

A Systematic Review of Research Articles on Measurement Invariance in Education and Psychology

Betul Alatli ^{1,*}

¹Department of Educational Sciences, Faculty of Education, Balıkesir University, Balıkesir, Turkey

ARTICLE HISTORY

Received: May 17, 2020

Revised: Aug. 28, 2020

Accepted: Sep. 25, 2020

KEYWORDS

Measurement invariance,
Confirmatory Factor
Analysis,
Content analysis,
Bibliometric analysis

Abstract: This study aims to reveal the trends in the related field by examining the researches evaluating the measurement invariance in education and psychology between 2008-2019. Accordingly, 99 articles published in three journals that were selected using the purposive sampling method among the journals indexed on Social Sciences Citation Index (SSCI) were analyzed within the scope of the study. As a result of the content analysis, in the studies investigating the measurement invariance, typical response tests were observed to be the most frequently employed tests, sample sizes often included 1501 or greater number of subjects, and data were mostly collected from students. The measurement invariance of the tests was mostly analyzed in terms of the gender variable. According to the results of the bibliometric analysis, on the other hand, only Multi-Group Confirmatory Factor Analysis was mostly conducted on the Mplus software package. In the studies, the most cited article was "Cheung and Rensvold (2002)", the author was "Cheung, G. W.", and the journal was "Structural Equation Modeling: A Multidisciplinary Journal". According to the results of the analysis, studies, references, and keywords including factor analysis were among the most commonly used group, which denotes that factor analysis has a crucial role in invariance measurement analyses.

1. INTRODUCTION

Science is defined as a whole consisting of systematic information; yet, the validity of this information should be accepted (Karasar, 2016). Accordingly, the validity of the information must be recognized so that it can be evaluated scientifically. In order for the validity of the information to be accepted, measurement results are needed. The branches of science have two dimensions: theoretical and experimental. In the theoretical dimension, facts and factual relationships are conceptually explained. Experimental/observational studies make it possible to observe the phenomenon in question or relationships under suitable conditions and to quantify or qualify the results of observation. The two dimensions have intertwined processes. In other words, when making theoretical explanations according to the results of an observation, we need observation results again to confirm these theoretical explanations. Precisely at this point, science introduces the importance of measurement. The branches of science can maintain their development in parallel with the development of measurement processes. Measurement is the process of observing a characteristic and displaying the results of observation using symbols

CONTACT: Betül ALATLI ✉ betulkarakocalatli@gmail.com 📍 Department of Educational Sciences, Faculty of Education, Balıkesir University, Balıkesir, Turkey

ISSN-e: 2148-7456 /© IJATE 2020

or numbers (Turgut, 1977). Measurement is the process of determining whether an individual, an object, or a phenomenon has certain characteristics, and showing the degree of the feature using numbers or symbols if the property sought is available (Tekin, 2012). A measurement tool is as important as the measured feature and the thing measured. The validity and reliability of the measurement results are directly correlated with the validity and reliability of the measurement tool. For this reason, it is considered quite important for science that measurement tools can make valid and reliable measurements.

Measurement tools vary by the fields of science. Fields of science are classified into three basic fields: natural sciences, social sciences, and mathematics. Mathematics is absolute, but natural, and social sciences have a relative nature. While social sciences focus on social phenomena, natural sciences address natural events. The desire of human beings to have knowledge and skills has come out to understand and control primarily their environment and then themselves. For this reason, natural sciences date back to much older times than social sciences. Social sciences concentrate human and human behaviors and interactions (Karasar, 2016). Therefore, compared to natural sciences, the difficulty in making and controlling objective observations in social sciences have a significant impact on the measurements made in this field.

Measurements made in social sciences often involve human characteristics. Measurement tools used in these types of measurements are generally called a test. Some features can also be measured with non-test techniques. While it is more convenient to observe some features with non-test techniques, others can be observed with tests. Tests are measurement tools that consist of stimuli (items) for a certain characteristic to be measured. The status of the individual regarding this characteristic is determined by the response shown to these items. The extent to which the tests or the results obtained from these tests serve the purpose and the error-free nature of the tests are highly important in making decisions based on these results. Special studies are required to obtain evidence for such features called validity and reliability. One of these studies is the measurement invariance. Measurement invariance is defined as a condition where individuals in different groups who have had the same observed score in terms of a specific implicit structure get the same score at the subscale and item levels. According to the statements in the Test Adaptation Guidelines (International Test Commission-ITC, 2005) and the Standards of Measurement in Education and Psychology (American Educational Research Association-AERA, American Psychological Association-APA, and National Council on Measurement in Education-NCME, 1999), evidence of measurement invariance must be obtained for tests which aim to make intergroup comparisons. Accordingly, to make decisions about individuals and the groups that they belong to for many features and to make comparisons between individuals and groups in social sciences, measurement invariance analyses are considered to be highly important in making fair and appropriate decisions.

Although there are primarily invariance analyses at the test level under the name of measurement invariance, studies for determining the Differential Item Functioning also aim to determine the measurement invariance at item level (Holland & Wainer, 1993). As in many study areas, measurement invariance studies, too, can vary in different aspects. Many variables such as the test under investigation, measured feature, study group, and statistical technique and statistical software used increase the variety of measurement invariance studies. In this sense, determining the trends in the field by reviewing the measurement invariance studies is considered important and necessary.

New developments in a particular field of study can be followed by reviewing scientific studies such as projects, theses, and articles obtained as a result of a literature review. For this purpose, periodical literature reviews help determine trends in a given field and guide new studies (Chang, Chang & Tseng, 2010; Cohen, Manion & Morrison, 2007; Falkingham & Reeves, 1998; Keselman et al., 1998; Kilbourne & Beckmann, 1998; Lee, Wu, & Tsai, 2009). In Turkey, the

number of educational researches is known to show a huge increase after the 2000s (Karadağ, 2010; Göktaş et al., 2012; Vega Arce et al., 2019). To reveal the quality of educational research, information about the quality and quantity of studies should be questioned (Bacanak, Değirmenci, Karamustafaoğlu & Karamustafaoğlu, 2011; Fazlıoğulları & Kurul, 2012).

When the literature is reviewed, it can be seen that studies conducted in Turkey for reviewing the literature in the field of education consist of many review studies on Science Education (Arıcı, Yıldırım, Çalıklar & Yılmaz, 2019, Chang, Chang & Tseng, 2010; de Jong 2007; Lin, Lin & Tsai, 2014; Ören & Sarı, 2019; Sırakaya & Alsancak Sırakaya, 2020; Sözbilir & Kutu, 2008; Lee, Wu & Tsai, 2009; Tsai & Wen, 2005; White, 1997), Mathematics Education (Aztekin & Taşpınar Şener, 2015, Baki, Karataş, Akkan & Çakıroğlu, 2011; Çiltaş, Güler & Sözbilir, 2012; Hart, Smith, Swars & Smith, 2009; Ulutaş & Ubuz, 2008), Social Studies Education (Tarman, Güven & Aktaşlı, 2011) Pre-School Education (Yılmaz & Altınkurt, 2012), chemistry education (Eybe & Schmidt, 2001; Ulutaş, Üner, Turan Oluk, Yalçın Çelik & Akkuş, 2015) special education (Aslan & Özkubat, 2019), Classroom Teacher Education (Küçükkoğlu & Ozan, 2013; Akaydın & Çeçen, 2015), Educational Sciences (Arık & Türkmen, 2009; Doğan & Tok, 2018; Erdem, 2011; Erdem Aydın, Kaya, İşkol & İşcan, 2019; Hsu, 2005; Karadağ, 2009; Selçuk, Kandemir, Palancı & Dündar, 2014; Tavşancıl et al., 2010). Psychological Counseling and Guidance (Seçer, Ay, Ozan & Yılmaz, 2006), Curriculum and Instruction (Saracaloğlu & Dursun, 2010; Hazır Bıkmaz, Aksoy, Tatar & Atak Altinyüzük, 2013; Ozan & Köse, 2014), Educational Administration (Aydın, Erdağ & Sarier, 2010; Aypayet et al., 2010; Turan, Karadağ, Bektaş & Yalçın, 2014; Murphy, Vriesenga & Storey, 2007), Educational Technology (Bozkut et al., 2015; Erdem Aydın, Bozkaya & Genç Kumtepe, 2019; Erdoğan & Çağiltay, 2009; Göktaş et al., 2012; Gülbahar & Alper, 2009; Hew, Kale & Kim, 2007; Özyurt & Özyurt, 2015; Zainuddin, et al., 2019).

Studies that investigate the relevant literature can be conducted to determine trends in the field while the review of studies on a particular topic in the field allows a detailed examination of that area. One example of this situation involves studies on leadership (Özkan, 2016) or special education in early childhood in Turkey (Öncül, 2014). The field of study of this research is measurement and evaluation in psychology and education. According to the content analysis studies in journals of education, the field of measurement and evaluation ranks fourth and sixth in journals overseas and Turkey, respectively (Hsu, 2005; Selçuk et al., 2014; Yalçın, Yavuz ve İlgün Dibek, 2015). Accordingly, it is necessary to increase studies on research trends by conducting content analysis studies in an important field of educational sciences such as measurement and evaluation.

The review of trend studies in the field of measurement and evaluation shows that many content analysis studies have been conducted on scale development and adaptation (Acar Güvendir & Özer Özkan, 2014; Bastos, Celeste, Faerstein & Barros, 2010; Boztunç Öztürk, Eroğlu & Kelecioğlu, 2015; Çüm & Koç, 2013; Hinkin, 1995; Kapuscinski & Masters, 2010; Ladhari, 2010; Morgado, Meireles, Neves, Amaral & Ferreira, 2017; Sveinbjornsdottir & Thorsteinsson, 2008; Şahin & Boztunç Öztürk, 2019; Tavşancıl, Güler & Ayan, 2014; Worthington & Whittaker, 2006). In a trend study, 40 papers and 49 researches related to the "Measurement-Evaluation" dimension of the 2004 primary education program were examined (Kazu & Aslan, 2013). In another study conducted by Şenyurt and Özer Özkan (2017) master's theses on measurement and evaluation in education were examined methodologically and thematically. Kazu and Deniz (2019) evaluated the studies investigating teachers in terms of using measurement and evaluation techniques. Gotch and French (2014) reviewed studies on teachers' evaluation literacy. Apart from these, trend researches are required in many other areas of measurement and evaluation. Kieffer, Reese and Thompson (2001) investigated statistical techniques used in education and psychology studies. Yalçın (2016) studied 584 articles for many aspects from the field of measurement and evaluation indexed in the Social Science

Citation Index (SSCI). According to this study, measurement invariance took place in 41 areas. In this sense, measurement invariance can also be said to have a crucial role in the field of measurement and evaluation. Vandenberg and Lance (2000) analyzed 14 studies published between 1971 and 1998 aiming to define and develop measurement invariance theoretically and within the framework of Confirmatory Factor Analysis (CFA) and 67 studies published between 1982 and 1999 which tested hypotheses intending to study measurement invariance. In studies in both groups, reviews focused on how the CFA procedure is carried out. Another trend research that examined studies published between 2000 and 2007 and analyzed measurement invariance of measurement tools in terms of different groups by using CFA was conducted by Schmitt and Kuljanin (2008). This study addressed the study areas (intelligence, depression, etc.), groups that were compared (gender, age, etc.), whether the scale was translated, and the analysis steps of measurement invariance (Schmitt & Kuljanin, 2008). Accordingly, when the related literature is analyzed, it can be seen that the measurement invariance trend studies focus on the framework of CFA. Although measurement invariance studies have been handled in terms of variables such as the subject area, compared groups, and translation of the scale, it can be said that these variables are limited. The measurement invariance studies usually deal with the CFA steps. Nevertheless, reviews examining the studies conducted in 2008 and on have not been found. The analysis of measurement invariance studies published in recent years in several aspects and in terms of several variables is considered important in this respect.

For inter-group comparisons to be made fairly and appropriately, the tests must meet the measurement invariance assumption. Studies conducted to determine the measurement invariance are known to differ in many aspects. There is a need for literature reviews to determine trends in the related area. As with similar studies in many fields, trend studies can also be conducted on measurement invariance. For this reason, this study aims to analyze studies conducted on measurement invariance in the last 12 years and reveal the trends in terms of several variables such as what groups were involved in measurement invariance, the type of measurement tool in terms of the measured feature, the size of the study group, the statistical technique and software used, and the bibliometric analysis regarding most frequently used keywords, cited articles, authors, and journals.

2. METHOD

This study used the review model since it aimed to examine the measurement invariance articles that were published between 2008 and 2019 in three journals, which are reviewed on SSCI, according to some specific criteria, and to reveal trends in this area (Karasar, 2017).

2.1. Population and Sample

The universe of the study consisted of articles on measurement invariance published in journals reviewed on SSCI. Sampling was conducted in two stages in this study, which used the criterion sampling method, which is a purposive sampling method (Patton, 2002). Accordingly, first, the journals were selected. The criteria in the selection of journals were as follows: the journal should be in the field of educational sciences, it should include the words "measurement" or "evaluation", and it should have a high impact factor. Accordingly, the following journals were included in the sample when they were sorted according to impact factors.

- Educational and Psychological Measurement
- Journal of Psychoeducational Assessment
- Applied Measurement in Education

In the second stage of sampling, measurement invariance studies that were published between 2008 and 2019 and whose full-text version could be accessed were determined. However, simulation studies that were not appropriate for the variables to be examined concerning the

purpose of the research were excluded from the sample. Accordingly, the distribution of the articles included in the present study by years and journals is given in Table 1.

Table 1. Distribution of articles examined within the scope of the study by journals and years

Journal	Years												Total
	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	
Applied Measurement in Education	0	0	1	0	0	0	0	1	0	1	0	0	3
Educational and Psychological Measurement	5	7	4	1	1	2	1	1	0	0	3	0	25
Journal of Psychoeducational Assessment	2	0	2	4	2	5	8	9	12	4	17	6	71
Total	7	7	7	5	3	7	9	11	12	6	20	6	99

As seen in Table 1, 71 articles were included from the "Journal of Psychoeducational Assessment", 25 articles from the "Educational and Psychological Measurement" journal, and three articles from the "Applied Measurement in Education" journal. Regarding the distribution of the articles by years, the majority of the articles (20) on measurement invariance were published in 2018. A total of 99 articles published between 2008 and 2019 in these three journals were reviewed.

2.2. Data Collection and Analysis

To collect data an article review form specific to this study was developed by examining the article or thesis review forms previously used in similar studies in the literature (Tavşancıl et al., 2010; Yalçın, 2016). The form consisted of a total of 10 sections including the code of the article (M1, M2..); the name of the journal in which the article was published; the year of the publication; the name of the article; the test whose invariance was analyzed; the type of the test, whose measurement invariance was analyzed, in terms of the measured feature; the groups between which invariance was analyzed; sample group; sample size; data analysis method; and the statistical software package used.

Content analysis and bibliometric analysis methods were used for the analysis of the data obtained. Content analysis can be done in the form of systematic coding of qualitative or quantitative data within the framework of certain themes and classifications (Cohen et al., 2007; Fraenkel et al., 2007). According to Falkinham and Reeves (1998), content analysis is specified as a method used in studies in which piles of publications are analyzed or evaluated. In this study, each article was analyzed according to each variable in the evaluation form, and the data obtained were classified according to these variables. Themes were created for the qualitative data obtained for each category. For example, themes such as *maximum performance or typical response test* for the measurement tool whose invariance was analyzed or *ethnicity, gender* for groups for which invariance was analyzed were created. The coding process in determining the validity and reliability categories in content analysis depends on the comprehensibility, clarity, and overlapping (Tavşancıl & Aslan, 2001). For this purpose, 20 randomly selected articles from the article group were re-analyzed by the researcher for intra-rater and inter-rater reliability. Three different raters also reviewed another 20 randomly selected articles for inter-rater reliability. No discrepancy was determined in both the intra-rater and inter-rater reliability study.

Another data analysis technique used in the study was bibliometric mapping analysis. Bibliometrics is defined as the statistical analysis of articles, books, and other publications (Oxford Dictionary, 2017). With bibliometric methods, the image of the related science field can be obtained from bibliographic data obtained from databases (Zupic, 2015). Bibliometric mapping method, on the other hand, is a widely used method in which visual and quantitative findings obtained from the relationship between the data of certain fields in terms of certain variables can be obtained (Small, 1999; Wang et al., 2016). In the present study, VOSViewer 1.6.14 software package was used for bibliometric analysis. The steps followed for the analysis were as follows. The articles examined within the scope of the study were accessed on the WoS and Ebscohost databases. The database widely used in bibliometric studies, especially for social sciences, is the Web of Science (WoS) database (Yang et al., 2015). WoS database was preferred in this study because it is a user-friendly database with easy-to-use rich content. A folder containing the articles investigated within the scope of the study was obtained on the Ebscohost and WoS databases. The data related to the source file obtained can be downloaded by selecting the "Full Record and Cited References" option in the "Tab-delimited" file format, which is a suitable file type for VOSViewer. Afterward, the related file is transferred onto the program, the criterion (for example, the most used keywords) is determined, and the analysis is completed (van Eck & Waltman, 2020). For this purpose, the following criteria were used for the review: five matches for determining the most frequently used keywords, at least 20 views for the most cited articles and authors, and at least 50 views for the most cited source. To determine the criteria the criteria, the structure of the data was taken into account so that the findings could be interpreted.

3. RESULT / FINDINGS

This section deals with findings and interpretations. For this purpose, firstly, content analysis findings, and then bibliometric analysis findings were discussed.

3.1. Content Analysis Findings related to Studies and Trends in the Field of Measurement Invariance

In the study, each article was examined under predetermined categories. Accordingly, while examining the measurement invariance articles, the following categories were taken into consideration: the type of the test, whose invariance was examined, in terms of the measured feature; groups between which the invariance was examined; the study group; sample size; data analysis method; and the software package used. Accordingly, Table 2 presents the distribution of studies in the field of measurement invariance by maximum performance or the typical response test variables regarding "the type of the test, whose invariance was analyzed, in terms of the measured feature".

Table 2. *The distribution of tests analyzed in studies in terms of measured feature*

Test type	f	%
Typical Response	75	75.76
Maximum Performance	24	24.24
Total	99	100.00

As seen in Table 2, 75.76% of the articles analyzed in the study were typical response tests ($f = 75$) while 24.24% of the tests were determined to be the maximum performance tests. When the tests were examined, it is noteworthy that nine of the maximum performance tests were IQ tests, which outnumbered others. Table 3 presents the distribution of the measurement invariance studies in terms of study group characteristics.

Table 3. *The distribution of the measurement invariance studies in terms of study group characteristics*

Study Groups	f	%
Students	84	84.85
Teachers	9	9.09
Adults	8	8.08
All age groups	2	2.02
Total	103	100.00

As seen in Table 3, students were the most frequent study groups selected in measurement invariance studies with 84.85%. On the other hand, studies selecting teachers as the study group ranked second with 9.09%. Measurement invariance studies based on data obtained from adults constituted 8.08% of all studies. Also, measurement invariance studies using all age groups made up 2.02%. Under this classification, there were a total of 103 study groups. There were four studies using more than one group. For example, the study coded M3 was found to use data obtained from students and teachers. The distribution of measurement invariance studies by sample size is given in Table 4.

Table 4. *The distribution of measurement invariance studies by sample size*

Size of the Study Group	f	%
1- 500	24	24.24
501 -1000	19	19.19
1001 – 1500	20	20.20
1501 or larger	36	37.37
Total	99	100.00

As seen in Table 4, the size of the study group was 1501 or larger in 37.37% of the 99 articles, which examined the measurement invariance. Also, studies with a study group size in the range of 1-500, 501-1000, and 1001-1500 made up 24.24%, 19.19%, and 20.20% of all studies, respectively. The study coded M1 was conducted with a group of 144 people. On the other hand, the study coded M46 was carried out with 42.163 people. This difference was observed to arise from the number of groups used in the studies according to certain variables. For example, in the study coded M1, the time-dependent invariance of the test was examined based on data from a single group, while in the study coded M46, there were students from 10 different grade levels. Table 5 shows the distribution of invariance studies by groups between which invariance was examined.

Table 5. *The distribution of invariance studies by groups between which invariance was examined*

Variable	f	%	Variable	f	%
Gender	42	42.42	Place of residence	3	3.03
Ethnicity	15	15.15	Socio-economic level	3	3.03
Age	15	15.15	Experiment-Control	2	2.02
Country	9	9.09	Paper Pen Test- Computer Based Test	2	2.02
Culture	9	9.09	Status of Learning Difficulty	2	2.02
Education	8	8.08	Self-Peer Assessment	2	2.02
Types of school	6	6.06	Use of Tests in Low-High Risk Exams	1	1.01
Grade level	6	6.06	With-Without Diagnosis of ADHD*	1	1.01
Time	6	6.06	Learning difficulty due to ADHD*	1	1.01
Inter-rater	4	4.04	Student-teacher	1	1.01
Language	3	3.03	Face-to-Face / Online Course	1	1.01
Total			f=142		%=100.00

*ADHD: Attention Deficit Hyperactivity Disorder

The examination of the variables in [Table 5](#) indicated that the tests were mostly analyzed regarding whether they showed measurement invariance for the gender variable. Accordingly, in 42.42% of the articles examined, the invariance of measurement was examined in terms of gender variable. Gender was followed by some demographic characteristics such as age (15.15%), socio-economic level (3.03%), and place of residence (3.03%). Also, cultural variables are important in the measurement invariance studies analyzed. Accordingly, the variables such as ethnicity with 15.15%, country with 9.09%, culture with 9.09%, and language with 3.03% drew attention as cultural variables. On the other hand, education-related variables such as education level (8.08%), school type (6.06%), and grade level (6.06%) also appeared to be among the most used variables in testing invariance of tests. The learning difficulty was another variable employed in the analysis of the measurement invariance of tests. Accordingly, the measurement invariance was also analyzed in terms of learning difficulty with 2.02%, with–without a diagnosis of ADHD with 1.01%, and learning difficulty due to ADHD with 1.01%. Also, the rate of articles analyzing the time-dependent invariance of the tests was 6.06%. In terms of the application methods of the tests, the invariance of the Paper-Pen Tests-Computer-Based Tests (2.02%) was also analyzed in the articles examined. In addition the invariance of tests was analyzed in terms of their use in Low-High Risk Exams (1.01%). Inter-rater invariance was another variable that was analyzed in the studies. For example, inter-rater invariance studies accounted for 4.04% of the total measurement invariance studies, and the self-peer assessment made up 2.02%. Also, the invariance of inter-group tests was studied in the experimental studies. Accordingly, the measurement invariance of the tests applied in terms of experiment-control groups (2.02%) and students taking face-to-face / online courses (1.01%) was examined, too. The examination of the statistical techniques used for measurement invariance analysis in the articles which examined the measurement invariance indicated that the "Multi-Group Confirmatory Factor Analysis" technique was used in all studies. [Table 6](#) presents the distribution of studies according to the statistical software packages used to conduct this analysis.

Table 6. *The distribution of studies according to the statistical software packages used*

Statistical Software	f	%
Mplus	54	54.55
LISREL	14	14.14
Amos	11	11.11
R	7	7.07
Not specified	6	6.06
EQS	6	6.06
SAS/STAT® software	1	1.01
Total	99	100.00

The examination of [Table 6](#) showing the distribution by statistical software packages indicated that the most frequently used statistical software package in measurement invariance studies was Mplus with 54.55%. Also, LISREL, Amos, and EQS, which are often used for structural equation modeling, were also used in studies which investigate measurement invariance with 14.14%, 11.11%, and 6.06%, respectively. R, which is the latest launched software package, was used in 7.07% of the studies. On the other hand, SAS/STAT® was observed to be used in just one of the studies. Besides, 6.06% of the measurement invariance studies were found to not specify the statistical software employed.

3.2. Bibliometric Analysis Findings Regarding Studies and Trends in the Field of Measurement Invariance

With bibliometric mapping analysis, the findings obtained for 99 articles which investigate the measurement invariance were included in this study are presented under this heading. The articles were examined in terms of the most frequently used keywords, the most cited publications, the most cited authors, and the most cited journals. Most frequently used keywords, with at least five matches, were examined in the measurement invariance articles. Accordingly, the map obtained as a result of the analysis is given in Figure 1 and the frequency values for each keyword are given in Table 7.

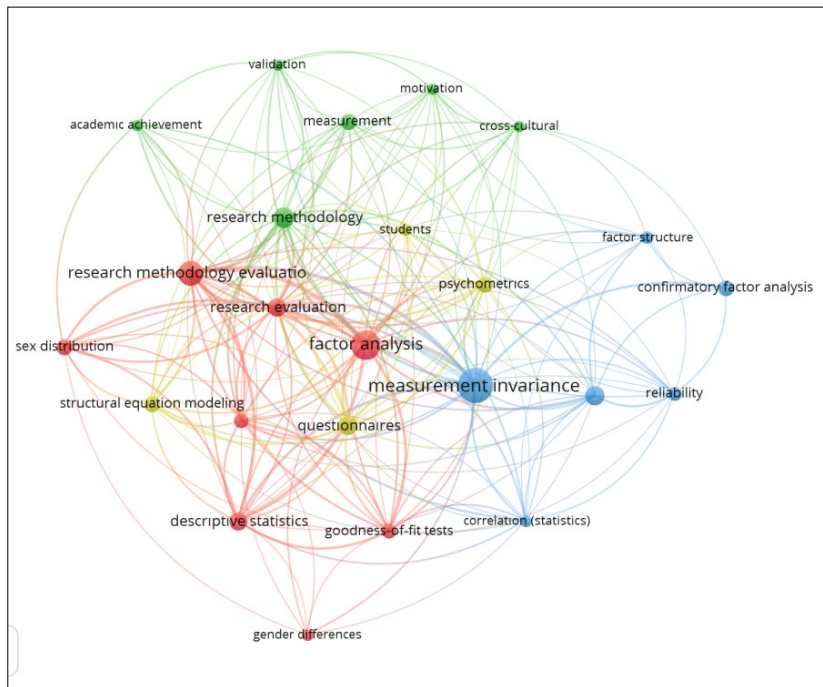


Figure 1. The most frequently used keywords in studies in the field of measurement invariance

Table 7. The most frequently used keywords in studies in the field of measurement invariance

Keywords	f	Keywords	f
Measurement invariance	46	Confirmatory factor analysis	9
Factor analysis	31	Goodness-of-fit tests	8
Research methodology evaluation	23	Self-evaluation	8
Research methodology	15	Reliability	7
Questionnaires	14	Factor structure	6
Validity	14	Correlation	5
Descriptive statistics	13	Students	5
Research evaluation	12	Academic achievement	5
Measurement	10	Cross-cultural	5
Psychometrics	9	Motivation	5
Structural equation modeling	9	Validation	5
Sex distribution	9	Gender differences	5

As seen in Figure 1 and Table 7 showing the most frequently used keywords in measurement invariance studies, as expected, the most frequently used keyword was "measurement invariance" (f=46). "Factor analysis" (f=31), "Structural equation modeling" (f=9),

modeling statistics, was among the most cited publications. Figure 3 shows the map obtained according to the results of the analysis done for determining the most cited authors with at least 20 citations, and Table 9 shows the frequency values of the citations.

Table 8. The most cited publications in studies in the field of measurement invariance

Study Code	Name of The Author	Publication Title	f
Y1	Cheung & Rensvold (2002)	Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance	65
Y2	Hu & Bentler (1999)	Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives	51
Y3	Vandenberg & Lance (2000)	A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research	43
Y4	Meredith (1993)	Measurement invariance, factor analysis and factorial invariance	27
Y5	Byrne, Shavelson & Muthén (1989)	Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance	27
Y6	Chen (2007)	Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance	26
Y7	Horn & Mcardle (1992)	A practical and theoretical guide to measurement invariance in aging research	14
Y8	Hu & Bentler (1998)	Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification	14
Y9	Muthén and Muthén (2007)	Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén	13
Y10	Widaman & Reise (1997)	Exploring the measurement invariance of psychological instruments: Applications in the substance use domain	12
Y11	Kline (2005)	Methodology in the social sciences. Principles and practice of structural equation modeling (2nd ed.). Guilford Press.	12
Y12	Little (1997)	Mean and Covariance Structures (MACS) Analyses of Cross-Cultural Data: Practical and Theoretical Issues	11
Y13	Browne & Cudeck (1993)	Alternative ways of assessing model fit. In K. A. Bollen and J. S. Long (Eds.), Testing structural equation models (pp. 136-162)	11
Y14	Satorra & Bentler (2001)	A scaled difference chi-square test statistic for moment structure analysis	10
Y15	Steenkamp & Baumgartner (1998)	Assessing Measurement Invariance in Cross-National Consumer Research	10
Y16	Brown (2015)	Confirmatory factor analysis for applied research. The Guilford Press.	10

As seen in Table 9 and Figure 3, the most cited author was Cheung, G. W. (f=80). Cheung, G. W was also one of the authors of the most cited publications (f_{Y1}=65). According to the findings, the most cited authors were sequenced similar to the most cited publications. Also, the references to some authors who did not have a publication among the most cited publications were quite high. This might be because the total number of citations to different publications of the related authors was high. These authors included Marsh, H.W. (f=47), Millsap, R.E. (f=29), Martin, A. (f=26), Reynolds, C.R. (f=22), Hancock, G. (f=20), and Bandura, A. (f=21). On the other hand, the book titled "Statistical Approaches to Measurement Invariance" which was written by Millsap, R.E. (f=29) was seen as an important source that addresses measurement invariance. Regarding the investigation of references to measurement invariance studies, Figure 4 shows the bibliographic map obtained considering the references to each journal according to

the results of the analysis done for determining the most cited journals that had at least 50 views, and Table 10 shows the frequency values of the citations.

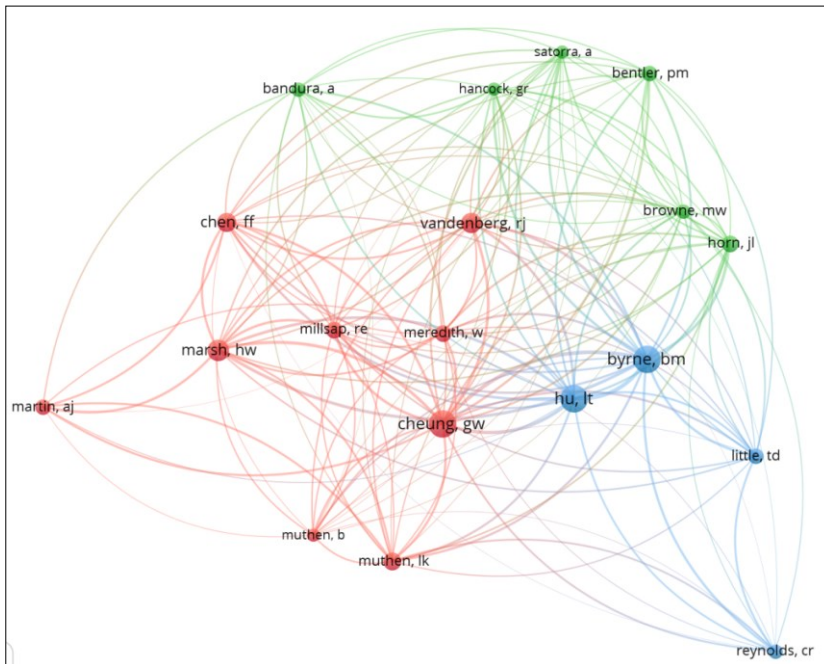


Figure 3. The most cited authors in studies in the field of measurement invariance

Table 9. The most cited authors in studies in the field of measurement invariance

Author	f	Author	f
Cheung, G. W.	80	Bentler, P.	27
Byrne, B. M	77	Martin, A.	26
Hu, L.	75	Little, T.	24
Marsh, H. W.	47	Browne, M. W.	22
Vandenberg, R. J	45	Reynolds, C. R.	22
Chen, F. F.	39	Bandura, A	21
Muthen, L.	36	Muthen, B.	20
Meredith, W.	30	Satorra, A.	20
Millsap, R. E.	29	Hancock, G.	20
Horn, J. L.	28		

As seen in Figure 4 and Table 10, the reference to the journal named "Structural Equation Modeling: A Multidisciplinary Journal" was quite high (f=235). This might be because structural equation modeling has an important place in the investigation of measurement invariance. On the other hand, journals that use the word "psychology" or its derivations in their name were also observed to be cited Journals such as "Psychometrika" (f=103), "Psychological Bulletin" (f=96), "Educational and Psychological Measurement" (f=84), "Psychological Methods" (f=79), "Journal of Psychoeducational Assessment" (f=75), "Journal of Educational Psychology" (f=73), and "Psychological Assessment" (f=51) can be given as examples. This reveals that measurement invariance is important especially in psychological measurements. Also, parallel to the fact that the most used keywords were related to the research methodology, these journals were found to contain the word "research" e.g. "Multivariate Behavioral Research" (f=57), "Organizational Research Methods" (f=54). Considering the importance of

measurement invariance for measurements conducted in the field of education, journals such as "Educational and Psychological Measurement" (f=84), "Journal of Psychoeducational Assessment" (f=75), and "Journal of Educational Psychology" (f=73) were observed to have educational content. On the other hand, journals such as "Journal of Psychoeducational Assessment" and "Journal of Educational Psychology", which contained the articles analyzed in the study, were also among the most cited journals.

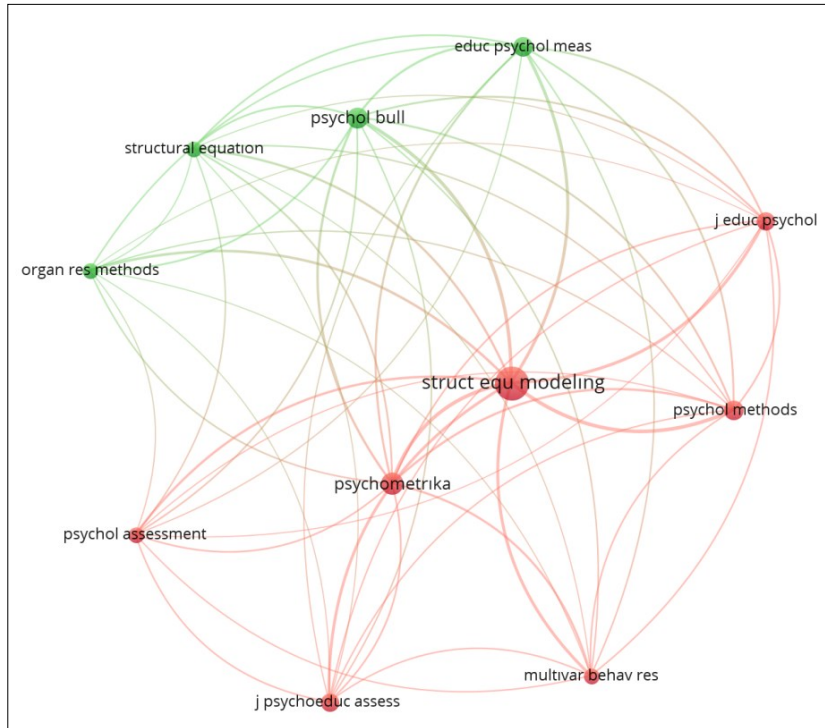


Figure 4. The most cited sources in studies in the field of measurement invariance

Table 10. The most cited sources in studies in the field of measurement invariance

Source	f
Structural Equation Modeling: A Multidisciplinary Journal	235
Psychometrika	103
Psychological Bulletin	96
Educational and Psychological Measurement	84
Psychological Methods	79
Journal of Psychoeducational Assessment	75
Journal of Educational Psychology	73
Multivariate Behavioral Research	57
Organizational Research Methods	54
Structural Equation	54
Psychological Assessment	51

4. DISCUSSION and CONCLUSION

This study examined 99 articles in the field of measurement invariance published between 2008 and 2019 using content and bibliometric analyses. Accordingly, the majority of the articles were found to address the measurement invariance of typical reaction tests. Also, maximum performance tests were among the tests whose measurement invariance was examined. Studies handling inter-group measurement invariance regarding characteristics such as interest,

perception, attitude, and personality measured by typical reaction tests were found to outnumber studies investigating measurement invariance of features such as intelligence or success which were measured by maximum performance tests. Similarly, typical reaction tests were found to be widely used in trend surveys related to the field of science education in Turkey (Erdem, 2011; Göktaş et al., 2012; Selçuk et al., 2014). Also, according to the results of a content analysis study on educational journals with the highest impact factor, the most used tests were achievement tests (Yalçın, et al., 2015). However, in another study in which journals in the field of measurement and evaluation with high impact factors were examined, achievement tests were found to be used in more than half of the studies (Yalçın, 2016). Accordingly, unlike studies in educational sciences and especially in the field of measurement and evaluation, the invariance of typical reaction tests can be said to be analyzed more in measurement invariance studies.

In studies investigating the measurement invariance, the study group or sample was found to consist mostly of students. On the other hand, the study group or samples were also observed to include teachers, adults, and all age groups in measurement invariance studies. This is because students are the focus group in education, and measurement tools developed for students are more than other elements of education. In trend surveys conducted in Turkey in the field of educational sciences, the research group or sample was determined to often consist of students (Arık & Türkmen, 2009; Göktaş, et. al., 2012; Selçuk et al., 2014; Şenyurt & Özer-Özkan, 2017; Yalçın, et al., 2015).

Since structural equation modeling is used in measurement invariance analysis, comparisons between models are known to be based on model fit indexes. For this reason, since many studies have revealed that the group size affects the model fit indexes (Fan & Sivo, 2007; Fan, Thompson & Wang, 1999; Hu & Bentler, 1998; Lei & Lomax, 2005; Mahler, 2011), the size of the group has an important place in measurement invariance studies. Particular attention should be paid to the size of the compared groups. In articles investigating the measurement invariance, the size of study groups or samples was mostly 1501 and above. The sample size increases according to the number of groups for which the invariance is examined. Yet, the average study group or sample size was determined as 3436.39. In the studies found in educational journals with a big impact factor, sample sizes were much higher (Yalçın, 2016; Yalçın, et al., 2015). For example, the average sample size was 81.008 according to the content analysis related to journals that had the highest impact factor and which were reviewed on SSCI in the field of educational sciences (Yalçın, et al., 2015). This is associated with simulation studies and the availability of data from large-scale applications. However, since this study did not include simulation studies, the sample sizes of studies reported here were lower. Sample sizes were even smaller in studies investigating research in the field of educational sciences held in Turkey (Arık & Türkmen, 2009; Göktaş et al., 2012; Selçuk et al., 2014).

In this study, which examined measurement invariance studies, the invariance of measurement tools was found to be mostly analyzed in terms of gender variable. Also, demographic variables such as age, socio-economic level, and place of residence were among the variables that were analyzed for the invariance of measurement tools. Besides, according to the articles examined within the scope of the study, the invariance of the tests was also analyzed in terms of cultural variables such as ethnicity, country, culture, and language. However, some variables related to education were also considered important for the invariance of the tests. These variables can be listed as education level, school type, and grade level. Although the rate was not high, the measurement invariance of measurement tools was also analyzed according to the diagnostic variables of the learning difficulties of individuals. There were also time-dependent invariance analyses that provided important evidence of validity and reliability. The invariance of measurement tools was also analyzed in terms of exam types including paper-pen, computer-based, or low-high risk tests. Finally, measurement invariance analyses were carried out based

on the rater (such as self-peer) or groups created for experimental studies (such as experiment-control). This can be explained by the fact that measurement invariance analyses are performed in terms of many variables since human behaviors measured by measurement tools are complex and abstract. While developing measurement tools, many variables are taken into account from writing items to the selection of groups in experimental studies. However, evidence of whether the measurement tool shows invariance in terms of some variables cannot be obtained during the development stage of the measurement tool. For this reason, conducting measurement invariance studies regarding the important variables is considered to be very important primarily for the validity and reliability of the measurement tool. In their trend analyses on measurement invariance articles, Schmitt and Kuljanin (2008) conducted measurement invariance analyses considering the sub-groups created according to gender, ethnic, cultural, linguistic, and other demographic variables. Indeed, cultural variables have great importance. It is especially necessary to analyze invariance for cultural variables in tests adapted to different cultures. This may explain the abundance of invariance studies conducted in terms of cultural variables in this study. In their trend study investigating measurement invariance studies, which was conducted by Schmitt and Kuljanin (2008), 20 out of 75 articles were tests translated from another language. Similarly, it can be concluded that in invariance studies, invariance analyses are conducted for demographic variables and adapted tests.

In this study, which discusses the measurement invariance of some measurement tools in terms of some variables, the measurement tools and the variables for which the invariance is analyzed show variety, as stated earlier. All of these studies have one aspect in common; that is, they use “Multi-Group Confirmatory Factor Analysis” as the statistical analysis technique. Accordingly, 75 out of 88 articles which were published between 2000 and 2007 and which used the term “measurement invariance” were found to employ confirmatory factor analysis (Schmitt & Kuljanin, 2008). Studies in which measurement invariance studies are analyzed were also determined to use measurement invariance analyses based on the analysis of the difference in variance and covariance matrices, but they were not preferred as much as CFA (Vandenberg & Lance, 2000; Schmitt & Kuljanin, 2008).

The examination of the measurement invariance studies according to the statistical programs used indicated that more than half of the studies had used Mplus statistical software package, and this was followed by LISREL, Amos, R, EQS, and SAS/STAT® according to the frequency of use. Also, some studies had not specified the statistics program employed. In a study investigating articles from the field of measurement and evaluation which had a high impact factor and were reviewed on SSCI, the most frequently used statistical software packages in the descending order were R, Mplus, and SAS (Yalçın, 2015). In another study in the field of educational sciences analyzing articles from journals that had a high impact factor and which were reviewed on SSCI, the frequently used statistical software packages were Mplus, SPSS, and SAS, respectively. On the other hand, the statistical software packages mostly used in trend research in Turkey were SPSS and LISREL, respectively (Arık & Türkmen, 2009; Doğan & Uluman, 2015). The findings obtained in this study were almost similar to other studies conducted in the field of educational sciences. In some articles, the name of the statistical software was not given. This can be due to concerns about avoiding the advertisement in the case of paid software.

As a result of the bibliometric analyses of the measurement invariance studies, the analysis of most frequently used keywords, most cited publications, authors, and sources was conducted. The most frequently used keyword was found to be "measurement invariance". In addition to this, the words specific to factor analysis were used also widely. Among them were the keywords such as "factor analysis", "confirmatory factor analysis", "factor structure", and "goodness of fit tests". Another group of frequently used keywords was related to research methodology such as "evaluation of research methodology", "research methodology", and

"research evaluation", which indicate that measurement invariance studies have an important place in research methodology. Similarly, "validity", "reliability", and "validation" keywords were also among the frequently used keywords. Accordingly, it has been re-established that the measurement invariance is directly related to validity and reliability. Validity and reliability cannot be achieved unless evidence regarding measurement invariance is obtained (Vanderberg & Lance, 2000). Some basic concepts of measurement and evaluation such as "surveys", "descriptive statistics", "measurement", "psychometrics", "self-evaluation" and "correlation" were also included in the measurement invariance studies. Another set of frequently used keywords was observed to include "gender", "gender distribution", "gender differences", and "intercultural". In line with the results of the content analysis, the keyword "students" was found to be among the frequently used keywords. Moreover, "academic achievement" and "motivation" keywords, which are frequently measured concepts in education and psychology, were among the most used keywords in measurement invariance studies.

The article named "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance" written by Cheung and Rensvold (2009) was determined to be the most frequently cited publication among measurement invariance studies published in three journals in the field of measurement and evaluation, which had a high impact and which were reviewed on SSCI. This publication was followed by "Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives", which was written by Hu and Bentler (2009). Apart from these, nine other publications, which were among the most cited were related to statistical processes for analyzing measurement invariance. Vandenberg and Lance's (2000) publication named "A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research" was another most frequently cited study that addressed both statistical processes and trend research regarding measurement invariance comprehensively. Schmitt and Kuljanin (2008) referred to the study of Vandenberg and Lance (2000) in their trend research on measurement invariance and stated that they aimed to examine studies published after this study. Similar to the study of Vandenberg and Lance (2000), the studies of Horn and Mcardle (1992), Widaman and Reise (1997), and Steenkamp and Baumgartner (1998), which dealt with the theoretical, application and evaluation aspects of measurement invariance, were also among the most cited publications. The publications of Horn and Mcardle (1992) and Steenkamp and Baumgartner (1998) were included in the analysis group during reviewing the study of Vandenberg and Lance (2000), which aimed to define and develop measurement invariance theoretically. According to the content analysis findings, the most used statistical software package was "Mplus". The user manual ($f_{Y9}=13$) of the Mplus statistical software package, which is widely used in structural equation modeling statistics, was also among the most cited publications. This user manual was written by Muthén and Muthén (2007) and named "Mplus User's Guide".

According to the results of bibliometric analysis of the measurement invariance studies, the most cited author was identified as Cheung, G. W. In parallel to the most cited publications, the authors of the publications were also in the most cited authors list. This list also included the following authors: Marsh H. W., Millsap R. E., Martin A, Reynolds C. R, Hancock G., and Bandura A. Moreover, the author of the publication that addressed statistical methods regarding measurement invariance, Millsap, R. E. (2011) was among the most cited authors.

When the measurement invariance studies were examined in terms of the most cited sources (journals, publishing houses, etc.), the journal with the highest reference was the "Structural Equation Modeling: A Multidisciplinary Journal". In addition to this, journals from the field of psychology such as "Psychometrics", "Psychological Bulletin", "Educational and Psychological Measurement", "Psychological Methods", "Journal of Psychoeducational Assessment", "Journal of Educational Psychology", and "Psychological Assessment" were also found to be among the most cited journals. This may be associated with the fact that the

measurement invariance of typical response tests was analyzed more. On the other hand, journals such as "Multivariate Behavioral Research" and "Organizational Research Methods" were also frequently cited, which were considered to show up associated with research methodology related keywords. Concerning the place and importance of measurement invariance in the field of measurement and evaluation in education, the following journals from the field of educational research were among the most cited journals: "Educational and Psychological Measurement", "Journal of Psychoeducational Assessment", and "Journal of Educational Psychology".

4.1. Recommendations

According to the findings, the measurement invariance of the maximum performance tests was analyzed less frequently compared to the typical response tests. However, the results of maximum performance tests can provide insights into making important decisions about individuals such as placement in an educational institution or recruitment for a job. For this reason, we recommended that the measurement invariance analysis of maximum performance tests should be increased in terms of many variables to make fairer and more appropriate decisions about individuals.

According to the results obtained from the study, the majority of measurement invariance analyses were found to be done considering the gender variable. Also, many other variables were handled in measurement invariance studies. Accordingly, more reviews and studies should be carried out to guide the developers of measurement tools in terms of showing which variables should be analyzed for a given measurement tool. Considering the importance of measurement invariance in terms of validity and reliability, measurement invariance analyses should be carried out within the scope of validity and reliability studies to emphasize this significance for measurement tool developers, especially for the item writing process.

Given the finding that measurement invariance studies are mostly based on data obtained from students, the measurement invariance studies of measurement tools used in the field of education and psychology targeting groups such as teachers, administrators, and parents can also be carried out. According to the results obtained from the study, many statistical software packages were used in measurement invariance studies. In terms of measurement invariance analysis, the advantages and disadvantages of the related software packages over each other can be investigated.

In this study, measurement invariance studies were examined in terms of various variables. However, we could not analyze the findings obtained as a result of the studies conducted. With the investigation of studies which focus on the measurement invariance of certain measurement tools through certain variables, significant interpretations can be put forward regarding the validity and reliability of the related measurement tool. In this study, measurement invariance studies published in three high impact factor measurement and evaluation journals that were reviewed on SSCI were analyzed. Measurement invariance studies from different databases, journals, years and countries can also be analyzed.

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 author(s).

ORCID

Betul ALATLI  <https://orcid.org/0000-0003-2424-5937>

5. REFERENCES

- Acar, G. M. & Özkan, Ö. Y. (2015). Türkiye'deki eğitim alanında yayımlanan bilimsel dergilerde ölçek geliştirme ve uyarılama konulu makalelerin incelenmesi. *Elektronik Sosyal Bilimler Dergisi*, 14(52), 23-33. <http://dx.doi.org/10.17755/esosder.54872>
- Akaydın, Ş. & Çeçen M. A. (2015). Okuma becerisiyle ilgili makaleler üzerine bir içerik analizi. *Eğitim ve Bilim*, 40(178), 183-198. <http://dx.doi.org/10.15390/EB.2015.4139>
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education [AERA/APA/NCME]. (1999). *Standards for educational and psychological testing*. Washington: American Psychological Association.
- Arıcı, F., Yıldırım, P., Çalıklar, Ş., & Yılmaz, R. M. (2019). Research trends in the use of augmented reality in science education: Content and bibliometric mapping analysis. *Computers & Education*, 142(December), 103647. <http://dx.doi.org/10.1016/j.compedu.2019.103647>
- Arık, R. S., & Türkmen, M. (2009). *Eğitim bilimleri alanında yayınlanan bilimsel dergilerde yer alan makalelerin incelenmesi*. Retrieved November 11, 2019, from <http://www.eab.org.tr/eab/2009/pdf/488.pdf>
- Aslan, C., & Özkubat, U. (2019). Ulusal özel eğitim kongresi bildirilerindeki araştırma eğilimleri: Bir İçerik analizi. *Türkiye Sosyal Araştırmalar Dergisi*, 23(2), 535-554.
- Aydın, A., Erdağ, C., & Sarier, Y. (2010). Eğitim yönetimi alanında yayınlanan makalelerin konu, yöntem ve sonuçlar açısından karşılaştırılması. *Eurasian Journal of Educational Research*, 39, 37-58.
- Aypay, A., Coruk, A., Yazgan, D., Kartal, O., Çağatay, M., Tuncer, B., & Emran, B. (2010). The status of research in educational administration: An analysis of educational administration journals, 1999-2007. *Eurasian Journal of Educational Research*, 39, 59-77.
- Aztekin, S., & Taşpınar Şener, Z. (2015). Türkiye'de matematik eğitimi alanındaki matematiksel modelleme araştırmalarının içerik analizi: Bir meta-sentez çalışması, *Eğitim ve Bilim*, 40(178), 139-161. <http://dx.doi.org/10.15390/EB.2015.4125>
- Bacanak, A., Karamustafaoğlu, S., Değirmenci, S., & Karamustafaoğlu, O. (2011). E-dergilerde yayınlanan fen eğitimi makaleleri: Yöntem analizi. *Türk Fen Eğitimi Dergisi*, 8(1), 119-132.
- Baki, A., Güven, B., Karataş, İ., Akkan, Y., & Çakıroğlu, Ü. (2011). Türkiye'deki matematik eğitimi araştırmalarındaki eğitimler: 1998 ile 2007 yılları arası. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 40, 57-68.
- Bastos, J. L., Celeste, R. K., Faerstein, E., & Barros, A. J. D. (2010). Racial discrimination and health: a systematic review of scales with a focus on their psychometric properties. *Social Science and Medicine*, 70(7), 1091-1099. <https://doi.org/10.1016/j.socscimed.2009.12.20>
- Bozkurt, A., Akgün-Özbek, E., Yılmazel, S., Erdoğan, E., Uçar, H., Güler, E., Sezgin, S., & Dincer, G.D. (2015). Trends in distance education research: A content analysis of journals 2009-2013. *The International Review of Research in Open and Distributed Learning*, 16(1), 330-363. <https://doi.org/10.19173/irrodl.v16i1.1953>
- Boztunç Öztürk, N., Eroğlu, M. G., & Kelecioğlu, H. (2015). Eğitim bilimleri alanında yapılan ölçek uyarılama makalelerinin incelenmesi, *Eğitim ve Bilim*, 40(178) 123-137. <http://dx.doi.org/10.15390/EB.2015.4091>
- Brown, T. A. (2015). *Methodology in the social sciences. Confirmatory factor analysis for applied research (2nd ed.)*. New York: Guilford Press.
- Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, 21(2), 230-258. <https://doi.org/10.1177/0049124192021002005>

- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466. <https://doi.org/10.1037/0033-2909.105.3.456>
- Chang, Y., Chang, C., & Tseng, Y. (2010). Trends of science education research: An automatic content analysis. *Journal of Science Education and Technology*, 19(4), 315-331.
- Chen, F., F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464-504. <https://doi.org/10.1080/10705510701301834>
- Cheung G., W., & Rensvold R., B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9(2), 233-255. https://doi.org/10.1207/S15328007SEM0902_5
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education*. London and New York, NY: Routledge Falmer.
- Çiltaş, A., Güler, G., & Sözbilir, M. (2012). Türkiye’de matematik eğitimi araştırmaları: Bir içerik analizi çalışması. *Kuram ve Uygulamada Eğitim Bilimleri*, 12(1), 565-580.
- Çüm, S., & Koç, N. (2013). Türkiye’de psikoloji ve eğitim bilimleri dergilerinde yayımlanan ölçek geliştirme ve uyarlama çalışmalarının incelenmesi. *Eğitim Bilimleri ve Uygulama*, 12(24), 115-135.
- de Jong O (2007) Trends in western science curricula and science education research: A bird’s eyeview. *Journal of Baltic Science Education*, 6(1). 15–22.
- Doğan C., D., & Uluman M. (2016). İstatistiksel Veri Analizinde R Yazılımı ve Kullanımı. *İlköğretim Online*, 15(2), 615-634., <https://doi.org/10.17051/io.2016.24991>
- Doğan, H., & Tok, T. N. (2018). Türkiye’de eğitim bilimleri alanında yayımlanan makalelerin incelenmesi: Eğitim ve Bilim Dergisi örneği. *Current Research in Education*, 4(2), 94-109.
- Erdem Aydın İ., Kaya S., İşkol S., & İşcan A., (2019). Anadolu Üniversitesi uzaktan eğitim bölümünde yayınlanmış yüksek lisans ve doktora tezlerinin içerik analizi. *Journal of Higher Education and Science*, 9(3), 430-441. <https://doi.org/10.5961/jhes.2019.343>
- Erdem Aydın, İ., Bozkaya, M., & Genç Kumtepe, E. (2019). Research trends and issues in educational technology: Content analysis of TOJET (2012–2018). *The Turkish Online Journal of Educational Technology*, 18(4), 46-61.
- Erdem, D. (2011). Türkiye’de 2005–2006 yılları arasında yayımlanan eğitim bilimleri dergilerindeki makalelerin bazı özellikler açısından incelenmesi: Betimsel bir analiz. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2(1), 140-147.
- Erdoğan, F., U., & Çağiltay, K. (2009, Şubat). *Türkiye’de eğitim teknolojileri alanında yapılan master ve doktora tezlerinde genel eğilimler*. Paper presented at the XI. Akademik Bilişim Konferansı, Harran Üniversitesi, Şanlıurfa. Retrieved May 15, 2019, from https://ab.org.tr/ab09/kitap/erdogmus_cagiltay_AB09.pdf
- Eybe, J., & Schmidt, H.-J. (2001). Quality criteria and exemplary papers in chemistry education research. *International Journal of Science Education*, 23, 209-225. <https://doi.org/10.1080/09500690118920>
- Falkingham, L. T. & Reeves, R. (1998). Context analysis a technique for analysing research in a field, applied to literature on the management of R and D at the section level. *Scientometrics*, 42(2), 97-120. <https://doi.org/10.1007/bf02458351>
- Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, 42(3), 509-529. <https://doi.org/10.1080/00273170701382864>

- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, 6(1), 56–83. <https://doi.org/10.1080/10705519909540119>
- Fazlıoğulları, O., & Kurul, N. (2012). Türkiye'deki eğitim bilimleri doktora tezlerinin özellikleri. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 12(24), 43-75.
- Fraenkel, J.R. & Wallen, N. (2005). *How to design and evaluate research in education*. New York, NY: McGrawHill.
- Gotch, C. M., & French, B. F. (2014). A systematic review of assessment literacy measures. *Educational Measurement: Issues and Practice*, 33, 14-18. <http://dx.doi.org/10.1111/emip.12030>
- Göktaş, Y., Küçük, S., Aydemir, M., Telli, E., Arpacık, Ö., Yıldırım, G., & Reisoğlu, I. (2012). Türkiye'de eğitim teknolojileri araştırmalarındaki eğilimler: 2000-2009 dönemi makalelerinin içerik analizi. *Kuram ve Uygulamada Eğitim Bilimleri*, 12(1), 177-199.
- Gülbahar, Y., & Alper, A. (2009). Öğretim teknolojileri alanında yapılan araştırmalar. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 42(2), 93-111. https://doi.org/10.1501/Egifak_0000001178
- Hart, L. C., Smith, S. Z., Swars, S. L., & Smith, M. E. (2009). An examination of research methods in mathematics education: 1995–2005. *Journal of Mixed Methods Research*, 3(1) 26–41. <https://doi.org/10.1177/1558689808325771>
- Hazır Bıkmaç, F., Aksoy, E., Tatar, Ö., & Atak Altınyüzük, C. (2013). Eğitim programları ve öğretim alanında yapılan doktora tezlerine ait içerik çözümlemesi (1974-2009). *Eğitim ve Bilim*, 38(168), 288-303.
- Hew, K. F., Kale, U., & Kim, N. (2007). Past research in instructional technology: Results of a content analysis of empirical studies published in three prominent instructional technology journals from the year 2000 through 2004. *Journal of Educational Computing Research*, 36(3), 269-300. <https://doi.org/10.2190/K3P8-8164-L56J-33W4>
- Hinkin, T. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, 21(5), 967-988. [https://doi.org/10.1016/0149-2063\(95\)90050-0](https://doi.org/10.1016/0149-2063(95)90050-0)
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale, NJ: Lawrence Erlbaum.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3-4), 117-144. <https://doi.org/10.1080/03610739208253916>
- Hsu, T. (2005). Research methods and data analysis procedures used by educational researchers. *International Journal of Research & Method in Education*, 28(2), 109–133. <http://dx.doi.org/10.1080/01406720500256194>
- Hu L-T. & Bentler, M., P. (1999). Cut off 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>
- Hu, L.T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to under parameterized model misspecification. *Psychological Methods*, 3(4), 424–453. <https://doi.org/10.1037/1082-989X.3.4.424>
- ITC (2005). *International test commission guidelines for test adaptation*. London: Author.
- Ivanović L., & Ho Y. S. (2019) Highly cited articles in the education and educational research category in the Social Science Citation Index: A bibliometric analysis, *Educational Review*, 71(3), 277-286. <https://doi.org/10.1080/00131911.2017.1415297>
- Kapuscinski, A. N., & Masters, K. S. (2010). The current status of measures of spirituality: A critical review of scale development. *Psychology of Religion and Spirituality*, 2(4), 191–205. <https://doi.org/10.1037/a0020498>

- Karadağ, E. (2009). Eğitim bilimleri alanında yapılmış doktora tezlerinin tematik açıdan incelemesi. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi*, 10(3), 75-87.
- Karasar N. (2017). *Bilimsel araştırma yöntemi: Kavramlar ilkeler teknikler*. Ankara: Nobel Yayıncılık.
- Kazu, H., & Aslan, S. (2013). 2004 ilköğretim programının ölçme-değerlendirme boyutu ile ilgili yapılan araştırmaların değerlendirilmesi. *İlköğretim-Online*, 12, 87-108.
- Kazu, İ., Y., & Deniz, E. (2019). Öğretmenlerin ölçme ve değerlendirme tekniklerini kullanma durumlarını inceleyen araştırmaların değerlendirilmesi. *Journal of History School*, 12(XXXVIII) 174-195. <https://doi.org/10.14225/Joh1527>
- Keselman, H. J., Huberty, C. J., Lix, L. M., Olejnik, S., Cribbie, R. A., Donahue, B., & Levin, J. R. (1998). Statistical practices of educational researchers: An analysis of their ANOVA, MANOVA and ANCOVA analyses. *Review of educational research*, 68(3), 350-386. <https://doi.org/10.3102/00346543068003350>
- Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in AERJ and JCP articles from 1988 to 1997: A methodological review. *Journal of Experimental Education*, 69(3), 280-309. <https://doi.org/10.1080/00220970109599489>
- Kilbourne, W. E., & Beckmann, S.C., (1998). Review and critical assessment of research on marketing and the environment. *Journal of Marketing Management* 14(6), 513-32. <https://doi.org/10.1362/026725798784867716>
- Kline, R. B. (2005). *Methodology in the social sciences. Principles and practice of structural equation modeling (2nd ed.)*. New York: Guilford Press.
- Küçüköğlü, A., & Ozan, C. (2013). Sınıf öğretmenliği alanındaki lisansüstü tezlere yönelik bir içerik analizi. *Uluslararası Avrasya Sosyal Bilimler Dergisi*, 4(12), 27-47.
- Ladhari, R. (2010). Developing e-service quality scales: A literature review. *Journal of Retailing and Consumer Services*, 17, 464-477. <https://doi.org/10.1016/j.jretconser.2010.06.003>.
- Lee, M.H., Wu, Y.T. & Tsai, C.C. (2009). Research trends in science education from 2003 to 2007: A content analysis of publications in selected journals. *International Journal of Science Education*, 31(15), 1999-2020. <https://doi.org/10.1080/09500690802314876>
- Lei, M., & Lomax, R. G. (2005). The effect of varying degrees of nonnormality in structural equation modeling. *Structural Equation Modeling*, 12(1), 1-27. https://doi.org/10.1207/s15328007sem1201_1
- Lin, T. C., Lin, T. J., & Tsai, C. C. (2014). Research trends in science education from 2008 to 2012: A systematic content analysis of publications in selected journals. *International Journal of Science Education*, 36(8), 1346-1372. <http://dx.doi.org/10.1080/09500693.2013.864428>
- Little, T., D. (1997). Mean and Covariance Structures (MACS) analyses of cross-cultural data: Practical and theoretical issues, *Multivariate Behavioral Research*, 32(1), 53-76. https://doi.org/10.1207/s15327906mbr3201_3
- Mahler, C. (2011). *The effects of misspecification type and nuisance variables on the behaviors of population fit indices used in structural equation modeling*. (Unpublished master dissertation The Faculty of Graduate Studies, University of British Columbia, Vancouver, Canada). Retrieved January 17, 2020, from <https://open.library.ubc.ca/cIRcle/collection/s/ubctheses/24/items/1.0105120>
- Meredith W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543. <https://doi.org/10.1007/BF02294825>
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge: Taylor & Francis Group.
- Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2017). Scale development: Ten main limitations and recommendations to improve future

- research practices. *Psicologia: Reflexão e Crítica*, 30(3). <https://doi.org/10.1186/s41155-016-0057-1>
- Murphy, J, Vriesenga, M., & Storey, V. (2007). Educational administration quarterly, 1979-2003: An analysis of types of work, methods of investigation, and influences. *Educational Administration Quarterly*, 43(5), 612-628. <https://doi.org/10.1177/0013161X07307796>
- Muthén, L. K., & Muthén, B. O. (2008). Mplus (Version 5.1) [Computer software]. Los Angeles: Muthén & Muthén.
- Oxford Dictionary (2017). Retrieved January 15, 2020, from <https://en.oxforddictionaries.com/definition/bibliometrics>
- Öncül, N. (2014). Türkiye'de erken çocuklukta özel eğitim ile ilgili yapılmış makalelerin gözden geçirilmesi. *International Journal of Early Childhood Special Education*, 6(2), 247-284.
- Ören F. Ş. & Sarı, K. (2019). Web of Science veri tabanında fen eğitimi üzerine yapılan araştırmaya dayalı öğrenme stratejisi konulu çalışmaların değerlendirilmesi. *İlköğretim Online*, 18(4), 1875-1901. <https://doi.org/10.17051/ilkonline.2019.639353>
- Özkan, M. (2016). Liderlik hangi sıfatları, nasıl alıyor? Liderlik konulu makalelerin incelenmesi. *Gaziantep Üniversitesi Sosyal Bilimler Dergisi*, 15(2), 615-639. <https://doi.org/10.21547/jss.256732>
- Özyurt, Ö., & Özyurt, H. (2015) Learning style based individualized adaptive e-learning environments: Content analysis of the articles published from 2005 to 2014. *Computers in Human Behavior*, 52, 349-358. <https://doi.org/10.1016/j.chb.2015.06.020>
- Patton, M. Q. (2002). *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage.
- Saracaloğlu, A. S., & Dursun, F. (2010, Mayıs). Türkiye'de eğitim programları ve öğretim alanındaki lisansüstü tezlerinin incelenmesi. Paperpresented at the 1. Ulusal Eğitim Programları ve Öğretim Kongresi, Balıkesir Üniversitesi, Ayvalık. Retrieved October, 9, 2019, from https://www.pegem.net/akademi/kongrebildiri_detay.aspx?id=117909
- Satorra, A., & Bentler, P., M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507-514. <https://doi.org/10.1007/BF02296192>
- Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review*, 18(4), 210-222. <https://doi.org/10.1016/j.hrmr.2008.03.003>
- Seçer, İ., Ay, İ., Ozan, C., & Yılmaz, B. Y. (2014). Rehberlik ve psikolojik danışma alanındaki araştırma eğilimleri: Bir içerik analizi. *Turkish Psychological Counseling and Guidance Journal*, 5(41), 49-60.
- Selçuk, Z., Palancı, M., Kandemir, M. & DüNDAR, H. (2014). Eğitim ve bilim dergisinde yayınlanan araştırmaların eğilimleri: İçerik analizi. *Eğitim ve Bilim*, 39(173), 430-453. <https://doi.org/10.15390/eb.v39i173.3278>
- Sırakaya M., & Alsancak Sırakaya, D. (2020). Augmented reality in STEM education: A systematic review. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2020.1722713>
- Small, H. (1999). Visualizing science by citation mapping for Information Science. *Journal of the American Society*, 50, 799-813. [https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:93.3.CO;2-7](https://doi.org/10.1002/(SICI)1097-4571(1999)50:93.3.CO;2-7)
- Sözbilir, M., & Kutu, H. (2008). Development and currentstatus of scienceeducationresearch in Turkey. *Essays in Education, Special Issue*, 1-22.
- Steenkamp, J., & Baumgartner, H. (1998). Assessing Measurement Invariance in Cross-National Consumer Research. *Journal of Consumer Research*, 25(1), 78-107. <https://doi.org/10.1086/209528>

- Sveinbjornsdottir, S., & Thorsteinsson, E. B. (2008). Adolescent coping scales: A critical psychometric review. *Scandinavian Journal of Psychology*, 49(6), 533-548. <https://doi.org/10.1111/j.1467-9450.2008.00669.x>
- Şahin, M. G., & Boztunç Öztürk, N. (2018). Eğitim alanında ölçek geliştirme süreci: Bir içerik analizi çalışması. *Kastamonu Eğitim Dergisi*, 26(1), 191-199. <https://doi.org/10.24106/kefdergi.375863>
- Şenyurt, S., & Özer Özkan, Y. (2017). Eğitimde ölçme ve değerlendirme alanında yapılan yüksek lisans tezlerinin tematik ve metodolojik açıdan incelenmesi. *Elementary Education Online*, 16(2), 628-653. <https://doi.org/10.17051/ilkonline.2017.304724>
- Tarman, B., Güven, C., & Aktaşlı, İ. (2011). Türkiye’de sosyal bilgiler eğitimi alanında yapılan doktora tezlerinin değerlendirilmesi ve alana katkıları. *Selçuk Üniversitesi Ahmet Keleşoğlu Eğitim Fakültesi Dergisi*, 32, 391-410.
- Tavşancıl, E., & Aslan, E. (2001). *Sözel, yazılı ve diğer materyaller için içerik analizi ve uygulama örnekleri*. İstanbul: Epsilon Yayınları.
- Tavşancıl, E., Çokluk, Ö., Çıtak, G., Kezer, F., Yıldırım, Ö., Bilican, S., Büyükturan, E., Şekercioğlu, G., Yalçın, N., Erdem, D., & Özmen, T. (2010). *Eğitim Bilimleri Enstitülerinde Tamamlanmış Lisansüstü Tezlerin İncelenmesi (2000-2008)*. Ankara Üniversitesi Bilimsel Araştırma Projeleri. Retrieved June 9, 2019, from <https://dspace.ankara.edu.tr/xmlui/handle/20.500.12575/68960>
- Tavşancıl, E., Güler, G., & Ayan, C. (2014, June). 2002-2012 yılları arasında Türkiye’de geliştirilen bazı tutum ölçeği geliştirme çalışmalarının ölçek geliştirme süreci açısından incelenmesi. Paperpresented at the IV. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı) 9-13 Haziran, Hacettepe Üniversitesi, Ankara.
- Tekin, H. (2008). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı Yayınevi.
- Tsai, C.-C. & Wen, L.M.C. (2005). Research and trends in science education from 1998 to 2002: A content analysis of publication in selected journals. *International Journal of Science Education*, 27, 3–14. <https://doi.org/10.1080/0950069042000243727>
- Turan, S., Karadağ, E., Bektaş, F., & Yalçın, M. (2014). Türkiye’de eğitim yönetiminde bilgi üretimi: Kuram ve Uygulamada Eğitim Yönetimi Dergisi 2003-2013 yayınlarının incelenmesi. *Kuram ve Uygulamada Eğitim Yönetimi Dergisi*, 20(1), 93-119. <https://doi.org/10.14527/kuey.2014.005>
- Turgut, M., F., (2006). *Eğitimde ölçme ve değerlendirme*. Ankara: Saydam Yayıncılık.
- Ulutaş, B., Üner, S., Turan Oluk, N., Yalçın Çelik, A., & Akkuş, H. (2015). Türkiye’deki kimya eğitimi makalelerinin incelenmesi: 2000-2013. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 16(2), 141-160.
- Ulutaş, F. & Ubuz, B. (2008). Matematik eğitiminde araştırmalar ve eğilimler: 2000 ile 2006 yılları arası. *İlköğretim Online*, 7(3), 614-626.
- vanEck, N. J., & Waltman, L. (2020). *VOSviewer manual. Manual for VOSviewer version 1.6.14, software documentation*. Leiden: Univeristeit Leiden.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–69. <https://doi.org/10.1177/109442810031002>
- Vega-Arce, M., Salas, G., Núñez-Ulloa, G., Pinto-Cortez, C., Fernandez, I. T., & Ho, Y.S. (2019). Research performance and trends in child sexual abuse research: A Science Citation Index Expanded-based analysis. *Scientometrics*, 121(3), 1505-1525. <https://doi.org/10.1007/s11192-019-03267-w>

- Wang, Y., Lai, N., Zuo, J., Chen, G., & Du, H. (2016) Characteristics and trends of research on waste-to-energy incineration: A bibliometric analysis, 1999–2015. *Renewable and Sustainable Energy Reviews*, 66, 95–104. <http://doi.org/10.1016/j.rser.2016.07.006>
- White, R. (1997). Trends in research in science education. *Research in Science Education*, 27(2), 215-221. <https://doi.org/10.1007/BF02461317>
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substanceuse domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (p. 281–324). American Psychological Association. <https://doi.org/10.1037/10222-009>
- Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist*, 34(6), 806–838. <https://doi.org/10.1177/0011000006288127>
- Yalçın, S. (2016). Content analysis of research articles in measurement and evaluation journals. *Ankara University Journal of Faculty of Educational Science*, 49(1), 65-84. https://doi.org/10.1501/Egifak_0000001375
- Yalçın, S., Yavuz, H. Ç. ve İlgün Dibek, M. (2015). An examination of articles published in educational journals having highest impact factors: Content analysis. *Eğitim ve Bilim*, 40(182), 1-28.
- Yang, Y. T., Iqbal, U., Ching, J. H. Y., Ting, J. B. S., Chiu, H. T., Tamashiro, H., & Hsu, Y. H. E. (2015). Trends in the growth of literature of telemedicine: A bibliometric analysis. *Computer Methods and Programs in Biomedicine*, 122(3), 471-479. <https://doi.org/10.1016/j.cmpb.2015.09.008>
- Yılmaz, K., & Altinkurt, Y. (2012). An examination of articles published on preschool education in Turkey. *Kuram ve Uygulamada Eğitim Bilimleri*, 12(4), 3227-3241.
- Zainuddin, Z., WahChu, S. K., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research*, 30. <https://doi.org/10.1016/j.edurev.2020.100326>
- Zupic, I. (2015). Bibliometric methods in management and organization. *Organizational Research Methods*, 18(3), 429-472. <https://doi.org/10.1177/1094428114562629>