissettiğimde kontrolden çıkarım.					
5. Kendimi kötü hissettiğimde uzun süre böyle kalacağına inanırım.					
6. Kendimi kötü hissetmenin yoğun depresif duyguyla sonuçlanacağına inanırım.					
7. Kendimi kötü hissederken başka şeylere odaklanmakta zorlanırım.					
8. Kendimi kötü hissederken kontrolden çıktığım korkusu yaşarım.					
9. Kendimi kötü hissettiğimde bu duygumdan dolayı kendimden utanırım.					
10. Kendimi kötü hissettiğimde zayıf biri olduğum duygusuna kapılırım.					
11. Kendimi kötü hissettiğimde davranışlarımı kontrol etmekte zorlanırım.					
12. Kendimi kötü hissettiğimde daha iyi hissetmem için yapabileceğim hiçbir şey olmadığına inanırım.					
13. Kendimi kötü hissettiğimde böyle hissettiğim için kendimden rahatsız olurum.					
14. Kendimi kötü hissettiğimde kendimle ilgili olarak çok fazla endişelenmeye başlarım.					
15. Kendimi kötü hissettiğimde başka bir şey düşünmekte zorlanırım.					
16. Kendimi kötü hissettiğimde duygularım dayanılmaz olur.					

Digital and online tools employed by prospective teachers to access information

N. Tayyibe Ateş [iD][1*], **Nurcan Tekin** [iD][2]

[1]Selçuk University, Faculty of Education, Department of Turkish Language and Social Sciences Education, Konya, Türkiye
[2]Aksaray University, Faculty of Education, Department of Mathematics and Science Education, Aksaray, Türkiye

Abstract: The proliferation of digital information technologies has underscored the heightened necessity for online information search. This study is directed towards the analysis of the strategies adopted by prospective teachers in the selection of digital content employed for accessing information, their preferences concerning the use of digital and online tools, and the rationale that underpins these preferences. To achieve this aim, the case study method was used. Employing the purposive sampling method, the study involved a participant pool of 72 teacher candidates in the 3rd and 4th grades within the faculties of education. To collect data, a personal information form, an online interview form, a focus group interview form, and a word association test were employed. The data were analyzed using the content and descriptive analysis methods. At the end of the study, the prospective teachers expressed a common tendency that when it comes to scientific research, they often prefer to search for articles whereas for more general inquiries, they have a propensity for utilizing internet searches. Most of the prospective teachers follow a step-by-step research process when accessing information. In addition, they employ both goal-oriented and general-content resources while in pursuit of scientific information. They pay attention to data reliability, accurate information, speed, and access to the Internet in the process of obtaining information. They consider the tools they use adequate and reliable. Teacher candidates establish complex conceptual networks among the array of tools they employ to access information. A comprehensive discussion concerning the outcomes of the research is also presented.

1. INTRODUCTION

Today's world is marked by the pervasive influence of digitalization, particularly within the domain of information and communication technologies. This shift has led to a near-total transformation of people's habits and lifestyles. Over the past few years, the emergence of digital online content, programs, and applications has replaced conventional face-to-face lectures and scientific activities, largely attributable to the COVID-19 outbreak. Hence, the utilization of digital and online tools, characterized by their potential to yield both advantageous

and detrimental effects, has gained substantial prevalence today, thus accentuating the critical import of "digital literacy" as a pivotal concept.

In the digitalized global landscape, the skill of accessing information through the processes of interpretation and evaluation stands out as an important skill. However, even though it is often seen as a fast and easy medium, extracting accurate information from the extensive array of information in the digital environment proves to be an intricate undertaking (Topal & Süner, 2020). During daily affairs, the practices of seeking and reaching information employed to fulfill a majority of requirements associated with both personal and corporate work contexts assume substantial significance (Berget & MacFarlane, 2020). The expansion of digital information technologies has elevated the necessity of online information search, and today, students are expected to manage online information. However, not every student is adept at searching and finding information effectively and efficiently. Students can apply different search strategies to obtain information (Zhou & Lam, 2019) and exhibit different information-seeking practices in the digital environment.

Digital literacy skills are necessary to confirm the reliability of the information source and the accuracy of the information encountered on the internet (Hamutoğlu et al., 2017). Adapting to the digitalizing world requires the acquisition of skills encompassing digital information search, along with the proficiency to evaluate and interpret the obtained information. Digital literacy skills extend beyond merely utilizing information and communication technologies or conducting information searches through search engines. Students need to be supported in order to develop the knowledge, skills, and understanding they need (Ministry of National Education [MoNE], 2020). Teachers who are adept at digital literacy skills are capable of offering guidance to students as they navigate the application of these skills. It therefore holds significance to provide prospective teachers (PTs) with digital literacy skills at universities prior to embarking on their careers (Boyacı, 2019; Öztürk, 2020). Consequently, it is essential to establish whether teacher candidates have the requisite knowledge and skills concerning digital literacy.

1.1. Conceptual Framework

It is posited that digital literacy intersects with other interconnected terms such as information literacy, media literacy, computer literacy, network literacy, and e-literacy (Koltay, 2011). Media literacy, computer literacy, and e-literacy constitute skills-oriented literacies that foster the cultivation of information literacy (Martin & Grudziecki, 2006). These 'new literacy skills', including critical and digital literacy types, require new forms of strategic knowledge as they are multifunctional and versatile (Leu et al., 2017). Digital technology, a fundamental aspect of globalization within the educational context, necessitates the acquisition of emerging literacy skills. Teachers are expected to develop their own literacy skills and to guide their students in developing theirs. For this reason, teachers and PTs need to have up-to-date digital literacy skills in order to use online digital tools, artificial intelligence, and virtual applications, as well as traditional literacy skills.

The concept of "digital literacy" was initially formulated by Paul Gilster (1997), who characterized it as the integration of diverse knowledge and skills by students so that they can proficiently engage with the internet environment and its associated educational content. Although Gilster primarily associates this concept with education, Park et al. (2020) state that the concept of digital literacy can also be associated with fields as diverse as science, health, language education, and information and communication technologies. In addition, digital literacy is characterized by Polizzi (2020) as a process involving the adept use of functional and digital skills for accessing information, facilitated by utilizing multiple sources and establishing contextual relationships. For this process to work, individuals need to be able to use a multitude of skills such as scanning, researching, questioning, criticizing, analyzing, synthesizing, evaluating, and decoding.

1.2. Literature Review

The current literature involves several studies that undertake a comparative analysis of the digital literacy levels exhibited by teachers and PTs, with a focus on certain variables (Erdem et al., 2022; Kara, 2021; Korkmaz, 2020). Research focusing on the digital literacy level of teacher candidates (Kozan & Bulut-Özek, 2019; Méndez et al., 2023; Üstündağ et al., 2017) suggests that their digital literacy levels are generally good. However, there are also studies (Altıner, 2019; Boyacı, 2019; Kaya-Özgül et al., 2023; Onursoy, 2018) concluding that their level of digital literacy is insufficient.

The literature contains various studies addressing the process of accessing and evaluating information through online and digital content. Examining students' strategies for searching and interpreting information on the internet, scholarly investigations (Çelen & Seferoğlu, 2017; Enochsson, 2019; Geçer & İra, 2015; Gürsoy, 2019; Kurulgan & Argan, 2007; Yalçınalp & Aşkar, 2003) indicate that students generally exhibit an acceptable level of internet competence and that a connection exists between strategies used for information search and strategies for analysis and synthesis. Further studies carried out on PTs (Kolburan Geçer et al., 2017; Sırakaya & Çakır, 2014; Tekin & Polat, 2017) emphasize that their assessment and problem-solving strategies stand at a moderate level, underscoring the need for improvement in their abilities to select information online, engage in comparisons, and provide solutions. The results of studies (Chen et al., 2019; Turan et al., 2015; Wu & Tsai, 2007) undertaking a comparative analysis of teachers and PTs in various aspects (experience, demographic features, etc.) demonstrate that current teachers do in-depth research and use advanced online search strategies more than teacher candidates.

In some studies, participants are compared according to the difficulty level of the tasks assigned to them. Reisoğlu et al. (2022) contend that changes in PTs' online information-seeking behaviors depend on the complexity of tasks. According to this view, experienced PTs use the strategies of problem-solving, determining the main idea, evaluating, and purposeful thinking in a shorter time when dealing with challenging tasks, while less experienced PTs resort to using irrelevant keywords or clicking on irrelevant links. According to Sendurur et al. (2019), even successful students face difficulties navigating online information searches for complex tasks and are inclined to the practice of copy-pasting. It has been observed that as assigned tasks transition into higher levels of complexity, there is a noticeable shift in information search strategies, leading to a surge in the number of opened tabs and websites, the employment of keywords, and the duration dedicated to research. Lai (2020) also confirms that PTs use both complex (e.g., multiple sources and content) and simple assessment standards (e.g., authority and interactive interface) when searching for information online.

1.3. Rationale

The distinctive aspects that set this study apart from previous research in the literature involve the participation of the PTs pursuing their studies in different departments across two separate universities. Particularly noteworthy is the pivotal role of distance education in shaping the majority of their educational experiences, leading to a more profound acquaintance with digital information technologies. Due to the development of technology, the digitalization of the world, and some extraordinary circumstances, the internet has been the most preferred source for individuals to access information. Cultivating attitudes and perceived competencies to effectively teach and apply technology within practical contexts becomes a crucial endeavor in harnessing the potential of information and communication technologies (Pozas & Letzel, 2023). In the context of mindful internet usage, it is necessary to engage in information search and interpretation as essential steps toward accessing accurate information. For this reason, there is an increasing need to identify PTs' strategies for searching and interpreting digital and online information to determine their information search strategies on the internet, to ensure the development of these strategies, and to obtain more efficient results.

In addition to examining strategies employed by teacher candidates in the selection of digital content for information access in response to the digitalizing world, this study elucidates their preferences for using digital and online tools and explores the motivations underpinning these predilections. In line with research endeavors dedicated to technological progress, this study also holds a dual significance both in the background and forefront for its potential contribution to academic research in the field, as well as its utility in the realm of education. Considering its potential to serve as a valuable resource for future academic endeavors and its role in aiding PTs in their personal development and guiding students in their educational journeys, it carries a two-fold importance.

1.4. Purpose of the Study

Nowadays, it is inevitable for PTs to utilize online and digital information, particularly within the educational environment, and to share this information with their students. Given that teacher candidates will interact with a multitude of students from the outset of their professional careers, their influence on these students becomes a pivotal consideration (Öztürk, 2020). While the concept of distance has nearly been eliminated through the internet, there are still advanced processes that need to be carried out to access information, such as using correct search strategies, selecting, and evaluating information, and performing analysis and synthesis. For instance, PTs may have the opportunity to access a library's database miles away from their location or visit a museum located far away through online tools. However, the question of whether they possess the competence to utilize these opportunities is debatable. Although PTs have the potential to question from which sources they can access scientific knowledge, whether certain information is found in verifiable and reliable sources, and while they take courses at university that should enable them to distinguish scientific knowledge (research techniques), access information from online environments (information technologies), and evaluate the reliability of information sources (information literacy), it cannot be said that these courses are highly applicable in academic life. The low performance of PTs in their assignments and research involving online and digital content throughout their academic careers indicates both the problem and the need for such work. Investigating teacher candidates' digital literacy skills, their preferred digital tools for accessing information, the motivations influencing their preferences, and their strategies for assessing the reliability of information and conducting information evaluations is a critical endeavor. Therefore, this study aims to assess the digital literacy behaviors exhibited by teacher candidates enrolled in different undergraduate programs, describe the tools they employ for accessing information, and elucidate their strategies for selecting these tools. In keeping with the central purpose of the study, the research questions are as follows:

1. What approach do PTs adopt when seeking information during a scientific study?
2. What tools do PTs employ when seeking information during a scientific study?
3. What are the opinions of PTs regarding the significance of the process of accessing information?
4. What are the opinions of PTs regarding the sufficiency and reliability of the digital tools they use during a scientific study for scientific knowledge?
5. How do PTs associate words related to the process of accessing information with other concepts?

2. METHOD

2.1. Design

This research, which investigates the digital and online tool preferences of teacher candidates in accessing information, has been structured based on the case study method as one of the qualitative research designs. The case study method involves examining the phenomenon in

question within its real-world setting (Yin, 2018). In this study, the identified case pertains to the "digital and online tools employed by PTs to access information."

2.2. Participants

A total of 72 teacher candidates currently pursuing their studies within faculties of education located in two cities in Türkiye took part in this research. During the initial phase of involving teacher candidates in the study, the purposive sampling method was employed. Accordingly, interviews were conducted with the PTs who were in their 3rd and 4th grades in various departments of two faculties of education. Participants were included in the study on a voluntary basis. In the faculties of education, there are classes on Information Technologies and Instructional Technologies offered as compulsory courses in the first or second year. All the participants in this study have taken these courses. Table 1 presents the personal information of the participating teacher candidates (with 51 of them reporting they have received training on internet-based information search and 21 stating they have not).

Table 1. *Personal information of PTs.*

Feature		*n*	%
Gender	Female	44	61.11
	Male	28	38.89
Grade level	3rd grade	28	38.89
	4th grade	44	61.11
Major	Science Education	4	5.56
	Mathematics Education	5	6.94
	English Language Education	5	6.94
	Special Education	20	27.78
	Guidance and Psychological Counseling	4	5.56
	Elementary School Teaching	3	4.17
	Social Sciences Education	3	4.17
	Turkish Language Education	28	38.89
Grade point average	<1.50	-	-
	1.51-2.00	-	-
	2.01-2.50	6	8.33
	2.51-3.00	15	20.83
	3.01-3.50	45	62.5
	3.51-4.00	6	8.33
Internet usage in a day	Less than 1 hour	-	-
	1-3 hours	15	20.83
	3-5 hours	30	41.67
	More than 5 hours	27	37.5
Frequency of internet usage	None	-	-
	Rarely	-	-
	Sometimes	10	13.89
	Generally	42	58.33
	Always	20	27.78
Constant internet access	Available	51	70.83
	Unavailable	21	2917
Feature		**f**	**%**
Purpose of internet usage	Research	40	30.53
	Corporate actions	11	8.40
	Game	10	7.63
	Social media	64	48.86
	Other*	6	4.58

* Other purposes: lesson listening-watching (*f*=3), series-movies (*f*=2) and communication (*f*=1).

Based on the data provided in Table 1, it can be observed that the participating teacher candidates in this study consist of 44 females (61.11%) and 28 males (38.89%). The PTs are currently enrolled in either their third year (*n*=28) or fourth year (*n*=44) of study. The teacher candidates from a variety of departments participated in the research, with representation from Turkish Language Education (*n*=28), Special Education (*n*=20), Mathematics Education (*n*=5), English Language Education (*n*=5), Science Education (*n*=4), Guidance and Psychological Counseling (*n*=4), Elementary School Teaching (*n*=3), and Social Sciences Education (*n*=3). The academic grade point average (GPA) of more than half of the participants (*n*=45) falls in the 3.01-3.50 range, and none of the PTs included in the study have a grade point average of 2.00 or below. Among the participating candidates, 15 (20.83%) use the internet for 1-3 hours daily, 30 (41.67%) for 3-5 hours, and 27 (37.5%) for over five hours each day. Most of the participants (*n*=51) always have access to the internet. The internet is most used by the PTs for social media (*f*=64) and research activities (*f*=40). The teacher candidates also report that their preferred device to connect to the internet is a smartphone.

2.3. Data Collection Tools

To collect data, a personal information form, an online interview form, a focus group interview form, and a word association test were employed. The research was approved by Selçuk University Faculty of Education Ethical Committee Report No. 54, dated 05.06.2023. The details of the ethical committee report are kept confidential in accordance with the rules of blind review. Comprehensive details regarding the data collection methods are presented below. The development of data collection instruments involved obtaining the opinions of three experts for each data collection tool. The readability and comprehensibility of the data collection instruments were reviewed and confirmed by a Turkish language education specialist. Based on the input received from a computer education specialist for the online interview form, focus group interview form, and word association test, questions related to digital and online tools and the words to be applied in the word association test were adjusted. Input from a science education specialist was also obtained regarding conducting research in an online environment. The final versions of the data collection instruments were prepared in accordance with the recommendations provided by the experts.

2.3.1. *Personal information form*

The personal information form consists of questions related to gender, department, university affiliation, and grade, among other details. This form was developed by the researchers and made available to the participants in an online format.

2.3.2. *Online interview form*

The interview consists of five open-ended questions as a structured questionnaire, which aims to reveal PTs' preferences for digital and online tools used in accessing information. The research questions focus on how PTs approach a scientific study (such as assignments or exam preparation), the sources they use, the key aspects they consider important in the process of accessing information, and whether the tools they use are adequate and reliable for assignments or scientific knowledge. The interview questions were developed by the researchers based on a review of the literature. The implementation was conducted electronically via Google Forms.

2.3.3. *Focus group interview form*

Focus group interviews are conducted within the framework of a semi-structured interview plan created by researchers for the purpose of providing an overall plan for the interviews, including the date, time, location, participant names, and an overview of the interview stages (Krueger & Casey, 1994). This interview form consists of introductory questions, transition questions, key questions, research questions, and closing questions. The interview form developed for this research was utilized to gain in-depth insights in line with the research aim. Accordingly, the

responses of PTs to the questions in the online interview form, which were reached through easily accessible sampling in the initial phase of the data collection process, were examined. Subsequently, researchers found it necessary to conduct in-depth research to obtain clearer results for certain questions. In this process, focus group interviews were conducted with 15 PTs selected through purposive sampling. The focus group interview form was created to clarify points where responses to interview questions were not understood or where there was a dilemma.

2.3.4. *Word association test*

This test was used to support data obtained from the online interview form and focus group interview form. In the word association test, PTs participating in the focus group interview process were given terms such as internet, digital tools, online tools, artificial intelligence, and information reliability and were asked to generate related words for each keyword within 1 minute. In the literature, it is seen that various periods are given for writing the words to be written for the keyword at different education levels. For example, Özkaral and Akdoğan (2022) in a study conducted with middle school students gave 1 minute. Additionally, Bahar and Özatlı (2003) recommended 30 seconds for high school students; Alaca et al. (2020) used 30 seconds for teacher candidates. In our study, since the data were collected online, it was deemed appropriate to allow 1 minute to avoid any disruption in communication with all participants. The students were informed before the word association test was conducted that the time to be given was determined according to this situation. The word association test aimed to identify PTs' associations with digital and online tools and to determine their conceptual networks/relationships. It is believed that this will contribute to identifying the perceptions of PTs participating in the research regarding fundamental concepts related to accessing information through digital and online tools.

2.4. Data Collection Process

The research data were gathered in two consecutive stages. During the period spanning 2020 to 2022, as was the case worldwide, Türkiye also transitioned from in-person classes and scientific activities to digital online content, programs, and applications as a response to the COVID-19 pandemic. In 2023, Türkiye experienced an earthquake that necessitated the reintegration of digital and online content instead of some face-to-face classes and scientific gatherings. For this reason, the data collection process for this study took place online. The initial phase of the study involved sending the personal information form and interview questions to the PTs through Google Forms. A time frame of around one month was allocated for them to complete the interview questions. The second phase involved conducting word association tests and focus group interviews with the teacher candidates selected through purposive sampling via the Zoom platform. Before the start of the word association process, the participants were given a brief reminder and explanation about the word association test. In the word association process, which took approximately 15 minutes, each participant was tasked with writing down related words to the provided key terms within a 1-minute timeframe. Subsequently, the focus group interview process commenced. During focus group interviews, the key priority is placed on understanding rather than interpreting, identifying diversity instead of generalizing, and elucidating how participants perceive the situation rather than providing explanations (Kreuger & Casey, 1994). As an illustration, in the study, a teacher candidate (PT 25) was included in the focus group process after reporting during the initial interview phase that they conducted general research on the methods of accessing information. The outcome of the focus group interview revealed that general research diversified into conducting article searches, exploring scientific websites, and utilizing textbooks. Hence, the focus group interview served as a valuable asset in accomplishing the goals of this study. In focus group interviews, the required time is usually determined based on the number of participants, and it is possible to conduct the interviews in two rounds to ensure that all participants have the opportunity to provide their in-

depth views. In this study, the focus group interview process was divided into two parts, each lasting 25 and 35 minutes, to allow for an equal share of speaking time for all participants and to facilitate a comprehensive exploration of any unclear aspects. During the interviews, one of the researchers assumed the role of moderator, while the second researcher was responsible for serving as the reporter. The moderator's primary task was to foster a thorough examination of the research topic while directing the questions to the participants, and the reporter played a role in ensuring an equal share of participation time, providing an opportunity for non-speaking participants to contribute, keeping track of time, and documenting key points highlighted during the interview.

2.5. Data Analysis and Validity-Reliability

The data obtained from interview questions and focus group interviews were analyzed using the content analysis method, in which the collected data are converted into codes and then categorized and organized into themes to facilitate interpretation (Yıldırım & Şimşek, 2021). Since there was no coding key in the relevant literature, an open coding method was adopted. Categories were derived from the generated codes. Once the planned number of participants was secured to obtain sufficient data, the initial stage of data analysis was set in motion. In this stage, both authors initially analyzed the responses of the first 10 PTs for each question. Following a comparative examination of the findings obtained by the researchers through their analysis, a 100% agreement was calculated for questions 2, 3, 4, and 5, while question 1 showed a 97% agreement (Miles & Huberman, 1994). Subsequently, the researchers separately analyzed the responses of the remaining teacher candidates. At the end of this process, the initial stage of data analysis was then finalized with an inter-rater agreement of 100% for all questions. In the second stage, the researchers worked together to analyze the word association test and focus group interview data. The data obtained from the word association test were subjected to descriptive analysis by the researchers, and a cut-off point (CP=3) was determined. The cut-off point technique proposed by Bahar et al. (1999) was utilized in determining the cut-off point. The related words with frequencies above the cut-off point were presented in a conceptual network figure. Microsoft PowerPoint was used to create the conceptual network of words. Considering that the data derived from the focus group interviews consistently contributes to either supporting or deepening the interview questions, the researchers integrated their queries into the analysis tables. Examples obtained from the analysis of the focus group interview questions were also added to the relevant sections to ensure clarity. For example, PT 23, who was identified as providing answers in the online interview form that were not clearly understood, and was consequently included in the focus group interview, provided a clearer explanation of their expression regarding scientific research as a result of the focus group interview. Similarly, since the responses of PTs did not clearly indicate the adequacy or reliability of the tools they used for scientific knowledge, the focus group interview was used to provide clarity. Following the focus group interview, PTs provided more detailed information about the adequacy and reliability of these tools. The word association test also provides data supporting how information is sought, the tools used, and perceptions regarding the adequacy and reliability of the information access process.

To ensure internal validity in the study, participant confirmation was obtained at the beginning and continuously throughout the data collection process. Additionally, in the initial part of the data collection process, the data collection tool was directed to PTs in all departments of two universities. This allowed the researchers to conduct the research process without bias or subjectivity since they had no prior knowledge about the participants. To establish external validity in the study, participants were introduced in detail, and detailed information about the data collection tools was provided to them. In the second part of the data collection process, participants were included in the study according to certain inclusion criteria based on purposive sampling to clarify the data. To ensure the reliability of the study, a review of the

literature was conducted in the development of data collection tools, and continuous expert opinions were sought for each data collection tool.

3. RESULTS

3.1. Views on the approaches PTs adopt for searching scientific information

This section provides the teacher candidates' opinions concerning the methods they employ when seeking information during a scientific research study (Table 2).

Table 2. *The approaches the PTs employ in accessing information.*

Categories	Codes	*f*	PT
Scientific research	Article searches	24	2, 5, 8, 12, 13, 14, 15, 16, 19, 21, 23, 24, 31, 38, 42, 55, 58, 61, 63, 65, 68, 69, 72, 25
	Research on scientific websites	12	6, 11, 18, 19, 28, 30, 32, 33, 35, 41, 51, 25
	Thesis search	8	6, 8, 14, 19, 31, 41, 65, 72
General research	General internet searches	29	1, 3, 4, 7, 8, 15, 16, 17, 20, 22, 27, 37, 40, 41, 43, 44, 45, 46, 49, 50, 53, 54, 56, 57, 58, 60, 66, 70, 71
	General research	11	10, 25, 26, 29, 34, 36, 47, 48, 57, 59, 62
Traditional research	Research from textbooks	13	12, 19, 20, 27, 28, 30, 37, 44, 49, 60, 66, 71, 25
	Using lecture notes	6	4, 12, 21, 33, 37, 70
	Research from libraries	3	22, 44, 72
	Using printed materials	1	40
Safe research	Researching trusted websites	6	6, 16, 23, 28, 61, 67
	Using reliable sources	3	9, 28, 39
	Searching on corporate sites	2	14, 68
Contemporary-technological research	Watching videos on the Internet	3	19, 21, 65
	Using artificial intelligence	3	46, 65, 68
	Review distance education resources	1	24
Uncategorized answer	Homework	2	52, 64

According to Table 2, the teacher candidates are conducting their research in the following categories: scientific research (*f*=42), general research (*f*=40), traditional research (*f*=23), safe research (*f*=11), and contemporary-technological research (*f*=7). Among these research types, they mostly access information through general internet searches (*f*=29), article searches (*f*=24), research from textbooks (*f*=13), and general research (*f*=11). The subsequent section offers a selection of distinctive findings.

Several of the PTs stated that they do not adopt a step-by-step approach to accessing information; instead, they rely on a single method. A few illustrative statements are as follows:

> *"I look up articles related to it on the internet"* (PT 13, general internet search/general research)

> *"I use the Google search engine. I usually make use of articles there."* (PT 58, article search/scientific research; general internet search/general research)

Additionally, most of the PTs reported using the step-by-step search process presented in this table. The statements below serve as examples, illustrating how the teacher candidates used this gradual process to verify or integrate information:

> *"I consult the internet for additional information after reviewing the content in my books or course materials. I cross-reference the information with one another. I generally pay attention*

to reading articles on the internet so that the information I gather aligns with previously researched and watched information and instructions." (PT 12, article search/scientific research; research from textbooks/traditional research; using lecture notes/traditional research)

"I start by conducting research on multiple websites to build a foundational understanding of the subject. Then, I verify it using articles and trusted websites." (PT 16, article search/scientific research; general internet search/general research; searching on trusted websites/safe research)

"1) First, I break down the topic into subheadings. 2) I identify reliable sources for research. 3) I carry out sequential research on the subheadings. 4) After the research, I review the information I have gathered. 5) I organize and consolidate the reviewed information into a coherent form." (PT 39, using reliable sources/safe research).

Some teacher candidates have summarized the research process in general terms. The following expressions can serve as examples:

"Understanding the topic - researching sources - evaluating sources - categorizing the information - using the information." (PT 26, general research/general research)

"I conduct research on how to do the assignment, gather information about the assignment - try to access content that supports the assignment." (PT 29, general research/general research)

Some of the PTs sorted the articles into categories. As demonstrated in the following examples, there are statements suggesting that the articles can be sourced from national or international outlets:

"I conduct a relevant source search, read articles from peer-reviewed international journals, examine publications from universities and institutions, conduct a Google Scholar search, and then categorize and analyze the data I obtain according to a specific plan. Or I make use of artificial intelligence tools." (PT 68, article search/scientific research; using artificial intelligence/contemporary-technological research)

"I typically start my research process by reading articles in foreign languages, followed by those written in Turkish, and conclude by conducting research on trusted websites." (PT 23, article search/scientific research)

However, during the focus group interview, the same participant expressed that when it comes to countries other than Türkiye, they select resources specific to the respective foreign country but opt for Turkish-language sources when conducting research related to their field:

"For example, I believe that accessing detailed and accurate information about an uprising that occurred in a country other than Türkiye should be done using the sources of that country. However, when it comes to research related to my own field, I first look at Turkish articles. In the field of science education (my own field), there are experts in Türkiye. I prioritize them, and then I also look at articles from other countries and synthesize the information." (PT 23, article search/scientific research, focus group interview)

"I pay attention to articles that are in Turkish, not English... I believe that translations from English are not suitable for assignments in their full sense. Because I can't translate from English myself, I can't make sense of it..." (PT 65, Article search/scientific research, focus group interview)

One PT's statement below, in which they outlined their article research process as akin to that of an "academician," has truly excited us as researchers:

"I will search on Google Scholar for articles related to the topic. First, I download them one by one, then select the most suitable ones to store on Google Drive. I skim through them briefly to identify the main arguments. I enter my arguments in the search bar of Drive and note how they are discussed in each article on a Word document. Finally, I compile a synthesis." (PT 63, article search/scientific research)

As a result of the focus group interview, it was understood that the PTs who mentioned benefiting from articles used the term "article" in a broad sense, encompassing any piece of writing related to the subject matter rather than strictly referring to a scientific article. It was

also noted that they used the term to indicate content based on current or societal issues. Below are expressions from the participants:

"An article is suitable for me if it contains the information I'm looking for in my assignment in terms of diversity and if it provides a comprehensive explanation..." (PT 25, article search/scientific research, focus group interview)

"It should consist of introduction, development, and conclusion sections. It's important that the article is recent, and its content is relevant to my research or assignment." (PT 65, article search/scientific research, focus group interview)

"For example, I believe that accessing detailed and accurate information about an uprising that occurred in a country other than Türkiye should be done using the sources of that country." (PT 23, article search/scientific research, focus group interview)

3.2. Views on the tools PTs employ for searching scientific information

This section provides the teacher candidates' opinions concerning the tools they employ when seeking information during a scientific research study (Table 3).

Table 3. *Tools used by PTs when searching for scientific information.*

Categories	Codes	f	PT
Purpose-specific sources	Textbooks	15	1, 4, 9, 19, 20, 22, 25, 29, 30, 32, 34, 37, 49, 50, 57
	Google Scholar	15	2, 6, 12, 14, 18, 21, 22, 24, 33, 39, 45, 52, 64, 70, 49
	CoHE*	10	2, 3, 6, 8, 12, 18, 36, 39, 41, 49
	Dergipark	8	3, 12, 14, 28, 36, 42, 66, 68
	TR Index	3	12, 39, 42
	ULAKBİM***	1	39
	Articles and theses with unspecified sources of access	14	8, 10, 17, 26, 27, 31, 32, 34, 37, 38, 47, 55, 68, 69
	Open-course wares	4	11, 30, 68, 70
	PPSE** books	10	13, 15, 27, 28, 30, 40, 41, 53, 54, 61
	Academia	1	25
	Field-specific websites	4	15, 16, 58, 65
	Lecture notes	3	29, 30, 37
	Scientific-academical websites	2	26, 35
	Presentation files	1	37
	University publications	2	68, 70
	Artificial intelligence applications	2	25, 65
General content resources	İnternet	12	1, 4, 5, 20, 29, 31, 47, 49, 50, 53, 56, 71
	YouTube	8	15, 21, 23, 44, 53, 54, 55, 57
	Trusted internet websites	3	28, 62, 68
	Wikipedia	2	8, 43
	Web 2.0 tools	2	44, 65
	Social media	2	65, 70

*CoHE [Council of Higher Education's Thesis Center]: Thesis center in Türkiye where postgraduate theses that are prepared at universities are collected and opened for students and researchers.

**PPSE [Public Personnel Selection Exam]: It is an exam used by the Student Selection and Placement Center for appointments to some staff and positions of public institutions and organizations (SSPC, 2023). Teachers are expected to take this exam and meet certain other requirements in order to work in government institutions.

***ULAKBİM: Turkish Academic Network and Information Center

According to Table 3, the teacher candidates use purpose-specific sources (f=95) and general content sources (f=29) for seeking scientific information. They have stated that they utilize sources such as textbooks (f=15), Google Scholar (f=15), articles and theses for which the source of access has not been specified (f=14), the internet (f=12), and PPSE (Public Personnel

Selection Exam) books (*f*=10) the most. Below are sample statements from teacher candidates regarding the sources they use:

> *"I examine that topic in my teaching principles and methods book from the educational sciences set. To delve deeper, I conduct research online. However, there are very few websites that can be considered reliable sources for educational sciences on the internet. The website I trust the most in this regard is https://dergipark.org.tr/tr/."* (PT 28, PPSE books/purpose-specific sources; trusted internet websites/general content sources; Dergipark/purpose-specific sources)

> *"I generally use YouTube videos as a source. I prefer listening to professors who give lectures on YouTube. Additionally, I purchase source books from some publishing houses."* (PT 57, YouTube/general content sources; textbooks/purpose-specific sources)

> *"For articles, I use Google Scholar, or I benefit from the PDF documents on Telegram. The name of the Telegram group is 'Babilkütüphanesi,' and on the internet, I use the open university's and Gazi University's documents."* (PT 70, Google Scholar/purpose-specific sources; social media/general content sources; open-course wares/purpose-specific sources; university publications/purpose-specific sources).

> *"I usually benefit from articles published in this field. More specifically, I use platforms such as Dergipark and CoHE."* (PT 36, Dergipark/purpose-specific sources; CoHE/purpose-specific sources)

Some of the PTs have provided further specifications of the sources they use:

> *"Lecture notes, Open Education Faculty books, PPSE books, etc. If I were to specify an example source of reference: Genç, F. (2020). Öğretim ilke ve yöntemleri konu öğretimi. İsem Yayıncılık."* (PT 30, textbooks/purpose-specific sources; lecture notes/purpose-specific sources; PPSE books/purpose-specific sources; open-course wares/purpose-specific sources)

> *"PPSE educational sciences books (Pegem, Benim Hocam), http://www.kpsskonu.com/egitim-bilimleri/ogretim-yontem-ve-teknikleri, YouTube videos."* (PT 15, PPSE books/purpose-specific sources; YouTube/general content sources; field-specific websites/purpose-specific sources)

> *"1) CoHE: https://tez.yok.gov.tr/UlusalTezMerkezi/*
>
> *2) Google Scholar: https://scholar.google.com/schhp?hl=tr*
>
> *3) ULAKBİM: https://ulakbim.tubitak.gov.tr/*
>
> *4) The TR Index: https://trdizin.gov.tr/"* (PT 39, CoHE/purpose-specific sources; Google Scholar/purpose-specific sources; the TR Index/purpose-specific sources; ULAKBİM/purpose-specific sources)

> *"Wordwall https://wordwall.net/tr, sosyalbilgilerbiz https://www.sosyalbilgiler.biz/, Pinterest application."* (PT 65, field-specific websites/purpose-specific sources; Web 2.0 tools/general content sources; social media/general content sources)

As a result of the focus group interview, the PTs expressed that they conduct research using artificial intelligence applications. Here is a sample statement:

> *"I believe that the assignments I've completed using artificial intelligence applications are consistent, and they are adequate sources of information for me."* (PT 25, artificial intelligence applications/purpose-specific sources, focus group interview)

3.3. Views on the significance of the information access process

This section provides the PTs' opinions concerning the importance of the process of accessing information (Table 4).

According to Table 4, teacher candidates have expressed the importance of reliability (*f*=63), content (*f*=22), access (*f*=20), time (*f*=15), self-efficacy (*f*=5), and technological infrastructure (*f*=5) in the process of accessing information. Among these categories, information reliability/security (*f*=28), accurate information (*f*=23), speed (*f*=12), and internet access (*f*=9) stand out the most. A few sample responses are presented below.

Table 4. *The PTs' views on the significance of the information access process.*

Categories	Codes	f	PT
Reliability	Information reliability/security	28	2, 3, 6, 8, 9, 11, 12, 13, 16, 22, 23, 28, 29, 34, 35, 37, 39, 45, 47, 49, 52, 55, 58, 59, 62, 63, 66, 67
	Accurate information	23	4, 8, 9, 11, 14, 15, 16, 18, 20, 24, 31, 36, 43, 48, 49, 50, 52, 54, 60, 69, 70, 72, 27
	Comparison of sources	4	6, 10, 72, 27
	Verified information	1	14
	Objectivity	3	22, 28, 29
	Scientific-academical resource	3	30, 38, 64
	Official source	1	39
Access	Internet access	9	1, 7, 27, 32, 40, 61, 65, 70, 71
	Information access	2	19, 33
	Accessibility	1	42
	Source diversity	4	10, 19, 32, 41
	Library access	2	44, 65
	Information diversity	2	12, 26
Content	Intelligibility	4	12, 15, 45, 51
	Simplicity	2	11, 27
	Clarity	3	11, 12, 45
	Inclusion of references	1	12
	Inclusion of in-depth information	1	12
	Brief/concise information	3	15, 48, 47
	Obviousness	2	20, 45
	Consistency	3	22, 25, 26
	Inclusivity	1	33
	Qualification	2	51, 70
Time	Speed	12	2, 4, 5, 8, 12, 21, 32, 34, 40, 53, 61, 66
	Practicality	2	5, 12
	Functionality	1	33
Technological infrastructure	Tools and resources	1	25
	Technological support	1	7
	User-friendly interface	1	12
	Technological equipment	1	70
	Artificial intelligence	1	65
Self-efficacy	Accurate scanning of information	2	24, 54
	Ability to use media tools	2	65, 68
	Foreign language knowledge	1	68

While some teacher candidates emphasized the importance of a single aspect, most of them highlighted the significance of various aspects in the process of accessing information:

> *"To ensure that the content of the accessed source is reliable, I usually click on links with official domain extensions." (PT 3, information reliability/reliability)*

> *"Gathering information from verified and widely recognized websites is the most important issue." (PT 14, verified information/reliability)*

"I prioritize the reliability of sources, user-friendly and clear website interfaces, and books and websites containing the most in-depth information with proper references. I mean, sometimes a website may have limited information about a topic. It might be better not to waste time collecting information individually and gather it from fewer sources." (PT 12, information reliability / security; information diversity/access; clarity/content; inclusion of references/content; inclusion of in-depth information/content; speed/time; practicality/time; user-friendly interface/technological infrastructure)

"The reliability of sources, their consistency with each other, and the fact that the information is not based on belief but on objectivity and experimentation are important to me." (PT 22, information reliability/reliability; objectivity/reliability; consistency/content)

"Inclusivity, accessibility, and functional as well as applicable resources are important." (PT 33, information access/access; inclusivity/content; functionality/time)

"Trying to access more sources and testing the accuracy of information by comparing them is quite important because information can become outdated or there is a possibility of acquiring incorrect information." (PT 72, accurate information/reliability; comparison of sources/reliability)

"I prioritize conducting research from various sources and then compiling, of course." (PT 10, information diversity/access; comparison of sources/reliability)

"The availability of an internet network and access to a good nearby library are important. Having good software and media literacy is also essential. Effective use of artificial intelligence and being able to transform it hold significance as well." (PT 65, internet access/access; library access/access; artificial intelligence/technological infrastructure; ability to use media tools/self-efficacy)

Furthermore, after the focus group interview, one teacher candidate expressed the need to verify information obtained from artificial intelligence due to its virtual nature.

"I have some reservations concerning artificial intelligence. It seems like when it's written by artificial intelligence, it's not real. It's necessary to verify the information from other sources. Looking at just one place is not sufficient for me. I compare all the content that comes up on the internet. I check out more reliable sources." (PT 27, simplicity/content, focus group interview)

3.4. Views on the sufficiency and reliability of the tools PTs use for searching scientific information

This section provides the teacher candidates' opinions concerning the sufficiency and reliability of the tools they employ when seeking information during a scientific research study (Table 5).

Table 5. *PTs' views on the sufficiency and reliability of the tools they use for searching scientific information.*

| Sufficiency/reliability of tools	*f*	%
Sufficient	33	45.83
Generally sufficient	13	18.06
Partly sufficient	9	12.50
Insufficient	17	23.61
Reliable	36	50.00
Generally reliable	13	18.06
Partly reliable	17	23.61
Unreliable	6	8.33

According to Table 5, teacher candidates commonly regard the tools they use as largely sufficient and reliable. However, because the responses of PTs do not clearly specify in what respect they find them sufficient or reliable, these issues were revisited and discussed with the participants during the focus group interview process. According to this, the candidates

generally find the sources they use to be sufficient and reliable for the following reasons: accessing websites recommended by professors, consistency among sources, being based on scientific principles, ensuring consistency in a study, validity and reliability, richness of references, using cult sources (those written by well-known people in the field), including sources with "*edu*" and "*gov*" extensions, including university publications, up-to-dateness of sources, the number of citations for a source, and ease of access.

3.5. Word Association Test Results

The results of the word association test carried out among teacher candidates are delineated below (Figure 1). The PTs wrote a total of 132 words for "internet". Among these, the words social media (*f*=18) and information (*f*=9) were repeated the most. For "digital tools," they wrote a total of 101 words. Among these, the words telephone (*f*=12), computer (*f*=10), smart technologies (*f*=9), and tablet (*f*=8) were repeated the most. For "online tools," they wrote a total of 79 words. Among these, the words social media (WhatsApp, Twitter, Facebook, Instagram, Messenger, and Snapchat) (*f*=24), online meetings (Zoom, Google Meet, and Microsoft Teams) (*f*=11), and convenience (*f*=7) were repeated the most. For "artificial intelligence," they wrote a total of 90 words. Among these, the words robot (*f*=10) and devices (smartphone, microphone, computer, Togg, and 3D glasses) (*f*=7) were repeated the most. For the word "information reliability," they wrote a total of 73 words. Among these, the words precaution (*f*=10), scientific resource (article, thesis, encyclopedia, and book) (*f*=10), and references (*f*=9) were repeated the most. For "scientific resource," they wrote a total of 84 words. Among these, the words article (*f*=10), thesis (*f*=8), and research (*f*=8) were repeated the most.

Figure 1. *Conceptual network of teacher candidates regarding the concepts in the process of accessing information.*
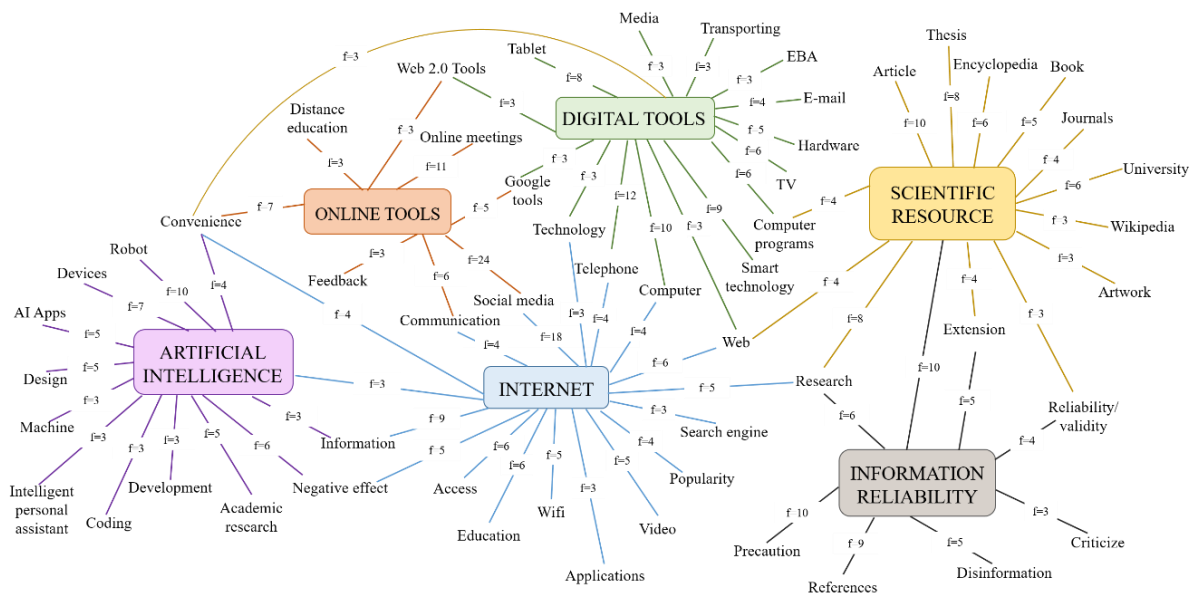
It is also noteworthy that the teacher candidates associated some keywords with multiple common words that are related to one another. For example, they associated the words "internet", "digital tools", "online tools", and "artificial intelligence" with convenience; "internet" and "online tools" with communication and social media; "internet", "digital tools", and "scientific resource" with Web; and "internet", "scientific resource," and "information reliability" with research.

The word association test supports findings for the first research question in a qualitative manner (articles, books, theses, web); it supports findings for the second research question in a

qualitative manner (Wikipedia, books, journals, articles, theses, web, university, web 2.0 tools, Google tools, artificial intelligence); it supports findings for the third research question in a qualitative manner (reliability/validity, convenience, access, devices, references, technology, artificial intelligence, media, scientific resource, information reliability); and it supports findings for the fourth research question in a qualitative manner (references, convenience, scientific resource, reliability/validity, extension, university, popularity, access).

Upon a more in-depth examination of the results, the majority of participants (n=51) mentioned receiving education related to internet information searching. However, among the participants included in the focus group and word association processes for clarification purposes, only 2 of them (PT 25 and PT 54) had received education specifically related to internet information searching. The fact that the majority of participants in the second research process had not received education on this topic aligns with the researcher's decision to include these participants in the process to clarify their responses since they did not provide clear statements in the initial research process.

4. DISCUSSION

This study basically highlights the tools that pre-service teachers use in the process of accessing information and their strategies for selecting these tools. The analysis of the ways that pre-service teachers follow while searching for information during a scientific study shows that they manage the process on the basis of scientific research and general research. In the course of conducting general research, pre-service teachers frequently perform general searches on the internet. However, they stated that they focused on searching for articles in the scientific research process. This result is similar to both the findings obtained by Kurulgan and Argan (2007) and the results obtained by Gürsoy (2019) in terms of the participants' general information search behavior on the internet and their use of search engines. On the other hand, we concluded that the pre-service teachers who stated that they researched articles meant any article with up to date/social content rather than a scientific article through deeper research applied by us. Despite a pre-service teacher summarised the process like a scholar in the article research opens a significant door for us as researchers, the fact that only one pre-service teacher among all participants explained the research process in an academic way revealed that pre-service teachers do not receive adequate support in terms of academic behavior in accessing information. This result highlights the digital literacy skills that need to be developed in pre-service teachers as stated by Sırakaya and Çakır (2014). On the other hand, Çaka et al. (2016 stated that students are not aware of their search orientations, they have incomplete information about how to start searching for information, where they are and where to go, and that the strategy of using appropriate keywords in the searching process should also be developed. Toquero (2021) concluded that multicultural pre-service teachers have novice research skills and that the application of research skills in the real world improves research competencies. Due to the participation of pre-service teachers from various departments, our study reveals the existence of points that need to be developed in the research skills of pre-service teachers in line with the results of this study.

Even though most of the pre-service teachers follow a gradual research process in accessing information, some of them access information over a single source. In our study, pre-service teachers who reached knowledge in a gradual manner stated that they used this process to confirm or integrate the knowledge. Wu and Tsai (2007) concluded that students used the strategy of organizing information, and Geçer and İra (2015) concluded that students used the strategies of integrating information on different sites and organizing them. Zimmermann et al. (2022) concluded that pre-service teachers were competent in using multiple search engines, adapting or reconstructing search terms and selecting at least one Web item from advanced search strategies on the internet. Enochsson (2019) observed that students are capable of finding information on a site when they are guided, but they are unable to establish relationships

between information. The results obtained from this study show that students' awareness levels should be increased in terms of using high-level strategies such as searching, interpreting, selecting, and combining information. This need is supported by the fact that even though the participants in our study stated that they used progressive research strategies to integrate information, they could not even distinguish the content (scientific, social, political, etc.) of the information obtained from the information source as a result of in-depth research. McGarr and Ó Gallchóir (2020) also suggest a greater emphasis on professional development. Despite the statement by Zipke et al. (2019) that the participants gained knowledge and confidence by using educational technologies, yet there were quite large differences between what they learned at university and their classroom experiences, it can be said that there is still a need for this support in the course of university education as a result of our study. As a result of their research, Reisoğlu and Çebi (2020) stated that pre-service teachers should be trained in areas such as information and data literacy or security and problem-solving. Alelaimat et al. (2020) and Avcı and Candan (2023) suggested that teacher education programmes and curricula should be strengthened and improved in order to enable future teachers to provide reliable teaching in technology-integrated classrooms.

As they search for information during a scientific study, pre-service teachers use purpose-oriented and generalised sources. The goal-oriented research results of pre-service teachers overlap with the skill of being goal-oriented, one of the individual skills suggested by List et al. (2020) to be necessary for digital literacy in their study. Yalçınalp and Aşkar (2003) stated that students who used goal-directed information-seeking strategies used planned and purposeful search styles, whereas students who followed data-directed information-seeking strategies used more superficial information-seeking types in scanning styles. Çelen and Seferoğlu (2017) also revealed that students with a data-driven search strategy did not start their research with a specific purpose but used all the information they randomly accessed without making a selection. These students, who have deficiencies in their information-seeking styles, make superficial searches without any purpose. The result obtained from this study is in line with the findings of both Yalçınalp and Aşkar (2003) and Çelen and Seferoğlu (2017). Textbooks, Google Scholar, and articles and theses with unspecified access sources are among the sources most frequently used by pre-service teachers. This result coincides with the results of the studies of Kurulgan and Argan (2007) and Zimmermann et al. (2022) in terms of the use of Google and Yahoo among internet search engines. Sendurur et al.'s (2019) observation that students tend to use websites such as wikipedia.org also supports this result.

Pre-service teachers find reliability, content, access, and time important in the process of accessing information. In particular, information reliability/security, accurate information, speed, and access to the internet stand out. While certain pre-service teachers mentioned the importance of only one feature, most of the pre-service teachers discussed the importance of the process of accessing information considering many features. Aşkar and Mazman (2013) emphasised that the main reason why the Internet is preferred so much is that it eliminates time and space limitations and Ackerman et al. (2020) emphasised the importance of the reliability of the website and consistency in terms of content. Çebi et al. (2022) stated that information and data literacy skills such as accessing accurate information and ways to verify information were used by the participants in the transition from digital competencies to technology integration. In our study, the fact that pre-service teachers found both information-data literacy and multiple features essential in accessing information coincides with this study.

Pre-service teachers find the tools they use adequate and reliable to a great extent. In general, they associated the conditions such as the tools they used being recommended by a reliable person, being consistent, containing rich sources, being based on scientific foundations, and being available on reliable internet pages with the adequacy and reliability of the sources. The fact that pre-service teachers find these tools reliable differs from the results compiled by

Torres-Hernández and Gallego-Arrufat (2022). On the one hand, they stated that the education of individuals about data protection and privacy, ethics, and responsible use of technology should be emphasised, on the other hand, our study shows that pre-service teachers find the sources they use reliably. However, since the pre-service teachers in our study based the reasons for this reliability mostly on the recommendations of the lecturers who helped them in accessing the source, the fact that it was published by experts in the field and reliable internet links, the difference between the results can be seen as normal. In such a case, Reisoğlu and Çebi (2020) suggest that lecturers should structure their courses with technology support, as pre-service teachers may perceive their lecturers as role models. This situation also raises the concern that pre-service teachers may accept all the information given without questioning it. For this reason, it is thought that pre-service teachers should receive more detailed and comprehensive instructions on academic information-seeking behaviors in their education life.

The analysis of the conceptual network of pre-service teachers regarding the process of access to information showed that the concepts of social media, telephone-telephone, computer, online meetings, robots, security, scientific sources, articles and thesis stand out. The analysis of the relationships between these concepts showed that the pre-service teachers associated the words internet, digital tools, online tools and artificial intelligence with "convenience"; internet and online tools with "communication" and "social media"; internet, digital tools and scientific source with "Web"; internet, scientific source and information reliability with "research", indicating that complex conceptual connections were established between these concepts. Especially the feature of being appropriate/favourable among the associated concepts led us to the conclusion that pre-service teachers attach importance to research-oriented to their purposes. The fact that PTs provided examples such as articles, books, theses, etc., when making word associations, illustrates their understanding of the avenues for accessing information. Likewise, their examples such as university, Google tools, artificial intelligence, etc., indicate the tools commonly used when searching for scientific knowledge. The results of Erdoğan and Bozkurt's (2023) study on the metaphors produced by PTs regarding the perceptions of artificial intelligence also show that PTs' perceptions of artificial intelligence point to similar conceptual categories. Additionally, their examples such as reliability/validity, convenience, access, devices, references, technology, artificial intelligence, media, scientific resources, and information reliability highlight the significance of the information access process. Lastly, their use of terms like references, scientific resource, reliability/validity, extension, and convenience largely support their views on reliability in the information access process.

Despite most participants mentioning that they received training related to web searching, the researchers included participants in the second research process to clarify certain expressions, regardless of this background. The fact that most of these participants had not received education on this topic aligns with the idea that the subject needed to be investigated in-depth.

5. LIMITATIONS, FUTURE DIRECTIONS AND CONCLUSION

In this study, data was collected through an online data collection process due to a natural disaster in the country. The inability to conduct face-to-face data collection in this research is a limitation, as it could have allowed participants to access different sources while answering the questions. In studies that could be conducted face-to-face, the results obtained from the simultaneous responses of teacher candidates could be compared with the results of this study in terms of possible different aspects.

Even though most of the participants mentioned searching for articles on the internet, there were very few responses in which the process was properly explained. The implication of this finding extends to offering guidance for both current researchers of this study and future researchers who will address this field in their upcoming studies. Providing participants with

specific tasks in the process of accessing information (such as tasks where researchers determine the information to be accessed, the subject area, and the access method) can allow for a detailed examination of how teacher candidates manage the process. In addition to recommending digital tools that students can use to access information, training should also be provided on strategies for selecting these sources of information and improving their digital skills. It was found that teacher candidates who claimed to search for articles within the scope of scientific research were not accurately aware of what exactly scientific articles mean. Therefore, it is recommended to provide teacher candidates with hands-on training concerning the proper methods of scientific research.

All the participants took courses in Information Technologies and Instructional Technologies, and most of them indicated that they received training in searching for information on the Internet. However, given that Media Literacy is an elective course, not all teacher candidates chose to enroll in it. Hence, it is reasonable to anticipate differences among participants concerning their approaches to information access and their perceptions of information source reliability. It is recommended that the development of skills such as media literacy, digital literacy, information literacy, and technology literacy be encouraged for all teacher candidates.

The integration of the digital world into our daily lives has catalyzed a shift in our information-seeking behaviors. As a result, individuals have developed certain skills such as understanding the source of information, harnessing a variety of access channels as well as digital resources, and being able to connect and contextualize information. The process of evaluating data within the vast expanse of the internet is achievable through the concurrent use of various skills.

In digital and online media, there are numerous options with different content, making it crucial to question the accuracy of acquired information and the reliability of information sources with respect to information literacy and digital literacy. The proper use of digital technologies plays a key role in terms of enabling remote access to information. However, the indiscriminate use of digital content without questioning and awareness can lead to information pollution. Therefore, especially in terms of accessing information, it is important to consider not only the information obtained from various sources but also the scientific validity, reliability, and validity of these sources.

The results of this study indicate that PTs preferred articles for scientific research and internet searches for general research. During their pursuit of scientific information, they tended to use goal-oriented and general content sources. However, they exhibited a limited grasp of the specific definition and extent of what qualifies as an "article". In light of this, it is worth noting that while they professed to adhere to a structured research process for accessing information, they exhibited resemblances to children who still require training wheels on their bicycles. Additionally, it is demonstrated that PTs prioritized information reliability, accuracy, speed, and internet accessibility during the information access process. Nonetheless, participants tended to accept information from trusted individuals without verifying it. Finally, teacher candidates displayed their capacity to establish complex conceptual networks among information access tools and processes.

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**: Selçuk University, 06.06.2023, E-16343714-605.02-531014.

Contribution of Authors

N. Tayyibe Ateş: Investigation, Resources, Software, Formal Analysis, and Writing-original draft. **Nurcan Tekin**: Methodology, Visualization, Supervision, Validation, Formal Analysis, and Writing.

Orcid

N. Tayyibe Ateş ⓘD https://orcid.org/0000-0002-3596-5473
Nurcan Tekin ⓘD https://orcid.org/0000-0002-2848-9739

REFERENCES

Ackerman, R., Yom-Tov, E., & Torgovitsky, I. (2020). Using confidence and consensuality to predict time invested in problem solving and in real-life web searching. *Cognition, 199.* https://doi.org/10.1016/j.cognition.2020.104248

Alaca, M.B., Yaman, H., & Er-Nas, S. (2020). Using the word association test to examine life skills cognitive structures of pre-service science teachers. *Journal of Science Learning, 4*(1), 69-79. https://doi.org/10.17509/jsl.v4i1.25441

Alelaimat, A.M., Ihmeideh, F.M., & Alkhawaldeh, M.F. (2020). Preparing preservice teachers for technology and digital media integration: Implications for early childhood teacher education programs. *International Journal of Early Childhood, 52*, 299-317. https://doi.org/10.1007/s13158-020-00276-2

Altıner, S. (2019). Developing digital literacies of pre-service EFL teachers through engagement with research [Unpublished doctoral dissertation]. University of Bahçeşehir.

Aşkar, P., & Mazman, S.G. (2013). Adaptation of online information searching strategy inventory into Turkish. *Education and Science, 38*(168), 167-182. http://egitimvebilim.t ed.org.tr/index.php/EB/article/view/1288

Avcı, Ü. & Candan, Ö. (2023). Öğretmenlerin bilgi teknolojileri okuryazarlık düzeylerine göre endüstri 4.0 farkındalıklarının incelenmesi [Examination of industry 4.0 awareness of teachers according to their information technology literacy levels]. *Journal of Necmettin Erbakan University Ereğli Faculty of Education, 5*(1), 160-178. https://dergipark.org.tr/t r/pub/neueefd/issue/78300/1287446

Bahar, M., Alex H., Johnstone, & Sutcliffe, R. (1999). Investigation of students' cognitive structure in elementary genetics through word association tests. *Journal of Biological Education, 33*(3), 134-141. https://doi.org/10.1080/00219266.1999.9655653

Bahar, M., & Özatlı, N.S. (2003). Kelime iletişim test yöntemi ile lise 1. sınıf öğrencilerinin canlıların temel bileşenleri konusundaki bilişsel yapılarının araştırılması [Investigation of the cognitive structure of first grade high school students on the basic components of living things using the word communication test method]. *Journal of Balıkesir University Institute of Science and Technology, 5*(2), 75-85. https://dergipark.org.tr/tr/pub/baunfbe d/issue/24783/261831#article_cite

Berget, G., & Macfarlane, A. (2020). What is known about the impact of impairments on information-seeking and searching?. *Journal of the Association for Information Science and Technology, 71*(5), 596-611. https://doi.org/10.1002/asi.24256

Boyacı, Z. (2019). Öğretmen adaylarının yaşam boyu öğrenme eğilimleri ile dijital okuryazarlık düzeyleri arasındaki ilişki (Düzce Üniversitesi Örneği) [The relationship between lifelong learning tendencies and digital literacy levels of pre-service teachers (A Düzce University sample)] [Unpublished master dissertation]. University of Sakarya.

Chen, Y.J., Chien, H.M., & Kao, C.P. (2019). Online searching behaviours of preschool teachers: A comparison of pre-service and in-service teachers' evaluation standards and searching strategies. *Asia-Pacific Journal of Teacher Education, 47*(1), 66-80. https://doi.org/10.1080/1359866X.2018.1442556

Çaka, C., Doğan-Barut, E., & Şahin, Y. (2016). Sosyal ağ kullanan öğrencilerin çevrimiçi bilgi arama stratejilerinin çeşitli değişkenler açısından incelenmesi [The examination of online information searching strategies of students that use social network]. *Trakya Üniversitesi Eğitim Fakültesi Dergisi [Trakya University Journal of Education], 6*(1), 1-13. https://dergipark.org.tr/tr/download/article-file/200421

Çebi, A., Bahçekapılı-Özdemir, T., Reisoğlu, İ., & Çolak, C. (2022). From digital competencies to technology integration: Re-formation of pre-service teachers' knowledge and understanding. *International Journal of Educational Research, 113*. https://doi.org/10.1016/j.ijer.2022.101965

Çelen, F.K. & Seferoğlu, S.S. (2017). İnternet ortamında öğrencilerin bilgi arama biçimleri: Stratejiler, tarzlar, tutumlar, gözlemler ve sorunlarla ilgili bir değerlendirme [Students' information search types on the Internet: An evaluation of strategies, styles, attitudes, observations and problems.]. In Ö. Demirel & S. Dinçer (Eds.). Eğitim bilimlerinde yenilik ve nitelik arayışı [The search for innovation and qualifications in educational sciences] (pp. 1147-1160). Pegem Academy.

Enochsson, A.B. (2019). Teenage pupils' searching for information on the internet. *Information Research, 24*(1). http://InformationR.net/ir/24-1/isic2018/isic1822.html

Erdem, C., Oruç, E., Atar, C., & Bağcı, H. (2022). The mediating effect of digital literacy in the relationship between media literacy and digital citizenship. *Education and Information Technologies, 28*, 4875-4891. https://doi.org/10.1007/s10639-022-11354-4

Erdoğan, Ş. & Bozkurt, E. (2023). Fizik öğretmen adaylarının "yapay zekâ" kavramına ilişkin algılarının incelenmesi: Bir metafor çalışması [Examining the Perceptions of Prospective Physics Teachers on the Concept of "Artificial Intelligence": A Metaphor Study]. *Medeniyet ve Toplum Dergisi [The Journal of Civilization and Society], 7*(2), 55-65. https://dergipark.org.tr/en/pub/metder/issue/81808/1412874

Geçer, A., & İra, N. (2015). Examining information in web environment searching and commitment strategies of university students according to demographic variables. *Education and Science, 40*(179), 383-402. http://dx.doi.org/10.15390/EB.2015.3131

Gilster, P. (1997). *Digital literacy*. John Wiley.

Gürsoy, D. (2019). A critical review of determinants of information search behavior and utilization of online reviews in the decision making process. *International Journal of Hospitality Management, 76*, 53-60. https://doi.org/10.1016/j.ijhm.2018.06.003

Hamutoğlu, N., Güngören, Ö.C., Uyanık, G.K., & Erdoğan, D.G. (2017). Dijital okuryazarlık ölçeği: Türkçe'ye uyarlama çalışması [Adapting digital literacy scale into Turkish]. *Ege Journal of Education, 18*(1), 408-429. https://doi.org/10.12984/egeefd.295306

Kara, S. (2020). Öğretmen adaylarının dijital okuryazarlık düzeyleri ile web ortamında bilgi arama ve yorumlama stratejileri arasındaki ilişkinin incelenmesi [Examination of the relationship between digital literacy levels of teacher candidates and information search and interpretation strategies in the web environment] [Unpublished master dissertation]. University of Necmettin Erbakan.

Kaya-Özgül, B., Aktaş, N., & Çetinkaya-Özdemir, E. (2023). Sınıf öğretmenlerinin ve sınıf öğretmeni adaylarının dijital okuryazarlık düzeylerinin çeşitli değişkenlere göre incelenmesi [Examination of digital literacy levels of primary school teachers and primary school teacher candidates according to various variables]. *Cumhuriyet International Journal of Education, 12*(1), 204-221. https://dx.doi.org/10.30703/cije.1191366

Kolburan-Geçer, A., İra, N., Yenal, H., Bozcan, E., Yalçınkaya, M., & Dinçer, A.T. (2017). Sınıf öğretmenliği bölümü ve beden eğitimi bölümü öğrencilerinin web ortamında bilgi arama-yorumlama stratejilerinin belirlenmesi [The determination of the web-information search and interpret strategies of the students at class-teaching and pe teaching departments]. *International Journal of New Trends in Arts, Sports &Science Education, 6*(3), 16-33. http://www.ijtase.net/index.php/ijtase/article/view/155

Koltay, T. (2011). The media and the literacies: Media literacy, information literacy, digital literacy. *Media, Culture & Society, 33*(2), 211-221. https://doi.org/10.1177/0163443710393382

Korkmaz, M. (2020). *Sınıf öğretmenlerinin dijital okuryazarlık seviyelerinin belirlenmesi* [*Determining digital literacy levels of primary school teachers*] [Unpublished master dissertation]. University of Eskişehir Osmangazi.

Kozan, M., & Bulut-Özek, M. (2019). BÖTE bölümü öğretmen adaylarının dijital okuryazarlık düzeyleri ve siber zorbalığa ilişkin duyarlılıklarının incelenmesi [Examination of department of CEIT teacher candidates' digital literacy levels and cyberbullying sensitivities]. *Fırat University Journal of Social Sciences, 29*(1), 107-120. https://doi.org/10.18069/firatsbed.538657

Krueger, R.A., & Casey, M.A. (1994). *Focus groups: A practical guide for applied research.* Sage.

Kurulgan, M., & Argan, M. (2007). Anadolu Üniversitesi öğrencilerinin internet üzerinden bilgi arama davranışları [Information seeking behavior of Anadolu University students on the internet]. *Journal of Graduate School of Social Sciences, 9*(1), 291-304. https://dergipark.org.tr/tr/pub/ataunisosbil/issue/2819/38001

Lai, C.L. (2020). From organization to elaboration: Relationships between university students' online information searching experience and judgements. *Journal of Computers in Education, 7*, 463-485. https://doi.org/10.1007/s40692-020-00163-8

Leu, J.D., Kinzer, K.C., Coiro, J., Castek, J., & Henry, A.L. (2017). New literacies: A dual-level theory of the changing nature of literacy, instruction, and assessment. *Journal of Education, 197*(2), 1-18. https://doi.org/10.1177/002205741719700202

List, A., Brante, E.W., & Klee, H.L. (2020). A framework of pre-service teachers' conceptions about digital literacy: Comparing the United States and Sweden. *Computers & Education, 148*, 1-20. https://doi.org/10.1016/j.compedu.2019.103788

Martin, A., & Grudziecki, J. (2006). Digeulit: Concepts and tools for digital literacy development. *Innovation in Teaching and Learning in Information and Computer Sciences, 5*(4), 1-19. https://doi.org/10.11120/ital.2006.05040249

McGarr, O., & Ó Gallchóir, C. (2020). Examining supervising field instructors' reporting and assessment of technology use by pre-service teachers on school placement. *Computers & Education*, *146*, 1-11. https://doi.org/10.1016/j.compedu.2019.103753

Méndez, V.G., Suelves, D.M., Méndez, C.G., & Ramon-Llin Mas, J.A. (2023). Future teachers facing the use of technology for inclusion: A view from the digital competence. *Education and Information Technologies, 28*, 9305-9323. https://doi.org/10.1007/s10639-022-11105-5

Miles, M.B., & Huberman, A.M. (1994). *Qualitative data analysis: An expanded sourcebook.* Sage.

Ministry of National Education (MoNE). (2020). *Digital literacy teacher's guide.* http://cdn.eba.gov.tr/kitap/digital/#p=1

Onursoy, S. (2018). Üniversite gençliğinin dijital okuryazarlık düzeyleri: Anadolu Üniversitesi öğrencileri üzerine bir araştırma [Digital literacy levels of university youth: Research on the students of Anadolu University]. *Gümüşhane University e-Journal of Faculty of Communication, 6*(2), 989-1013. https://doi.org/10.19145/e-gifder.422671

Özkaral, T.C., & Akdoğan, V. (2022). Ortaokul öğrencilerinin sivil toplum kuruluşlarına yönelik bilişsel yapılarının kelime ilişkilendirme testiyle incelenmesi [Examining the cognitive structures of secondary school students towards nongovernmental organizations through word associatıon test]. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi, 4*(2), 398-425. https://doi.org/10.38151/akef.2022.25

Öztürk, Y. (2020). Dijital okuryazarlık hakkında lise öğrencilerinin kendilerine ve anne-babalarına yönelik görüşleri Kırıkkale ili örneği [High school students' opinion on digital literacy on self and mother-faters example of Kırıkkale] [Unpublished master dissertation]. University of Gazi.

Park, H., Kim, H.S., & Park, H.W. (2020). A scientometric study of digital literacy, ICT literacy, information literacy, and media literacy. *Journal of Data and Information Science, 6*(2), 116-138. https://doi.org/10.2478/Jdis-2021-0001

Polizzi, G. (2020). Digital literacy and the national curriculum for England: Learning from how the experts engage with and evaluate online content. *Computers & Education, 152,* 1-13. https://doi.org/10.1016/j.compedu.2020.103859

Pozas, M., & Letzel, V. (2023). "Do you think you have what it takes?" – Exploring predictors of pre-service teachers' prospective ICT use. *Technology, Knowledge and Learning, 28,* 823-841. https://doi.org/10.1007/s10758-021-09551-0

Reisoğlu, İ., & Çebi, A. (2020). How can the digital competencies of pre-service teachers be developed? Examining a case study through the lens of DigComp and DigCompEdu. *Computers & Education, 156*, 1-6. https://doi.org/10.1016/j.compedu.2020.103940

Reisoğlu, İ., Çebi, A., & Bahçekapılı, T. (2022). Online information searching behaviours: Examining the impact of task complexity, information searching experience, and cognitive style. *Interactive Learning Environments 30*(3), 417-434. https://doi.org/10.1080/10494820.2019.1662456

Sendurur, E., Efendioğlu, E., Senturk, H., & Calıskan, N. (2019). High achievers' web searching behaviors and patterns in two different task types. *Journal of Educational Multimedia and Hypermedia, 28*(2), 217-238. https://www.learntechlib.org/primary/p/182423/

Sırakaya, M., & Çakır, H. (2014). Determination of teacher candidates' online searching strategies. *Kırşehir Education Faculty Journal (KEFAD), 15*(2), 191-206. https://dergipark.org.tr/tr/pub/kefad/issue/59460/854348

Student Selection and Placement Centre (SSPC) (2023). PPSE: Public Personnel Selection Examination. Turkish Republic Organisation of Student Selection and Placement Centre. https://www.osym.gov.tr/TR,8837/hakkinda.html

Tekin, A., & Polat, E. (2017). Investigation of digital empowerment levels and online information searching strategies of teacher candidates in terms of various variables. *Trakya Journal of Education, 7*(2), 635-658. https://doi.org/10.24315/trkefd.304174

Toquero, C.M.D. (2021). "Real-world:" preservice teachers' research competence and research difficulties in action research. *Journal of Applied Research in Higher Education, 13*(1), 126-148. https://doi.org/10.1108/JARHE-03-2019-0060

Topal, A.D., & Süner, M. (2020). Information searching and commitment strategies of maritime faculty students on the web. *Information Development, 37*(3), 1-13. https://doi.org/10.1177/0266666920903782

Torres-Hernández, N., & Gallego-Arrufat, M.J. (2022). Indicators to assess preservice teachers' digital competence in security: A systematic review. *Education and Information Technologies, 27*, 8583–8602. https://doi.org/10.1007/s10639-022-10978-w

Turan, Z., Reisoğlu, İ., Özçelik, E., & Göktaş, Y. (2015). The examination of teacher's online information-searching strategies in terms of different variables. *Kastamonu Education Journal, 23*(1), 1-16. https://dergipark.org.tr/tr/pub/kefdergi/issue/22600/241445

Üstündağ, M.T., Güneş, E., & Bahçivan, E. (2017). Turkish adaptation of digital literacy scale and investigating pre-service science teachers' digital literacy. *Journal of Education and Future,* (12), 19-29. https://dergipark.org.tr/tr/pub/jef/issue/30777/332813

Wu, Y.T., & Tsai, C.C. (2007). Developing an information commitment survey for assessing students' web information searching strategies and evaluative standards for web materials. *Educational Technology & Society, 10*(2), 120-132. https://www.learntechlib.org/p/75300/

Yalçınalp, S., & Aşkar, P. (2003). Öğrencilerin bilgi arama amacıyla İnterneti kullanım biçimlerinin incelenmesi [Examination of students' use of the Internet for information

seeking]. *The Turkish Online Journal of Educational Technology, 2*(4), 100-107. http://www.tojet.net/articles/v2i4/2415.pdf

Yıldırım, A., & Şimşek, H. (2021). *Sosyal bilimlerde nitel araştırma yöntemleri [Qualitative research methods in the social sciences]* (12nd ed.). Seçkin Publishing.

Yin, R.K. (2018). *Case study research and applications: Design and methods* (6th ed.). Sage.

Zimmermann, M., Engel, O., & Mayweg-Paus, E. (2022). Pre-service teachers' search strategies when sourcing educational information on the Internet. *Frontiers in Education, 7*, 1-15. https://doi.org/10.3389/feduc.2022.976346

Zipke, M., Ingle, J.C., & Moorehead, T. (2019). The effects of modeling the use of technology with pre-service teachers. *Computers in the Schools, 36*(3), 205-221. https://doi.org/10.1080/07380569.2019.1640038

Zhou, M., & Lam, K.K. (2019). Metacognitive scaffolding for online information search in K-12 and higher education settings: A systematic review. *Education Tech Research, 67*, 1353-1384. https://doi.org/10.1007/s11423-019-09646-7

Research Article

Examining the validity and reliability of the academic entitlement scale in Turkish culture

Adem Kantar [1*], **Büşra Aslan Cevheroğlu** [1], **Ayşegül Bakır** [2], **Serel Akdur Çiçek** [3]

[1]Erzurum Technical University, Department of Psychology, Erzurum, Türkiye
[2]Hacettepe University, Division of Guidance and Psychological Counseling, Ankara, Türkiye
[3]Independent Researcher and Clinical Psychologist (PhD), Ankara, Türkiye

Abstract: This study aims to adapt the Academic Entitlement Scale to Turkish culture and examine the scale's measurement invariance according to gender. The study was conducted with 372 university students aged 18 and 54 ($I = 20.90$). The study used the Academic Entitlement Scale, Rosenberg Self-Esteem Scale, Narcissistic Personality Inventory, and Locus of Control Scale as data collection tools. Confirmatory factor analysis was performed to identify the construct validity of the scale. The measurement invariance of the scale according to gender was examined using multi-group confirmatory factor analysis. Correlations between research variables were analyzed with the Pearson Correlation Test. Cronbach's Alpha was performed to measure the internal consistency reliability of the study. As a result of the analyses, it was determined that academic entitlement was positively correlated with self-esteem and external locus of control. In addition, it was revealed that the externalized responsibility sub-dimension was positively correlated to narcissism. The research findings indicate that the scale is a valid and reliable measurement tool within Turkish culture.

1. INTRODUCTION

Academic entitlement (AE) is a problem that has recently increased among university students (Singleton-Jackson et al., 2011). Many students expect to achieve high scores with a low effort from exams or tasks that require performance. Instead of focusing on their work and performance, students with AE concentrate more on what instructors or university management should and should not do for them (Chowning & Campbell, 2009). They expect to leave class early or to be tolerated for being late for class, to be given additional exams to compensate for the exams they cannot take, and to postpone the deadlines for the assignments (Greenberger et al., 2008). These attitudes are also reflected in the students' relationships with the instructors. Students with an inflated sense of entitlement may be more likely to complain about their grades or to demand special treatment from their instructors (Kopp et al., 2011). Accordingly, they place the responsibility for their learning on the trainers rather than themselves (Lippman et al., 2009).

---

*CONTACT: Adem KANTAR ✉ adem.kantar@erzurum.edu.tr 🖳 Erzurum Technical University, Faculty of Letters, Department of Psychology, Erzurum, Türkiye

The effects of the increase in AE are reflected in the educational environment as incompatible characteristics. Students with high levels of academic entitlement tend not to see cheating in exams as an ethical problem (Elias, 2017) and tend to argue with the instructors against their grades due to their performance (Ciani et al., 2008). Studies have shown that AE is positively related to the use of mobile phones during courses (Boswell, 2012). It has been found that AE is associated with the use of mobile phones in the classroom much more strongly than nomophobia (Bhattacharya et al., 2019), which is defined as the fear of disconnection from communication established through mobile phones which considerably affects the use of mobile phones (Reysen et al., 2020). As AE and academic achievement are negatively correlated (Bonaccio et al., 2016; Yan et al., 2021), entitled students tend to exploit and manipulate the efforts of other students (Morrow, 1994; Greenberger et al., 2008). This characteristic of AE, which disrupts the classroom teaching environment, student-instructor interaction, and student-student communication, can potentially harm the value of university education (Morrow, 1994). It is concerning that AE attitudes erode the value of achievement in postsecondary degree attainment through various mechanisms, such as oversimplifying course content or awarding points for non-achievement outcomes such as course attendance. This way, a college degree is made accessible to a wide range of students who may not be deserving. The involvement of people without qualified education in business life deteriorates the quality of work and negatively affects people's well-being (Jack & Donnellan, 2010). To better understand academic entitlement, it seems necessary to consider the concept of entitlement.

1.1. Entitlement and Academic Entitlement (AE)

Most social situations are reciprocity-based (West et al., 2007). Efforts and investments are made to achieve the desired outcome. Entitlement is one's expectation of receiving more than what one gives (Harvey & Martinko, 2008). Entitlement means "unreasonable expectations, especially regarding favorable treatment, or the expectation of automatic submission to one's expectations" (Neville & Fisk, 2019; Harvey & Martinko, 2008). Entitlement is not a prediction that one will achieve a particular outcome but the belief that one should "already" reach a particular outcome (Kopp et al., 2011). Individuals' expectations of high levels of rewards, regardless of their ability and performance levels, is a relatively stable and universal phenomenon (Harvey & Harris, 2010).

AE is a structure that exclusively reflects the manifestation of entitlement in academic situations even though it relates to the idea of non-contextual entitlement (Chowning & Campbell, 2009). Researchers do not distinguish AE from non-contextual entitlement and see entitlement as a component of narcissism (Solomon & Leven, 1987). Accordingly, self-entitled individuals have a sense of superiority and do not need to strive for what others have to strive for because of their excellence (Greenberger et al., 2008). Self-entitled individuals frequently react with anger rather than dissatisfaction when they anticipate a specific result but don't obtain it (Greenberger et al., 2008). Accordingly, it has been determined that entitlement has a positive relationship with hostility, domination, relationship issues, assault, malicious purpose, greed, and stealing other people's property (Campbell et al., 2004). People who have a high sense of entitlement are less likely to forgive and more prone to harbor resentments (Exline et al., 2004). Researchers considering AE as a different structure than non-contextual entitlement suggest that context is essential (Chowning & Campbell, 2009). Students who expect AE may not internalize entitlement in their relationships with their families and peers (Chowning & Campbell, 2009). While people who expect to be entitled in a way that is not context-specific may behave in this way due to the sense of superiority they feel towards others, people who are entitled in an academic context believe that they or their families deserve tolerance because they pay tuition fees (Kopp et al., 2011). AE is positively correlated with perceived competence in the course (Boswell, 2012), self-regulated learning skills (Bonaccio et al., 2016), high socioeconomic status (Côté et al., 2021), external locus of control (Sohr-Proston & Boswell, 2015; Bonaccio

et al., 2016), and academic unreliability (Sohr-Preston & Boswell, 2015). It was observed that AE was higher among students in private universities than in public universities (Kumari & Gautam, 2022). This situation is explained as the creditor-customer understanding brought about by tuition fee payment (Kopp et al., 2011).

1.2. Measurement of Academic Entitlement

Due to the unfavorable psychological and interpersonal consequences of AE, there was a need to measure the AE structure with high validity and reliability. For this reason, scales measuring AE were developed by various researchers (Achacaso, 2006; Aksoy & Coban-Sural, 2022; Greenberger et al., 2008; Chowning & Campbell, 2009; Kopp et al., 2011; Wasieleski et al., 2014). These scales are terms regarding of the definition and conceptualization of AE and their validity and reliability levels. For example, not all items in the Academic Entitlement Scale developed by Achacaso (2006) are attitudes and behaviors that reflect academic entitlement. Additional research should be conducted to assess the suitability of the items. Wasieleski et al. (2014) and the scale developed by Greenberger et al. (2008), the construct validity of the scale was not analyzed. Therefore, it is not known whether the items reflect one or more dimensions. Another scale, Kopp et al. (2011), is theoretically well structured and has good construct validity. However, the authors of the scale state that the scale needs an additional dimension within the scope of student's incivilization. The scale, developed by Kopp et al. (2011), was adapted to Turkish by Kurtyılmaz (2019). In this adaptation, the Cronbach Alpha value of the scale was found to be 0.69, which indicates that the scale is moderately reliable in terms of consistency (Kılıç, 2016). In addition, three items in the original scale were removed from the adapted scale, and the original structure of the scale was changed. Measurement experts do not find it appropriate to remove items from the original scale in adaptation studies (Yalçın, 2021). Therefore, it is unknown whether the scale adapted by Kurtyılmaz (2019) accurately measures the AE structure. In addition, the Academic Entitlement Expectation Scale was developed by Aksoy and Coban-Sural (2022) in the Turkish sample. The validity and reliability values of the scale are good (between .68 and .83). The Academic Entitlement Expectation Scale is a four-dimensional, 21-item scale. The dimensions of the scale include academic narcissism, individual entitlement expectations, general entitlement expectations and externalized responsibility dimensions. This scale has been newly introduced into Turkish and has not yet been used in other Turkish samples other than the sample for which it was developed. The measurement tool Chowning and Campbell (2009) developed is a frequently used scale in the international literature to measure the concept of academic entitlement. The scale is structured in the context of student's incivilization. Its internal validity is high, and its construct validity has been achieved. Its structure in Turkish has not yet been confirmed.

Gender is among the demographic variables frequently used in the field of educational sciences and psychology. In studies conducted on the concept of academic entitlement, gender differences have been frequently examined in the literature, and it was determined that academic entitlement differs between genders (Boswell, 2012; Sohr-Preston & Boswell, 2015; Kumari & Gautam, 2022). Many studies indicate that entitlement is more common and high in male students than female students (Aksoy & Coban-Sural, 2022; Hill & Fischer, 2001; Ciani et al., 2008; Boswell, 2012; Sohr-Preston & Boswell, 2015; Kumari & Gautam, 2022). To see whether the level of academic entitlement will differ significantly according to gender in future studies, this study aims to examine the measurement invariance of the scale according to gender. Measurement invariance refers to the examination of whether the characteristics of a scale are invariant in different groups. Besides, a scale with measurement invariance can objectively measure the characteristics of subgroups. For this reason, on a scale with measurement invariance, it can be seen that the differences obtained between the groups are not caused by the measurement tool but rather by individuals (Wicherts, 2007). In this respect, measurement invariance is a validity test (Basusta & Gelbal, 2015). When it comes to the concept of academic

entitlement, it is important whether the mean scores obtained from the scale differ according to gender. Studies in the literature report that men have higher academic entitlement score averages than women (Keith et al., 2008; Sohr-Preston & Boswell, 2015). When the literature on this subject was examined, it was seen that no study exists examining the measurement invariance related to AE by gender. For this reason, it is necessary to examine whether AE has a measurement invariance regarding gender differences.

1.3. Current Study

When the concept of academic entitlement was first examined, it was thought that it was a phenomenon belonging to North America and the West. However, recent studies have started to reveal the universal nature of academic entitlement (Blincoe & Garris, 2017). Studies conducted in different countries on this subject indicate that AE is not only in a certain group or field; indicates that it can be seen universally and in all educational environments (e.g., Aksoy & Sural, 2022; Kantar et al., 2023; Pilotti et al., 2022; Reysen et al., 2022; Yan et al., 2021; Zarei, 2022). This shows that the measured construct has the same meaning in all cultures compared. The Academic Entitlement Scale (Chowning & Campbell, 2009), which is planned to be adapted within the scope of this study, was developed in the US culture. The USA is the place where the concept of equity was first examined in terms of education (Dubovsky, 1986) and is rich in academic equity literature. In other words, studies shaping the structure of the concept of AE were conducted in the USA (Achacaso, 2006; Greenberger et al., 2008; Chowning & Campbell, 2009; Kopp et al., 2011; Wasieleski et al., 2014). The Academic Entitlement Scale developed by Kopp et al. (2011), which was previously adapted from the US culture to Turkish, was also developed in the US culture, and in the adaptation study (Kurtyılmaz, 2019), it was seen that it also has an equivalent in Turkish culture. In summary, since the structure to be measured is the same in both cultures, it is appropriate to conduct adaptation studies (Hambleton & Patsula, 1999). Besides, the sample for the scale was selected from a university in Oklahoma, USA. The participants in the sample are international, representing many different races and cultures, such as Caucasian, Black, Native American, Asian, and Hispanic. Intercultural examinations enable understanding the structure of AE in all aspects and identifying the sources of intercultural differences. Detailed examinations are needed to determine the extent to which AE is seen in Türkiye and the extent to which the proper functioning of the education system is affected. In addition, the Academic Entitlement Expectation Scale was developed by Aksoy and Coban-Sural (2022) in the Turkish sample. The dimensions of the scale include the academic narcissism dimension in addition to the entitled expectations and externalized responsibility dimensions in the scale of Chowning and Campbell (2009). However, comparing cross-cultural characteristics with the same measurement tool structure is more effective in making cultural inferences (Hambleton & Patsula, 1999). Additionally, researchers have more confidence in a well-known measurement tool than a newly developed one (Hambleton & Patsula, 1999). For this reason, there is a need to adapt a measurement tool frequently used in the literature to Turkish. However, due to significant differences and deficiencies in the measurement and conceptualization of AE in existing scales, it is unclear whether AE is assessed appropriately or accurately. Therefore, this study aimed to adapt the Academic Entitlement Scale (Chowning & Campbell, 2009) to the Turkish culture. In this regard, the research questions of the study are as follows:

1. Validity of Factor Structure and Item Loadings:
    - How valid are the Academic Entitlement Scale's factor structures and item loadings in the context of its adaptation to Turkish culture?
    - To what extent do the correlations between the scale and criterion variables demonstrate statistical significance?

2. Reliability Across Sub-Dimensions:
    - What is the level of internal consistency within the adapted Academic Entitlement Scale

- How does internal consistency fluctuate across various sub-dimensions of the scale?

3. Measurement Invariance and Structural Consistency Across Gender:

- To what degree is the adapted Academic Entitlement Scale measurement invariant across different gender groups, namely females and males?

- How does structural consistency within the scale's sub-dimensions vary between these gender groups?

2. METHOD

This section provides information on the research design, research sample, data collection tools, data collection process, and data analysis procedures.

2.1. Research Design and Participants

This study was conducted with a quantitative research method and a relational research design. The study participants were convenience sampled. Convenience sampling selects participants based on their availability and proximity to the researcher. Convenience sampling prioritizes participant accessibility over study objectives (Campbell et al., 2020). When determining the study's sample size, the researchers considered the following rules that are generally accepted in confirmatory factor analysis studies: (1) a minimum of 10 participants per item (Bentler & Chou, 1987), or (2) a minimum sample size of 200 (Kline, 1994). Considering that the scale used in the study consists of 15 items, the sample size selected following these rules increases the study's statistical power.

372 university students participated in the research. Of these, 36% ($n = 135$) were first-year undergraduate students, 28% ($n = 104$) were second-year undergraduate students, 21% ($n = 79$) were language preparatory class students, and 15% ($n = 55$) were third- and fourth-year undergraduate students. The participants' ages ranged from 18 to 54 ($\bar{X} = 20.90$, $s = 2.85$). 303 participants (81.5%) were female, and 69 (18.5%) were male. The majority of participants (63.4%) self-identified as socioeconomically middle class.

2.2. Process

We followed the steps outlined by Hambleton and Patsula (1999) for the translation of the scale items from English to Turkish and then from Turkish to English, culminating in a final translation back to Turkish. This process involved consulting with a team comprising experts in measurement, language, educational sciences, and psychology, all proficient in Turkish and English. Initially, we translated the scale items from English to Turkish. During this phase, we collectively requested revisions from the experts related to the scale's second, third, and fifth items. Following the consensus of the experts, we updated the items. In the second phase, we performed the reverse translation of the scale from the target language back to the original language. At this stage, we asked the experts to evaluate whether there were any differences in meaning, clarity, and grammar between the first and second versions of the scale for each item. We asked the experts to score the differences between the versions on a scale of 0 to 10. The experts unanimously agreed that the two forms of the scale were similar. After this phase, we conducted a final translation process from English to Turkish, akin to the first step. Experts in language, psychology, educational sciences, and measurement independently reviewed the final Turkish version of the scale. The experts collectively reported that the items were consistent with the original items regarding meaning, clarity, and grammar. We have presented the final version of the scale in Appendix. To ensure clarity, understandability, and grammatical correctness, the trial form of the Turkish version of the scale was presented to 10 undergraduate students before the data collection phase began. After the students' feedback was evaluated, the data collection phase commenced to assess the scale's construct validity. The research participants were university students in Türkiye, the same population as the scale's original sample (Chowning & Campbell, 2009). Participants participated in the study on an online

survey platform. Participants were informed about the study with an informed online consent form before they participated in the study. The research data were obtained in the fall semester of the 2021-2022 academic year. It took approximately 10 minutes for each participant to complete the research form. Permission was obtained from the researchers who developed the scale to adapt the Academic Entitlement Scale to Turkish. Ethics committee approval of the study was obtained from a state university in Türkiye (approval number: 10/5). The collection, analysis, interpretation, and storage of research data adhered to the Helsinki Declaration on Human Rights.

2.3 Data Collection Tools
2.3.1. Personal information form

The personal Information Form was developed by the researchers in order to determine the demographic characteristics of the participants, such as gender, socioeconomic level, age, and class. The participants filled in the personal information form after the informed consent form during the data collection process.

2.3.2. Academic entitlement scale (AES)

AES was developed by Chowning and Campbell (2009), and it measures a person's propensity to expect success without a sense of personal responsibility. AES is a seven-point Likert-type scale consisting of two factors and 15 items. The total score of Academic Entitlement is obtained by summing the scores of both factors. According to the confirmatory factor analysis findings performed to test the construct validity of the scale, it was reported that the structure consisting of two factors with 15 items was confirmed ($\chi^2 = 410.08$, *GFI = .938, CFI = .897, RMSEA = .064*). Regarding the reliability analysis of the scale, the internal consistency Cronbach Alpha coefficient value for the 10-item Externalized Responsibility sub-dimension was found to be .81 and for the five-item Entitled Expectations sub-dimension, .62 (Chowning & Campbell, 2009). The externalized responsibility sub-dimension was found to be positively related to narcissism, grandiosity, and entitlement (Chowning & Campbell, 2009).

2.3.3. Rosenberg self-esteem scale (RSES)

The scale developed by Rosenberg (1965) is used to determine the self-esteem levels of individuals. The scale was adapted to Turkish by Çuhadaroğlu (1986). This study used the 10-item Self-Esteem subscale of the scale consisting of 63 items and 12 subscales. The Cronbach Alpha internal consistency coefficient for the self-esteem subscale was found to be .785 in the current study.

2.3.4. Narcissistic personality inventory (NPI)

NPI was developed by Raskin and Terry (1998) as a 40-item scale. Later, the revised scale was rearranged to 16 items to reduce the errors caused by the participants and the time taken to complete it (Ames et al., 2006). Higher scores on the scale indicate higher narcissistic tendencies. The last version of the NPI was adapted to Turkish by Atay (2009). Cronbach Alpha coefficient of the scale was reported as .66 by Atay (2009). In this study, the Cronbach Alpha coefficient was found to be .71.

2.3.5. Locus of control scale (LCS)

The LCS scale was developed by Dağ (2002) to test the validity and reliability of university students. The scale consists of 47 items and five sub-dimensions. A total score can be obtained from the scale. While the high scores obtained from the scale represent an increase in external locus of control, the low scores obtained from the scale indicate an increase in internal locus of control. The Cronbach Alpha internal consistency coefficient of the locus of control scale was found to be .842 for the present study sample.

2.4. Preparing Data and Applying Statistical Tests

Analysis of the adaptation of the scale to Turkish culture was carried out in several stages. First, it was examined whether the research data were suitable for statistical analysis, such as confirmatory factor analysis. For this reason, missing values, outliers, and normality assumptions were examined, respectively (Çokluk et al., 2018; Tabachnick & Fidell, 2013). In this context, outliers were examined via Mahalanobis Distance Value, Z standard score, and box-line graphs. Normal distribution examinations of scale items were examined through Skewness/Kurtosis coefficients and histogram graphs (Hair et al., 2014; Harrington, 2009; Pallant, 2005). The findings of the research affirm that all variables exhibit a normal distribution. The obtained correlation matrix substantiates the multicollinearity assumption by indicating significant correlations among variables. Additionally, examinations through scatter plots also validate that the data set fulfills the assumptions of homoscedasticity and linearity. In light of all these evaluations, the data set has been deemed suitable for confirmatory factor analysis (CFA) and has met the prerequisites for conducting the analyses.

Secondly, validity and reliability analysis of the AES were performed. For the reliability analysis of the scale, the Cronbach Alpha internal consistency coefficient was calculated, and the binary correlations of the scale items with the scale total scores were examined. The validity analysis of the scale examined whether the structure consisting of fifteen items and two factors was confirmed by confirmatory factor analysis (CFA). In addition, the scale of measurement invariance based on gender was tested by Multiple Group Confirmatory Factor Analysis (MG-CFA). The relationship between the Academic Entitlement Scale and the criterion variables was examined with the Pearson Correlation Test.

SPSS 26.0 and RStudio statistical programs were used to analyze the data. SemTools (Jorgensen et al., 2021) and lavaan (Rosseel, 2012) packages were used for CFA and MG-CFA.

3. FINDINGS

During our study, we found some important results. These outcomes, based on careful analysis, highlight the main goals of our research. Below are the key findings from our study:

3.1. Descriptive Statistics Regarding the Academic Entitlement Scale

Descriptive statistics such as mean, standard deviation, and Skewness/Kurtosis coefficients related to the academic entitlement scale's items are presented in Table 1.

Table 1. *Descriptive statistics of the academic entitlement scale items.*

Items	n	$\bar{X} \pm s$	Skewness	Kurtosis
I1	372	4.58 ± 1.85	-.55	-.67
I2	372	4.41 ± 1.86	-.49	-.75
I3	372	4.85 ± 1.68	-.70	-.11
I4	372	3.78 ± 2.35	.23	-1.58
I5	372	3.77 ± 2.29	.20	-1.54
I6	372	4.34 ± 1.74	-.45	-.62
I7	372	4.69 ± 1.75	-.68	-.21
I8	372	3.69 ± 2.52	.28	.-1.68
I9	372	3.95 ± 2.47	.13	-1.72
I10	372	4.58 ± 1.80	-.53	-.71
I11*	372	4.36 ± 1.81	-.33	-.86
I12	372	3.85 ± 2.12	-.05	-1.40
I13	372	4.59 ± 1.64	-.41	-.49
I14	372	4.14 ± 2.52	.02	-1.77
I15	372	4.51 ± 1.67	-.63	-.35

Note. * = reverse item.

I4, I5, I8, I9, and I14, related to the items of AES presented in Table 2, represent the self-entitled expectations factor, while the remaining items represent the externalized responsibility factor. Accordingly, while the item with the highest average was I3 among the scale items, the item with the lowest average was determined as I8.

When the Skewness and Kurtosis coefficient values taken for the normal distribution of the scale items were examined, it was determined that the items I4, I5, I8, I9, I12, and I14 were not in the ± 1 range (Hair et al., 2014; Harrington, 2009; Pallant, 2005). In this context, it was decided to perform CFA analysis with the diagonally weighted least squares (*DWLS*) estimation method used in samples that do not conform to normal distribution (Kline, 2015; Mindrila, 2010; Schumacker & Beyerlein, 2000).

3.2. Findings Regarding the Structural Validity of AES

Confirmatory factor analysis was performed to determine whether AES had a structure similar to the original scale in Turkish culture. The analyses were carried out through *semTools* (Jorgensen et al., 2021) and *lavaan* (Rosseel, 2012) packages in the RStudio program. DWLS (diagonally weighted least squares) was used as the estimation method in CFA. The DWLS method is used when the multivariate normality requirement is not met for scale items, and better parameter estimates are made with the DWLS method under these conditions (Kline, 2015; Mindrila, 2010; Schumacker & Beyerlein, 2000). Therefore, the DWLS method was preferred as the estimation method in CFA because the Skewness and Kurtosis coefficients of the scale items were not in the range of ± 1 (Hair et al., 2014; Harrington, 2009; Pallant, 2005). In the literature, many goodness of fit indices are used in the evaluation of CFA results. The criterion values of the goodness of fit indices used are presented in Table 2.

Table 2. *Fit indices and standard fit criteria for the proposed model.*

| Fit Indices	Excellent	Acceptable	Estimated Model
χ^2/df [1]	≤3	≤5	3.21
RMSEA [2]	≤.05	≤.08	.07
SRMR [3]	≤.05	≤.08	.07
CFI [3]	≥.95	≥.90	.98
TLI [3]	≥.95	≥.90	.98
GFI [3]	≥.95	≥.90	.98
AGFI [3]	≥.95	≥.90	.98

Note. [1] = Kline (2015), [2] = Browne & Cudeck (1993), [3] = Baumgartner & Homburg (1996); Marsh et al., (2006); Sumer (2000); Byrene (2010).

CFA analysis was performed to confirm the two-factor structure of the Academic Entitlement Scale consisting of 15 items. When the first results were examined, $\chi^2 = 289.548$ ($p < .05$), $df = 89$ and $\chi^2/df = 3.25$ were found. However, *CFI, GFI, AGFI*, and *TLI* values were found to be above .90. It can be stated that the relevant values are acceptable according to the criteria specified in the literature. In contrast, SRMR and RMSEA values were found to be above .08. Since the relevant values were not within acceptable ranges, model modification indices were examined.

The RStudio program proposed the creation of covariance between items I2 (reverse matter) and I15. When the items in the scale were examined, it was seen that item I2 was "It is my responsibility to reach the course notes if I miss the lesson." and item I15 was "Teachers are only workers who get paid to teach." When the items were examined, a modification was applied based on the fact that both items were similar in structure, could be interpreted similarly, perceived similarly, and the responses to these items could be similar. The values after modification were found as $\chi^2 = 283.008$ ($p < .05$), $df = 88$ and $\chi^2/df = 3.21$. However, as

presented in Table 2, while RMSEA and SRMR values were in the acceptable range, *CFI, GFI, AGFI,* and *TLI* values were found to be in the perfect fit range. In light of what was reported, it can be said that the structure examined by confirmatory factor analysis was confirmed. Standardized factor loads of the verified structure are presented in Figure 1.

Figure 1. *Academic entitlement scale standardized factor loads.*
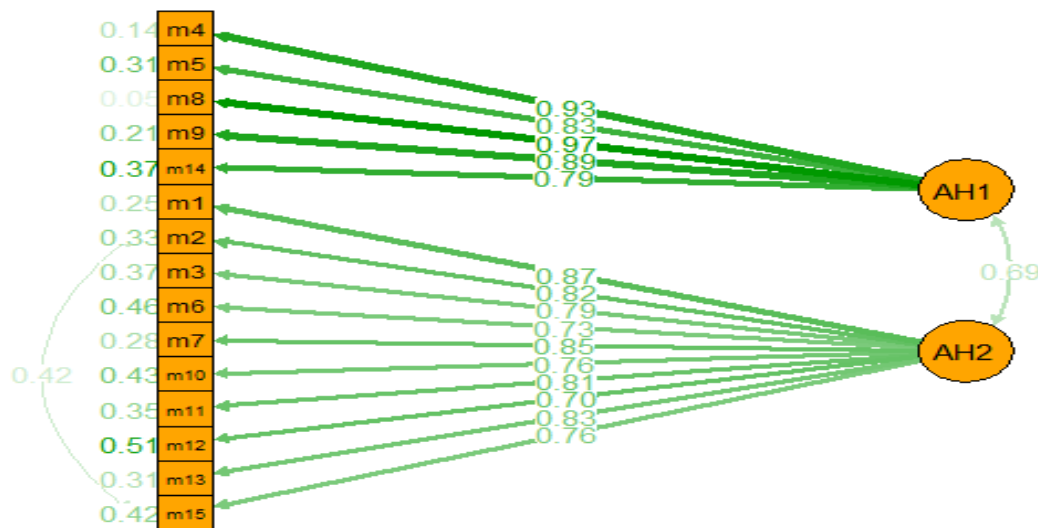
When Figure 1 was examined, it was seen that the standardized factor loads of AES vary between .93 to .70. However, it was observed that the item that contributed the most to the externalized responsibility factor was I1. On the other hand, it was determined that the item that contributed the least to the externalized responsibility factor was I12.

It was determined that the item that contributed the most to the sub-dimension of Self-entitled Expectations was I4, while the item that contributed the least was I12. However, between the factors of Externalized Responsibility (AH2) and Entitled Expectations (AH1), there was a positive, moderate, significant covariance factor load coefficient (.69).

3.3. Measurement Invariance Results

After verifying the structure of academic entitlement, whether this model is invariable for female and male university students was tested with MG-CFA. The results of the tested invariance stages are presented in Table 3.

Table 3. *Compliance statistics regarding measurement invariance stages.*

	Formal Invariance	Metric Invariance	Scale Invariance	Solid Invariance
χ^2	317.23	336.34	340.20	344.07
χ^2/df	1.78	1.76	1.66	1.57
TLI	.990	.990	.991	.993
RMSEA	.065	.064	.060	.056
SRMR	.071	.073	.074	.074
CFI	.991	.991	.992	.992
ΔCFI	-	.000	.001	.001
ΔSRMR	-	.003	.000	.000
Δχ^2	-	*p >.05*	*p >.05*	*p >.05*

Table 3 presents the findings of the MG-CFA analysis. In light of these MG-CFA findings, measurement invariance was examined in a four-stage process. Accordingly, the goodness of fit values of the model established at each stage should comply with acceptable criteria. The

acceptable goodness of fit indices stated in the literature are presented in Table 2. In addition, in order to provide evidence for the measurement invariance steps, the difference between the more limited models and the formal models is compared, and the difference regarding the fit coefficients is examined (*ΔCFI, ΔSRMR*). Cheung and Rensvold (2002) recommend examining the *ΔCFI* value when comparing between models. According to this proposition, it is desired that the comparison *ΔCFI* value should be between -0.01 and +0.01. Similarly, Chen (2007) stated that 0.030 changes can be accepted for the *ΔSRMR* value as well as the 0.01 change in the *ΔCFI* value.

In light of what has been transferred, the values in Table 3 can be interpreted as follows at each stage. As a result of the MG-CFA analysis conducted to test formal invariance, fit indices showed that this stage met the necessary criteria (χ^2/df = 1.78, *CFI* = .99, *TLI* = .99, *RMSEA* = .065 and *SRMR* = .071). The achievement of formal invariance means that the measured structures were the same between the groups. In other words, it shows that female and male university students used the same conceptual perspectives when answering scale items.

The metric invariance step was examined by interpreting the fit indices of MG-CFA and the ΔCFI value obtained as a result of the CFI difference test. The goodness of fit indices for metric invariance (χ^2/df = 1.76, *CFI* = .99, *TLI* = .99, *RMSEA* = .065 and *SRMR* = .071) adjust well. To obtain evidence that metric invariance was achieved, ΔCFI, ΔSRMR, $\Delta\chi^2$ difference values between the formal invariance model were examined. It was revealed that $\Delta\chi^2$ value should not be significant, ΔCFI value should be between -0.01 and +0.01, and ΔSRMR value should be between - 0.03 and +0.03 to ensure metric invariance (Chen, 2007; Cheung & Rensvold, 2002). When Table 4 was examined, it was seen that the calculated values met the specified criteria. In this context, it can be interpreted that metric invariance was provided. By providing metric invariance, it can be said that the factor loadings of the academic entitlement scale were equal in the gender subgroups. This scale can be interpreted similarly in the gender subgroups.

Following the metric invariance, the values related to the scale invariance were examined. Similar to the previous step, the fit indices of the model established for scale invariance were first evaluated. It was determined that the goodness of fit indices calculated during the scale invariance model stage were well adjusted. In addition, Δ*CFI* and Δ*SRMR* values were examined. After the evaluations, it was concluded that there was no bias based on items for male and female university students. More specifically, it was accepted that the fixed number in the regression equations created for the items was invariant between the groups. In the final stage of measurement invariance, evidence of solid invariance was examined. For this purpose, the goodness of fit indices and Δ*CFI* and Δ*SRMR* values of the established solid invariance model were examined. In the examinations, it was determined that the indices of the established model fit well, and the Δ*CFI* and Δ*SRMR* difference fit values of the strict invariance model comply with the criteria specified in the literature (Chen, 2007; Cheung & Rensvold, 2002).

When the findings obtained are evaluated as a whole, the average scores obtained from the academic entitlement scale do not show a bias according to gender. In other words, the differences between the mean scores of male and female university students on the scale are not the structure of the scale but the differences arising from individuals. Therefore, the mean scores obtained from the scale can be compared significantly by gender.

3.4. Examination of AES Reliability Analysis

The reliability analysis of AES was performed with the Cronbach Alpha internal consistency coefficient. Accordingly, the Cronbach Alpha value for the entitled expectations (AH1) factor was determined as .95. In contrast, the Cronbach Alpha value for the externalized responsibility factor was found to be .94. Similarly, the internal consistency analysis for the whole scale was found as .94. In this regard, it can be understood that AES is a perfectly reliable scale with its sub-dimensions (Fraenkel et al., 2012).

In the AES item-total test correlation examinations, the total score of the scale items was found to be correlated in the range of .70 to .83. Thus, it can be stated that each item has a strong positive relationship with the scale total score.

3.5. Theoretical Framing of Academic Entitlement

This section includes the binary relationship of the concept of academic entitlement with the frequently studied criterion variables in the literature. The relationships between the variables were examined with the Pearson Correlation test. The findings are presented in Table 4.

Table 4. *Examination of binary relationships between academic entitlement and criterion variables.*

Criterion Variables	Entitled Expectations	Externalized Responsibility	Academic Entitlement
Self-esteem	-.035	-.215*	-.152*
Narcissism	-.010	.216*	.131
Locus of Control	.033	.105*	.106*

*Note. *$*p < .05$. Externalized responsibility represents the individual's level of avoidance of responsibility in the education process. On the other hand, entitled expectations express the individual's high expectations for the courses and professors in their favor. The total score of Academic Entitlement is obtained by summing the scores of both factors.

Table 4 shows the binary relationship of the Academic Entitlement variable with the criterion variables. Accordingly, self-esteem is negatively related to academic entitlement (r= -.152, p< .05). In addition, externalized responsibility is significantly negatively related to self-esteem (r= -.215, p< .05). There was no significant relationship between self-esteem and entitled expectations (p> .05).

While narcissism was significantly positively related to the externalized responsibility sub-dimension (r= .216, p< .05), it was not significantly related to entitled expectations and academic entitlement (p> .05). While the locus of control was significantly positively related to externalized responsibility (r= .105, p< .05) and academic entitlement (r= .106, p<.05) it did not have a significant relationship with entitled expectations.

4. DISCUSSION and CONCLUSION

In this study, AES developed by Chowning and Campbell (2009) was adapted to Turkish. For many reasons, scale adaptation was preferred instead of scale development in this study. First of all, Chowning and Campbell's (2009) AES is a well-known measurement tool used in many studies in the literature (e.g., Boswell, 2012; Cain et al., 2012; El-Alayli et al., 2018). Measurement tools that are well-known in the literature provide a greater sense of security than newly created measurement tools (Hambleton & Patsula, 1999). Second, adapting an instrument is often much cheaper and faster than developing one (Hambleton & Patsula, 1999). In this respect, it aimed to quickly translate Chowning and Campbell's (2009) AES into Turkish and to bring a comprehensive view to the problem by examining current problems in Turkish culture. Finally, since it was aimed to determine the universal structure of the concept of academic entitlement and thus to be able to make cross-country evaluation, scale adaptation was preferred instead of scale development in the study.

There are many positive aspects in adapting this scale. First, in the data analysis, it was found that the internal consistency of the Academic Entitlement Scale was excellent (Cronbach alpha: .94). In addition to the total scale, the consistency of the subscales is also quite high (entitled expectations Cronbach alpha: .95; externalized responsibility Cronbach alpha: .94). In the original scale, the Cronbach alpha internal consistency coefficient for the *externalized responsibility* sub-factor was found to be .81. For *the entitled expectations* sub-factor, it was .64. In this respect, it can be stated that the internal consistency of the adapted scale is similar to the original scale. Second, in the confirmatory factor analysis, after the modification procedures recommended by the RStudio program, it was found that the model fit indices of the 15-item,

two-dimensional Academic Entitlement Scale showed acceptable fit (χ^2/df: 3.21; RMSEA: .07; SRMR = .07; CFI: .98; GFI: .98; AGFI: .98; TLI: .98). These results showed that the adapted scale was consistent with the number of items and dimensions in the original scale. Third, the validity of the measurement invariance of the adapted scale in terms of gender was also examined. It was found that the scale met the condition of measurement invariance according to the gender variable. Accordingly, it can be said that the differences between the mean scores of female and male university students are not the structure of the scale but the differences arising from the individuals; therefore, it can be understood that the mean scores obtained from the scale can be compared significantly according to gender.

Other scales measure academic entitlement in Turkish culture. These scales are the Academic Entitlement Questionnaire (AEQ) adapted by Kurtyılmaz (2018) and the Academic Entitlement Expectation Scale (AEES) developed by Aksoy and Coban-Sural (2022). Chowning and Campbell's (2009) AES, adapted in this study, showed that, like other academic entitlement scales in Turkish, the academic entitlement structure is also present in Turkish culture. For example, the structure of AEQ is RMSEA=0.04, SRMR=0.03, GFI=0.99, AGFI=0.98, CFI=0.99, NNFI=0.98; The structure of AEES has fit index values of RMSEA 0.054, NFI 0.93, NNFI 0.95, CFI 0.96, SRMR 0.06, GFI 0.91, AGFI 0.89. The standardized pattern coefficients of the scale structure of AEQ are between 0.38 and 0.73, and of AEES are between .38 and .81. In this study, AES's standardized pattern coefficients are between .70 and .93. In addition, the Cronbach alpha internal reliability coefficients of the AES (entitled expectations Cronbach alpha: .95; externalized responsibility Cronbach alpha: .94) adapted within the scope of this study are higher than those of AEQ (.66) and AEES (between .68 and .83). All three scales determined that academic entitlement can be measured validly and reliably in Turkish culture.

In this study, it was found that academic entitlement was negatively correlated with self-esteem (r= -.152, p< .05). There are some findings in the literature that reveal that self-esteem and self-efficacy and academic entitlement are positively related (Boswell, 2012; Whatley et al., 2019). However, many studies also state that narcissism masks low self-esteem and that high self-esteem and narcissism are two opposite measures of a positive self-image (Baumeister et al., 2000; Bosson et al., 2008). From this point of view, the negative relationship between low self-esteem in this study and academic entitlement seems to be compatible with the narcissism literature. According to the other finding in this study, narcissism was significantly positively correlated with the externalized responsibility sub-dimension (r= .216, p< .05). At the same time, it was not significantly related to entitled expectations and academic entitlement (p> .05). Similar studies have also revealed the relationship between narcissism and academic entitlement (Turnipseed & Cohen, 2012; Whatley et al., 2019). It is thought that the relationship between narcissism and academic entitlement with dishonest academic behaviors may be due to attitudes that exploit people (Menon & Sharland, 2011). While the locus of control was significantly positively related to externalized responsibility (r= .105, p< .05) and academic entitlement (r= .106, p<.05), it did not have a significant relationship with entitled expectations. According to research on self-serving biases, many people tend to attribute success to internal factors such as the abilities they possess, and the perseverance they show, while attributing failure to external factors such as other people and coercive conditions (Zuckerman, 1979). The externalized responsibility factor of academic entitlement is related to the self-serving judgments of entitled students. The scale includes items that attribute responsibilities to factors such as university administration, lecturers, and classmates for their academic failure. The lack of a relationship between entitled expectations and narcissism may be because people's attitudes may be inconsistent with their behaviors (Wilson et al., 1989). While the participants disagree with the items at the behavioral level associated with narcissistic tendencies, they may tend to agree with the items related to attitude. From this point of view, while the externalized responsibility factor explaining academic entitlement through behavioral items was related to narcissism, the entitled expectations were not. The concept of entitled expectations refers to

individuals' rights regarding the course and rules. Accordingly, since the study participants stated that they did not expect any privilege for themselves in teaching the lessons at the attitude level, there may not have been a relationship with narcissism in the dimension of entitled expectations. However, since they may show a pattern of holding others responsible for the consequences of their behavior, they may have found a relationship with narcissism in the externalized responsibility dimension.

AE is associated with decreasing academic success in the university student population, damaging the belief in justice in academic environments, and deteriorating the quality of education (Bonaccio et al., 2016; Greenberger et al., 2008; Yan et al., 2021). In addition, it is known that students with high levels of AE have unrealistic expectations in their business lives (vanWieirngen et al, 2013). This study on the quantitative measurement of AE can potentially improve the academic environment on a class and country basis. Because this measurement tool can be used to reveal a student's beliefs about academic rights. Defining and evaluating AE in the university student population can also enable students to develop functional strategies to positively affect their success.

The study has some limitations as well as its contributions to the field. First of all, of the 372 university students participating in the study, 69 were male, and the remaining 303 were female. The participant groups of the study are not equal or close to each other in number by gender. Despite our efforts to achieve balanced gender representation, the voluntary nature of participation and the specific demographics of the population from which we drew our sample resulted in an imbalance. To mitigate the impact of this imbalance, we have conducted statistical adjustments, including an analysis of measurement invariance by gender. Based on the measurement invariance analyses, we have concluded that the scale can be used in both the female and male groups.

In addition, Cronbach Alpha internal consistency analysis was calculated for the reliability analyses. In addition, the sub-dimensions of the academic entitlement scale and their binary correlations with the criterion variables were examined. However, pre-and post-test reliability data were not collected and analyzed for the reliability analysis of the scale. Future studies may analyze the test-retest reliability of the scale. Finally, the AES measurement invariance according to gender was examined. Future studies may focus on measurement invariance for other demographic variables.

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**: Erzurum Technical University Scientific Research and Publication Ethics Committee, 10 Number of Decision: 5 Date: 20.04.2021.

Authorship Contribution Statement

Each author has made an equal contribution to the research.

Orcid

Adem Kantar https://orcid.org/0000-0001-8486-9376
Büşra Aslan Cevheroğlu https://orcid.org/0000-0001-8730-0441
Ayşegül Bakır https://orcid.org/0000-0002-9182-9492
Serel Akdur Çiçek https://orcid.org/0000-0003-3981-1188

REFERENCES

Achacoso, M.V. (2006). *"What do you mean my grade is not an A?": An investigation of academic entitlement, causal attributions, and self-regulation in college students.* Retrieved from https://www.proquest.com/dissertations-theses/what-do-you-mean-my-grade-is-not-investigation/docview/304984244/se-2

Aksoy, N., & Sural, Ü.Ç. (2022). Academic Entitlement Expectations of Preservice Primary School Teachers. *International Journal of Progressive Education*, *18*(4), 1-20. https://doi.org/10.29329/ijpe.2022.459.1

Barton, A.L., & Hirsch, J.K. (2016). Permissive parenting and mental health in college students: Mediating effects of academic entitlement. *Journal of American College Health*, *64*(1), 1-8. https://doi.org/10.1080/07448481.2015.1060597

Baumeister, R.F., Bushman, B.J., & Campbell, W.K. (2000). Self-esteem, narcissism, and aggression: Does violence result from low self-esteem or from threatened egotism? *Current Directions in Psychological Science*, *9*(1), 26-29. https://doi.org/10.1111/1467-8721.00053

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

Bhattacharya, S., Bashar, M.A., Srivastava, A., & Singh, A. (2019). Nomophobia: No mobile phone phobia. *Journal of Family Medicine and Primary Care*, *8*(4), 1297. https://doi.org/10.4103/jfmpc.jfmpc_71_19

Blincoe, S., & Garris, C.P. (2017). Challenging the assumption of a Western phenomenon: Academic entitlement in Saudi Arabia. *The Journal of Experimental Education*, *85*(2), 278-290. https://doi.org/10.1080/00220973.2016.1199009

Bonaccio, S., Reeve, C.L., & Lyerly, J. (2016). Academic entitlement: Its personality and general mental ability correlates, and academic consequences. *Personality and Individual Differences*, *102*, 211-216. https://doi.org/10.1016/j.paid.2016.07.012

Bosson, J.K., Lakey, C.E., Campbell, W.K., Zeigler-Hill, V., Jordan, C.H., & Kernis, M.H. (2008). Untangling the links between narcissism and self-esteem: A theoretical and empirical review. *Social and Personality Psychology Compass*, *2*(3), 1415-1439. https://doi.org/10.1111/j.1751-9004.2008.00089.x

Boswell, S.S. (2012). "I deserve success": Academic entitlement attitudes and their relationships with course self-efficacy, social networking, and demographic variables. *Social Psychology of Education*, *15*(3), 353-365. https://doi.org/10.1007/s11218-012-9184-4

Cain, J., Romanelli, F., & Smith, K.M. (2012). Academic entitlement in pharmacy education. *American Journal of Pharmaceutical Education*, *76*(10). https://doi.org/10.5688/ajpe7610189

Campbell, S., Greenwood, M., Prior, S., Shearer, T., Walkem, K., Young, S., Bywaters, D., & Walker, K. (2020). Purposive sampling: complex or simple? Research case examples. *Journal of Research in Nursing: 25*(8), 652-661. https://doi.org/10.1177/1744987120927206

Campbell, W.K., Bonacci, A.M., Shelton, J., Exline, J.J., & Bushman, B.J. (2004). Psychological entitlement: Interpersonal consequences and validation of a self-report measure. *Journal of Personality Assessment, 83*, 29-45. https://doi.org/10.1207/s15327752jpa8301_04

Chowning, K., & Campbell, N.J. (2009). Development and validation of a measure of academic entitlement: Individual differences in students' externalized responsibility and entitled expectations. *Journal of Educational Psychology*, *101*(4), 982. https://doi.org/10.1037/a0016351

Ciani, K.D., Summers, J.J., & Easter, M.A. (2008). Gender differences in academic entitlement among college students. *The Journal of Genetic Psychology*, *169*(4), 332-344. https://doi.org/10.3200/GNTP.169.4.332-344

Côté, S., Stellar, J.E., Willer, R., Forbes, R.C., Martin, S.R., & Bianchi, E.C. (2021). The psychology of entrenched privilege: High socioeconomic status individuals from affluent backgrounds are uniquely high in entitlement. *Personality and Social Psychology Bulletin*, *47*(1), 70-88. https://doi.org/10.1177/0146167220916633

Dubovsky, S.L. (1986). Coping with entitlement in medical education. *New England Journal of Medicine*, *315*(26), 1672-1674. https://doi.org/10.1056/NEJM198612253152609

El-Alayli, A., Hansen-Brown, A.A., & Ceynar, M. (2018). Dancing backwards in high heels: Female professors experience more work demands and special favor requests, particularly from academically entitled students. *Sex Roles*, *79*, 136-150. https://doi.org/10.1007/s11199-017-0872-6

Elias, R.Z. (2017). Academic entitlement and its relationship with perception of cheating ethics. *Journal of Education for Business*, *92*(4), 194-199. https://doi.org/10.1080/08832323.2017.1328383

Exline, J.J., Baumeister, R.F., Bushman, B.J., Campbell, W.K., & Finkel, E.J. (2004). Too proud to let go: narcissistic entitlement as a barrier to forgiveness. *Journal of Personality and Social Psychology*, *87*(6), 894. https://doi.org/10.1037/0022-3514.87.6.894

Fletcher, K.L., Pierson, E.E., Speirs Neumeister, K.L., & Finch, W.H. (2020). Overparenting and perfectionistic concerns predict academic entitlement in young adults. *Journal of Child and Family Studies*, *29*(2), 348-357. https://doi.org/10.1007/s10826-019-01663-7

Greenberger, E., Lessard, J., Chen, C., & Farruggia, S.P. (2008). Self-entitled college students: Contributions of personality, parenting, and motivational factors. *Journal of Youth and Adolescence*, *37*(10), 1193-1204. https://doi.org/10.1007/s10964-008-9284-9

Hambleton, R.K. and Patsula, L. (1999). Increasing the validity of adapted tests: myths to be avoided and guidelines for improving test adaptation practices. *Journal of Applied Testing Technology, 1*, 1-16.

Harvey, P., & Harris, K.J. (2010). Frustration-based outcomes of entitlement and the influence of supervisor communication. *Human Relations*, *63*(11), 1639-1660. https://doi.org/10.1177/0018726710362923

Hill, M.S., & Fischer, A.R. (2001). Does entitlement mediate the link between masculinity and rape-related variables?. *Journal of Counseling Psychology*, *48*(1), 39. https://doi.org/10.1037/0022-0167.48.1.39

Jack, G., & Donnellan, H. (2010). Recognising the person within the developing professional: Tracking the early careers of newly qualified child care social workers in three local authorities in England. *Social Work Education*, *29*(3), 305-318. https://doi.org/10.1080/02615470902984663

Kılıç, S. (2016). Cronbach'ın alfa güvenirlik katsayısı *[Cronbach's alpha reliability coefficient]. Journal of Mood Disorders*, *6*(1), 47-48. https://doi.org/10.5455/jmood.20160307122823

Kopp, J.P., Zinn, T.E., Finney, S.J., & Jurich, D.P. (2011). The development and evaluation of the academic entitlement questionnaire. *Measurement and Evaluation in Counseling and Development*, *44*(2), 105-129. https://doi.org/10.1177/0748175611400292

Kumari, D., & Gautam, M.K. 648. "I deserve higher Grades": An Analysis of Academic Entitlement Attitude among Students of Higher Education. *Journal of Dialogues on Education, 11*(1), 1-10.

Lippmann, S., Bulanda, R.E., & Wagenaar, T.C. (2009). Student entitlement: Issues and strategies for confronting entitlement in the classroom and beyond. *College Teaching, 57*, 197–204. https://doi.org/10.1080/87567550903218596

McLellan, C.K., & Jackson, D.L. (2017). Personality, self-regulated learning, and academic entitlement. *Social Psychology of Education*, *20*(1), 159-178. https://doi.org/10.1007/s11218-016-9357-7

Menon, M.K., & Sharland, A. (2011). Narcissism, exploitative attitudes, and academic dishonesty: An exploratory investigation of reality versus myth. *Journal of Education for Business*, *86*(1), 50-55. https://doi.org/10.1080/08832321003774772

Moeller, S.J., Crocker, J., & Bushman, B.J. (2009). Creating hostility and conflict: Effects of entitlement and self-image goals. *Journal of Experimental Social Psychology*, *45*(2), 448-452. https://doi.org/10.1016/j.jesp.2008.11.005

Morrow, W. (1994). Entitlement and achievement in education. *Studies in Philosophy and Education*, *13*(1), 33-47. https://doi.org/10.1007/BF01074084

Peirone, A., & Maticka-Tyndale, E. (2017). "I bought my degree, now I want my job!" Is academic entitlement related to prospective workplace entitlement?. *Innovative Higher Education*, *42*(1), 3-18. https://doi.org/10.1007/s10755-016-9365-8

Pilotti, M.A., Al Ghazo, R., & Al Shamsi, S.J. (2022). Academic entitlement amid social change in the Kingdom of Saudi Arabia. *Journal of Applied Research in Higher Education*, *14*(4), 1718-1730. https://doi.org/10.1108/JARHE-05-2021-0168

Reysen, R., Reysen, M., & Reysen, S. (2021). Academic entitlement predicts smartphone usage during class. *College Teaching*, *69*(1), 52-57. https://doi.org/10.1080/87567555.2020.1803192

Singleton-Jackson, J.A., Jackson, D.L., & Reinhardt, J. (2011). Academic Entitlement: Exploring Definitions and Dimensions of Entitled Students. *International Journal of Interdisciplinary Social Sciences*, *5*(9). https://doi.org/10.18848/1833-1882/CGP/v05i09/51883

Sohr-Preston, S., & Boswell, S.S. (2015). Predicting academic entitlement in undergraduates. *International Journal of Teaching and Learning in Higher Education*, *27*(2), 183-193.

Solomon, I., & Leven, S. (1975). Entitlement. *Psychotherapy: Theory, Research & Practice*, *12*(3), 280. https://doi.org/10.1037/h0086444

Turnipseed, D.L., & Cohen, S.R. (2015). Academic entitlement and socially aversive personalities: Does the Dark Triad predict academic entitlement?. *Personality and Individual Differences*, *82*, 72-75. https://doi.org/10.1016/j.paid.2015.03.003

vanWieirngen, K.A., Formicola, A., Peirone, A., & Falsetta, T. (2013, May 30). *Academic entitlement leads to negative outcomes, student researchers find*. Retrieved from http://www.uwindsor.ca/dailynews/2013-05-29/academic-entitlement-leadsnegative-consequences-student-researchers-find

Wasieleski, D.T., Whatley, M.A., Briihl, D.S., & Branscome, J.M. (2014). Academic entitlement scale: Development and preliminary validation. *Psychology Research*, *4*(6). https://doi.org/10.17265/2159-5542/2014.06.004

Whatley, M.A., Wasieleski, D.T., Breneiser, J.E., & Wood, M.M. (2019). Understanding academic entitlement: Gender classification, self-esteem, and covert narcissism. *Educational Research Quarterly*, *42*(3), 49-71.

Wicherts, J.M. (2007). *Group differences in intelligence test performance*. Unpublished dissertation, Faculty of Social and Behavioural Sciences, University of Amsterdam. Retrieved from https://hdl.handle.net/11245/1.280565

Wilson, T.D., Dunn, D.S., Kraft, D., & Lisle, D.J. (1989). Introspection, attitude change, and attitude-behavior consistency: The disruptive effects of explaining why we feel the way we do. In *Advances in Experimental Social Psychology* (Vol. 22, pp. 287-343). Academic Press. https://doi.org/10.1016/S0065-2601(08)60311-1

Yalçın, S (2021). Ölçek geliştirme ve uyarlama süreci bilgi notu. [Scale development and adaptation process information note]. Retrieved from http://egitim.ankara.edu.tr/wp-

content/uploads/sites/347/2021/06/Olcek-Gelistirme-ve-Uyarlama.Bilgi-notu.-Doc.-Dr.-Seher-Yalcin.pdf.

Yan, D., Xi, L., Shu, S.U., & Li, L. (2021). The Online Learning Academic Achievement of Chinese Students during the COVID-19 Pandemic: the Role of Self-Regulated Learning and Academic Entitlement. *International Journal of Psychology and Educational Studies*, *8*(3), 116-127. Retrieved from https://dergipark.org.tr/tr/pub/pes/issue/64305/976295

Zarei, S. (2022). The Mediating role of academic entitlement in the relationship between Permissive Parenting and Psychological Distress among Students. *Psychological Studies*, *18*(1). https://doi.org/10.1080/07448481.2015.1060597

Zuckerman, M. (1979). Attribution of success and failure revisited, or: The motivational bias is alive and well in attribution theory. *Journal of Personality*, *47*(2), 245-287. https://doi.org/10.1111/j.1467-6494.1979.tb00202.x

APPENDIX

Turkish Version of the Academic Entitlement Scale

Aşağıda 1'den 7'ye kadar derecelendirilmiş çeşitli ifadeler vardır. İfadeleri okuyup, size en uygun gelen seçeneği yuvarlak içine alınız.

1.	Hoca öğretmek için para aldığına göre derslere aktif olarak katılım göstermek benden çok onun sorumluluğudur. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
2.	Dersi kaçırırsam ders notlarına ulaşmak benim sorumluluğumdur. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
3.	Grup çalışmalarında çok fazla çaba harcamaya istekli değilim, çünkü illa ki gruptan birileri gerekeni yapacaktır. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
4.	Hocalarım sınava hazırlanmamda bana yardım etmekle yükümlüdür. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
5.	İyi bir hoca eğlenceli olmalıdır. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
6.	Üniversitenin bana başarılı olmak için ihtiyacım olan kaynakları sağlamadığına inanıyorum. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
7.	Hocalar tam olarak neden bahsettiklerini bilmiyorlar. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
8.	Eğer istediğim nota yakın bir not aldıysam dersin hocası notumu yeniden değerlendirmelidir/değerlendirebilir. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
9.	Verdiğim bir ödevden asla sıfır almamalıyım. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
10.	Eğer bir derste kötü bir not alırsam ve dersin hocasının ofis saatlerine denk gelemezsem, hata dersin hocasına aittir. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
11.	Üniversitede başarılı olmak için gerekli kaynakları bulmak benim sorumluluğumdadır. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
12.	Eğer meşgulsem grup ödevlerinde geriye çekilip diğerlerinin çoğu şeyi yapmasını beklemek kabul edilebilir bir şeydir. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
13.	Grup ödevlerinde gösterdiğim çabadan bağımsız olarak diğer grup üyeleriyle aynı notu almalıyım. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
14.	Eğer bir üst harf notuna yakın bir not aldıysam hoca notumu yuvarlamalıdır. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**
15.	Hocalar yalnızca öğretmek için para alan işçilerdir. **Kesinlikle katılmıyorum** 1 2 3 4 5 6 7 **Kesinlikle katılıyorum**

Note. Items 2 and 11 in the scale are reverse ones. Those items have to be recoded. High scores obtained from the scale signify higher levels of academic entitlement modes.