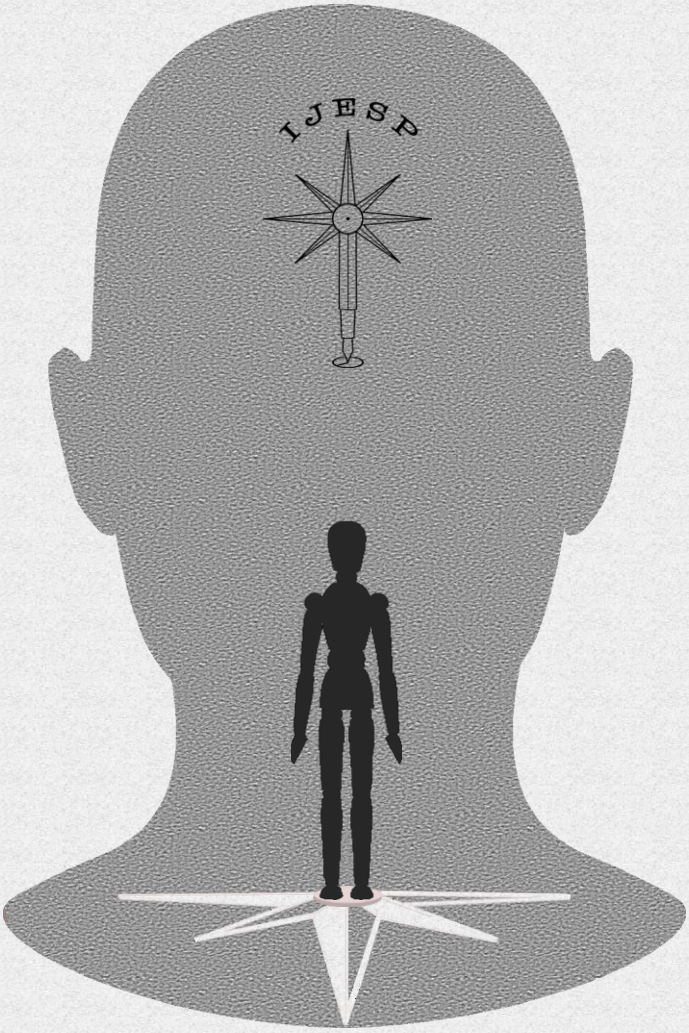
I J E S P

International Journal of Educational Studies and Policy

ISSN: 2718-1022

Volume: 5, Issue: 2, November 2024



www.ijesp.net

International Journal of Educational Studies and Policy (IJESP)

Volume:5, Issue:2, November 2024

International Journal of Educational Studies and Policy (IJESP) focuses on teacher education, applied studies, and research methods within the scope of research, review and conceptual studies in the field of educational sciences. It is an international peer-reviewed journal that publishes two issues a year, around May and November, and its publication language is English.

Web: www.ijesp.net

E-mail for correspondence: info@jesp.net

CONTENTS

IJESP Volume:5, Issue:2, November 2024

International Journal of Educational Studies and Policy (IJESP)

Volume: 5, Issue: 2, November 2024

Bibliometric Analysis of COVID-19 Publications in Education Between 2020 and 2022*

Mehmet Erdoğan[1], Zeynep Meral Tanrıöğen[2]

ABSTRACT

The COVID-19 pandemic has spurred significant scholarly interest, particularly within the realm of education, prompting researchers to investigate its multifaceted impacts. Despite the global urgency, prior literature reviews have identified a scarcity of bibliographic research addressing the pandemic's effects on education. To address this gap**,** a meticulous literature review employing bibliometric analysis was conducted, scrutinizing 1,659 publications from the Web of Science database between 2020 and 2022. The study uncovered various characteristics of these publications, including publication type, language, contributing countries, and prevalent keywords. Notably, the United States and China emerged as the leading contributors to research on this subject during the pandemic. Furthermore, all publications predominantly comprised original articles, with Taylor and Francis being the most prolific journal in terms of publication count and citations. The key topics focused on areas like COVID-19 and its impact on education. Additionally, discussions included remote learning, online education platforms, and the psychological toll of the pandemic on both students and educators. The study explored diverse topics, such as distance learning modalities and quality of life concerns. Importantly, the study's findings offer valuable insights for educational administrators, potentially guiding future research endeavors and facilitating informed decision-making within academic institutions. These results not only highlight current trends but also suggest new areas for exploration, making this study a crucial step for continued research on education during the pandemic.

Keywords: Bibliometric analysis, content analysis, COVID-19, education, VOSviewer, Web of Science

DOI: https://doi.org/10.5281/zenodo.14012411

Article Info:

Received: 07.03.2024 **Accepted: 16.09.2024** **Article Type: Research Article**

Cite as: Erdoğan. M. & Tanrıöğen, Z. M. (2024). Bibliometric analysis of COVID-19 publications in education between 2020 and 2022. *International Journal of Educational Studies and Policy*, *5*(2), 98-118.

[1]Corresponding Author: Mehmet Erdoğan, Ministry of Education, mehmeterdogan640@outlook.com, ORCID: 0000 0001 7731 0718

[2] Assoc. Prof. Dr. Zeynep Meral Tanrıöğen, Pamukkale Universty, zatogen@gmail.com, ORCID: 0000 0001 6300 9457

*This study was presented as an oral presentation at the International Education Management Form EYFOR-14 on May 03-07, 2023.

Introduction

COVID-19, originating in 2019, was designated as a global health emergency by the World Health Organization (WHO) in the same year. This disease has spread worldwide, leading to high rates of psychological disorders and deaths. Following the declaration of the pandemic, countries have created emergency action plans and implemented various measures to combat the outbreak. As highlighted by Korkmaz and Altuntaş (2022), this phenomenon has amplified research attention towards COVID-19. An illustrative instance of the extensive influence of the pandemic on societal domains like education includes the widespread adoption of remote learning due to the suspension of in-person classes across numerous countries during the 2019-2020 academic period. As of 2020, schools in many countries worldwide were closed for face-to-face education (Huber and Helm, 2020). Selwyn (2012) and Watson & Murin (2014) note that this period coincided with a time of new technologies and rapid digital developments in education. Therefore, advancements in the digital realm of education were already prominent before the pandemic (Gewerkschaft Erziehung and Wissenschaft, 2020).

The pandemic period has led to shifts towards new education methods (Lake and Dusseault, 2020). However, individual differences in access to educational technologies and the unfamiliarity of educators and students with distance education have initially caused confusion. Additionally, the extended break from in-person instruction has forced families to take on more responsibility for caring for their students (Harris, 2020). The educational impact of COVID-19, coupled with broader social disruption including economic hardship, unemployment, protests against racial injustice, and health threats, has led to more people dropping out of college. This creates a complex situation that makes it difficult to assess student success. As academic publications rapidly increase, staying up-to-date becomes challenging, and accessing existing knowledge and synthesizing past research becomes a complex process. Therefore, literature reviews play a critical role in advancing research fields and supporting evidence-based practices (Rousseau, 2012). Systematic literature reviews, involving both quantitative and qualitative approaches, are a common technique researchers use to examine past studies (Aria and Cuccurullo, 2017).

Bibliometrics has the potential to statistically evaluate activities in the scientific field. This methodology analyzes the production and interaction of scientific knowledge through scientific publications, citations, and other academic indicators. Thus, it enables an understanding of researchers' work, developments in scientific fields, and research trends. Bibliometric analyses provide an objective evaluation of scientific activities and offer a transparent, systematic, and replicable review process. This procedure is crucial for recognizing research that contributes to the advancement of science and for comprehending the factors affecting the spread of scientific knowledge (Broadus, 1987; Diodato, 1994; Pritchard, 1969).

Bibliometrics, as a discipline, enables the impartial assessment of scientific endeavors and offers a more objective and dependable analysis compared to alternative approaches. Alongside new knowledge and conceptual advancements, the increasing availability of data allows for various advantages, such as determining trends over time, defining research areas, monitoring interdisciplinary interactions, and identifying the most effective scientists and institutions (Donthu, Kumar, Mukherjee, Pandey, & Lim, 2021).

Examinations conducted on extensive datasets offer a more inclusive perspective on advancements within the scientific community. This constitutes a valuable instrument for grasping

the evolution of scientific knowledge and offering an outline of ongoing research endeavors (Crane, 1972). The use of bibliometric methods emerges as an important tool for identifying, understanding, and addressing issues arising in the context of a pandemic. This research emphasizes the utilization of bibliographic analysis to evaluate the worldwide repercussions of the pandemic, with a particular emphasis on articles investigating educators and the pandemic's effects on education. Bibliometric analyses conducted on published articles serve as a guide for education administrators and researchers to make informed decisions and anticipate future developments. The results of these analyses will contribute to the effective evaluation of studies on pandemics and education, as well as the development of more effective preparation and response strategies in education against pandemics.

This study delves into the significant impact of the COVID-19 pandemic on the education sector and assesses scholarly research in this field through the application of bibliographic analysis techniques. Aimed at understanding the current status and evolution of COVID-19-related publications in education, this study seeks to identify gaps in the literature and provide recommendations. It encourages academics in the education field to easily access current information on COVID-19 and stay updated on developments in this area. Additionally, it aims to reach a broader audience by summarizing the findings of existing scientific publications and raising awareness on this topic. In line with these objectives, the research question is formulated as "What is the current knowledge structure and development of publications related to education and COVID-19?"

1. What is the breakdown of research conducted from 2020 to 2022?
2. In which nations are the highest number of publications originating?
3. Which journals, publications, authors, and countries receive the most citations?
4. Which journal boasts the highest number of publications?
5. To what extent do authors and countries collaborate?
6. What are the predominant themes and concepts explored through keyword analysis in research?

Method

The bibliometric methodology employs quantitative techniques to systematically analyze bibliometric data, encompassing publications and citations. Key among these techniques is citation analysis, which examines how often articles are cited within a field to assess their impact and influence (Broadus, 1987; Pritchard, 1969). This analysis helps identify seminal works, influential authors, and prominent journals, offering insights into the dissemination and reception of research findings over time. Co-authorship analysis explores collaborative patterns among researchers, revealing networks of collaboration and the geographic distribution of scientific contributions (Huang & Chang, 2011; Talan, 2021). Additionally, co-word analysis examines relationships between terms in publications, uncovering thematic trends and the evolution of research discourse (Bağış, 2021). Together, these bibliometric techniques provide a robust framework for understanding the dynamics of scientific research, facilitating insights into research trends, collaborative networks, and emerging topics within a field.

Research Model

This research employed bibliographic analysis methods alongside document review methodology to accomplish its research objectives and address the research inquiries. Document review is particularly valuable when extensive interviews and observations are impractical. It involves a comprehensive examination of documentary sources, offering insights into the phenomenon or event under investigation. The study integrated bibliographic analysis techniques with document review methodology to provide a robust framework for analysis.

Bibliometric Analysis and Document Review

Document review is considered an important research method in cases where concrete data collection is not possible. In this context, written sources were meticulously examined to obtain and analyze the necessary data for the research objectives (Güçlü, 2014). According to Şimşir (2021), bibliometrics is a type of analysis that provides significant facilitation in identifying studies representing a particular topic. By scrutinizing the attributes of publications within a specific field, this approach unveils diverse insights concerning scientific output. These findings encompass elements like publication volume, researched topics, affiliations of contributors, and keywords (Çiftçi et al., 2016). Bibliometric research serves as an important tool for understanding developments in scientific fields and research trends. Donthu et al. (2021), Çiftçi et al. (2016), and Şimşir (2021) note that bibliometrics offers an effective way to understand the qualitative and quantitative characteristics of scientific production in a specific field. By analyzing factors influencing the dissemination of scientific knowledge, this method provides researchers with a broader perspective. Various bibliographic methods were employed in this study. Citation analysis assesses the influence of articles, authors, journals, and countries based on the number of citations received (Bağış, 2021; Erdoğan, 2021). Coauthorship analysis investigates instances where multiple authors contribute to a publication, shedding light on collaboration among authors (Huang & Chang, 2011; Talan, 2021). Conversely, co-word analysis explores relationships between terms found in titles, abstracts, or keywords of studies within a research field (Bağış, 2021). This form of analysis is specifically suited for keyword co-occurrence examination, offering deeper insights into conceptual trends and relationships within the research domain.

Data Collection/ Research material

The study utilized data sourced from the Web of Science Core Collection (WoSCC) database, encompassing pertinent publications spanning from January 1, 2020, to December 31, 2022. A literature search was conducted employing the keywords "education" (all fields), "Novel Coronavirus Infection (COVID-19)," and "Pandemic". The selection of these keywords facilitated a focused examination of the relationship between education and the pandemic. The research results indicate access to 1659 publications, including original articles indexed by the Social Science Citation Index (SSCI) between 2020 and 2022, published in English. All bibliographic data were transferred to an Excel spreadsheet from the Web of Science database. This electronic spreadsheet facilitated easier analysis of the data and served as a foundation for comprehensive bibliometric analysis of the research.

Data Analyses

This research employed two distinct methodologies to analyze the data: descriptive content analysis techniques and bibliographic content analysis. Descriptive content analysis was conducted on the Web of Science Core Collection (WoSCC) platform, followed by an in-depth examination of the gathered outcomes. This analysis aimed to uncover descriptive characteristics by examining various attributes of the publications (such as publication years, languages, most cited first publications or journals, etc.). On the other hand, bibliometric analysis was carried out using the VOSviewer software tool for mapping and visualization. In this analysis process, descriptive characteristics such as the distribution of publications by years, languages, and most cited first publications or journals were examined, and tables and figures were created based on this information. These two distinct analytical approaches facilitated a thorough organization of the study and an in-depth analysis of the acquired data. The VOSviewer software was employed to visually represent relationships and collaborative networks among publications, journals, authors, and countries. Additionally, this software was used to visualize the frequency of co-occurring terms in keywords. VOSviewer is an effective tool for visualizing the relationships and collaboration networks among publications, journals, authors, countries, and keywords, as well as the frequency of co-occurring terms in keywords.

Applied Procedures

The procedures applied in this research encompassed several critical steps. Initially, data retrieval was conducted through a systematic search of the Web of Science Core Collection (WoSCC) database, utilizing specific keywords to extract bibliographic data from January 1, 2020, to December 31, 2022. During the data processing stage, the retrieved bibliographic data were transferred into an Excel spreadsheet for preliminary analysis. The data were then organized by attributes such as publication years and languages, followed by thorough cleaning and validation to ensure accuracy and completeness. Descriptive content analysis was performed using the WoSCC platform to gain an overview of publication trends. This involved identifying key attributes, including publication volume, languages, and highly cited journals. For a more in-depth bibliometric analysis, VOSviewer was employed. This software facilitated citation analysis to pinpoint influential articles, authors, and journals; coauthorship analysis to explore collaborative patterns, and co-word analysis to investigate keyword co-occurrence and emerging conceptual trends. Visualization of the data involved creating visual maps with VOSviewer to illustrate the relationships and networks among publications, authors, and keywords. These visualizations helped provide a comprehensive understanding of the collaborative and conceptual landscapes within the research domain. During the interpretation phase, these visual maps and tables were

analyzed to derive insights into publication trends, collaboration networks, and research themes, thereby offering a deeper understanding of scientific output and its dissemination patterns. Ethical considerations throughout the research were paramount, including ensuring the accuracy and integrity of the data, respecting intellectual property rights, and properly citing all sources used in the analysis. Adhering to these ethical principles was crucial for maintaining the credibility and reliability of the research findings. Overall, these meticulous procedures ensured a thorough and systematic examination of the bibliometric data, enabling the study to effectively achieve its research objectives.

Ethics committee approval process

Since this study involves bibliometric analysis, ethical committee approval was not required. Bibliometric studies are conducted through the systematic analysis of existing literature and publications, and do not entail direct research involving human or animal subjects. Consequently, ethical committee approval is typically not necessary for such studies (Aria & Cuccurullo, 2017).

Results

Descriptive Results

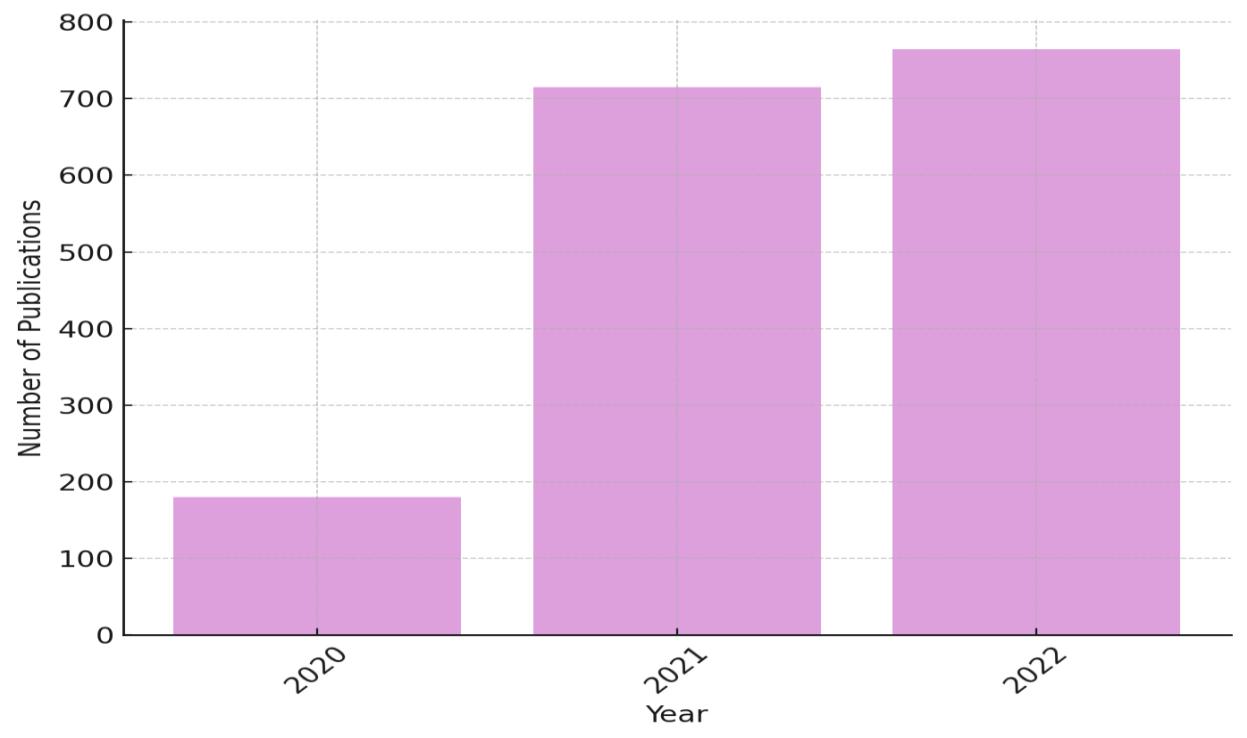


Figure 1. Publication graph by years (Source: VOS Viewer)

This figure illustrates the distribution of relevant publications across different years within the Web of Science Core Collection platform. Accordingly, there were 180, 715, and 764 publications in 2020, 2021, and 2022 respectively, totaling 1659 original articles written in English. The number of publications has increased since 2020.

The Global Perspective on Publications

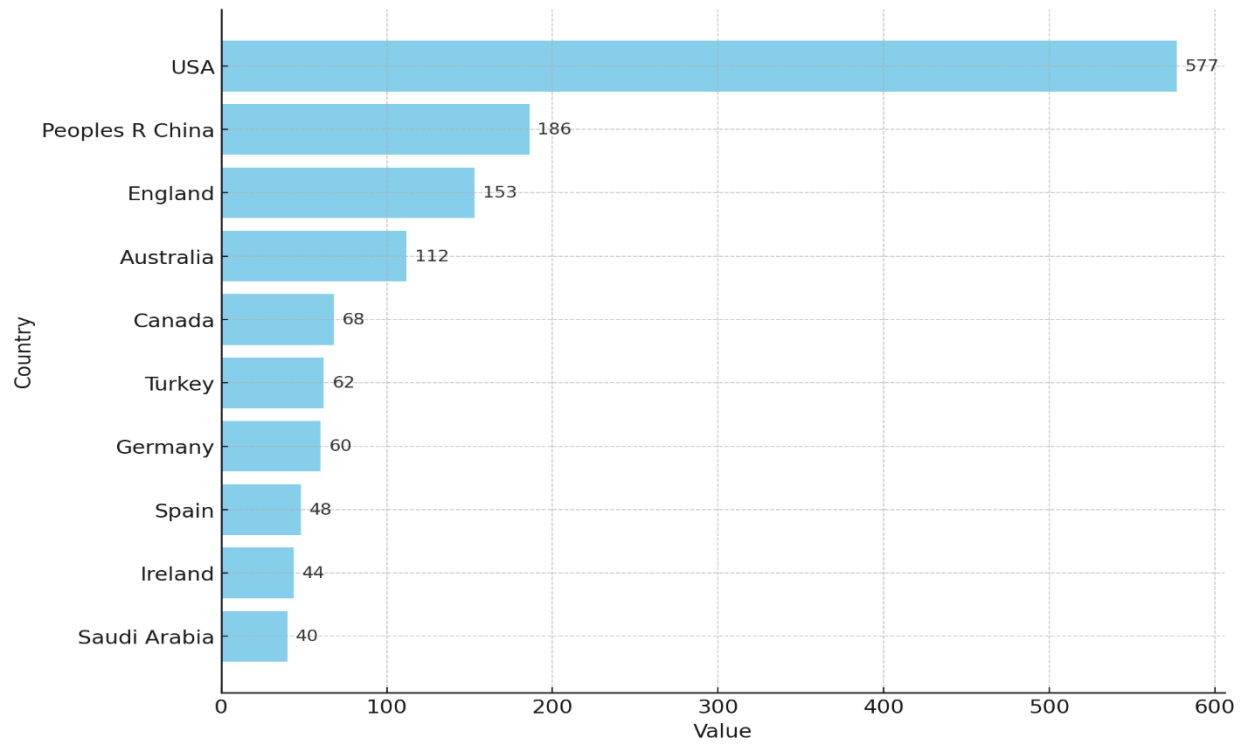


Figure 2. Universal publication graph (Source: VOS Viewer)

The analysis in the graph focuses on the top 10 countries in terms of studies conducted on education during the pandemic. According to this analysis, the United States ranks first with 577 publications, followed by China with 186 publications. The United Kingdom ranks third with 153 publications, while Australia has 112 publications. Canada follows with 68 publications, and Turkey with 62 publications. During the specified period, the United States recorded the highest number of COVID-19 cases at 45,406,263, followed by India with 34,175,468 cases and Brazil with 21,723,559 cases (News Google, 2021). These statistics highlight significant disparities between the number of publications on COVID-19 in education and the reported cases of the pandemic across different countries.

The Bibliometric View of the Data

The citation view of publications by author, journal, and country

Table 1. Table of publications with the most citations

Number	Title	Author(s)-Year	Journal	Total Citation
1.	Adapting to online teaching during COVID-19 school closure: teacher education and teacher competence effects among early career teachers in Germany	(Konig, J; Jager-Biela, DJ and Glutsch, N,2020)	European Journal Of Teacher Education	338
2.	Projecting the Potential Impact of COVID-19 School Closures on Academic Achievement	(Kuhfeld, M; Soland, J; (...); Liu, J,2020)	Educational Researcher	303
3.	COVID-19 and digital disruption in UK universities: afflictions and affordances of emergency online migration	(Watermeyer, R; Crick, T; (...); Goodall, J, 2021)	Higher Education	303
4.	Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic	Almaiah, MA; Al-Khasawneh, A and Althunibat, A,2020)	Education And Information Technologies	302
5.	The sudden transition to synchronized online learning during the COVID-19 pandemic in Saudi Arabia: a qualitative study exploring medical students' perspectives	(Khalil, R; Mansour, AE; (...); Al-Wutayd, O, ,2020)	Bmc Medical Education	245
6.	Distance learning in clinical medical education amid COVID-19 pandemic in Jordan: current situation, challenges, and perspectives	(. Al-Balas, M; Al-Balas, HI; (...); Al-Balas, B,2020)	Education And Information Technologies	219
7.	The impact of the COVID-19 pandemic on final year medical students in the United Kingdom: a national survey	Choi, B; Jegatheeswaran, L; (...); Mutengesa, E,2020)	Bmc Medical Education	192
8.	COVID-19 and schooling: evaluation, assessment, and accountability in times of crises-reacting quickly to explore key issues for policy, practice, and research with the school barometer	(Huber, SG and Helm, C, 2020)	Educational Assessment Evaluation And Accountability	145
9.	Adaptations to a face-to-face initial teacher education course 'forced' online due to the COVID-19 pandemic	(Moorhouse, BL.,2020)	Journal Of Education For Teaching	144
10.	Factors Contributing to Teacher Burnout During COVID-19	(Pressley, T,2021)	Educational Researcher	142

The table includes two articles with the same number of citations.

The bibliographic analysis commenced with the citation evaluation of publications obtained through the Web of Science Core Collection (WoSCC) platform. Table 1 presents the articles with the most citations in priority order. Out of 1659 publications, 71 of them received at least 40 citations or more. At the top of the list is the study titled "Adapting Online Teaching During School Closure Due to COVID-19: Early Career Teachers' Experiences in Germany" (König et al., 2020; 338 citations). In the second place, there are two articles: "Projecting the Potential Impact of COVID-19 School Closures on Academic Achievement" (Kuhfeld et al., 2020; 303 citations) and (Watermeyer et al., 2021; 303 citations).

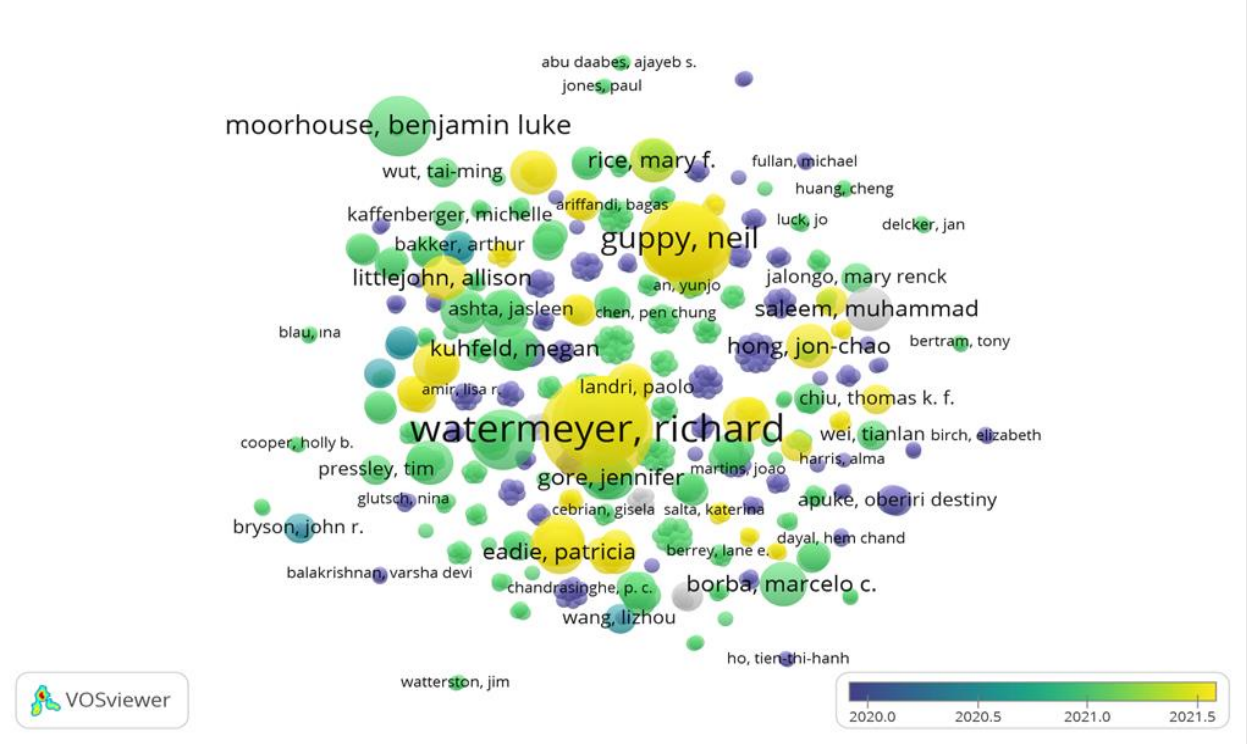


Figure 3. Network graph of researchers with the most citations

The related visual depicts a network graph of authors with the most citations. Out of the set of 5959 authors, 640 with 20 or more citations have been examined. Each distinct cluster is represented by its color. Larger and more prominent clusters belong to authors with higher citation counts. Clusters in yellow and green represent authors with a significant number of citations. Additionally, the legend in the bottom right corner indicates the colors according to the years and periods of publication by the authors. The author with the most citations is identified in the central circle, including names such as Watermayer, Richard, Guppy, Neil, Eadie, Patricia, and Thomas K.F. It is observed that these authors were predominantly published in 2021.

Table 2. Most prolific prestigious journals

Name	Number of publications
Taylor and Francis	642
Springer Nature	504
Sage	143
Wiley	113
Elsevier	96
Emerald Group	34
Oxford University Press	20
Athabasca Univ Press	11
Amer Physical Soc	8
Guilford Publications	7

Source: VOS Viewer

This table presents prominent journals with the highest volume of publications. Among the 10 specialized journals listed, 1,659 articles were published across various domains. Notably, Taylor and Francis lead with 642 publications, establishing itself as the most prolific journal in the field. Following closely, Springer Nature secures the second position with 504 publications, while Sage ranks third with 143 publications according to Table 2.

Analysis of Collaborative Work by Authors and Countries

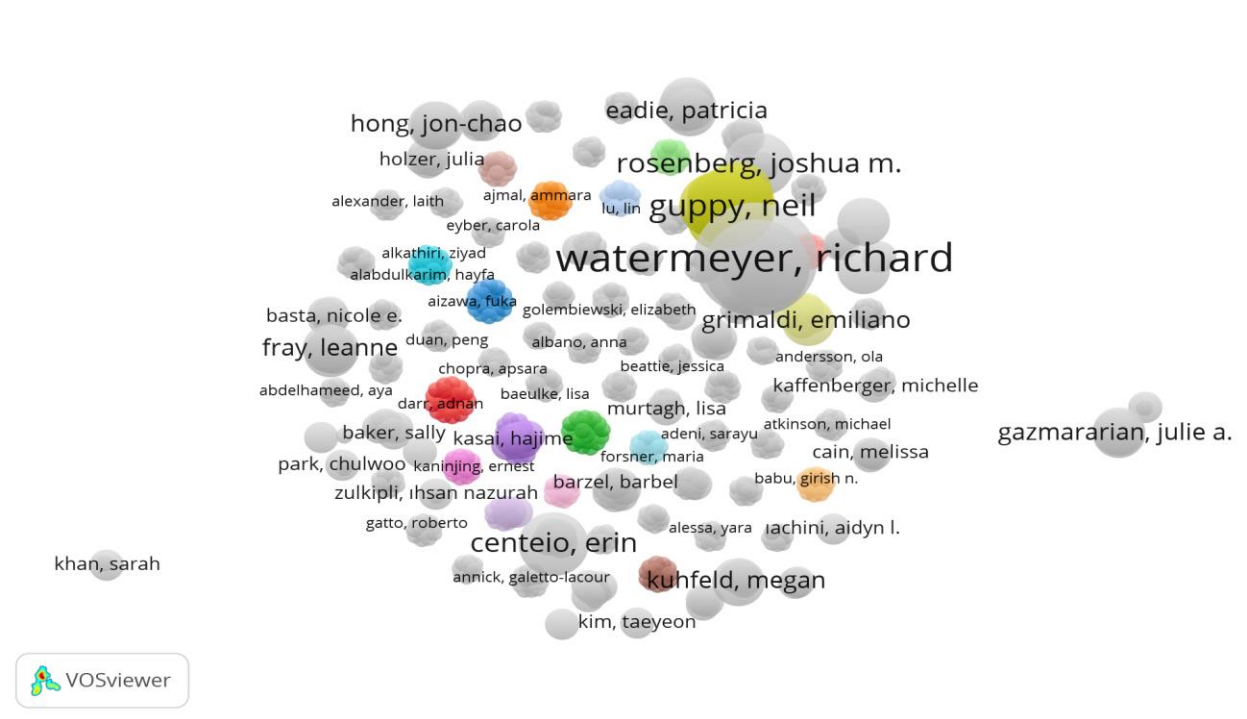


Figure 4. Collaborative authorship table graph

Figure 4 displays the graph obtained from the VOSviewer application, showing collaborative authors and citation networks among them. Out of 5959 authors, 640 authors with at least 20 citations were selected. Networks with the highest collaboration and denser publications

are represented by larger and denser circle networks of the same color. In Figure 4, larger circles represent more publications, while smaller circles represent fewer publications. In this context, it can be stated that authors in the particularly gray-colored cluster have more publication and authorship collaborations.

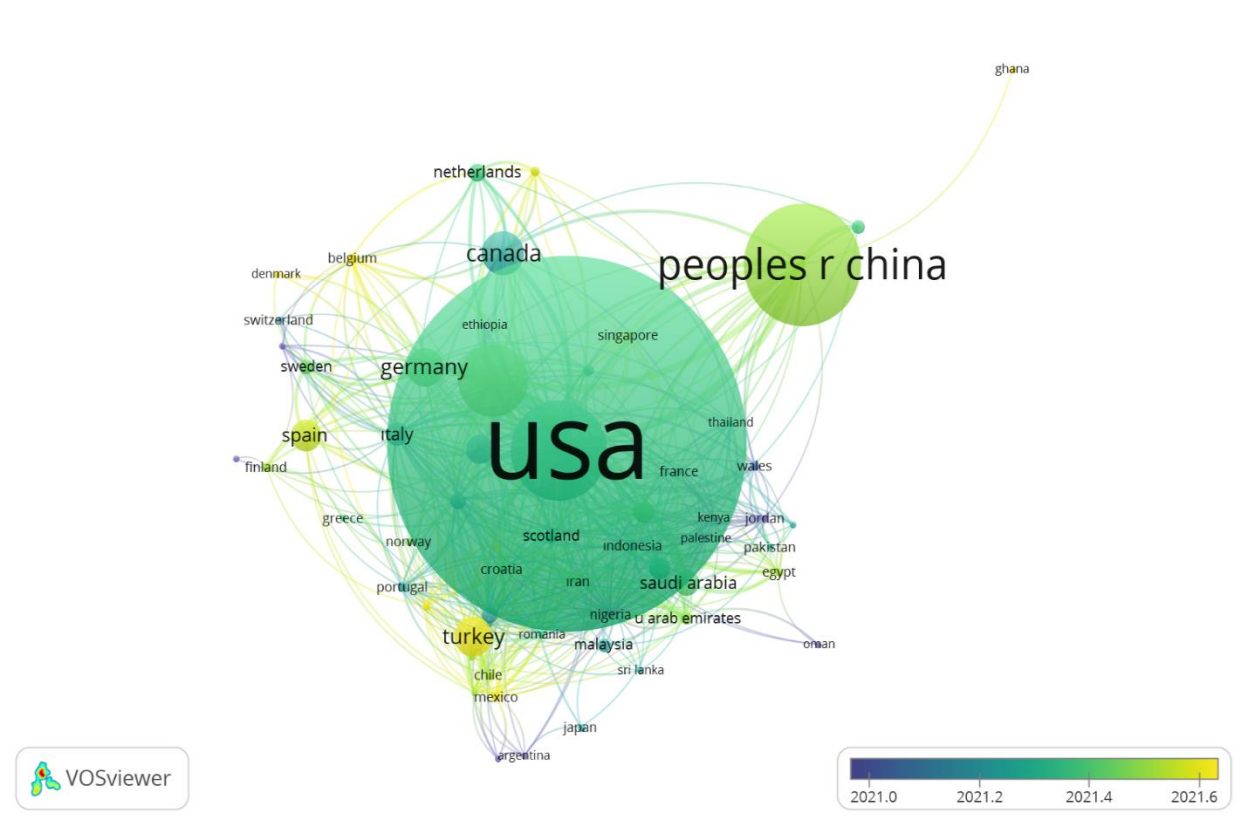


Figure 5. Citation status and relationships among countries graph

Figure 5 depicts a network graph illustrating the citation performance of each country based on the number of papers published. Notably, 62 out of 110 countries have contributed five or more papers. The United States emerges as the leading country in terms of publication citations, followed by China, Germany, Turkey, and Canada. The graph showcases clusters of various colors, indicating connections between countries in terms of publications. This highlights a robust publishing network particularly evident among countries such as the United States, Canada, the Netherlands, Italy, and Germany within the same cluster. Essentially, Figure 5 serves as a network diagram illustrating the collaborative and co-authorship analysis between countries concerning educational research during the pandemic. These clusters of different colors signify the proximity between nations, reflecting enhanced cooperation and communication within the field. Moreover, countries with the highest citation rates are represented by larger clusters compared to others. In this context, countries such as the United States, China, Germany, Turkey, and Canada are among those with higher citation rates. While closer and deeper collaborations are observed between China, Thailand, Singapore, and Wales, similar collaborations exist between the United States, Canada, the Netherlands, and New Zealand. Similarly, closer and tighter collaborations are observed among countries such as Turkey, Greece, Finland, Norway, Brazil, and Portugal. It is noticeable that publications from these countries were particularly concentrated in the year 2021.

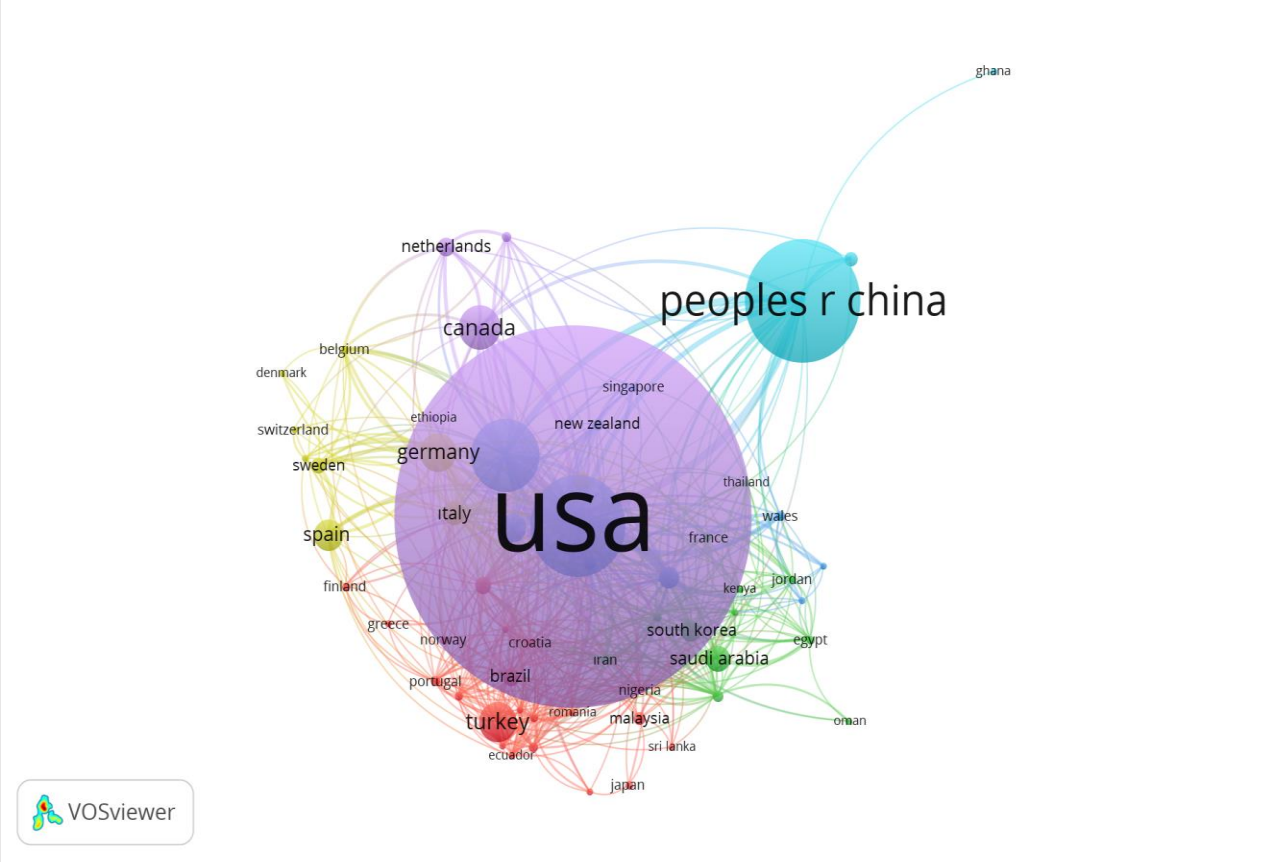


Figure 6. International collaboration network graph

The graph in Figure 6 illustrates the density of international collaboration networks obtained from the VOSviewer application. A total of 60 countries with at least 5 studies and at least 20 citations are represented in the graph. Networks between the countries with the highest collaboration are depicted as larger and denser. Hence, stronger and deeper collaboration is observed among China, Thailand, Singapore, and Wales, while a similar level of collaboration exists among the United States, Canada, the Netherlands, and New Zealand. Similarly, closer and tighter collaboration is observed among countries like Turkey, Greece, Finland, Norway, Brazil, and Portugal.

Frequently Used Keyword Analysis

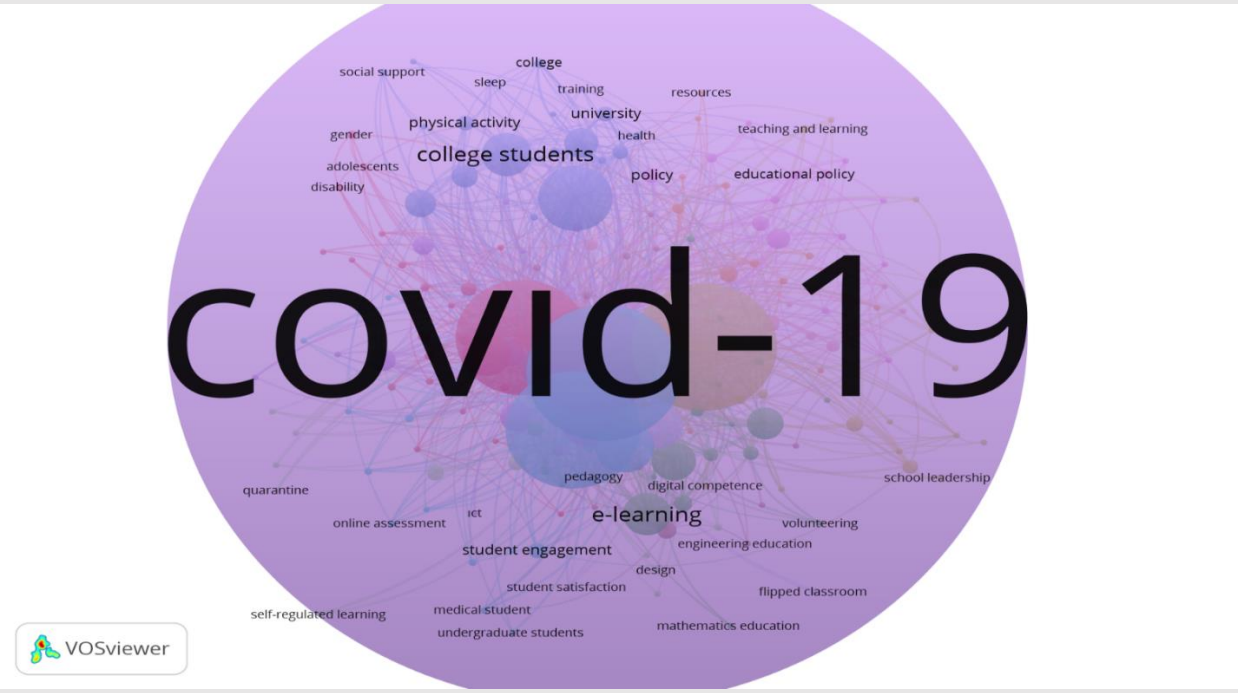


Figure 7. Network graph of keywords

Figure 7 illustrates the density of the keyword network based on co-occurrences, along with word clouds generated using the VOSviewer application. This diagram provides insights into the most frequently used keywords and their interrelationships. The size of each circle denotes the prominence of the respective keyword, with the purple region indicating the current topic. The proximity of words within the graph signifies their co-occurrence within studies, while lines connecting words indicate their association within the same study. Keywords appearing at least five times were included in the analysis. Notably, the central position of the term " COVID-19" underscores its prevalence as the most frequently used keyword. Other notable keywords include "online learning," "quarantine," "pandemic," and "digital literacy," among others, which are analyzed alongside other clusters. Current topics encompass areas such as "Online Learning," "Adapting to the Process," "Impact on Teachers and Students," "Impact of Closure on Academic Achievement," "Digital Disruption," "Online Transition," and "Opportunities," among others. Noteworthy topics include "Challenges in distance education" and "The impact of COVID-19 on teacher burnout." Additionally, a word cloud was generated to visualize the frequency of synonym usage, with the central placement of " COVID-19" indicating its high frequency of occurrence.

Discussion, Conclusion, and Suggestions

Due to the sudden emergence and swift worldwide dissemination of the COVID-19 pandemic, researchers across various disciplines, including education, have undertaken numerous studies to examine its impact on the field of education (Gewerkschaft Erziehung and Wissenschaft, 2020). As the pandemic escalated starting in the year 2020, the number of publications has steadily increased. It is evident that more research was conducted in this field in the years 2021 and 2022 (Huber and Helm, 2020). Researchers have shown particular interest in topics such as the COVID-19 pandemic, online learning, quarantine, and digital literacy (Balaman and Tiryaki, 2021). As the pandemic prolonged and its impact intensified, scholars in the field of education embarked on research endeavors to analyze the effects. In this context, it is understood that research in the field of education has continued to increase between 2020 and 2022 (Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Erdoğan (2021) emphasizes that the most common publications of countries, authors, or journals provide valuable information about their productivity. The United States leads in terms of publications regarding the pandemic's impact on education, followed by China (UNESCO, 2021; Huber & Helm, 2020; Lake & Dusseault, 2020; National Science Board, 2022). The heightened volume of publications from these nations could stem from their heightened vulnerability to the pandemic's effects. The extensive morbidity and mortality rates witnessed in the United States during the pandemic are also reflected in scientific research (World Health Organization [WHO], 2020; National Science Board, 2022). For instance, by October 2021, the United States had reported 45,406,263 confirmed COVID-19 cases and 735,964 deaths, making it the most severely affected country at that time (CDC, 2021; WHO, 2021). According to the World Health Organization's data as of February 16, 2024, there have been 103,436,829 confirmed cases globally, resulting in 1,144,877 deaths (WHO, 2024). This situation may explain the higher volume of research conducted in countries with acute case and mortality rates. Moreover, the substantial budget allocation for research and development (R&D) in countries like the United States and China, coupled with their capacity for innovation and technological advancement, influences the volume of research produced in these countries (National Science Board, 2022; Gewerkschaft Erziehung und Wissenschaft [GEW], 2020)

Citation analysis offers valuable insights into the most renowned and frequently referenced research findings, authors, and publishing institutions. Hence, our study prioritizes the examination of the most frequently cited publications (Garfield, 1972; Van Raan, 2005; Waltman & Van Eck, 2012). Accordingly, the most cited publication on the subject of pandemics and education is the study conducted by Konig et al. (2020). This research delved into teachers' perspectives on distance learning activities during the period of class suspension due to the pandemic (Reimers & Schleicher, 2020; Bao, 2020; Hodges et al., 2020). Another notable study by Kuhfeld et al. (2020) explored the impact of school closures on academic performance. Additionally, a study by Watermayer et al. on managing digital disruption, challenges, and opportunities during the pandemic emerged as a highly cited topic. Based on the findings of citation analysis, several authors have garnered significant attention for their contributions to education-related research on the pandemic, including Watermayer, Richard, Guppy, Neil, Eadie, Patricia, Zhongchao, and Thomas K.F. These studies generally address the significant changes brought about by the pandemic, which forced schools to close and transition to distance learning activities, presenting various challenges alongside opportunities (Al Lily et al., 2020). For example, they discuss the significant functional impairment and discomfort experienced by academics due to the transition to online platforms (Tondeur et al., 2023; Kop, 2011).

They also emphasize that the transition poses significant challenges in terms of student enrollment, market sustainability, the academic workforce market, and local economies (Watermayer et al., 2021). In terms of both citations and publications, Taylor & Francis leads among the journals where the papers are published, with Springer Nature following closely behind. These journals are prestigious publications indexed in the SSCI, known for their high citation and publication rates. It is believed that the analysis of these journals will contribute to raising awareness among researchers about the implications of the pandemic on education.

Consequently, the journal analysis is expected to guide our study in new research directions. Given the global scale of the COVID-19 pandemic and its widespread impact, collaboration between countries is vital (World Health Organization, 2020; Rosenbaum, 2020; Kupferschmidt & Cohen, 2020). In this context, collaborative research conducted and to be conducted is crucial in overcoming the challenges. Joint authorship analysis shows that authors from China, the United States, the Netherlands, Turkey, Brazil, Canada, and New Zealand have engaged in the most intensive collaborations with other countries. Global collaboration in educational research related to the pandemic is valuable. The topics that educational researchers focused on during this period can be understood through keyword analysis (Otte & Rousseau, 2002; Martín-Martín et al., 2018). According to the keyword analysis, education researchers initially concentrated on the core impacts of the COVID-19 pandemic (e.g., COVID-19, online learning, quarantine, pandemic, digital literacy, etc.) (Rasmitadila et al., 2020; Viner et al., 2020). Subsequently, attention shifted towards examining its psychosocial effects during the pandemic (such as adapting to online learning, the impact of closures, digital disruption, opportunities, teacher burnout, etc.) (Fegert et al., 2020; Kwok et al., 2020). These results are consistent with findings in the existing literature. The advancements in remote learning applications and digital literacy, among other topics, indicate that researchers likely focused on other aspects of the pandemic as well (Yılmaz and Toker, 2022; Doğan and Birişçi, 2022). The results we obtained will shed light on different themes and current issues for researchers, guiding for new and much-needed bibliometric education research. Although this analysis evaluates the research conducted on education during the pandemic through bibliometric methods, there may be some limitations. This analysis represents only a specific aspect of academic discourse. The study solely relied on the Web of Science database and did not incorporate other databases like Scopus, which could be a consideration for future research endeavors. Moreover, the sample exclusively encompasses original articles, excluding other types such as conference papers and editorials from the analysis. Despite these limitations, this study offers a comprehensive evaluation of global educational research about COVID-19, presenting a range of insights and features. For instance, it is believed that identifying the topics focused on in published research, will guide education researchers and encourage them to explore yet unexplored topics during and after the pandemic. Moreover, it assists education researchers in gaining awareness about countries open to collaboration and facilitates the establishment of new collaborations.

The results enable education researchers to assess their journal and publication performance and select research topics. Additionally, this study can pave the way for more comprehensive studies involving different databases such as PubMed and Scopus, thereby laying the groundwork for advanced research. The findings can help education administrators identify and clarify issues related to COVID-19 to be investigated in their institutions (Owusu-Fordjour et al., 2020). Furthermore, they can guide in selecting appropriate journals, identifying suitable countries for collaboration, and obtaining information on studies related to the subject. This study stands out by offering valuable insight into countries' research and publication performance in

education and COVID-19. It can motivate researchers to conduct more research on pandemic education globally, thereby positively contributing to publication performance, both worldwide and on a country-specific basis (Kostoff, 2007; Gargouri et al., 2010).

The comprehensive bibliometric analysis conducted on education-related research during the COVID-19 pandemic has unveiled critical insights into the evolving landscape of scholarly discourse (González-Padilla & Tortolero-Blanco, 2020; UNESCO, 2020). As education systems globally continue to navigate the challenges posed by the pandemic, policymakers are urged to implement evidence-based strategies to address emerging issues and facilitate educational resilience. To this end, it is recommended that policymakers prioritize the following initiatives: Enhancing Digital Infrastructure: Policymakers should invest in robust digital infrastructure and technological resources to support online learning platforms and ensure equitable access to education during periods of disruption (Organisation for Economic Co-operation and Development, 2020; Selwyn et al., 2020). Supporting Educator Well-being: Measures should be implemented to prioritize the well-being of educators, including providing adequate support for professional development, mental health resources, and workload management strategies (Hargreaves & Fullan, 2015; Maslach & Leiter, 2016). Promoting Research Collaboration: Encouraging international collaboration among researchers and educational institutions can foster knowledge exchange, innovation, and the development of best practices in pandemic education response (Galea et al., 2020; Viner et al., 2020). Addressing Equity Concerns: Policymakers should address equity concerns by implementing policies that mitigate disparities in access to education technology and resources among vulnerable populations (UNESCO, 2020; García & Weiss, 2020). Strengthening Health and Safety Protocols: Collaborative efforts should be made to strengthen health and safety protocols in educational settings, ensuring the well-being of students, educators, and staff. These recommendations aim to inform policymaking efforts and guide future research initiatives aimed at addressing the multifaceted impacts of the COVID-19 pandemic on education (Fitzpatrick et al., 2020; Lau et al., 2010). Moreover, to enrich the scholarly discourse on this topic, it is essential to expand the reference base by incorporating a wider range of relevant studies and publications. By leveraging additional references, researchers can provide a more comprehensive analysis and strengthen the theoretical underpinnings of their findings.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

Conflicts of Interest

There is no conflict of interest between authors.

Ethics

Since this study involves bibliometric analysis, ethical committee approval was not required. Bibliometric studies are conducted through analyses of existing literature and publications, and do not involve direct research on human or animal subjects. Therefore, ethical committee approval is not necessary for such studies (Aria & Cuccurullo, 2017).

References

Al Lily, A. E., Ismail, A. F., Abunasser, F. M., & Alqahtani, R. H. A. (2020). Distance education as a response to pandemics: Coronavirus and Arab culture. *Technology in Society, 63*, 101317. https://doi.org/10.1016/j.techsoc.2020.101317

Al-Balas, M., Al-Balas, H. I., Jaber, H. M., Obeidat, K., Al-Balas, H., Aborajooh, E. A., Al-Taher, R., & Al-Balas, B. (2020). Distance learning in clinical medical education amid COVID-19 pandemic in Jordan: Current situation, challenges, and perspectives. *BMC Medical Education, 20*, 341. https://doi.org/10.1186/s12909-020-02257-4

Almaiah, M. A., Al-Khasawneh, A., & Althunibat, A. (2020). Exploring the critical challenges and factors influencing the E-learning system usage during COVID-19 pandemic. *Education and Information Technologies, 25*(6), 5261–5280. https://doi.org/10.1007/s10639-020-10219-y

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11*(4), 959–975. https://doi.org/10.1016/j.joi.2017.08.007

Asandaş, N., & Hacıcafereoğlu, S. (2021). Koronavirüs (Covid-19) döneminde uzaktan eğitim süreci. *Mustafa Kemal Üniversitesi Eğitim Fakültesi Dergisi, 5*(7), 213–223.

Bağış, M. (2021). Main analysis techniques used in bibliometric research. In *Bibliometric analysis as a literature review tool* (pp. 97-123).

Balaman, F., & Tiryaki, S. H. (2021). Corona virüs (Covid-19) nedeniyle mecburi yürütülen uzaktan eğitim hakkında öğretmen görüşleri. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi, 10*(1), 52–84.

Bao, W. (2020). COVID-19 and online teaching in higher education: A case study of Peking University. *Human Behavior and Emerging Technologies, 2*(2), 113-115. https://doi.org/10.1002/hbe2.191

Briner, R. B., & Denyer, D. (2012). Systematic review and evidence synthesis as a practice and scholarship tool. In D. Rousseau (Ed.), *Handbook of evidence-based management: Companies, classrooms and research* (pp. 112–129). Oxford University Press.

Broadus, R. (1987). Toward a definition of bibliometrics. *Scientometrics, 12*(5–6), 373–379. https://doi.org/10.1007/BF02017100

Centers for Disease Control and Prevention (CDC). (2021). *COVID Data Tracker*. U.S. Department of Health and Human Services. Retrieved from https://covid.cdc.gov/covid-data-tracker

Choi, B., Jegatheeswaran, L., Minocha, A., Alhilani, M., Nakhoul, M., & Mutengesa, E. (2020). The impact of the COVID-19 pandemic on final year medical students in the United Kingdom: A national survey. *BMC Medical Education, 20*(1), 206. https://doi.org/10.1186/s12909-020-02117-1

Çiçek Korkmaz, A., & Altuntaş, S. (2022). A bibliometric analysis of Covid-19 publications in nursing by visual mapping method. *Journal of Nursing Management, 30*(6), 1892–1902. https://doi.org/10.1111/jonm.1363

Çiftçi, Ş. K., Danişman, Ş., Yalçın, M., Tosuntaş, Ş. B., Ay, Y., Sölpük, N., & Karadağ, E. (2016). Map of scientific publication in the field of educational sciences and teacher education in Turkey: A bibliometric study. *Educational Sciences: Theory & Practice, 16*, 1097–1123. https://doi.org/10.12738/estp.2016.4.0007

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. University of Chicago Press.

Diodato, V. (1994). *Dictionary of bibliometrics*. Haworth Press.

Doğan, C., & Birişçi, S. (2022). Covid-19 süreciyle birlikte öğretmenlerin dijital okuryazarlık düzeylerinin incelenmesi. *Ege Eğitim Teknolojileri Dergisi, 6*(1), 53–76. https://doi.org/10.1016/j.joi.2017.08.007

Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). *How to conduct a bibliometric analysis: An overview and guidelines*. Journal of Business Research, 133, 285-296. https://doi.org/10.1016/j.jbusres.2021.04.070

Erdoğan, R. E. (2021). Analysis and mapping of computational design research in architecture with bibliometric methods. [Unpublished master's thesis]. Başkent University.

Fegert, J. M., Vitiello, B., Plener, P. L., & Clemens, V. (2020). Challenges and burden of the Coronavirus 2019 (COVID-19) pandemic for child and adolescent mental health: A narrative review to highlight clinical and research needs in the acute phase and the long return to normality. *Child and Adolescent Psychiatry and Mental Health, 14*, 1–11. https://doi.org/10.1186/s13034-020-00329-3

Fitzpatrick, K. M., Harris, C., & Drawve, G. (2020). How bad is it? Suicidality in the middle of the COVID-19 pandemic. *Suicide and Life-Threatening Behavior, 50*(6), 1241–1249. https://doi.org/10.1111/sltb.12655

Galea, S., Merchant, R. M., & Lurie, N. (2020). The mental health consequences of COVID-19 and physical distancing: The need for prevention and early intervention. *JAMA Internal Medicine, 180*(6), 817–818. https://doi.org/10.1001/jamainternmed.2020.1562

García, E., & Weiss, E. (2020). COVID-19 and student performance, equity, and U.S. education policy: Lessons from pre-pandemic research to inform relief, recovery, and rebuilding. *Economic Policy Institute*. https://www.epi.org/publication/the-consequences-of-the-covid-19-pandemic-for-education-performance-and-equity-in-the-us-what-the-pre-pandemic-research-says-we-can-expect-impacts-will-be-worse-for-students-of-color-and-students-from-low-income

Garfield, E. (1972). Citation analysis as a tool in journal evaluation. *Science, 178*(4060), 471–479. https://doi.org/10.1126/science.178.4060.471

Gargouri, Y., Hajjem, C., Larivière, V., Gingras, Y., Carr, L., Brody, T., & Harnad, S. (2010). Self-selected or mandated, open access increases citation impact for higher quality research. *PLOS ONE, 5*(10), e13636. https://doi.org/10.1371/journal.pone.0013636

Gewerkschaft Erziehung und Wissenschaft (GEW). (2020). *Digitalpakt Schule und Digitalisierung an Schulen* [Dijital Okul Paketi ve Okullarda Dijitalleşme]. GEW. https://www.gew.de

González-Padilla, D. A., & Tortolero-Blanco, L. (2020). Social media influence in the COVID-19 pandemic. *International Braz J Urol, 46*, 120–124. https://doi.org/10.1590/S1677-5538.IBJU.2020.S116

Güçlü, N. (2014). Türkiye'de lisansüstü turizm tezlerinin bibliyometrik profili (1990–2013). In *VII. Lisansüstü Turizm Öğrencileri Araştırma Kongresi*, Kuşadası, Aydın, April 4–5, 2014.

Hao, Y.-F., Peng, K., Mai, Q.-L., Meng, M.-Q., Wang, D., & Zhang, X.-Y. (2020). A bibliometric analysis of nursing research in COVID-19 in China. *Journal of Integrative Nursing, 2*(3), 116. https://doi.org/10.4103/jin.jin_32_20

Hargreaves, A., & Fullan, M. (2015). *Professional capital: Transforming teaching in every school*. Teachers College Press.

Harris, E. A. (2020, April 27). "It was just too much": How remote learning is breaking parents. *The New York Times*. https://www.nytimes.com/2020/04/27/nyregion/coronavirus-homeschooling-parents.html

Hodges, C., Moore, S., Lockee, B., Trust, T., & Bond, A. (2020). The difference between emergency remote teaching and online learning. *Educause Review*. https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Huang, M.-H., & Chang, Y.-W. (2011). A study of interdisciplinarity in information science: Using direct citation and co-authorship analysis. *Journal of Information Science, 37*(4), 369–378. https://doi.org/10.1177/0165551511407141

Huber, S. G., & Helm, C. (2020). COVID-19 and schooling: Evaluation, assessment and accountability in times of crises—reacting quickly to explore key issues for policy, practice and research with the school barometer. *Educational Assessment, Evaluation and Accountability, 32*, 237–270. https://doi.org/10.1007/s11092-020-09322-y

Khalil, R., Mansour, A. E., Fadda, W. A., Almisnid, K., Aldamegh, M., Al-Nafeesah, A., Alkhalifah, A., & Al-Wutayd, O. (2020). The sudden transition to synchronized online learning during the COVID-19 pandemic in Saudi Arabia: A qualitative study exploring medical students' perspectives. *BMC Medical Education, 20*(1), 1-10. https://doi.org/10.1186/s12909-020-02208-z

König, J., Jäger-Biela, D. J., & Glutsch, N. (2020). Adapting to online teaching during COVID-19 school closure: Teacher education and teacher competence effects among early career teachers in Germany. *European Journal of Teacher Education, 43*(4), 608-622. https://doi.org/10.1080/02619768.2020.1809650

Kop, R. (2011). The challenges to connectivist learning on open online networks: Learning experiences during a massive open online course. *International Review of Research in Open and Distributed Learning, 12*(3), 19–38. https://doi.org/10.19173/irrodl.v12i3.882

Kostoff, R. (2007). The difference between highly and poorly cited medical articles in the journal Lancet. *Scientometrics, 72*(3), 513–520. https://doi.org/10.1007/s11192-007-1762-9

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). Projecting the potential impact of COVID-19 school closures on academic achievement. *Educational Researcher, 49*(8), 549–565. https://doi.org/10.3102/0013189X20965918

Kupferschmidt, K., & Cohen, J. (2020). Can China's COVID-19 strategy work elsewhere? *Science, 367*(6482), 1061–1062. https://doi.org/10.1126/science.367.6482.1061

Kwok, K. O., Li, K. K., Chan, H. H. H., Yi, Y. Y., Tang, A., Wei, W. I., & Wong, S. Y. S. (2020). Community responses during the early phase of COVID-19 epidemic, Hong Kong. *Emerging Infectious Diseases, 26*(7), 1575. https://doi.org/10.3201/eid2607.200185

Lake, R., & Dusseault, B. (2020). Remote classes are in session for more school districts, but attendance plans are still absent. *Center for Reinventing Public Education*. https://www.crpe.org/thelens/remote-classes-are-session-more-school-districts-attendance-plans-are-still-absent

Lau, J. T., Griffiths, S., Choi, K. C., & Tsui, H. Y. (2010). Avoidance behaviors and negative psychological responses in the general population in the initial stage of the H1N1 pandemic in Hong Kong. *BMC Infectious Diseases, 10*, 1–13. https://doi.org/10.1186/1471-2334-10-139

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics, 12*(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.09.002

Maslach, C., & Leiter, M. P. (2016). Understanding the burnout experience: Recent research and its implications for psychiatry. *World Psychiatry, 15*(2), 103–111. https://doi.org/10.1002/wps.20311

Moorhouse, B. L. (2020). Adaptations to a face-to-face initial teacher education course 'forced' online due to the COVID-19 pandemic. *Journal of Education for Teaching, 46*(4), 609-611. https://doi.org/10.1080/02607476.2020.1755205

National Science Board. (2022). *Science and engineering indicators 2022. U.S. and global research and development*. National Science Foundation. https://ncses.nsf.gov/pubs/nsb20221/u-s-and-global-research-and-development

News Google. (2021). "Coronavirüs-19 (COVID-19)" (24.20.2021). *Google News*. https://news.google.com/covid19/map?hl=tr&gl=TR&ceid=TR:tr

Organisation for Economic Co-operation and Development. (2020). *Education responses to COVID-19: Embracing digital learning and online collaboration*. OECD Publishing. https://www.oecd.org/education/embracing-digital-learning-and-online-collaboration.htm

Otte, E., & Rousseau, R. (2002). Social network analysis: A powerful strategy, also for the information sciences. *Journal of Information Science, 28*(6), 441–453. https://doi.org/10.1177/016555150202800601

Owusu-Fordjour, C., Koomson, C. K., & Hanson, D. (2020). The impact of Covid-19 on learning—the perspective of the Ghanaian student. *European Journal of Education Studies*. https://doi.org/10.5281/zenodo.3753586

Pressley, T. (2021). Factors contributing to teacher burnout during COVID-19. *Educational Researcher, 50*(5), 325-327. https://doi.org/10.3102/0013189X211004138

Pritchard, A. (1969). Statistical bibliography or bibliometrics. *Journal of Documentation, 25*, 348.

Rasmitadila, R., Aliyyah, R. R., Rachmadtullah, R., Samsudin, A., Syaodih, E., Nurtanto, M., & Tambunan, A. R. S. (2020). The perceptions of primary school teachers of online learning during the COVID-19 pandemic period. *Journal of Ethnic and Cultural Studies, 7*(2), 90–109. https://doi.org/10.29333/ejecs/388

Reimers, F. M., & Schleicher, A. (2020). A framework to guide an education response to the COVID-19 pandemic of 2020. OECD. https://www.oecd.org/education/framework-guide-education-response-covid-19-pandemic

Rosenbaum, L. (2020). Facing Covid-19 in Italy—ethics, logistics, and therapeutics on the epidemic's front line. *New England Journal of Medicine, 382*(20), 1873–1875. https://doi.org/10.1056/NEJMp2005492

Rousseau, D. M. (Ed.). (2012). *The Oxford handbook of evidence-based management*. Oxford University Press.

Selwyn, N. (2012). *Education in a digital world: Global perspectives on technology and education*. Routledge.

Selwyn, N., Hillman, T., Eynon, R., Ferreira, G., Knox, J., Macgilchrist, F., & Sancho-Gil, J. M. (2020). What's next for Ed-Tech? Critical hopes and concerns for the 2020s. *Learning, Media and Technology, 45*(1), 1–6. https://doi.org/10.1080/17439884.2020.1694945

Talan, T. (2021). Bibliometric analysis of the research on seamless learning. *International Journal of Technology in Education, 4*(3), 428–442. https://doi.org/10.46328/ijte.113

Tondeur, J., Howard, S. K., Scherer, R., & Siddiq, F. (2023). Untangling the great online transition: A network model of teachers' experiences with online practices. *Computers & Education, 203*, 104866. https://doi.org/10.1016/j.compedu.2023.104866

UNESCO. (2021). *Education: From disruption to recovery*. Global Education Monitoring Report. https://unesdoc.unesco.org/ark:/48223/pf0000376153

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2020). *Global education monitoring report 2020: Inclusion and education: All means all*. https://unesdoc.unesco.org/ark:/48223/pf0000373718

Van Raan, A. F. (2005). Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods. *Scientometrics, 62*, 133–143. https://doi.org/10.1007/s11192-005-0008-6

Viner, R. M., Russell, S. J., Croker, H., Packer, J., Ward, J., Stansfield, C., & Booy, R. (2020). School closure and management practices during coronavirus outbreaks including COVID-19: A rapid systematic review. *The Lancet Child & Adolescent Health, 4*(5), 397–404. https://doi.org/10.1016/S2352-4642(20)30095-X

Waltman, L., & Van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology, 63*(12), 2378–2392. https://doi.org/10.1002/asi.22748

Watermeyer, R., Crick, T., Knight, C., & Goodall, J. (2021). The international impact of COVID-19 and emergency remote teaching on computer science education practitioners. In *2021 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1048-1055). Institute

of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/EDUCON46332.2021.9453971

Watson, J., & Murin, A. (2014). *Keeping pace with K-12 digital learning: An annual review of policy and practice*. Evergreen Education Group.

World Health Organization (WHO). (2021). *WHO Coronavirus (COVID-19) Dashboard*. Retrieved from https://data.who.int/dashboards/covid19/data

World Health Organization (WHO). (2024, February 16). *COVID-19 epidemiological update – 16 February 2024*. World Health Organization. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

World Health Organization. (2020). Rolling updates on coronavirus disease (COVID-19). *World Health Organization*. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen

Yılmaz, E. O., & Toker, T. (2022). Covid-19 salgını öğretmenlerin dijital yeterliliklerini nasıl etkiledi?. *Millî Eğitim Dergisi, 51*(235), 2713–2730.

International Journal of Educational Studies and Policy (IJESP)

Volume: 5, Issue: 2, November 2024

Item Response Theory Assumptions: A Comprehensive Review of Studies with Document Analysis

Mahmut Sami Yiğiter[1], Erdem Boduroğlu[2]

ABSTRACT

Item Response Theory (IRT), over its nearly 100-year history, has become one of the most popular methodologies for modeling response patterns in measures in education, psychology and health. Due to its advantages, IRT is particularly popular in large-scale assessments. A pre-condition for the validity of the estimations obtained from IRT is that the data meet the model assumptions. The purpose of this study is to examine the testing of model assumptions in studies using IRT models. For this purpose, 107 studies in the National Thesis Center of the Council of Higher Education that use the IRT model on real data were examined. The studies were analyzed according to sample size, unidimensionality, local independence, overall model fit, item fit and non-speedness test criteria. According to the results, it was observed that the unidimensionality assumption was tested at a high level (89%) and Factor Analytic approaches were predominantly used. Local independence assumption was not tested in 36% of the studies, unidimensionality was cited as evidence in 40% of the studies and tested in 24% of the studies. Overall model fit was tested at a moderate level (51%) and Log-Likelihood and information criteria were used. Item fit and Non-Speedness testing were tested at a low level (26% and 9%). IRT assumptions should be considered as a whole and all assumptions should be tested from an evidence-based perspective.

Keywords: Item response theory, assumption, unidimensionality, local independence, overall model fit, Item fit, non-speedness test, factor analysis.

DOI: https://doi.org/10.5281/zenodo.14016086

Article Info:

Received: 03.06.2024 **Accepted: 12.10.2024** **Article Type: Research Article**

Cite as: Yiğiter, M. S. & Boduroğlu, E. (2024). Item response theory assumptions: A comprehensive review of studies with document analysis. *International Journal of Educational Studies and Policy, 5*(2), 119-138.

[1]Corresponding Author: Mahmut Sami Yiğiter, Social Sciences University of Ankara, mahmutsamiyigiter@gmail.com, ORCID: 0000-0002-2896-0201

[2]Erdem Boduroğlu, Ministry of National Education, erdemboduroglu@gmail.com, ORCID: 0000-0001-8318-4914

* This article was presented as an oral presentation at the VII. International Congress on Measurement and Evaluation in Education and Psychology.

Introduction

Many models have been developed throughout the history to place scores obtained from educational and psychological measurements on a scale. Classical Test Theory (CTT) and Item Response Theory (IRT) are the most widely used, known and important among these models. Classical Test Theory models are often referred to as "weak models". The reason for this is that the assumptions of the models can be easily met with the test data. On the other hand, the test data are less likely to meet the assumptions since they are strict in Item Response Theory models and therefore they are called as “strong models” (Hambleton and Jones, 1993). Classical Test Theory is a theory based on observed score (X), true score (T) and error score (E). This theory has a simple linear equation expressed as X=T+E, consisting of the sum of the observed test scores (X), the unobservable and often latent true score (T) and the error score (E) (Novick, 1966). Since there are two unknown variables in the equation (T and E), the equation cannot be solved unless there are some assumptions. These assumptions of the CTT are (a) the true scores and error scores are uncorrelated, (b) the expected value of the error score is equal to zero, (c) the true scores and error scores in parallel tests are uncorrelated (Lord & Novick, 1968). The true score in the basic equation of the theory is the difference between the observed test score and the error score. The true score of the examinee is also defined as the expected score from the parallel forms. CTT models are focused on modelling at the test score level. These models relate the true score to the total score obtained from the test, not to the scores obtained from the items. The biggest advantage of CTT is that its assumptions are easily met and item parameters can be easily calculated (Fan, 1998). However, CTT has some limitations. Lord (1953) states that the true score in the CTT varies according to the difficulty of the test. For example, while an examinee will score low on a difficult test, he/she will score high on an easy test. While the examinee's ability level has not changed, the fact that the examinee's true score takes different values indicates that the examinee's true score in CTT is dependent on the test or the group he is in. Thus, different methods and models have been sought to overcome the limitations of CTT.

Item Response Theory (IRT) is one of the most popular methodologies used to model response patterns from measurements (Boduroğlu & Anil, 2023). IRT studies started with the modelling of latent variables with the work of Thorndike, Thurstone and Symond in the early 20th century until its foundations were laid with the studies of Lazarsfeld and Lord in the 1950s. In the 1960s, IRT developed with Rasch's studies on the "Rasch model" and Birnbaum's studies on the "logistical model" (Baker, 2001; Himelfarb, 2019). It remained as a theory with no practical application until the 1970s, as computational technology did not enable data analysis with IRT models. With the development of computers in computing technology, IRT applications and research have become widespread. Over time, more complex models have been developed on logistic models (De Boeck and Wilson, 2004; Reckase, 1997; Rijmen et al., 2003). In the 21st century, IRT has found a wide and fundamental application area, especially in large scale educational assessment. Today, IRT is used in the social sciences and behavioral sciences as well as education, psychology and medical sciences (Reise & Waller, 2009; Thissen & Steinberg, 2020; Zanon et al., 2016; Mutluer & Çakan, 2023).

IRT is a powerful scaling method used to determine the characteristics of the items and examinees based on the responses of examinees to the items in the test (Embretson & Reise, 2000; Selçuk & Demir, 2024; Sözer & Kahraman, 2021). In IRT, there is a parameter called ability, denoted by theta, which corresponds to the true score of the individual in CTT. In addition, IRT provides useful information about the contribution of the items in the measurement of the latent

construct, its quality and at which points of the ability scale it performs the best measurement. One of the important features of the IRT is that it places both examinees and items on the same scale. An examinee may have a high or low ability level, and an item may have high or low difficulty and be on the same scale. Having a common scale for examinees and items makes it possible to evaluate the amount of information that items provide in terms of latent structure and to match items in accordance with the ability level of the individual taking the test (Van der Linden & Glas, 2010). Another advantage of IRT is that both item parameters and ability parameters can be estimated without being dependent on the group or the test. That is, (a) examinees' ability parameters are independent of the test items they take, (b) item parameters are independent of examinees' ability distributions (Hambleton et al., 1991). Thanks to its advantages, IRT is actively used in large-scale tests, computerized adaptive testing, test equating, differential item functioning, cognitive diagnostic model and scale development applications (Aybek, 2023; Ayva Yörü, 2024; Doğan & Atar, 2024; Kılıç et al., 2023; Saatcioglu & Sen, 2023; Şahin, Yildirim & Boztunc Öztürk, 2023; Yiğiter & Doğan, 2023).

As previously mentioned, IRT models are strong due in part to the fact that their underlying assumptions are challenging to meet. To take advantage of the benefits of the above mentioned IRT, the assumptions of the model need to be tested and met. Estimation made without meeting assumptions will contain systematic error, and the validity of the obtained item and ability parameters will become doubtful (Hambleton & Swaminathan, 1985; Reckase, 2009). On the other hand, IRT needs large samples for the estimation of item and ability parameters. The minimum sample size differs according to the IRT model used for accurate estimation of parameters in IRT applications. As the IRT model used gets more complex, larger samples are required (Sireci, 1991). The assumptions of the Item Response Theory have been discussed in many different sources. Trabin and Weiss (1983) discussed the assumptions of IRT under three headings: (a) unidimensionality, (b) local independence, (c) item characteristic curve graph. Hambleton and Swaminathan (1985) stated that there are four assumptions: (a) dimensionality, (b) local independence, (c) item characteristic curve fit, (d) non-speedness test. Crocker and Algina (1986) express that there are two assumptions: (a) unidimensionality and (b) local independence. According to Embretson and Reise (2000), IRT has two basic assumptions: (a) item characteristic curve have a specified form and (b) local independence. Demars (2010), on the other hand, discussed this under the headings of (a) unidimensionality, (b) local independence and (c) fit. Stone and Zhu (2015) lists five different assumptions: (a) dimensionality, (b) local independence, (c) form of the IRT Model (Overall Model Fit), (d) non-speedness test and (e) Model Fit (Item and Person Fit).

In the following sections of the study, firstly, sample size in IRT is discussed. Then, the assumptions of IRT are described and the methods used to test these assumptions are explained under separate headings in line with the main sources and books in the literature.

Sample Size

It is difficult to accurately determine the sample size required for an accurate estimation of the item and individual parameters to be obtained from a test. In particular, as the IRT model used becomes more complex, both larger sample sizes and longer tests are needed to obtain accurate estimations (Hambleton, 1989). As an outgrowth, IRT models are used in large scale assessment. In addition, scaling and estimation can be made with IRT in tests which consist of fewer items and are applied to groups with a certain sample size (Emretson & Reise, 2000). In the literature, there are many studies on the sample size which is required to obtain accurate and stable parameter

estimations with IRT. Lord (1968) discusses that a sample size of more than 1000 is needed to estimate the item discrimination parameter accurately. It is stated that the Rasch or 1PL model with fixed item discrimination can be used with a sample size of 100 or 200 (DeMars, 2010). In models where item discrimination is estimated, it is seen that a larger sample size is required. Ree and Jensen (1980) state that a sample size of 500 or more is required in order to estimate the item discrimination and difficulty parameters accurately. According to Hulin et al. (1982), a sample size of 500 or more is required for the 2PL model, and a sample size of 1000 or more for the 3PL model. Swaminathan and Gifford (1983) state that a sample size of 1000 gives good results for the 3PL model. Harwell and Janosky (1991) suggested that the sample size should be more than 250 in order to estimate the parameters correctly. Demars (2010) says that the sample size should not be less than 500 for the 2PL and 3PL models.

Unidimensionality

Unidimensionality means that there is only one type of ability that affects a test taker's performance in a test subject (Lord & Novick, 1968). In other words, it is a single feature that keeps the items in the measurement tool together. Unidimensionality appears as the basic assumption of unidimensional IRT models. In order for an item group to be considered as a single dimension, these items must have a common characteristic and this item group must have a common variance that explains the variability among examinees. If unidimensionality is violated, the multidimensional structure of the latent trait space will not match one-to-one with the unidimensional IRT model. When unidimensionality is violated, scores obtained with the unidimensional IRT may be biased. On the other hand, it is very difficult to achieve pure unidimensionality in practice. Because examinees cannot be expected to act in line with only one trait while answering the items. Also, the measured trait may be a multidimensional construct. Multidimensional IRT models have been developed for multidimensional tests. These models assume that more than one trait underlies performance. Multidimensional IRT models can be used if more than one trait determines the examinee's performance (Reckase, 2009; Kartal & Mor Dirlik, 2021).

It is seen that Factor Analytical methods are generally used to test the unidimensionality assumption (Ziegler & Hagemann, 2015). Factor Analytical methods try to explain the relationships between responses to test items with fewer factors (Stone & Zhu, 2015). For this reason, with these methods, factor analysis is applied to the data obtained from the items and a dominant factor is sought (Erkus et al., 2017). Factor Analysis is categorized as Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). While EFA helps us to determine the possible factor structure underlying the observed variables based on examinees responses, CFA allows testing the hypothesis that the determined relationships exist (Suhr, 2006). Apart from factor analysis, many different methods have been developed to test dimensionality. Horn's Parallel Analysis (Watkins, 2006), Velicer's Map Test (Zwick & Velicer, 1986), DIMTEST (Stout, 1987; Nandakumar & Stout, 1993), hierarchical cluster analysis (HCA/CCPROX) (Roussos, Stout, & Marden, 1998) and DETECT index (Zhang & Stout, 1999) are other methods used to determine dimensionality. Principal Component Analysis on standardized residuals is another method used to test dimensionality (Chou & Wang, 2010). As noted above, Factor Analysis functions as a dimension reduction. Decision Tree Induction, Support Vector Machine (SVM), Naive Bayes Classifier and Random Forest Classifier methods from cluster analysis and Machine Learning (ML) algorithms as dimension reduction methods also appear in the literature as methods used in examining dimensionality (Hasan et al., 2021; Dogan & Basokcu, 2010).

Local Indepencence

Local independence means that an examinee's probability of answering an item is independent of their response behavior to other items. In other words, when the ability that affects test performance is kept constant, examinees' responses to items are statistically independent from each other (Hambleton & Swaminathan, 1985). Local independence comes from the basic rule of the probability function on which the IRT is based. Failure to meet this assumption causes the estimations obtained from statistical calculations to be incorrect (Looney & Spray, 1992). For example, in the estimation of ability with the Maximum Likelihood Function, the estimation of ability is estimated by multiplying the probabilities obtained from the responses of the examinees to the items. Likelihood functions calculate probability results by treating items as if they are independent of each other at a given ability level. In order for the probability of two events to occur at the same time to be equal to the product of the probabilities of the two cases, the cases must be independent of each other. It is stated that if the local independence assumption is not met, the test information and item discrimination parameters are overestimated (Chen & Thissen 1997; Embretson & Reise, 2000; Sireci et al., 1991). Junker (1991), on the other hand, states that ability parameters are strongly biased in case of local dependency. In addition, there are studies in the literature showing that the item difficulty parameters of local dependency are incorrectly estimated (Eckes, 2011; Min & He, 2014). There can be many factors affecting the assumption of local independence in educational tests. Response to an item in items with a common root may affect the responses to other items with the same common root (Chen & Thissen (1997). Also, cheating behavior, fatigue (Yen, 1993), students' different practice situations or the test being a speed test (Embretson & Reise, 1999). 2000) affects local independence. In cases where local independence is violated, three solutions are distinguished. First, one of the two items with local dependence between them can be excluded. Second, by creating an item group from local dependent items, this item group can be scaled with IRT models under multi-category models (Yen, 1993). Third, Testlet Response Theory models, which make predictions by considering grouped items, can be used, (Wainer et al., 2007).

A wide variety of methods have been developed to test local independence. Local independence is usually examined through the relationship between the items over the residual matrix calculated by the difference between the observed matrix and the produced matrix from the model. Analyses such as Yen's Q_3 (1984), G^2 squared (or χ^2) (Chen & Thissen, 1997), correlation between residuals (Linacre, 2009), JSI (Edwards et al., 2018) are the methods used to test local independence. It is also stated that local independence can be examined by categorizing the ability scale into different ability ranges and examining the correlation or covariance between the items (McDonald, 1981; Tucker et al., 1986).

Overall Model Fit

As with any modelling study, it is necessary to measure the misfit between the model and the data to determine which IRT model to use. Estimates and inferences made with an inappropriate IRT model will be invalid (Maydeu-Olivares, 2006). Evaluation of the overall fit of the model and the data can be done by comparing the total observed score distribution with the expected score distribution by the model. Evaluation of residuals from differences between observed and expected correct response rates at all skill levels provides precious information for overall model data fit. Model-data fit can be mentioned if the residuals are small and randomly distributed. Embretson and Reise (2000) state that residuals approaching zero for a model can be

taken as a measure of model-data fit. In the evaluation of the general model-data fit, information criterion values that exhibit the χ^2 distribution are generally used. The prominent ones are Log-Likelihood Test [-2*LL], Akaike Information Criteria [AIC], Consistent AIC [CAIC], Bayesian Information Criteria [BIC], sample size–adjusted BIC [SABIC], Hannan-Quinn Criterion (HQ) values (Antoniou et al., 2022; Hambleton & Swaminathan, 1985). The low statistics calculated for the models indicate a better model-data fit. The differences between these values obtained from the models allow comparison of the models with the chi-square statistics at the relevant degrees of freedom. Another test's goodness-of-fit statistic is M_2 (Maydeu-Olivares & Joe, 2006).

In addition, goodness of fit indexes (χ^2/sd, GFI, AGFI, CFI, NFI, TLI, SRMR, RMSEA, et al.) are used in model selection in order to determine which model fits the data better in IRT (Chalmers, 2012). (For more information on general model data fit, see Demars, 2010; De Ayala, 2013; Maydeu-Olivares, 2006).

Item Fit

In order to make inferences from a data set scaled with IRT, it is important to meet the fit of the items to the model. In terms of the accuracy of the estimations to be obtained from the model, the items in the test should fit with the model. Estimates from models with non-fit items will lead to biased estimation of ability parameters, unfair ranking of examinees, and incorrectly equalized scores (Wainer & Thissen, 1987; Yen, 1981). The Item Characteristic Curve (ICC) can be used to evaluate the fit of the items to the model. The indicator of the item's fit with the model is the similar distribution of the estimated ICC and the observed values across the throughout ability scale. In other words, the small difference between the ICC and the observed values will indicate the fit of the item to the model. The difference between the ICC and the observed values is called the residual. The fact that the residuals approach zero is a sign of good item fit (Embretson & Reise, 2000). Visual inspection of the residuals on the ICC is helpful, but it also draws criticism for the subjectivity of the evaluation. For this reason, many item fit indices have been developed. These indices are divided into two groups as Traditional and Alternative item fit indices. Traditional indices divide examinees into specific groups and examine the differences between the expected and observed mean values of these groups. Alternative item fit indices have been developed since it was stated in the traditional indices that if the item is misfit, the estimation made is also incorrect, and therefore the expected scores produced from the model will be incorrect. Traditional item fit indices can be listed as follows: OUTFIT, INFIT indices (for Rasch and 1PL models) (Wright & Panchapakesan, 1969), Bock's χ^2 index (Bock, 1972), Yen's Q_1 index (Yen, 1981), G^2 index (McKinley and Mills, 1985). Alternative item fit indices can be listed as S-χ^2 (Orlando & Thissen, 2000), scaling corrected fit statistics (χ^{2*}) (Stone, 2000), and adjusted chi-square/degrees of freedom ratio (χ^2/df ratio) (Drasgov et al., 1985).

Non-Speedness Test

It is an assumption of IRT models that the test is not performed under accelerated conditions. That is, if an examinee did not answer some test items, it must not be because he did not reach the test items or the time period has expired. This situation must be due to the insufficient level of talent of the examinee. This assumption is sometimes referred to in the unidimensionality assumption. When speed affects test performance; test performance is affected by at least two dimensions - measured talent and speed. This situation has a disruptive effect on unidimensionality. The non-speedness test assumption assumes that examinees should have enough time to answer the items they think they can answer. Many different methods have been proposed to test the non-speedness assumption. A few methods in the literature are as follows: The

first is to examine the relationship between the scores obtained by applying the same test form to the same group under a certain time limit and without a time limit. The second is the ratio of the variance of the number of items that each examinee left blank to the variance of the number of items they answered incorrectly. If this ratio is close to zero, it is stated that the test is a non-speedness test, and if it is close to one, it is a speed test (Gulliksen, 1950). The third is the examination of the marking rates of the items. It is expected that the percentage of those who reach the items and give correct or incorrect answers is high. The fourth is the evaluation of the percentage of "unreachable" items. In order to determine that the test is a non-speedness test, 80% of the examinees must complete the test by reaching all the items, and each examinee taking the test must reach at least 75% of the items (Swineford, 1956).

Aim and Significance of the Research

In this study, it is aimed to determine the status of examining the IRT assumptions of the studies using the IRT model in the literature. For this aim, master's theses and doctoral dissertations written using the IRT model in Türkiye are addressed. The methods by which the researchers tested the IRT assumptions were examined in detail. It is known that IRT models, which have been widely used recently, have many strengths. However, it should be taken into account that the estimated parameters and interpretations will be erroneous in cases where IRT assumptions are violated. When the literature is examined, it is seen that there is no study that deals with the examining of assumptions in detail. It is thought that this study will contribute to the field in terms of revealing the general framework of the IRT assumptions in the literature and presenting an awareness of testing of these assumptions.

Method

This study is a descriptive research as it reports the existing characteristics of the studies conducted using IRT models in terms of IRT assumptions. Descriptive research aims to report the characteristics of the situation examined in the research as it exists (Fraenkel & Wallen, 2011; Koyuncu & Kılıç, 2021). At the same time, document analysis method was used in this research, which was created by gathering information from the studies in the literature. Document analysis is a systematic process that enables the analysis of the information and content in the written elements considered for the purpose of the research (Ary, Jacobs, & Sorensen, 2010). In this study, document analysis method was preferred to examine the master's and doctoral theses written on Item Response Theory between 1993 and 2023 in Türkiye.

Many different and complex IRT models have been developed, such as Multidimensional IRT Models, Mixture IRT Models, Nonparametric IRT Models, Explanatory Item Response Models, etc. Studies conducted on these different IRT models were excluded from the context of this study due to the different scaling methods and assumptions. Therefore, in this study, only master's and doctoral theses prepared using unidimensional IRT models were analysed. This situation is a limitation of this study.

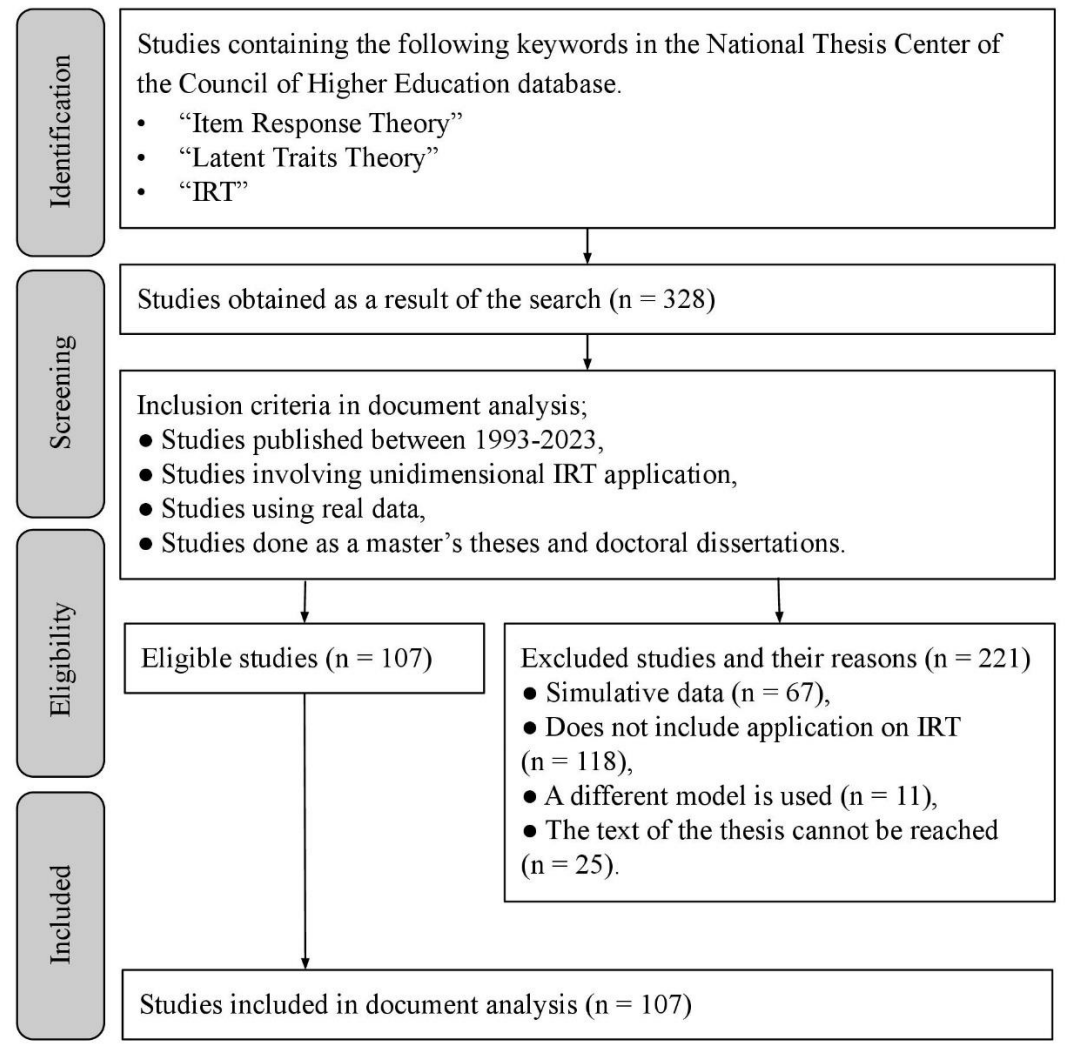


Figure 1. Literature review flow chart

Within the scope of the research, the database of the National Thesis Center of the Council of Higher Education was searched. Details of the literature review are presented in Figure 1.

107 studies that are in line with the inclusion criteria were examined within the scope of the research. In determining the criteria to be investigated, the researchers examined the books and articles that were the main sources in the development and dissemination of the IRT. Aligning with the literature review, unidimensional IRT assumptions are discussed under five headings in Figure 2.

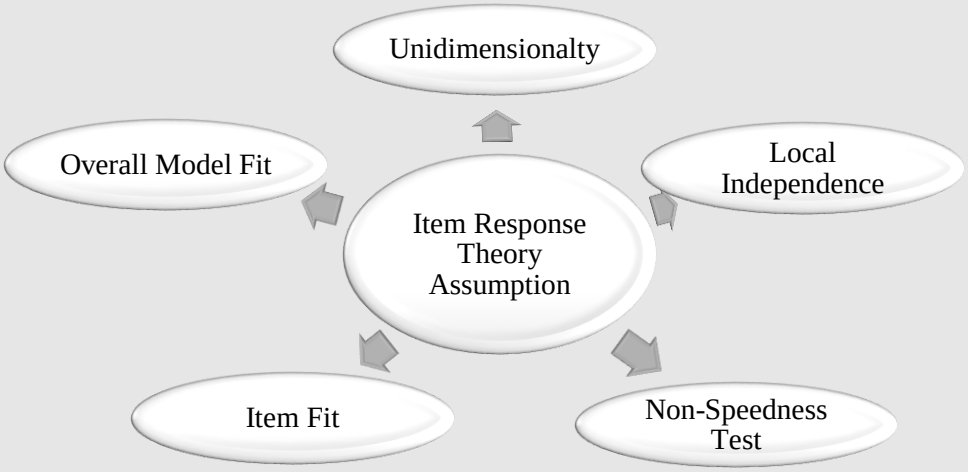


Figure 2. Unidimensional IRT assumptions

As noted in the Sample Size section, the sample must be of a certain size in IRT. For that reason, it was considered to analyze the sample size as a criterion. Then, 107 studies were analyzed according to the assumptions in Figure 1 (1) unidimensionality, (2) local independence, (3) overall model fit, (4) item fit and (5) non-speedness test.

After the included studies were reviewed within the scope determined by the researchers, the data were analysed using descriptive analysis. Results were reported using frequency and percentage.

Ethics

The ethics application for the study was made on 20/06/2021 and the research was carried out with the approval of Social Sciences University of Ankara Ethics Commission dated 06/08/2021 and numbered 14020.

Results

Distribution of Studies by Years

Table 1 shows the distribution of the studies in the study group by years.

Table 1. Distribution of IRT studies by years

Year	Number of Studies	Year	Number of Studies	Year	Number of Studies	Year	Number of Studies
1993	1	2001	0	2009	4	2017	6
1994	2	2002	1	2010	1	2018	11
1995	1	2003	1	2011	4	2019	10
1996	0	2004	0	2012	4	2020	5
1997	0	2005	2	2013	6	2021	8
1998	0	2006	3	2014	7	2022	2
1999	1	2007	0	2015	9	2023	6
2000	0	2008	5	2016	7	Total	107

When Table 1 is examined, it is seen that the studies written using IRT have increased significantly in the last 10 years.

Sample Size

The sample sizes of the studies were investigated in four classes as [0,200], [201,500], [500,1000], [1001+]. In order to evaluate the sample size of the polytomous IRT models, the dichotomous IRT model was coded as the corresponding models (Partial Credit Model -> Rasch or 1PL; Generalized Partial Credit Model, Graded Response Model -> 2PL) (Brzezińska, 2016).

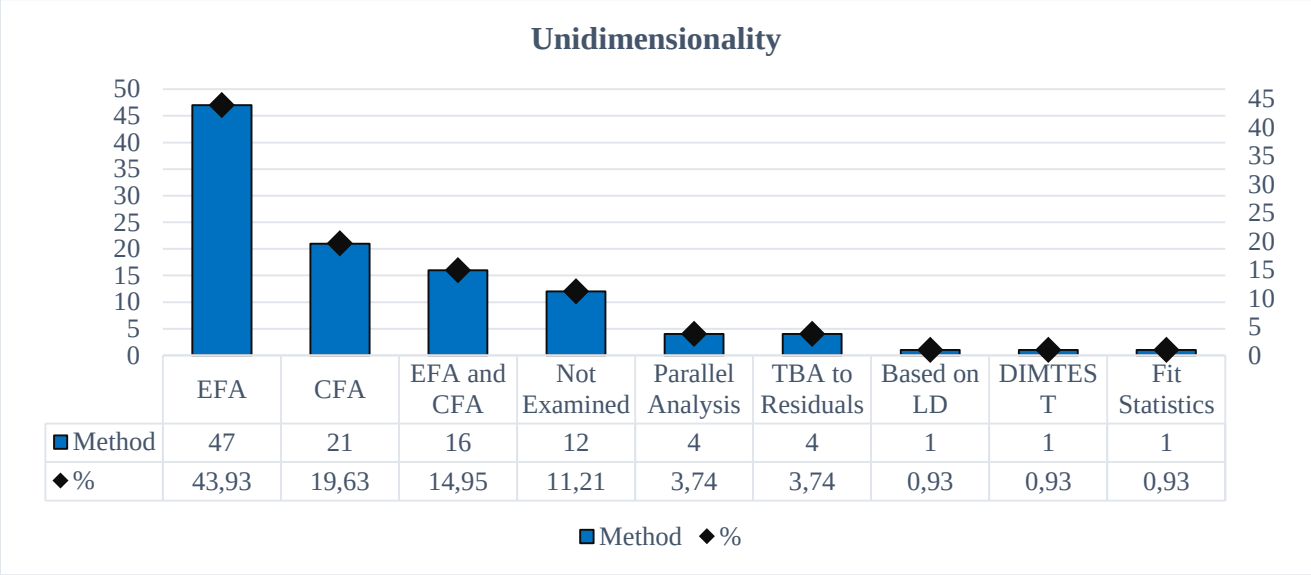
Table 2. Sample size

Model	Sample Size				Total
	[0,200]	[201,500]	[501,1000]	[1001+]	
Rasch or 1PL	3	9	3	16	31
2PL	3	10	14	25	52
3PL	0	1	3	20	24
Total	6	20	20	61	107

When Table 2 is examined, it is seen that there are few studies with low sample size. While there are no studies with a sample size of 200 or less in the 3PL model, there are some studies with low sample sizes in the Rasch-1PL and 2PL models.

Unidimensionalty

The distribution of the methods used in testing the unidimensionality assumption in 107 studies examined in the research is presented in the figure below.



**EFA: Explanatory Factor Analysis, CFA: Confirmatory Factor Analysis, PCA: Principal Component Analysis, LD: Local Dependence.*

Figure 3. Methods used in testing unidimensionality assumption

When the results are analyzed, it is seen that unidimensionality is tested at a high level (n=95, 88.79%). EFA (n=47,%=43.93), CFA (n=21, %=19.63) and EFA and CFA (n=16, %=14.95) are the most used methods for testing unidimensionality.

Local Independence

The distribution of the methods used in testing the local independence assumption in the studies examined in the research is presented in the Figure 4 below.

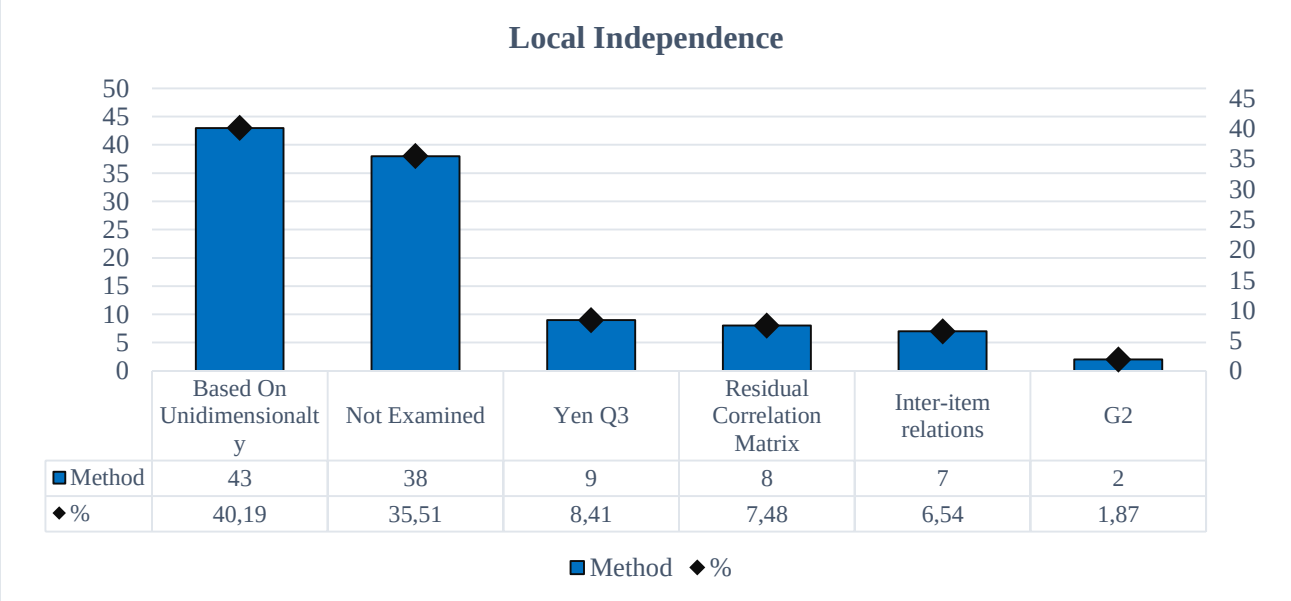


Figure 4. Testing the local ındependence assumption

When the results are analyzed, it is seen that the local independence is mostly handled on the basis of unidimensionalism and there is no additional testing (n=43,%=40.19). Yen Q_3 (n=9, %=8.41), Residual Correlation Matrix (n=8, %=7.48), Inter-item relations (n=7, %=6.54)

and G^2 (n=2, %=1.87) methods are used to test local independence. In many studies, local independence was not examined (n=38, %=35.51).

Overall Model Fit

The distribution of the methods used in testing the overall model fit assumption in the studies investigated in the research is presented in Figure 5.

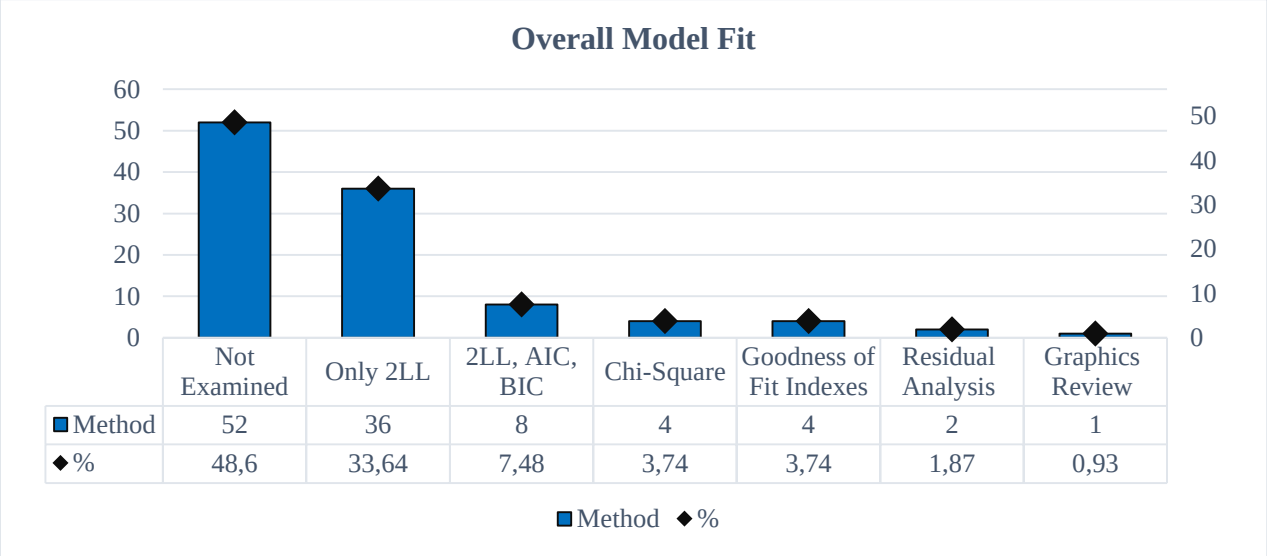


Figure 5. Testing of overall model fit

When the results are analyzed, it is seen that the overall model fit was tested at a moderate level (n=55, %51.4). In the testing of overall model fit, it is seen that Log Likelihood (2LL) (n=36, %=33,64) value is mostly analyzed. Information Criteria values (2LL, AIC, BIC) (n=8, %=7.48), Chi-Square (n=4, %=3.74%), goodness of fit indexes (n=4 , %=3.74), Residual Analysis (n=2, %=1.87) and Graphical Review (n=1, %=0.99) are other methods used to test the model data fit.

Item Fit

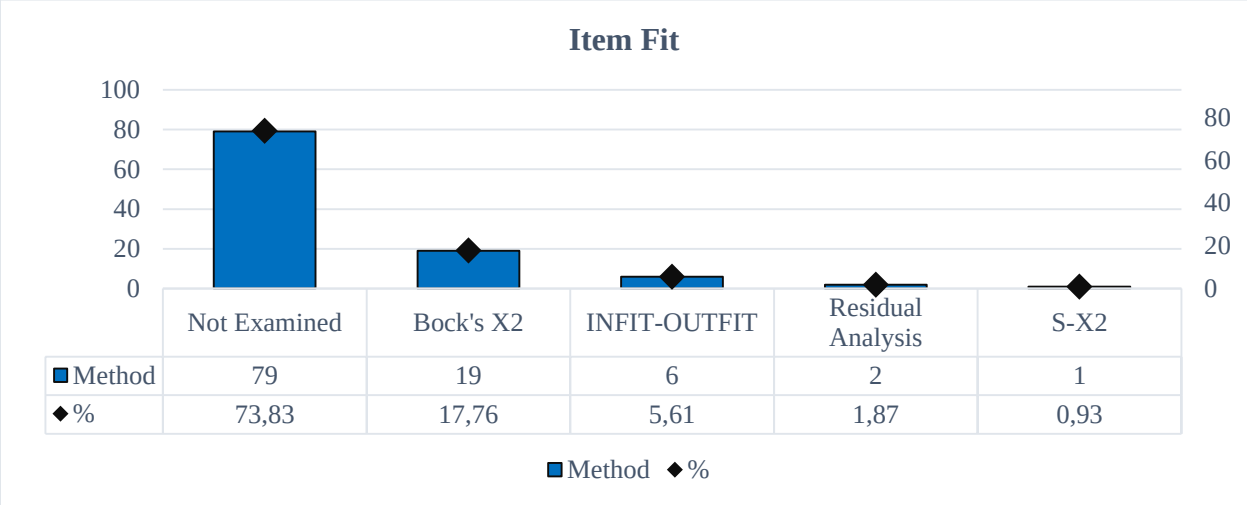


Figure 6. Testing of ıtem Fit

The distribution of the methods used in testing the item fit of the studies included in the study is presented in Figure 6.

When the results are analyzed, it is seen that the item fit is tested at a low level (n=28, %=26,16). It is seen that Bock's χ^2 (n=19, %=17.76) statistics are mostly used in the testing of item fit. INFIT-OUTFIT (n=6, %=5.61), Residual Analysis (n=2, %=1.87), and S-χ^2 (n=1, %=0.93) methods are other methods used to test item fit.

Non-Speedness Test

The distribution of the methods used in testing non-speedness test assumption is presented in Figure 7.

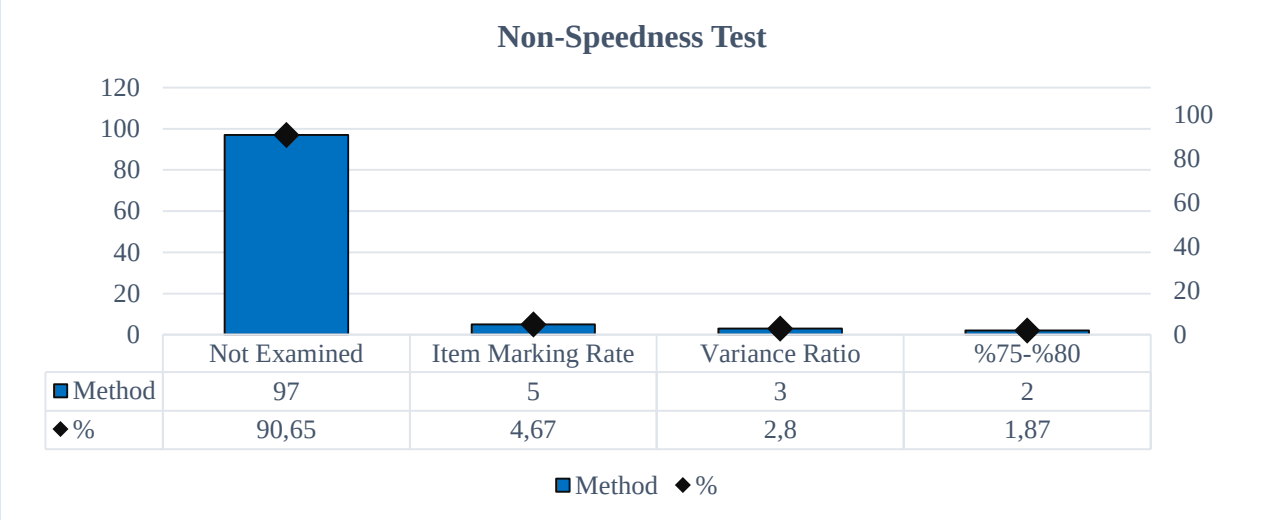


Figure 7. Testing of non-speeded test assumption

When the results are analyzed, it is seen that non-speedness test assumption is tested at a low level (n=10, %9.34). Item Marking Rate (n=5, %=4.67), variance ratio (n=3, %=2.80) and 75%-80% (n=2, %=1.87) methods were used in the testing of non-speedness test.

Discussion and Conclusion

Item Response Theory has been widely used by researchers in recent years thanks to the advantages it offers. In this study; the studies carried out with the IRT model in the last 30 years in Türkiye have been examined and it has been observed that approximately 93% of these studies have been carried out from 2005 to the present. This situation can be interpreted as an indication of the increasing interest in the IRT in recent years. It is known that model assumptions must be met in order to benefit from the strengths of IRT and to interpret test scores correctly. The studies examined in this study were limited to unidimensional IRT models. In this study, the testing of the dimensionality assumption was discussed primarily. Due to the difficulties in the unidimensional test development process, researchers are trying to obtain a dominant factor or component. Although there are many different methods to test the unidimensionality assumption, it is stated that factor analysis methods provide more effective results and are widely used (Erkus, 2006; Lumsden, 1961, 1976; Ziegler & Hagemann, 2015). As a result of the investigations made in this study, it was seen that 78% of the researchers resorted to factor analytical (EFA or CFA) methods in testing the unidimensionality assumption. It is stated that the unidimensionality assumption should be checked in studies using IRT models (Crocker & Algina, 1986; Demars, 2010;

Hambleton & Swaminathan, 1985; Lord & Novick, 1968). In approximately 11% of studies, it was observed that the unidimensionality assumption was not examined.

When the testing of the local independence assumption was analyzed, it was seen that approximately 40% of the researchers did not make a further test, referring to the fact that when the unidimensionality assumption is met, the local independence assumption will also be met. This situation is generally based on Lord's (1980) view that the correlation between the responses of individuals to the items for a given ability level in a one-dimensional test will be zero. So, when the unidimensionality assumption is met, the local independence assumption will also be met. Similarly, since Hambleton & Swaminathan (1985) stated that these two assumptions are equivalent when ϑ ability level is unidimensional, the researchers did not perform a local independence test other than unidimensionality. DeMars (2010), on the other hand, stated that in cases where the dependence between item pairs is at limited levels, it may not emerge as a separate dimension, for that reason, local independence may not be determined with unidimensionality tests and local independence should be tested with different methods. Local independence was tested in 24% of the studies reviewed. Yen's Q_3 test, Residual Correlation Matrix, Inter-item relations and G^2 methods were used to test local independence. In 36% of the studies, the assumption of local independence was not examined. Considering that the estimations obtained from the statistical calculations will be incorrect if the local independence assumption is not met, it is an important problem that there are many studies that do not examine this assumption.

The benefits of IRT for applications such as test development, item bank creation, differential item function (DIF), computerized adaptive testing (CAT), and test synchronization may not be realized unless a fit IRT model is used for a given dataset. The success of IRT applications requires a satisfactory fit between the model and data. The most critical problem caused by model-data misfit may be that parameter invariance, which is the hallmark of IRT, is no longer valid (Rupp & Zumbo, 2006; Shepard, Camilli, & Williams, 1984). Similarly, the items in the test should be fit with the model, that is, the values observed with the estimated ICC should exhibit a similar distribution across the throughout ability scale. When examining the overal model fit in the theses in this research, it was seen that this assumption was not tested in approximately 49% of the studies, and only the LogLikelihood (2LL) value was examined in 34% of the studies. Information criteria, Chi-Square, Goodness of Fit Indexes, Chart Review and Residual Analysis are other methods used to test the overall model fit.

It is seen that in 26% of studies, item fit is tested. Bock's χ^2, INFIT-OUTFIT, Residual Analysis, S-χ^2 indices are the methods used to test item fit. In approximately 74% of the studies, item fit was not examined. The fact that overall model fit and item fit were not tested at a high rate in the studies discussed makes the validity of the results obtained from the model questionable.

In IRT models, the failure of examinees to respond to test items should occur not because of their inability to reach the test items, but because of their limited abilities. In other words, the measurement tool in which IRT models are used should not be a speed test. There are many methods developed to test this assumption. However, it was observed that the assumption of non-speedness test was tested in only 9% of the studies examined. In 91% of the studies, this assumption was not examined. Evidence that the data included in the study was obtained from a measurement tool, which is a non-speedness test, should be presented.

As a result of the examinations carried out, it was seen that there is no certain standard for examining the assumptions in the studies. It is also thought that this situation is caused by the differences between the basic books in the literature in their handling of IRT assumptions. At the

same time, it was concluded that the researchers did not test many IRT assumptions with the necessary rigor. Assumptions must be met in order to benefit from the advantages provided by the IRT. However, in many studies, estimations were made without testing these assumptions. In this case, it should be considered that the measurement results obtained and the decisions made based on these results may be incorrect. Another point is that thanks to the many packages, programs and software developed with the advances in computer technologies, it has become easier to test the assumptions and access to many methods that can be used. It is thought that the use of alternative statistical methods in testing the assumptions in studies to be carried out with IRT models will contribute to the field.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethics

The ethics application for the study was made on 20/06/2021 and the research was carried out with the approval of Social Sciences University of Ankara Ethics Commission dated 06/08/2021 and numbered 14020.

References

Antoniou, F., Alkhadim, G., Mouzaki, A., & Simos, P. (2022). A Psychometric Analysis of Raven's Colored Progressive Matrices: Evaluating Guessing and Carelessness Using the 4PL Item Response Theory Model. *Journal of Intelligence 10*(1),6 MDPI AG. https://doi.org/10.3390/jintelligence10010006

Ary, D., Jacobs, L. C., Sorensen, C., & Razavieh, A. (2010). *Introduction to research in education (Eight).* Belmont: wadsworth Cengage Learning.

Aybek, E. C. (2023). The relation of item difficulty between Classical Test Theory and Item Response Theory: Computerized Adaptive Test perspective. *Egitimde ve Psikolojide Olcme ve Degerlendirme Dergisi*, *14*(2), 118–127. https://doi.org/10.21031/epod.1209284

Baker, F. B. (2001). *The basics of item response theory.* For full text: http://ericae. net/irt/baker..

Bock, R.D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika, 37*, 29-51.

Boduroğlu, E., & Anil, D. (2023). Examining group differences in mathematics achievement: Explanatory item response model application. *OPUS Toplum Araştırmaları Dergisi*, *20*(53), 385–395. https://doi.org/10.26466/opusjsr.1226914

Brzezińska, J. (2016). A polytomous item response theory models using R / Politomiczne modele teorii odpowiedzi na pozycje testowe w programie R. *Ekonometria*. https://doi.org/10.15611/ekt.2016.2.04

Chalmers, R., P. (2012). mirt: A multidimensional ıtem response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1-29. doi: https://doi.org/10.18637/jss.v048.i06

Chen, W.-H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics: A Quarterly Publication Sponsored by the American Educational Research Association and the American Statistical Association*, *22*(3), 265–289. https://doi.org/10.3102/10769986022003265

Chou, Y.-T., & Wang, W.-C. (2010). Checking dimensionality in ıtem response models with principal component analysis on standardized residuals. In educational and psychological measurement (Vol. 70, Issue 5, pp. 717–731). SAGE Publications. https://doi.org/10.1177/0013164410379322

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, 6277 Sea Harbor Drive, Orlando, FL 32887.

De Ayala, R. J. (2013). *The theory and practice of item response theory*. Guilford Publications.

De Boeck, P., & Wilson, M. (Eds.). (2004). *Explanatory item response models: a generalized linear and nonlinear approach.* New York: Springer.

DeMars, C. (2010). *Item response theory*. Oxford University Press.

Dogan, N., & Basokcu, T. O. (2010). İstatistik tutum ölçegi için uygulanan faktör analizi ve asamalı kümeleme analizi sonuçlarının karsilastirilmasi. *Journal of Measurement and Evaluation in Education and Psychology, 1*(2), 65-71.

Doğan, Ö., & Atar, B. (2024). Comparing differential item functioning based on multilevel mixture item response theory, mixture item response theory and manifest groups. *Egitimde ve Psikolojide Olcme ve Değerlendirme Dergisi*, *15*(2), 120–137. https://doi.org/10.21031/epod.1457880

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *The British Journal of Mathematical and Statistical Psychology*, *38*(1), 67–86. https://doi.org/10.1111/j.2044-8317.1985.tb00817.x

Eckes, T. (2011). *Introduction to many-facet Rasch measurement.* Frankfurt am Main: Peter Lang.

Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. *Psychological Methods, 23*(1), 138–149. https://doi.org/10.1037/met0000121

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Psychology Press.

Erkuş, A. (2006). S*ınıf öğretmenleri için ölçme ve değerlendirme: kavramlar ve uygulamalar*. Ekinoks Yayınları, Ankara.

Erkuş, A., Sünbül, Ö., Sünbül, S. Ö., Yormaz, S., & Aşiret, S. (2017). *psikolojide ölçme ve ölçek geliştirme-II ölçme araçlarının psikometrik nitelikleri ve ölçme kuramları.* Pegem Yayınları, Ankara.

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their ıtem/person statistics. *Educational and Psychological Measurement, 58*(3). 357–381. https://doi.org/10.1177/0013164498058003001

Gökçen Ayva Yörü, F. (2024). Thematic and metadological analysis of doctoral dissertations on measurement and. *International Journal of Education Technology and Scientific Researches*. https://doi.org/10.35826/ijetsar.721

Gulliksen, H. (1950). The reliability of speeded tests. *Psychometrika, 15*(3), 259-269.

Hambleton, R. K. (1989). *Principles and selected applications of item response theory. In R. L. Linn (Ed.), Educational measurement*. Washington, DC: American Council on Education and Macmillan.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38-47.

Hambleton, R. K., & Swaminathan, H. (1985). *A look at psychometrics in the Netherlands.*

Hambleton, R. K., & Swaminathan, H. (1985). Assumptions of item response theory. *Item response theory*, 15-31. Springer, Dordrecht.

Hambleton, R. K., Shavelson, R. J., Webb, N. M., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Harwell, M. R., & Janosky, J. E. (1991). An Empirical Study of the Effects of Small Datasets and Varying Prior Variances on Item Parameter Estimation in BILOG. *Applied Psychological Measurement, 15*(3), 279–291. https://doi.org/10.1177/014662169101500308

Hasan, B. M. S., & Abdulazeez, A. M. (2021). A review of principal component analysis algorithm for dimensionality reduction. *Journal of Soft Computing and Data Mining, 2*(1), 20-30. https://doi.org/10.30880/jscdm.2021.02.01.003

Himelfarb, I. (2019). A primer on standardized testing: History, measurement, classical test theory, item response theory, and equating. *The Journal of Chiropractic Education*, *33*(2), 151–163. https://doi.org/10.7899/jce-18-22

Hulin, C. L., Lissak, R. I., & Drasgow, F. (1982). Recovery of two- and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied Psychological Measurement, 6*(3), 249-260.

Junker, B. W. (1991). Essential independence and likelihood-based ability estimation for polytomous items. *Psychometrika, 56*(2), 255-278.

Kartal, S., & Mor Dirlik, E. (2021). Examining the dimensionality and monotonicity of an attitude dataset based on the item response theory models. *International Journal of Assessment Tools in Education*, *8*(2), 296–309. https://doi.org/10.21449/ijate.728362

Kılıç, A. F., Koyuncu, İ., & Uysal, İ. (2023). Scale development based on item response theory: A systematic review. *International Journal of Psychology and Educational Studies*, 10(1), 209–223. https://doi.org/10.52380/ijpes.2023.10.1.982

Koyuncu, İ., & Kılıç, A. F. (2019). The use of exploratory and confirmatory factor analyses: A document analysis. *TED EĞİTİM VE BİLİM*. https://doi.org/10.15390/eb.2019.7665

Linacre JM. (2009). Local independence and residual covariance: a study of olympic figure skating ratings. *Journal of Applied Measurement, 10*(2), 157-69.

Looney, M. A., & Spray, J. A. (1992). Effects of violating local independence on IRT parameter estimation for the binomial trials model. *Research Quarterly For Exercise and Sport, 63*(4), 356-359.

Lord, F. M. (1953). The relation of test score to the trait underlying the test. *Educational And Psychological Measurement, 13*(4), 517-549.

Lord, F. M. (1968). An Analysis of the Verbal Scholastic Aptitude Test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement, 28*, 989-1020.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.*

Lumsden, J. (1961). The construction of unidimensional tests. *Psychological Bulletin*, *58*(2), 122–131. https://doi.org/10.1037/h0048679

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*(4), 713–732. https://doi.org/10.1007/s11336-005-1295-9

McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of mathematical and statistical Psychology, 34*(1), 100-117.

Min, S., & He, L. (2014). Applying unidimensional and multidimensional item response theory models in testlet-based reading assessment. *Language Testing, 31*(4), 453–477. https://doi.org/10.1177/0265532214527277

Mutluer, C., & Çakan, M. (2023). Comparison of test equating methods based on Classical Test Theory and Item Response Theory. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, *36*(3), 866–906. https://doi.org/10.19171/uefad.1325587

Nandakumar, R., & Stout, W. (1993). Refinements of Stout's procedure for assessing latent trait unidimensionality. *Journal of Educational Statistics*, *18*(1), 41–68. https://doi.org/10.3102/10769986018001041

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal Of Mathematical Psychology*, *3*(1), 1-18.

Orlando, M., & Thissen, D. (2000). Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*, *24*(1), 50–64. https://doi.org/10.1177/01466216000241003

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, *21*(1), 25-36.

Reckase, M. D. (2009). *Multidimensional item response theory*. Springer, New York, NY.

Ree, M. J., & Jensen, H. E. (1980). E*ffects of sample size on linear equating of item characteristic curve parameters*. D. J. Weiss (Ed.), Proceedings of the 1979 computerized adaptive testing conference. Minneapolis: University of Minnesota.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual review of clinical psychology*, *5*(1), 27-48.

Rijmen, F., Tuerlinckx, F., De Boeck, P., & Kuppens, P. (2003). A nonlinear mixed model framework for item response theory. *Psychological Methods*, *8*(2), 185.

Rupp, A. A., & Zumbo, B. D. (2006). Understanding parameter invariance in unidimensional IRT models. *Educational and Psychological Measurement*, *66*(1), 63–84. https://doi.org/10.1177/0013164404273942

Saatcioglu, F. M., & Sen, S. (2023). The analysis of TIMSS 2015 data with confirmatory mixture item response theory: A multidimensional approach. *International Journal of Testing*, *23*(4), 257–275. https://doi.org/10.1080/15305058.2023.2214648

Selçuk, E., & Demir, E. (2024). Comparison of item response theory ability and item parameters according to classical and Bayesian estimation methods. *International Journal of Assessment Tools in Education*, *11*(2), 213–248. https://doi.org/10.21449/ijate.1290831

Shepard, L., Camilli, G., & Williams, D. M. (1984). Accounting for statistical artifacts in item bias research. *Journal of Educational Statistics*, *9*(2), 93–128. https://doi.org/10.3102/10769986009002093

Sireci, S. G. (1991, June). *Sample independent item parameters?An investigation of the stability of IRT item parameters estimated from small data sets*. Paper presented at the annual Conference of Northeastern Educational Research Association, New York, NY.

Sözer, E., & Kahraman, N. (2021). Investigation of psychometric properties of likert items with same categories using polytomous item response theory models. *Egitimde ve Psikolojide Olcme ve Degerlendirme Dergisi*, *12*(2), 129–146. https://doi.org/10.21031/epod.819927

Stone, C. A. (2000). Monte Carlo based null distribution for an alternative goodness-of-fit test statistic in IRT models. *Journal of Educational Measurement, 37*(1), 58–75. https://doi.org/10.1111/j.1745-3984.2000.tb01076.x

Stone, C. A., & Zhu, X. (2015). *Bayesian analysis of item response theory models using SAS*. SAS Institute Inc.

Stout, W. (1987). A nonparametric approach for assessing latent trait unidimensionality. *Psychometrika, 52*(4), 589-617.

Suhr, D. (2006). Exploratory or Confirmatory Factor Analysis? *Statistics and Data Analysis,* 1–17.

Swaminathan, H., & Gifford, J. A. (1983). *Esimation of parameters in the three-parameter latent trait model*. In D. J. Weiss (Ed.), New horizons in testing, (pp. 9-30). New York: Academic Press.

Swineford F. (1956). Technical Manual for Users of Test Analyses, *Statistical Report*, 56-42. Princeton, NJ: Educational Testing Service.

Şahin, M. G., Yildirim, Y., & Boztunc Öztürk, N. (2023). Examining the achievement test development process in the educational studies. *Participatory Educational Research*, *10*(1), 251–274. https://doi.org/10.17275/per.23.14.10.1

Thissen, D., & Steinberg, L. (2020). An intellectual history of parametric item response theory models in the twentieth century. *Chinese/English Journal of Educational Measurement and Evaluation, 1*(1). https://doi.org/10.59863/gpml7603

Trabin, T. E., & Weiss, D. J. (1983). *The person response curve: Fit of individuals to item response theory models*. New horizons in testing (pp. 83-108). Academic press.

Tucker, L. R., Humphreys, L. G., & Roznowski, M. A. (1986). C*omparative accuracy of five ındices of dimensionality of binary ıtems*.

van der Linden, W. J. (2010). *Elements of adaptive testing*. C. A. Glas (Ed.). New York, NY: Springer.

Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.

Watkins, M. W. (2006). Determining parallel analysis criteria. *Journal of Modern Applied Statistical Methods, 5*(2), 344-346.

Wright, B., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational And Psychological Measurement, 29*(1), 23-48.

Yen, W. M. (1993). Scaling performance assessments: Strategies for managing local item dependence. *Journal of Educational Measurement*, *30*(3), 187–213. https://doi.org/10.1111/j.1745-3984.1993.tb00423.x

Yiğiter, M. S., & Doğan, N. (2023). Comparison of different computerized adaptive testing approaches with shadow test under different test length and ability estimation method conditions. *Egitimde ve Psikolojide Olcme ve Degerlendirme Dergisi*, *14*(4), 396–412. https://doi.org/10.21031/epod.1202599

Zanon, C., Hutz, C. S., Yoo, H., & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica, 29*(1). https://doi.org/10.1186/s41155-016-0040-x

Ziegler, M., & Hagemann, D. (2015). Testing the Unidimensionality of Items. *European Journal of Psychological Assessment, 31(4),* 231–237. https://doi.org/10.1027/1015-5759/a000309

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*(3), 432.

International Journal of Educational Studies and Policy (IJESP)

Volume: 5, Issue: 2, November 2024

Impact of Artificial Intelligence on Assessment and Evaluation Approaches in Education

Hande Tanberkan[1] , Mahmut Özer[2] , Selahattin Gelbal[3]

ABSTRACT

Artificial Intelligence (AI) technologies are being applied commonly in all aspects of life. Education is one of the leading areas in this respect. AI applications offer significant opportunities for students, educators and education administrators. Students can benefit from these technologies for individualized education and addressing their deficiencies. A similar situation applies to educators. However, students are the most vulnerable group to the current and long-term risks posed by these technologies. While students fulfill a significant part of their responsibilities through the opportunities provided by AI technologies, they face two options: succeeding through ethical violations or addressing their deficiencies ethically. Students lacking AI ethical literacy often choose the first option, masking their failures and getting involved in ethical violations that will bring heavy burdens in the long run. This study discusses the benefits offered by AI technologies in education and the problems they cause in the context of measurement and evaluation. AI will have an important place in the measurement and evaluation as an auxiliary tool in producing texts, creating questions using the produced texts, scoring open-ended exams, solving problems and creating research reports in accordance with ethical rules. It is highlighted that developing AI technologies in education with a participatory approach involving all educational stakeholders and continuously monitoring potential risks during the implementation phase are crucial for establishing a responsible AI culture in education. Finally, considering the dramatic pace of developments in AI, the importance of dynamically updating the measures against ethical violations at the same pace is emphasized.

Keywords: Artificial intelligence, measurement, evaluation, ethics

DOI: https://doi.org/10.5281/zenodo.14016103

Article Info:

Received: 05.06.2024 **Accepted: 22.10.2024** **Article Type: Review Article**

Cite as: Tanberkan, H. Özer, M. & Gelbal, S. (2024). Impact of artificial intelligence on assessment and evaluation approaches in education. *International Journal of Educational Studies and Policy*, *5*(2), 139-152.

[1]Corresponding Author: Hande Tanberkan, Commission of National Education, Culture, Youth and Sports, TBMM, handetanberkan@gmail.com , ORCID: 0000-0001-7142-5397

[2]Mahmut Özer, Commission of National Education, Culture, Youth and Sports, TBMM, mahmutozer2002@yahoo.com , ORCID: 0000- 0001-8722-8670

[3] Selahattin Gelbal, Hacettepe University, Faculty of Education, sgelbal@gmail.com , ORCID: 0000-0001-5181-7262

Introduction

Artificial intelligence (AI) technologies are among the most significant technological disruptions in history, their capacity and effects have not been fully evaluated yet. Specifically, AI applications, which began to emerge in fields such as education, healthcare, and finance in the 1970s, have gained significant momentum over the past decade, elevating this technological disruption to a new phase compared to previous technological transformations (Acemoglu & Restrepo, 2018; Frank et al., 2019; İlikhan et al., 2024; Ozer, 2024; Perc et al., 2019; Septiandri et al., 2023). At this point, there are no areas untouched by AI technology, and new applications are being introduced in every field daily. Consequently, an AI ecosystem encompassing all areas of life is now being discussed (Ozer et al., 2024a; Stahl, 2023). In other words, there is a significant potential for jobs to fall under the dominance of AI (Ozer & Perc, 2024). Furthermore, the advancement of generative AI such as ChatGPT and Gemini has accelerated these developments (Hosseini et al., 2023; Lo, 2023).

The rapid proliferation of AI technologies is now drawing the attention of all segments of society, revealing significant risks alongside the advantages it provides (Suleyman, 2023). It is known that these technologies reproduce the biases present especially in the training data sets used during the learning process, thereby deepening inequalities (Ozer et al., 2024a). In this context, it has been shown that AI technologies in healthcare services deepen racial and ethnic disparities, further exacerbating the disadvantages of already socioeconomically disadvantaged groups (İlikhan et al., 2024; Obermeyer et al., 2019). Similar biases manifest themselves in various fields from education to security and law (Ozer et al., 2024a; 2024b). In short, awareness of AI is increases globally, discussions emerge on how to create an AI ecosystem that does not exacerbate social inequalities, respects social and ethical values, and particularly does not adversely affect employment (Acemoglu & Restrepo, 2018; Ozer & Perc, 2024; Ozer et al., 2024a; Varma et al., 2023).

The opportunities provided by AI increase the allure of AI utilization in also educational environments; however, discussions continue regarding the reliability of AI in education, whether these opportunities are offered within the framework of ethical principles, and whether they support the improvement of educational quality (Acemoglu, 2024). In other words, by utilizing educational data, AI offers opportunities to transform and enrich the learning and teaching processes on one hand, while on the other hand, it carries the risk of increasing existing inequalities among students and ethical misuse (Aquino, 2023; Silva-Jurado & Silva-Jurado, 2024; Ozer, 2024). Therefore, AI system needs to regulate the learning journey in a healthy manner while supporting student development at the same time. In this context, measurement and evaluation processes play a critical role. The development of AI systems signals the revision requirements of traditional assessment methods (Gardner et al., 2021). AI offers a process that encourages individual development, provides guidance, and supports fair assessment outcomes while also having the potential to reduce the workload of educators (Kamalov et al., 2023).

However, integrating AI into measurement and evaluation systems also brings along a series of risks (Cotton et al., 2024; Sok & Heng, 2023; Surahman & Wang, 2022; Verhoeven et al., 2023). AI's text generation capability leads to the risk of cheating and plagiarism in assessment processes. The text generation capabilities offered by generative AI tools like ChatGPT raise serious concerns about undetectable cheating and plagiarism (Kamalov et al., 2023; Ozer, 2024; Surahman & Wang, 2022). The inability of assessment mechanisms to be able to detect AI-generated content provides an unfair advantage to students who use AI for text generation,

adversely affecting the fundamental quality of fair evaluation in assessments. Another negative impact of this situation is the loss of opportunities for individuals to achieve the targeted learning outcomes (Lancaster, 2023). In this context, unethical use of AI for cheating and plagiarism distorts the true value of assessment results and disrupts academic development processes.

On the other hand, the capacity of generative tools like ChatGPT is limited by the dataset they are trained on (İlikhan et al., 2024; Ozer et al., 2024a). In other words, the outputs generated by ChatGPT are limited by the size, scope, and diversity of its training dataset. Since the datasets are derived from real-life scenarios where biases exist, there is a risk of obtaining biased results from AI. While the use of AI in text-related tasks such as translation, essay writing, and automatic content generation offers opportunities for saving time and managing the learning process, it also carries the risk of reinforcing and perpetuating these biases (Gardner et al., 2020; Rane et al., 2023). For example, when translating from a gender-sensitive language to a gender-insensitive language, biased results are observed. In Turkish, he/she is a nurse is translated as she is a nurse by referring to women, while he/she is a doctor is translated as he is a doctor by referring to men. In other words, AI-assisted translation perpetuates the societal biases, such as gender bias (Akgun & Greenhow, 2021; Johnson, 2020).

In addition, it is known that not all content produced by AI systems is accurate and often exhibits a behaviour called hallucination (Ji et al., 2023). When productive AI systems exhibit hallucinatory behaviour, they may produce context-irrelevant or non-existent content, although it may seem plausible. For example, it has been shown that most of the references provided by ChatGPT for use in scientific papers do not actually exist (Athaluri et al., 2023). More interestingly, it is known that once ChatGPT shows hallucinatory behaviour, it maintains this behaviour to ensure the consistency of the content it produces, thus leading to the snowball effect of hallucination (Zhang et al., 2023).

The impact of AI's capabilities on assessment and evaluation systems should be examined in terms of the opportunities it presents and the risks it entails. While maximizing the benefits of AI usage in assessment processes, risks should be minimized to a maximum as well. Enhancing the accuracy and reliability of assessment while using this technology to achieve maximum efficiency should be an essential responsibility for educators. Therefore, the purpose of this article is to identify the opportunities created by AI in measurement and evaluation processes, along with the ethical risks involved, and to propose solutions to mitigate these risks.

Method

This study employs a descriptive approach to evaluate the risks associated with the use of artificial intelligence in text generation and assessment processes in education and also to offer recommendations for managing these risks. The document analysis method is employed to conduct a detailed examination aimed at mitigating the risks associated with the use of artificial intelligence in text generation and assessment. As a qualitative method, document analysis involves conducting a comprehensive review of relevant documents including articles, books, and reports (Bowen, 2009). Accordingly, the literature on the use of artificial intelligence in text generation and assessment has been reviewed. The study illustrates application of artificial intelligence in text generation and assessment including a detailed examination of potential ethical risks. Additionally, a thorough analysis has been conducted to provide recommendations for addressing the risks associated with the use of artificial intelligence in text generation and assessment.

Text generation

Generative AI tools such as ChatGPT have the capability to generate coherent text tailored to assignments and exam questions using machine learning and natural language processing (Cotton et al., 2024; Salvagno et al., 2023). The biggest challenge encountered with AI in education is its predominant use in creating assignments and projects. By doing so, students put themselves into two problematic situations. Firstly, they engage in behavior where they present knowledge that is not their own as if it were theirs, potentially leading to significant long-term behavioral distortions. Secondly, they may mask their shortcomings by presenting deficiencies as competencies (Kasneci et al., 2023; Sok & Heng, 2023). In this case, the complementary support that artificial intelligence could provide to the measurement and evaluation processes could lead to a reverse effect and could enable students at all levels of education to progress successfully despite all their shortcomings. In other words, students may appear to succeed even though they are failing. This situation could deprive students of remedial training for their shortcomings, ultimately leading to their graduation from education without gaining the expected skills in human capital in the long term.

On the other hand, due to the widespread use of AI text generation among students, concerns have emerged that traditional evaluations must be revised (Khalil & Er, 2023). The high grades achieved by texts generated with ChatGPT validate these concerns (Stokel-Walker, 2022). As a strategy to address these concerns, some universities have developed policies prohibiting the use of ChatGPT (Sullivan et al., 2023). However, instead of banning it, universities should focus on integrating AI into their systems ethically and should provide regular trainings to their students on how to benefit from AI (Ozer, 2024; Sok & Heng, 2023; Yu, 2023).

In education, ethical concerns related to AI are not limited to K-12 levels but also encompass higher education institutions (Huallpa, 2023). The impact of AI technologies in higher education institutions is significantly higher for both the researchers and the students. In other words, AI's capabilities in generating texts extend beyond educational assignments, and offer substantial opportunities in scientific article production (Rane et al., 2023). Recently, there has been even discussion regarding whether ChatGPT should be considered a co-author in scientific publications (Stokel-Walker, 2023; Thorp, 2023). This is because ChatGPT's abilities in text generation, translation, and summarization enhance its potential utility in scientific writing (Verhoeven et al., 2023). Hence, AI is expected to support reseachers in organizing ideas, translating, drafting, providing feedback and proofreading in the process of producing scientific articles (Rane et al., 2024). This support makes the preparation of academic studies faster and easier and provides the advantage of increasing the quantity and quality of publications (Rane et al., 2023; Salvagno et al., 2023). However, in scientific articles, authors bear collective responsibility for the content (Stokel-Walker, 2023; Thorp, 2023). Computers are just tools and cannot assume responsibility for the content (Thorp, 2023). Therefore, it cannot be expected by AI to produce original, creative, and critical ideas like humans in scientific article production (Rane et al., 2023; Salvagno et al., 2023). AI-generated texts have not yet established significant credibility in scientific research due to risks such as containing biased content and hallucinations (Salvagno et al., 2023; Stokel-Walker, 2022; Verhoeven et al., 2023).

In summary, the main risks in creating text with AI are ethical concerns in general and the accuracy of the content produced in particular (Lo, 2023). In this context, using AI to generate entire texts rather than as a guide poses the greatest risk leading to fundamental ethical violations.

Therefore, AI's role in text writing should remain that of a helpful assistant to improve efficiency and performance (Verhoeven et al., 2023).

Automatic item generation

The preparation of measurement tools such as exams, assignments, and presentations to measure knowledge and skills increases the workload of educators due to the time involved in the item generation process. Additionally, questions created through traditional methods are constraint by the capabilities of the item writers. Therefore, developments in AI lead to the expectation of speed and increased quality in automatic item generation (AIG) (Bezirhan & von Davier, 2023; Cotton et al., 2024). As AI applications provide educators with the opportunity to easily create achievement tests for classroom assessments. In this context, it is seen that significant developments have been achieved in AI-supported creation of assessment tools by analyzing the educational content such as multiple-choice questions and open-ended questions. (Owan et al., 2023; Qi et al., 2020; Swiecki et al., 2022). In particular, generative AI tools such as ChatGPT and Gemini have the capacity to generate questions in line with the basic skills expected from students.

AIG with correct sentence structures has been partially achieved with existing AI algorithms; however, the content quality of these questions remains debated (Du et al., 2017, Swiecki et al., 2022). For instance, automatically generated items have been criticized for being difficult to understand and aligned with the objectives to be measured (Mulla& Gharpure 2023, Scialom et al., 2019). However, in the process of text generation with AI, studies are ongoing to produce questions with desired features by intervening in AI. For example, Sayın and Gierl (2024) standardise the item generation process by providing ChatGPT with templates containing limitations such as question types, word counts, and sentence structures during the process of item generation with AI. The quality of the generated items was evaluated by experts in the field and item statistics were also calculated. As a result, it was determined that AI increases efficiency in item development processes (Sayın& Gierl, 2024). On the other hand, Bezirhan and von Davier (2023) found that ChatGPT was effective in generating paragraph-based questions in a large-scale reading assessment (PIRLS), but it needed supervision to ensure the content quality of the questions in the AIG process.

The use of AIG studies by teachers in classroom assessments for getting ideas such as creating scenarios, preparing games, adding distractors to multiple-choice questions and improving assessment will increase the efficiency to be obtained from AI-supported AIG (Sherman et al., 2020). It is recommended that educators use these tools to guide them for the time being against the risks of the relevance, quality and content validity of an automatically generated measurement tool (Al-Worafi et al., 2023). In addition, efficiency will increase if the educators review the generated questions.

In summary, AIG with AI presents significant opportunities for educators. However, the generated items should not be used directly for measurement and evaluation purposes before being checked for accuracy, reliability and purposefulness by educators. In other words, educators should actively participate in the item generation process and assign AI an assistant role in supporting and complementing educators.

Automatic assessment and feedback

The concept of automatic text assessment was first introduced by Page (1966). This work laid the foundation for automatic assessment. As technology advanced, awareness of automatic assessment and feedback grew and led to the more widespread use of automatic assessment

platforms. Automated assessment platforms provide the opportunity to conduct exams and assignments electronically. These platforms are preferred because they allow question presentation with animation, video, and audio content that cannot be presented on paper. Additionally they produce objective results, perform automatic evaluations, and provide instant feedback by identifying students' strengths and weaknesses (Akgun & Greenhow, 2021; Kasneci et al., 2023; Owan et al., 2023; Swiecki et al., 2022; Wang et al., 2023). In these respects, they support student development, save time for the teacher, and enable rapid evaluation and feedback.

Advances in AI, have enhanced the potential to assess student-written texts, to provide feedback, and to reduce educators' grading time and, ultimately, have impacted classroom assessment processes (Huang et al., 2023; Jang, 2014). Classroom assessments are rooted in more diverse and enriched student responses compared to standardized tests. They aim to measure students' creative thinking, whether they have learned a topic in depth, and their written and oral communication skills (Warschauer & Grimes, 2008). For example, in essay assessments, although there is a strong relationship between machine and human scoring, machines focus on technical aspects such as word count and correct use of punctuation, while humans focus on skills such as fluency, completeness and creativity. Therefore, at this point, AI-supported automated assessment can increase the efficiency of human-led processes (Gardner et al., 2021). On the other hand, a teacher who realises that a student makes a mistake in a math problem knows that the source of this mistake may vary according to the student's visual defect, psychological state, and misconception. Therefore, teachers tailor the intervention method in the teaching process according to the student's situation. However, AI usually does not have the data to detect these differences while automating the decision-making process (Cardona et al., 2023). For this reason, it is recommended that AI-supported automated assessments should be used for formative assessment in schools until AI can assess similarly to humans (Cardona et al., 2023; Gardner et al., 2021).

On the other hand, despite the opportunities offered by AI-supported automated assessment software, it has the potential to provide unfair advantage or disadvantage to certain student groups. For example, in an exam canceled in the UK due to the Covid-19 pandemic, it is revealed that an algorithm that determines student results based on the success of schools in previous years provides an advantage to the students studying at private schools. This situation reveals that as a result of the decision making mechanism supported AI may let the unfair situations to continue (Akgun & Greenhow, 2021). For this reason, instead of fully automating the evaluation processes with support from AI, using automatic evaluation in situations where objective results will be obtained, such as evaluating multiple-choice questions and giving feedback, will increase the productivity of educators and students.

Responsible and participatory management

The ethical risks associated with the integrating AI in education extende beyond the issue of bias, cheating,and plagiarism in text creation, which fall under fairness. Data protection, data privacy, and accountability also stand as issues that should be confronted in measurement and evaluation processes (İlikhan et al., 2024; Huang, 2023; Lebovitz et al., 2021). In order to build measurement and evaluation processes that deal with these ethical risks, it is necessary to develop strategies to promote responsible AI practices (Theodorou & Dignum, 2020). Responsible AI discusses the question of who is responsible for the ethical use of AI. In this context, responsible AI encompasses AI developers, users, policy makers, societal norms, and even the system itself

(Ozer & Perc, 2024; Stahl, 2023). Therefore, in this section, the risks in the ethical dimension of measurement and evaluation processes are examined from a holistic perspective.

The ethical responsibility for using AI technologies, which develop as a dynamic ecosystem, is the responsibility of all stakeholders. Therefore, all stakeholders should take part in the integration of AI into measurement and evaluation processes, the functioning mechanism of the system should be open to users, and AI should be a system that is constantly monitored, evaluated and updated (Ozer et al., 2024a; Stahl, 2023). In this context, the ethical use encompasses both the development of AI technologies and the use of them. Therefore, both dimensions should be taken into consideration in order to create an AI ecosystem that is compatible with ethical and social values.

The rapid and profound advancement in AI technologies poses a risk that educators may lose control of their assessment processes related to AI. Since the control process is time-consuming and costly, there is a distance between educators' decisions and the evidence-gathering process on which these decisions should be based (Couldry, 2020; Swiecki et al., 2022). Additionally, the fact that AI algorithms are closed to users (blackbox) and controlled by a few companies raises concerns that AI will use student data through its ability to organize datasets, ultimately exacerbating concerns about trust in society (Khosravi et al., 2022; Sullivan et al., 2023; Stahl, 2023). For this reason, it is of great importance to work in a participatory way in which the development processes of AI systems to be used in education are open to all educational stakeholders and the system is constantly evaluated and updated (Ozer et al., 2024a). In other words, educators and educational administrators should be actively involved in the development processes of AI technologies used in education. This approach will also reduce the risk of educators being misled as it will increase their AI technology knowledge and awareness (Khosravi et al., 2022). In short, the participatory AI model will not only enable the development of more ethical AI practices by involving all stakeholders in the process, but also strengthen educators' immunity to the risks that these practices may pose (Ozer et al., 2024a). This approach is also important to ensure social acceptance of AI systems and to build trust among the society (Blasimme & Vayena, 2020; Huang, 2023; Ozer et al., 2024a).

On the other hand, it is necessary to rethink assessment systems to mitigate ethical risks in AI-influenced measurement and evaluation processes (Halaweh, 2023; Lancaster, 2023; Surahman & Wang, 2022). Existing measurement and evaluation approaches should be continuously updated in parallel with the advancements in AI. New tools to measure students' contribution to their work will increase academic honesty (Ozer, 2024). In this context, heterogeneous measurement and evaluation approaches should be integrated into the education systems to ensure fairness and objectivity (Yu, 2023). For example, measurement and evaluation approaches that reveal the student's contribution, such as oral presentations, laboratory activities, group work, and assignments with limited scope, will increase students' responsibility and significantly reduce the risk of using AI as a cheating tool (Sullivan et al., 2023).

Additionally, structured and clear instructions in measurement and evaluation processes encourage students to express their original thoughts and prevent them from engaging in unethical behaviors (Cotton et al., 2024). To reduce the risk of plagiarism and cheating, it is also recommended to compare current performence with previous ones (Sullivan et al., 2023). Assessors' close monitoring of students during the measurement and evaluation process is also an effective way to determine whether AI is being utilized (Cotton et al., 2024). Progress has been made in plagiarism detection with the development of tools that can detect how much of a text has

been written with AI. However, rapid developments in AI technologies often limit the reliability of these tools as well (Foltýnek et al., 2019; Khaled & Al-Tamimi, 2021). Therefore, such platforms are required to be closely monitored the developments in AI technology and be constantly updated.

On the other hand, the proliferation of AI increases the amount of data accumulated in AI systems every day, making it necessary to take serious steps to ensure data privacy (Yanisky-Ravid & Hallisey, 2018). In this sense, educational institutions and AI developers should determine data protection strategies together to ensure the security of students' data (Huang, 2023; Ozer et al., 2024a). At the same time, users of the AI system should have the freedom to protect their personal data and rights.

Certainly, the key approach to both the development and implementation of AI applications in education is to increase the AI literacy of students, educators and educational administrators (European Commission, 2022; Sok & Heng, 2023; Ozer, 2024). Thus, it will be possible to benefit from the advantages of these technologies with the least risk by increasing awareness of how to use them, their limitations and risks. Teachers' gaining competence in AI will support the more successful use of course contents and assessment and evaluation processes and will let them determine whether there is cheating in the assignments submitted by the students easily.

Discussion and Conclusions

AI technology affects and transforms all sectors in the context of the multifaceted opportunities it offers. Measurement and evaluation is one of the most important application areas of AI in education. The use of AI-supported text generation, individualised learning platforms, automatic item generation, automatic assessment and feedback applications is becoming widespread. Therefore, the effects of AI on students, educators and education administrators would increase swiftly (Ozer, 2024). In this context, the potential to produce content using AI in education raises ethical risks with texts being fully or partially authored by AI. In other words, the fact that AI tools using large language models such as ChatGPT offer the opportunity to perform textual tasks such as answering exam questions and writing essays poses the risk of deceiving assessment systems with masked performances. However, AI tools should play a role in supporting the development of students and teachers and organizing teaching processes (Owan et al., 2023). When assessment and evaluation systems lack the ability to detect ethical violations, AI is seen as a tool that supplies unfair advantage to students. An educational life sustained by AI-generated texts will distort the behavioural patterns of students in the short term and negatively affect the quality of human capital in the labour market in the long run.

On the other hand, AI-supported AIG promises to produce questions with high quantity and quality in a short time to be used in both large-scale exams and classroom assessment practices. Although there are still debates on the content quality of automatically generated items, progress is being made in producing items that are compatible with the skills targeted to be measured thanks to the natural language processing models. Although AIG in classroom assessments saves time for teachers, the items are still in need of teachers' supervision in terms of the appropriateness and accuracy of the content to the grade level. At the point where AI has reached today, the leadership of teachers cannot be given up in the process of AIG.

In addition, AI-supported automatic assessment applications also offer the potential to transfer assessment tasks to machines. Assessment processes that are completely left to machines have the potential to fail to monitor student development and produce unfair results due to the

limitations of AI such as algorithm biases, inability to be as effective as humans in assessing high-level skills (Acemoglu et al., 2023). An assessment process that integrates AI should offer a combination of automated and manual assessment methods to measure students' abilities with the least margin of error and support their learning processes. In this context, in automatic assessment processes, AI should be utilized in such a way that it can contribute to teachers and students such as giving feedback and making objective measurements strengthening indivualized learning.

Additionally, studies investigating the contributions of generative AI systems to employee performance have shown that these technologies most significantly enhance the performance of individuals with medium and low skill levels (Brynjolfsson et al., 2023; Noy& Zhang, 2023; Peng et al., 2023). In other words, the contribution of AI technologies to individual performance varies according to skill level, with the level of contribution decreasing as skill levels rise. These findings open up a significant opportunity for assessment and evaluation practices. First, these systems can significantly help reduce achievement gaps in schools by quickly addressing learning deficiencies, especially in low-skilled students. Secondly, considering that AI systems can rapidly elevate novices to targeted proficiency levels in workplaces (Alam et al., 2024; Korinek, 2023), teachers have the potential to quickly improve their measurement and evaluation skills and achieve a common convergence in measurement and evaluation.

The aspect of AI that supports student development such as generating ideas and providing guidance should be emphasized and the idea of using it as a cheating tool should be suppressed. In this context, the priority in AI systems should be given to educating the users (Khosravi et al., 2022; Sullivan et al., 2023). Because students need to be educated about academic honesty and awareness should be raised about the role of AI systems in this regard (Cotton et al., 2024). In other words, instead of deceiving individuals with short-term gains, the complementary aspects that constantly support them should be emphasized. Because those who can use this system efficiently will be ahead not only in their educational journey but also in the labour market (Ozer, 2024). When individuals make use of AI to organize their learning journey, they will gain a valuable companion in their educational journey (Lancaster, 2023).

Consequently, to deal with these ethical risks, AI should not go beyond being supportive and complementary tool for the individuals. To realize this goal, it is crucial to increase the AI literacy of all stakeholders of education. In order to achieve this goal, we suggest the development and use of processes of AI applications in a participatory way involving experts, students, teachers, administrators, and worker unions. On the other hand, it is critical to develop and continuously update digital platforms that will contribute to prevent ethical violations that may arise despite all precautions.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethics

In this study, ethical approval was not required as a review was conducted using previously published studies in the literature.

References

Acemoglu, D. (2024). *The simple macroeconomics of AI (No. w32487*). National Bureau of Economic Research. https://doi.org/10.3386/w32487

Acemoglu, D., & Restrepo, P. (2018). *Artificial intelligence, automation and work (No. w24196).* National Bureau of Economic Research. https://doi.org/10.3386/w24196

Acemoglu, D., Autor, D., & Johnson, S. (2023). Can we have pro-worker- AI? Choosing a path of machines in service of minds. *CEPR Policy Insight No.123*.

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, *2*(3), 431-440. https://doi.org/10.1007/s43681-021-00096-7

Alam, M. F., Lentsch, A., Yu, N., Barmack, S., Kim, S., Acemoglu, D., Hart, J., Johnson, S., & Ahmed, F. (2024). From automation to augmentation: Redefining engineering design and manufacturing in the age of NextGen-AI. *An MIT Exploration of Generative AI*. https://doi.org/10.21428/e4baedd9.e39b392d

Al-Worafi, Y. M., Hermansyah, A., Goh, K. W., & Ming, L. C. (2023). Artificial intelligence use in university: Should we ban ChatGPT? *Preprints* https://doi.org/10.20944/preprints202302.0400.v1

Aquino, Y.S.J. (2023). Making decisions: bias in artificial intelligence and data-driven diagnostic tools. *Australian Journal of General Practice*. *52* (7), 439-442. https://doi.org/10.31128/AJGP-12-22-6630

Athaluri, S. A., Manthena, S. V., Kesapragada, V. K. M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. S. (2023). Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus*, *15*(4). https://doi.org/10.7759/cureus.37432

Bezirhan, U., & von Davier, M. (2023). Automated reading passage generation with OpenAI's large language model. *Computers and Education: Artificial Intelligence*, *5*. https://doi.org/10.1016/j.caeai.2023.100161

Blasimme, A., & Vayena, E. (2020). What's next for COVID-19 apps? Governance and oversight. *Science*, *370*(6518), 760–762. https://doi.org/10.1126/science.abd9006

Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, *9*(2), 27-40.

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). *Generative AI at work (No. w31161)*. National Bureau of Economic Research.

Cardona, M. A., Rodríguez, R. J., & Ishmael, K. (2023). *Artificial intelligence and the future of teaching and learning: Insights and recommendations.* Office of Educational Technology, US Department of Education. https://www2.ed.gov/documents/ai-report/ai-report.pdf

Cotton, D. R., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in education and teaching international*, *61*(2), 228-239. https://doi.org/10.1080/14703297.2023.2190148

Couldry, N. (2020). Recovering critique in an age of datafication. *New Media & Society*, *22*(7), 1135-1151. https://doi.org/10.1177/1461444820912536

Du, X., Shao, J., & Cardie, C. (2017). Learning to ask: Neural question generation for reading comprehension. *arXiv preprint arXiv:1705.00106.* https://doi.org/10.48550/arXiv.1705.00106

European Commission. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators, *Publications Office of the European Union*. https://data.europa.eu/doi/10.2766/153756

Foltýnek, T., Meuschke, N., & Gipp, B. (2019). Academic plagiarism detection: a systematic literature review. *ACM Computing Surveys (CSUR), 52*(6), 1-42. https://doi.org/10.1145/3345317

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *PNAS, 116*(14), 6531-6539. https://doi.org/10.1073/pnas.190094911

Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment:'Breakthrough? Or buncombe and ballyhoo?'. *Journal of Computer Assisted Learning*, *37*(5), 1207-1216. https://doi.org/10.1111/jcal.12577

Halaweh, M. (2023). ChatGPT in education: Strategies for responsible implementation. *Contemporary Educational Technology*, 15(2), ep421. https://doi.org/10.30935/cedtech/13036

Hosseini, M., Rasmussen, L. M., & Resnik, D. B. (2023). Using AI to write scholarly publications. Accountability in Research, 1-9. https://doi.org/10.1080/08989621.2023.2168535

Huallpa, J. J. (2023). Exploring the ethical considerations of using ChatGPT in university education. *Periodicals of Engineering and Natural Sciences, 11*(4), 105-115. http://dx.doi.org/10.21533/pen.v11i4.3770

Huang, L. (2023). Ethics of artificial intelligence in education: Student privacy and data protection. *Science Insights Education Frontiers, 16*(2), 2577-2587. https://doi.org/10.15354/sief.23.re202

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society*, *26*(1), 112-131. https://doi.org/10.30191/ETS.202301_26(1).0009

İlikhan, SU., Ozer, M., Tanberkan, H., Bozkurt, V. (2024). How to mitigate the risks of deployment of artificial intelligence in medicine? *Turkish Journal of Medical Sciences, 54(*3), *483-492*. https://doi.org/10.55730/1300-0144.5814

Jang, E. E. (2014). *Focus on Assessment-Oxford Key Concepts for the Language Classroom*. Oxford University Press.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Yj., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1-38.

Johnson, M. (2020, April 22). *A scalable approach to reducing gender bias in Google Translate*. Google Research Blog. https://research.google/blog/a-scalable-approach-to-reducing-gender-bias-in-google-translate/

Kamalov, F., Santandreu Calonge, D., & Gurrib, I. (2023). New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. *Sustainability*, *15*(16), 12451. https://doi.org/10.3390/su151612451

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khaled, F., & Al-Tamimi, M. S. H. (2021). Plagiarism detection methods and tools: An overview. *Iraqi Journal of Science*, 2771-2783. https://doi.org/10.24996/ijs.2021.62.8.30

Khalil, M., & Er, E. (2023, June). Will ChatGPT get you caught? Rethinking of plagiarism detection. In *International Conference on Human-Computer Interaction* (pp. 475-487). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-34411-4_32

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074. https://doi.org/10.1016/j.caeai.2022.100074

Korinek, A. (2023). Generative AI for economic research: Use cases and implications for economists. *Journal of Economic Literature*, *61*(4), 1281-1317. https://doi.org/10.1257/jel.20231736

Lancaster, T. (2023). Artificial intelligence, text generation tools and ChatGPT – does digital watermarking offer a solution?. *International Journal for Educational Integrity, 19*(10). https://doi.org/10.1007/s40979-023-00131-6

Lebovitz, S., Levina, N., & Lifshitz-Assaf, H. (2021). Is AIground truth really true? The dangers of training andevaluating ai tools based on experts' know-what. *MISQuarterly*, *45*(3), 1501–1526. https://doi.org/10.25300/misq/2021/16564

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences*, *13*(410). https://doi.org/10.3390/educsci13040410

Mulla, N., & Gharpure, P. (2023). Automatic question generation: a review of methodologies, datasets, evaluation metrics, and applications. *Progress in Artificial Intelligence*, *12*(1), 1-32.

Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. *Science*, *381*(6654), 187-192. https://doi.org/10.1126/science.adh2586

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, *366*(6464), 447-453. https://doi.org/10.1126/science.aax2342

Owan, V. J., Abang, K. B., Idika, D. O., Etta, E. O., & Bassey, B. A. (2023). Exploring the potential of artificial intelligence tools in educational measurement and assessment. *EURASIA Journal of Mathematics, Science and Technology Education*, *19*(8), em2307. https://doi.org/10.29333/ejmste/13428

Ozer, M. (2024). Potential Benefits and Risks of Artificial Intelligence in Education. *Bartın University Journal of Faculty of Education*, *13*(2), 232-244. https://doi.org/10.14686/buefad.1416087

Ozer, M., & Perc, M. (2024). Human complementation must aid automation to mitigate unemployment effects due to AI Technologies in the labor market. *Reflektif Journal of Social Sciences,* 5(2), 503-514.

Ozer, M., Perc, M., & Suna, H.E. (2024b). Artificial Intelligence Bias and the Amplification of Inequalities in the Labor Market. *Journal of Economy Culture and Society*, 69, 159-168. https://doi.org/10.26650/JECS2023-1415085

Ozer, M., Perc,M., & Suna, HE. (2024a). Participatory management can help AI ethics adhere to the social consensus. *İstanbul University Journal of Sociology*, 44(1), 221-238. https://doi.org/10.26650/SJ.2024.44.1.0001

Page, E. B. (1966). The imminence of... grading essays by computer. *The Phi Delta Kappan*, *47*(5), 238-243.

Peng, S., Kalliamvakou, E., Cihon, P., & Demirer, M. (2023). The impact of AI on developer productivity: Evidence from GitHub Copilot. *arXiv preprint arXiv:2302.06590*. https://doi.org/10.48550/arXiv.2302.06590

Perc, M., Ozer, M., & Hojnik, J. (2019). Social and juristic challenges of artificial intelligence. *Palgrave Communications*, *5*(1). https://doi.org/10.1057/s41599-019-0278-x

Qi, P., Zhang, Y., & Manning, C. D. (2020). Stay hungry, stay focused: Generating informative and specific questions in information-seeking conversations. *arXiv preprint arXiv:2004.14530*. https://doi.org/10.48550/arXiv.2004.14530

Rane, N. L., Choudhary, S. P., Tawde, A., & Rane, J. (2023). ChatGPT is not capable of serving as an author: ethical concerns and challenges of large language models in education. *International Research Journal of Modernization in Engineering Technology and Science*, *5*(10), 851-874. https://www.doi.org/10.56726/IRJMETS45212

Salvagno, M., Taccone, F. S., & Gerli, A. G. (2023). Can artificial intelligence help for scientific writing?. *Critical care*, *27*(1), 75. https://doi.org/10.1186/s13054-023-04380-2

Sayın, A. & Gierl, M. (2024), Using OpenAI GPT to Generate Reading Comprehension Items. Educational Measurement: Issues and Practice, 43, 5-18. https://doi.org/10.1111/emip.12590

Scialom, T., Piwowarski, B., & Staiano, J. (2019, July). Self-attention architectures for answer-agnostic neural question generation. In *Proceedings of the 57th annual meeting of the Association for Computational Linguistics* (pp. 6027-6032), Florence, Italy. https://doi.org/10.18653/v1/P19-1604

Septiandri, A. A., Constantinidies, M., & Quercia, D. (2023). *The impact of AI innovations on US occupations.* Nokia Bell Labs, UK.

Sherman, A. T., Herman, G. L., Oliva, L., Peterson, P. A., Golaszewski, E., Poulsen, S., Scheponik, T., & Gorti, A. (2020). Experiences and lessons learned creating and validating concept

inventories for cybersecurity. In *National Cyber Summit Research Track 2020* (pp. 3-34). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-58703-1_1

Silva-Jurado, R. J., & Silva-Jurado, M. D. (2024). Educational innovation in the 21st century: gamification, artificial intelligence and art as transformative tools. *YUYAY: Educational, Methodologies & Didactics Strategies*, *3*(1), 35-52. https://doi.org/10.59343/yuyay.v3i1.60

Sok, S., & Heng, K. (2023). ChatGPT for education and research: A review of benefits and risks. *Cambodian Journal of Educational Research*, *3*(1), 110-121. https://doi.org/10.2139/ssrn.4378735

Stahl, B. C. (2023). Embedding responsibility in intelligent systems: from AI ethics to responsible AI ecosystems. *Scientific Reports*, *13*(1), 7586. https://doi.org/10.1038/s41598-023-34622-w

Stokel-Walker, C. (2022). AI bot ChatGPT writes smart essays - Should professors worry? *Nature*. https://doi.org/10.1038/d41586-022-04397-7

Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: Many scientists disapprove. *Nature, 613*(7945), 620-621. https://doi.org/10.1038/d41586-023-00107-z

Suleyman, M. (2023). *The coming wave: Technology, power and the 21st century's greatest dilemma*. Crown.

Sullivan, M., Kelly, A. & McLaughlan, P. (2023). ChatGPT in higher education: Considerations for academic integrity and student learning. *Journal of Applied Learning & Teaching*, *6*(1). https://doi.org/10.37074/jalt.2023.6.1.17

Surahman, E., & Wang, T.H. (2022). Academic dishonesty and trustworthy assessment in online learning: A systematic literature review. *Journal of Computer Assisted Learning, 38*(6), 1535–1553. https://doi.org/10.1111/ jcal.12708

Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S.,Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075. https://doi.org/10.1016/j.caeai.2022.100075

Theodorou, A., & Dignum, V. (2020). Towards ethical and socio-legal governance in AI. *Nature Machine Intelligence*, *2*(1), 10–12. https://doi.org/10.1038/s42256-019-0136-y

Thorp, HH. (2023). ChatGPT is fun, but not an author. *Science, 379*(6630), 313. https://doi.org/10.1126/science.adg7879

Varma, A., Dawkins, C., & Chaudhuri, K. (2023). Artificial intelligence and people management: A critical assessment through the ethical lens. *Human Resource Management Review*, *33*(1), 100923. https://doi.org/10.1016/j.hrmr.2022.100923

Verhoeven, F., Wendling, D., & Prati, C. (2023). ChatGPT: when artificial intelligence replace the rheumatologist in medical writing. *Annals of the Rheumatic Diseases, 82*(8), 1015-1017. https://doi.org/10.1136/ard-2023-223936

Wang, X., Gong, Z., Wang, G., Jia, J., Xu, Y., Zhao, J., Fan, Q., Wu, S.,Hu, W., & Li, X. (2023). ChatGPT performs on the Chinese national medical licensing examination. *Journal of Medical Systems, 47*(1), 86. https://doi.org/10.1007/s10916-023-01961-0

Warschauer, M., & Grimes, D. (2008). Automated writing assessment in the classroom. *Pedagogies: An International Journal*, *3*(1), 22-36. https://doi.org/10.1080/15544800701771580

Yanisky-Ravid, S., & Hallisey, S. (2018). 'Equality and privacy by design': Ensuring artificial intelligence (AI) is properly trained & fed: A new model of AI data transparency & certification as safe harbor procedures. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3278490

Yu, H. (2023). Reflection on whether ChatGPT should be banned by academia from the perspective of education and teaching. *Frontiers in Psychology, 14.* https://doi.org/10.3389/fpsyg.2023.1181712

Zhang, M., Press, O., Merrill, W., Liu, A., & Smith, N. A. (2023). How language model hallucinations can snowball. *arXiv preprint arXiv:2305.13534.* https://doi.org/10.48550/arXiv.2305.13534

International Journal of Educational Studies and Policy (IJESP)

Volume: 5, Issue: 2, November 2024

Perceptions of Preschool and Primary School Teachers About Blended Learning Experience

Taha Yasin Bacakoğlu[1] , Abdulhamit Kurupınar[2] , Mehmet Hilmi Koç[3]

ABSTRACT

In the current study, it was aimed to investigate the views and experiences of preschool and primary school teachers regarding blended learning. In the study, case study, one of qualitative designs, was used. In determining the working group, criterion-sampling method was used. The working group was made up of 29 teachers working in Ankara. The data which was collected with semi-structured interview questions was applied content analysed. The findings obtained through the analysis of the data showed that almost all the teachers felt themselves incompetent in planning teaching processes, "as in the nature of blended learning", with face-to-face and distance education. In addition, teachers pointed out that they needed education in such issues as preparing activities, producing digital content, using technology, classroom management, teaching methods and techniques in the part of distance education of blended learning. It was found that while the views of the primary school teachers teaching to third and fourth grades of primary education were quite positive regarding the applicability of the blended learning, first grade and second grade primary school teachers and preschool teachers were abstained. The results obtained from the findings of the research were discussed depending on the literature and some recommendations were made for the related partners. Given the qualitative nature and limited sample size of this study, future research should explore larger samples using experimental methods. Additionally, enhancing teacher training in technology integration for blended learning is recommended to better equip educators in applying these methods effectively.

Keywords: Preschool education, primary education, blended learning, technology integration, educational technology

DOI: https://doi.org/10.5281/zenodo.14016278

Article Info:

Received: 30.04.2024 **Accepted: 24.10.2024** **Article Type: Research Article**

Cite as: Bacakoğlu, T. Y. Kurupınar, A. & Koç, M. H. (2024). Perceptions of preschool and primary school teachers about blended learning experience. *International Journal of Educational Studies and Policy*, *5*(2), 153-173.

[1]Corresponding Author: Dr. Taha Yasin Bacakoğlu, Gazi University, tahayasinbacakoglu@gmail.com, ORCID: 0000-0001-8369-7840

[2]Dr. Abdulhamit Kurupınar, Ministry of National Education, hamit7406@gmail.com, ORCID: 0000-0001-5885-9131

[3]Associate professor Mehmet Hilmi Koç, Istanbul Metropolitan Municipality, kocmehmethilmi@gmail.com , ORCID: 0000-0001-6259-173X

This article was presented as an oral presentation at the 19th International Primary Teacher Education Symposium (IPTES 2021).

Introduction

Rapid changes in the field of technology have affected the field of education just as they have impacted many other sectors. One significant outcome of these technological advancements in education is the increased use of distance and blended learning. The importance of these methods surged during the Covid-19 pandemic, which had a profound global impact in 2020. As formal education was disrupted, distance and blended learning emerged as critical alternatives on an unprecedented scale.

Countries worldwide, including China, were compelled to close schools and universities in March 2020 due to the pandemic. To ensure continuity in learning, many adopted online and blended learning solutions (Cao et al., 2021). This global shift sets the stage for examining the status of the education system in Turkey during this period. In the academic year of 2019-2020, nearly 25.8 million students were enrolled in education in Turkey, with 1.6 million in preschool and 5.3 million in primary education. Preschool and primary education, considered the basic education stage, encompassed 6.9 million students, accounting for 26.7% of the total student population (Turkish Statistical Institute [TURKSTAT], 2022). With the World Health Organization (WHO) declaring a pandemic on March 11, 2020 (Ministry of National Health, 2022), Turkey's education system faced significant challenges, mirroring the global scenario (United Nations Educational Scientific and Cultural Organization [UNESCO], 2022). Each country, including Turkey, sought temporary solutions to cope with the crisis. Initially, education in Turkey was suspended from March 16, 2020, to March 30, 2020, followed by a mid-term break (Ministry of National Education [MoNE], 2022a). Education quickly resumed across all levels after the break (MoNE, 2022b). Subsequently, face-to-face education was intermittently conducted (MoNE, 2022c; MoNE, 2022d; Anadolu Agency [AA], 2022a), and a blended learning model was implemented for a period (AA, 2022b; MoNE, 2022e).

Blended learning model can be defined as the practice of face-to-face education and online/distance education together (Dağ, 2011). In other words, blended learning is the practice face-to-face and online/distance education together in order to make the learning process fruitful and flexible (Gilroy, 2001; Jumabaeva et al., 2020; Lago, 2000; Osguthorpe & Graham, 2003; Porter et al., 2014; Rasheed et al., 2020; Stein & Graham, 2014). Blended learning does not mean a limited case like using more technology in the class. According to Grupe (2002), it is the increase in distance education using various digital software by decreasing the number of face-to-face courses rather than increasing the technological possibilities of educational settings. In the blended learning is regarded as a modern teaching form because of its effectiveness in achieving flexible, timely, and uninterrupted learning (Porter et al., 2014). The use of blended learning has increased recently in every stage of education (Sloman, 2003), triggered the start of a new education period with the support of technology (Erdoğan, 2021) and according to some researchers of educational sciences, it has started to be regarded as “new normal in education” (Dziuban et al., 2018; Wut & Xu, 2021).

Blended learning has been extensively evaluated for its impact on various aspects of education. Studies consistently show that it enhances student success (Korkmaz & Karakus, 2009; Korkmaz & Kadirhan, 2020; Uluyol & Karadeniz, 2009; Yılmaz, 2018) improves athletic skills (Çakıt & Karadeniz, 2020), and increases active student participation in courses (Hastie & Curtner-Smith, 2006). Additionally, blended learning has been found to contribute positively to reading courses, facilitating deeper engagement in learning activities (Jumabaeva et al., 2020). The

primary focus of these studies has been on academic achievements and course effectiveness, often overlooking broader implications for younger learners.

Recent research further supports these findings and extends them to different educational contexts. Dikmen (2021) demonstrated that blended learning environments enhance teachers' social, cognitive, and instructional presence, particularly in programming education. Hiğde and Aktamış (2021) highlighted the positive impact of blended learning on student attitudes towards online learning, despite the presence of technical issues. Türker (2021) and Diana et al. (2022) emphasized blended learning's ability to provide personalized feedback, accommodate individual learning differences, and create a supportive communication environment between teachers and students. Finlay et al. (2022) found that sports students in the UK responded more favorably to blended learning, appreciating the opportunities it offered for practical application. Similarly, Prifti (2022) observed that blended learning improved students' self-efficacy and satisfaction, particularly in a higher education setting.

This study stands out by focusing on the largely unexplored area of blended learning in early education, particularly in preschool and primary settings. These stages are crucial for children's development, and the role of teachers in shaping foundational skills and attitudes is paramount. By examining the experiences of these educators, the research provides unique insights into the practical challenges and benefits of integrating technology and innovative teaching methods with young learners. This approach not only fills a gap in the existing literature but also offers valuable implications for educational practices and policies. To investigate the experiences and views of preschool and primary education teachers in blended learning, the following questions were addressed:

1. What do they think of the applicability of blended learning model in preschool / primary education?
2. What do they think of the advantages and disadvantages of blended learning model in terms of teachers?
3. What do they think of the advantages and disadvantages of blended learning model?
4. How do they practice blended learning model?
5. What kind of challenges did they encounter while practicing blended learning model?
6. Do they need education on blended learning, if they do, what must be the content of this education?

Method

The research was designed according to case study, a qualitative research approach. Case studies are carried out to make a deep investigation using the elements (place, person/s, process etc.), qualitative data collection tools (observation, interview etc.) affecting the case and to determine the details regarding the case (Creswell, 2014; Yıldırım & Şimşek, 2016). In this study, a case study approach is appropriate because it allows for an in-depth exploration of the specific context in which preschool and primary education teachers in Ankara experience blended learning. This method enables the collection of rich and detailed data on the practical applications, challenges, and benefits of blended learning as perceived by the teachers, providing valuable insights into their educational practices and the contextual factors influencing their experiences.

Participants

Which teachers would participate in the research was determined by using the criterion sampling model, which is one of the purposive sampling methods. Purposive sampling is to make a deep investigation by choosing rich cases in terms of information depending on the purpose of the study (Büyüköztürk et al, 2018). In accordance with the criterion that teachers are the primary school and preschool teachers having an experience in blended learning model, the working group was made up of 14 preschool teachers, and 15 primary school teachers teaching at 1-4 grades working in the city of Ankara. Preschool and primary school teachers were included in the study simultaneously because the early educational stages are foundational in children's development. This approach allows for a comprehensive understanding of how blended learning is applied across different age groups in early education. Professional seniority of the teachers varied from 7 to 25 years. Detailed personal information regarding the participants is given in Table 1.

Table 1. Personal information of participants

	n	**%**
Branch		
Preschool	14	48
Primary school	15	52
Gender		
Woman	21	72
Man	8	28
Age		
20-30	12	41
31-40	8	28
41-50	7	24
51 and over	2	7
Professional Seniority		
1-10 years	11	38
11-20 years	9	30
21-30 years	8	28
31 years and over	1	4
Status of Education		
Bachelor’s degree	25	86
Master’s degree	4	14

As given in Table 1, the working group was made up of 29 teachers, 48% in preschool (n=14), and 52% in primary school (n=15). As for the gender, 72% of the teachers were woman

(n=21) and 28% were men. Besides that, 41% were in the age group of 20-30 (n=12), 28% were in the age group of 31-40 (n=8) while 68% (n=20) had a professional seniority of 1-20 years with 86% were bachelor (n=25) teachers. In short, the distribution of the participant teachers was similar to each other, woman teachers were high in number, professional seniority outnumbered in the group of 1-20 years, and most of the teachers were a bachelor.

Data Collection Tools

Before the study began, ethical consent was obtained from Nevşehir Hacı Bektaş Veli University. The data was gathered using semi-structured interview questions developed by the researchers after a literature review on blended learning. Three experts, lecturers in primary and preschool teaching with experience in distance and hybrid learning and educational technologies, reviewed the interview form. Based on their feedback, two questions were removed, and three new ones were added. A pilot interview with three teachers revealed some questions were unclear, so probing questions were included.

Data Collection and Analysis

Participants were informed about the study and their consent was obtained. During the data collection process, necessary precautions were taken to ensure the data's security and confidentiality. Most of the data was collected through face-to-face or video calls, with a few conducted via telephone. The views of the teachers were coded as “P1, P2, P3, …” for preschool teachers and “C1, C2, C3, …” for primary school teachers to adhere to ethical rules. The teachers' views were shared in italic fonts.

The data obtained from the teachers were analyzed through content analysis. The data was coded in consistency with the nature of content analysis, themes were formed and evaluated (Yıldırım & Şimşek, 2016). To prevent bias in the data, it was examined by two researchers three times, significant parts were marked, and the codes obtained were noted down across the expressions. The repeating codes were evaluated with an inductive approach, and themes were formed.

Validity and Reliability

To ensure the reliability of the research, direct quotations from the obtained data were used, and multiple researchers participated in each phase of the study. Additionally, the Miles and Huberman formula was applied. This formula defines reliability as "agreement / (agreement + disagreement)." The calculations revealed that the reliability was 88%. Miles and Huberman (1994) consider reliability rates of 70% and above to be reliable. The high similarity in the codes (88%) indicates a high level of reliability in data coding. Although the initial coder similarity was 88%, discussions continued until full consensus was reached throughout the study. To ensure the validity of the research, expert opinions were obtained, data were collected through long-term and detailed interviews, and the stages of the research were described in detail.

Ethics committee approval process

The ethics application for the study was made on 07/05/2021 and the research was carried out with the approval of Haci Bektas Veli University Ethics Commission dated 14/06/2021 and numbered 2021.07.221.

Results

In this part, the findings of the research questions formed to investigate the views of preschool and primary education teachers regarding blended learning were given in order and evaluated. The first question of the research was *"*What do the preschool and primary education teachers think of *the applicability of blended learning model in preschool / primary education?* The findings obtained by means of the questions asked to the teachers to answer this question were gathered in the theme of "the applicability of blended learning model in basic education". Sub-themes and frequencies regarding the theme of applicability of the blended learning model in basic education were given in Table 2.

Table 2. Sub-themes and frequencies regarding the theme of applicability of the blended learning model in basic education

Main Theme	Sub-Themes	f
The Applicability of the Blended Learning Model in Basic Education	**Non-applicable**	
	It cannot be applied in preschool and 1st and 2nd grades of primary education	11
	Students in this age do not have basic skills yet	3
	It is necessary to teach the student by showing personally	2
	It cannot be applied without parent support	2
	Children in this age should learn by experiencing	2
	Children in this age cannot keep up with blended learning technology	2
	Games and activities are of vital importance in this period	2
	It is difficult to apply blended learning with old-age teachers	1
	Contact is important in preschool	1
	Distance education prevents socialization of students	1
	Applicable	
	It can be applicable in the 3rd and 4th grades of primary education	6
	It can be applied as face-to-face for four days and in distance form for one day	6
	It can be applied very easily in primary education	6
	It can be applied in preschool education	5
	Games and activities, face-to-face cognitive outcomes can be given online	3
	It can be applied as face-to-face for three days and in distance form for two days	2
	Students in 3rd and 4th grades have the skill of using technology adequately	2
	Duration should be adjusted according to age level	2
	It can be applied if teachers are trained in this case	2
	It can be applied easier with educated parents	1
	Parents should be trained on blended learning	1

As given in Table 2, some teachers pointed out that blended learning cannot be applied in preschool and primary education known as basic education while some others indicated that it can be applied. Those having the idea that it cannot be applied expressed that it would particularly be difficult to apply blended learning in preschool and first and second grades of primary education. As for the participant views regarding that it is difficult to apply blended learning in basic education, it was pointed out that students in this age do not have basic skills yet, it is necessary to teach the student by showing personally, children in this age should learn by experiencing, games and activities are of vital importance in this period, contact is important in preschool and that distance education prevents socialization of students. Some of the answers given by the participants regarding the "non-applicable" sub-theme of "can blended learning be applied in basic education" theme is given below:

"It is not possible to apply it in the first and second grades of primary education, since you should teach this age group by showing and making them experience personally" (C2).

"It might be applied at children, but I cannot do it much. I am an advanced age person. Some younger friends say that they can do it better. They are telling something, but I cannot understand it much" (C5).

"I do not find it applicable particularly in preschool and in the first years of primary education, because this young age group do not have the developmental characteristics to sit and listen in front of technological devices for minutes with a concentration. In addition, game is the most important method used in this period, and it would be more difficult to make the play games through distance education" (P17).

Some other participants argued that blended learning could be applied both in preschool and primary education. It was pointed out by the participants that the students particularly in the 3rd and 4th grades of primary education have a high technological adaptation and blended learning can easily be applied to the students in this age group. It was also indicated that planning it as a face-to-face education for three or four days a week and a distance education for one or two days a week would increase the applicability of blended learning. The participant also pointed out that giving education to both teachers and parents on blended learning would increase the applicability of this education model. Some of the answers given by the participants as to the fact that blended learning model can be applicable are given below:

"It can be applied in 4th grade. Children have the skill of using technology adequately" (C4).

"It can be applied in preschool period. It can be applied as a face-to-face education for four days and a distance education for one day. I would plan and inform parents at the very beginning of the educational year about blended learning process. Parent involvement is necessary, particularly the parents at medium and low economic level and those having a child with an attention deficit. Parents have responsibilities in blended learning" (P11).

"It can be applied as face-to-face for three days and distance education for two days" (P13).

As the second question of the research, the answers given to the question of "What do the teachers think of the advantages and disadvantages of blended learning model in terms of students and teachers?" were gathered under the theme of "evaluation of blended learning". Sub-themes and frequencies regarding the theme of evaluation of the blended learning were given in Table 3.

Table 3. Sub-themes and frequencies regarding the theme of evaluation of the blended learning

Main Theme	Sub-themes	f
Evaluation of the Blended Learning	Advantages	
	Using technology more functionally	10
	Providing the possibility of digital material development	9
	Encouraging using digital materials	9
	Saving time	7
	Being more economic	7
	More active parent involvement to education process	3
	It offers new alternatives in assessment and evaluation	3
	Flexibility in course hours	2
	Disadvantages	
	Increase in workload, carrying it to family life, expecting everything from teacher	7
	Disorganization, not knowing what to do	7
	Challenges in classroom management	6
	Challenges in knowing technology and using it	5
	Difficulty in assessment and evaluation	2
	Being unable to use non-verbal expressions (gesture, mimic)	1
	Unwillingness and weariness at teacher	1
	Not being able to produce digital materials	1
	Not being able to guide education according to individual differences	1
	Disadvantages resulting from non-involvement of parents to the education process	1

The participants evaluating blended learning process applied in the pandemic period indicated that the highest benefit they obtained in this process was technology use and developing digital education materials. In this sense, the expressions of *"I didn't think that I would be able to use technological devices in such a useful way"* (P4) and *"As far as I learned from my young colleagues, developing digital material was not so difficult"* (C10, with a higher seniority) could be commented such that blended learning applied in the pandemic period made a contribution to the professional development of teachers.

Another advantage expressed by most teachers was that they were able to spare more time for their daily affairs on the days when distance education was applied in blended learning model, they did not have to endure the traffic of the metropolitan city and that they were able to save both time and economy since they did not go to school every day. The participant "C14" pointed out that "*It is likely to say that it is more advantageous for teachers in terms of both time and transportation costs".* In addition, participants indicated that increased active involvement of parents, flexibility in the course hours and developing new alternatives in the assessment and evaluation were among other advantages of blended learning.

It was indicated that the increase in the workload and the discomfort felt from carrying work to family life were among the disadvantages concerning blended learning. Particularly woman teachers having a young child expressed that they got tired for that reason and that they could not spend qualitative time both to their job and to their children. Another disadvantage given by the teachers was that they found themselves insufficient regarding how to carry out blended learning in its most effective way. Besides that, the fact that teachers had difficulties in technology use and that they felt themselves insufficient in classroom management were among the frequently expressed disadvantages. In addition, not being able to produce digital materials, not being able to guide the education depending on individual differences and the challenges caused by the

involvement of parent in the education process in a wrong way were the disadvantages given by the teachers. Some of the answers given by the participants regarding the "disadvantages of blended learning" sub-theme is given below:

> *"… disorganisation, decrease in motivation based on the increase in workload and weariness come to the forefront" (P6).*
> *"… it is difficult for the married and for those having children. It brought me a serious workload" (C11).*
> *"I am in difficulty with what students learn and what they do not. I would not prepare for the courses before. I thought that I would teach it in any case. Thinking that everybody would watch me particularly in the distance education part of blended learning, I spend more time for the preparation. It brings us a serious workload" (C12).*

The third question of the research was "What do preschool and primary education teachers think of the advantages and disadvantages of blended learning model in terms of students?". The answers given to the interview questions were gathered under the theme of "blended learning in terms of students". The sub-themes and frequencies regarding the theme of blended learning in terms of students are given in Table 4.

Table 4. The sub-themes and frequencies regarding the theme of blended learning in terms of students

Main Theme	Sub-Themes	f
	Advantages	
Blended Learning in terms of Students	Increase in technology using skills	7
	Possibility of partly resting at home	4
	Being more comfortable at home	3
	Increase in the possibility to work with the family	3
	Missing the school setting	3
	Increase in the attendance to school	3
	Positive development at children with a slow learning pace and introverted ones	2
	Opportunity to learn from different sources	2
	Easier access to teacher	1
	Being able to support what is learned at school with distance education	1
	Increase in adaptation to school	1
	Being able to complete unfinished activities at school with distance education	1
	Carrying on education in the vacant periods because of interruption of education	1
	Disadvantages	
	Computer insufficiency and connection problems	7
	Lack of motivation at school attendance	4
	Lack of adoption by families for blended learning	4
	Screen addiction	2
	Spending more time with affairs out of courses	2
	Not being able to receive enough support from guidance service	1

Table 4 indicates that participants view the benefits of blended learning for students as outweighing the drawbacks. Benefits include enhanced tech skills, partial home rest, increased family collaboration, improved attendance, positive development for various student profiles, diverse learning sources, better teacher access, and support for in-school learning. Downsides encompass tech insufficiency, connectivity issues, reduced school motivation, family hesitancy

towards blended learning, screen addiction, non-academic time use, and limited guidance support. Some participant responses on "student-oriented blended learning" are as follows:

> *"Having one day distance education makes students be attached to school and increases missing. Insufficiency of computer at families with more children and connected might cause problems. Blended learning provides an advantage for children with a slow learning pace in cognitive sense." (P1).*
>
> *"They are happy as they carry school setting to home environment. They would miss the school when they are at home. It was hard for students to come to school every day. They felt easy in the distance part. Their wishes to go to school and attendance increased." P13.*
>
> *"Children play with the computer mostly and they do not listen what is taught. I tell the course on the book, but they would not listen to me much. I think they find it boring. I am talking about the distance education part. I do my course normally in face-to-face. Other days look like going for nothing." (C15).*

The fourth question of the research was "How do preschool and primary education teachers practice blended learning model?". The answers given to the interview questions were gathered under the theme of "the ways of practicing blended learning model by teachers". The sub-themes and frequencies regarding this theme are given in Table 5.

Table 5. The ways of practicing blended learning model by teachers

Main Theme	Sub-Themes	
	Primary	f
	I planned difficult courses (Turkish, mathematics, science) as face-to-face and easy courses (life sciences, physical education, art etc.) as distance education	5
	I did the teaching of the course as distance and its exercises and application as face-to-face	4
	I tried to take family support	4
	I did not make any different application	3
	I tried to do what I did at normal education as methods and techniques in distance education as well.	3
	I made use of visuals and videos more in distance education	3
	I used more digital application	2
	I made use of Power Point presentations	1
The Ways of Practicing Blended Learning Model by Teachers	Preschool	
	Family support is of quite importance	8
	Group activities could be made face-to-face while individual ones could be made in distance education	7
	I did concept teaching face-to-face, but application in distance education	6
	I informed the parents about the materials to be used one day before.	5
	Language and science activities could be done in distance education, but experiments, trip, and observation in face-to-face.	4
	I did not make any change in methods and techniques	3

As shown in Table 5, the application methods of blended learning by preschool and primary education teachers display both differences and similarities. Primary school teachers preferred distance teaching with exercises and face-to-face application, while primary school teachers favoured face-to-face concept teaching and distance application. Both primary and preschool education teachers indicated a preference for teaching subjects like Turkish language, mathematics, and science through distance education, but opted for face-to-face instruction in courses like life sciences, physical education, and art. Family support was considered crucial by both groups for blended learning success. Preschool teachers also emphasized informing parents about preparation before distance education days. Additionally, some teachers in both groups maintained their teaching methods and techniques from regular education during the distance education period, without making changes. Below are some participant responses related to "the ways of implementing the blended learning model by teachers" theme:

> *"… I tried to practice what I did in normal education in distance education as well. Perhaps, we were not able to learn at first how fast we carry on the course. However, we experienced which topic should be taught how long in time and made the program flexible" (C1).*
>
> *"I planned it as difficult courses in face-to-face education and easy courses in distance education. I made more use of visuals and videos in the distance education part of blended learning" (C3).*
>
> *"I would opt for methods and techniques that lend themselves well to distance education, like drama, material design, and painting, which can also be individually executed for planning. In terms of face-to-face teaching, I would select group games that are enjoyable when collaboratively undertaken. For me, executing group work in distance education might be more challenging, so I would consider group activities as classroom-based and individual tasks as suited for distance education" (P17).*
>
> *"I was supported by the parents. I had challenges at first in the adaptation to distance education. I improved myself but students had hard times in distance education. By getting the support of the parents, we carried on the process by training them. Then, we had a fruitful period. I would pay attention to teach practical courses at school but the theoretical ones as distance education. I used the methods that had used in the class in blended learning as well with the support of technology" (C25).*

The fifth question of the research was "What kind of challenges did teachers encounter while practicing blended learning model?". The answers given to the interview questions were gathered under the theme of "the challenges teachers encountered in blended learning". The sub-themes and frequencies regarding this theme are given in Table 6.

Table 6. The challenges teacher's encountered in blended learning

Main theme	Sub-themes	f
The Challenges Teachers Encountered in Blended Learning	Lack of technological devices for students	14
	Being insufficient by the teachers regarding technology use	9
	Problems experienced in classroom management	4
	Students' inability in motivation	4
	Challenges of students in adaptation to school (absenteeism)	3
	Discomfort felt in the reveal of teacher's privacy during the course	2
	Teacher's young child at home	1
	Long teaching hours	1
	Teacher's lack of motivation in courses	1
	Parental indifference	1
	Young students' not knowing what to do in full sense	1

As given in Table 6, the mostly encountered problem in the question where the problems experienced in the application process of blended learning by preschool and primary education teachers and the ones observed by their colleagues was that students did not have a computer to join the course or cuts in the internet connection that would prevent the flow of courses. Besides that, some teachers indicated that their computers were rather old-fashioned and it was hard for them to renew them or as they had children having distance courses at home, they had problems in sharing computers. Another problem that basic education teachers frequently encountered in the application process of blended learning was that they felt themselves insufficient in terms of technological literacy in the distance education part of blended learning. In particular, those who were near to retirement indicated that they had challenges both in using technology in the distance education part of blended learning and in double planning (face-to-face + distance= of this process.

Teachers working particularly in the preschool and first stages of primary education in the distance part of distance education of blended learning pointed out that they had problems in terms of classroom management and student motivations. They also expressed that they had worries about that parent were always listening the course they were teaching and that they couldn't obtain classroom management since they couldn't make harsh warnings to the students in terms of discipline. Some participants indicated that their privacy during the courses in blended learning was harmed and they were uncomfortable about that. Another problem experienced in this process was that the students could not adapt to the school as they attended to both school and distance education. Some of the answers given by the participants regarding the "the challenges teachers encountered in blended learning" theme are given below:

> *"Preschool children are young age children and they think everything in a concrete way, they cannot think in an abstract way" (P2).*
> *"… They could misunderstand the words and behaviours of the teacher. Families are sometimes lack of technological devices, particularly large families. As for distance education, it is hard to motivate children for the courses" (C5).*
> *"I am not so familiar with technology and my computer is old-fashioned. Brand-new computers are very expensive. I am not so good at it even though I buy it. It is very difficult to manage students in distance education courses" (C15).*

The sixth question of the research was "Do preschool and primary education teachers need education on blended learning, if they do, what must be the content of this education?". The answers given to the interview questions were gathered under the theme of "educations to be given to the teachers". The sub-themes and frequencies regarding this theme are given in Table 7.

Table 7. The sub-themes and frequencies regarding the theme of educations to be given to the teachers

Educations to Be Given to the Teachers	Content development	8
	Teaching methods and techniques	8
	Parent education	5
	Classroom management	5
	Use of Web 2.0 tools	5
	Family Involvement	4
	Communication	3

Twenty-six of the participants indicated that they wanted to have an education but three did not. Depending on this finding, it is likely to say that a great majority of the participants was willing to have an education on blended learning model and, they needed this education. The answers given to the question of "which items should be included in the education to be given?" by the participants were gathered under the theme of "the educations to be given to the teachers". The sub-themes and frequencies regarding this theme are given in Table 7.

Upon the revision of the education that the participants wanted to have regarding blended learning model, it was found that their technical information and skills in hand were not available for this model, and they needed to renew their technical information and skills. In addition, it is likely to say that the participants were aware of this case, and they expressed it. Some of the answers given by the participants regarding "educations to be given to the teachers" are given below:

> *"... I would like to learn blended learning better, since it is told that future education will in this way" (C9).*
>
> *"Yes, I would. Teachers need this education. Subjects, planning the courses, using technology" (C2).*
>
> *"Yes, I would. Teachers must have the education. Parents rather than teachers must have that education. It is because there are a great many parents who do not think positively over the distance part of blended learning. Teachers could be given education about technology use, managing the process, teaching methods, course planning, and access to children, and parent education." (P12).*

Discussion, Conclusion and Implications

With the analysis of the findings obtained in the research, it was found that teachers were confused about how to practice blended learning model, in other words, they did not know what to do exactly in this process. As an example, some teachers developed peculiar, blended learning models taking the age and grades of students, while some others did not do anything in this sense. It is thought that teachers needed education, particularly in the part of distance education of blended learning. It is because there is a need to knowledge of suitable planning (Haylock & Cockburn, 2014) and pedagogical methods and theoretical background (Van de Walle et al., 2010), and adoption of course and teaching design principles (National Council of Teachers of Mathematics [NCTM], 2000). However, because of urgency in the pandemic process, institutions were not able to find opportunities to plan this transition systematically to make an adaptation to new teaching and learning applications. As a matter of fact, what is ideal in a transition or a changing process is to invest in the precautions that would decrease challenges against professional development opportunities of institutions, research and data collection, capacity development and changing attempts (Rad et al., 2021). Upon the revision of the related literature, it is likely to see

that these challenges are not only experienced in basic education stages but also in higher education. In a study by Singh et al., (2021), it was found that Covid-19 pandemic process had significant challenges upon lecturers, students and administrator and it affected almost all the academic staff; technological fear, limited software knowledge, time management problems and feeling of loneliness created challenges for the lecturers at universities.

Upon the analysis of teacher views, it was found that family support is needed more to be able to apply blended learning at all stages of basic education successfully. For that reason, it is believed that parents must be given education to inform about blended learning. In their study, Barnett and Jung (2021) indicated that parents were not able to make necessary support to the educational process of their children, their efforts (e.g., the number of books they read) regarding the support they made for the personal development of their children decreased, and it led to possible learning losses in pandemic period.

Depending on the findings of study, it is likely to say that teachers try to increase their developments in technical fields such as technology literacy and producing digital content thanks to blended learning. In addition, providing with the opportunity of saving time and acting more flexibly in running the courses gives advantages to both teachers and students. On the other hand, students get accustomed to using technology for educational purposes and they have chances to learn from more different sources. This case that was determined in the study supports various studies in the literature. Studies show that blended learning increased the success of students (Gürdoğan & Bağ, 2019; Junco et al., 2011; Korkmaz & Kadirhan, 2020; Rovai & Jordan, 2004; Uluyol & Karadeniz, 2009; Ünsal, 2010; Usta & Mahiroğlu, 2008; Yılmaz, 2018), supported active involvement of students to learning processes (Junco et al., 2011; Yılmaz, 2018), provided easy access to information and enriched learning experience to students (Singh & Reed, 2001; Stein & Graham, 2014; Yıldırım & Vural, 2016), became more economic (Singh & Reed, 2001; Stein & Graham, 2014)), and was useful for the prospective teachers having difficulty in expressing themselves verbally (Alhan, 2019). It was determined in a study carried out with prospective teachers by Hiğde and Aktamış (2021) that participants found face-to-face learning settings more effective compared to online setting, and they also thought that having the chances of watching videos again and of saving time were the positive sides of distance education. Gürdoğan and Bağ (2019) pointed out that students found blended learning applications fun, it provided permanency in what they learned, varied learning and gave a chance to watch the videos again.

One of the positive experiences of teachers regarding blended learning was that some introverted and shy students mostly overcame these challenges in the distance education courses, in other words, they their social sides gained strength when they came to face-to-face education. This finding is parallel with the study by Xie et al. (2020). In this study, because online education was designed to focus on the educational demands of individual students by nature, it could be advantageous for the students who are socially weak and in cooperation. Singh et al. (2021) carried out research to determine the facilities, challenges, strong and weak sides of education software platforms for students. In this study, it was determined that self-sufficiency, self-awareness, and self-esteem of the student were affected positively thanks to blended learning experiences.

One of the results obtained through the analysis of the findings of the current study was that insufficiency of the infrastructure regarding the provision of technological equipment and the internet was one of the most important reasons making blended learning difficult. Both students and teachers have challenges in this sense. This case causes an important problem in providing

social justice in education (Cho, 2021). Xie et al. (2020) indicated that the insufficiency of technology and infrastructure with the compulsory process of distance education deepened inequalities in education more, and particularly underdeveloped countries were influenced badly in this fact. Similarly, Barnett and Jung (2021) found in a study carried out to investigate the effects of pandemics on preschool education in the USA that 51% of the children at the age of 3 and 71% of those at the age of 4 participated in preschool education before pandemic, but a 25% dropped was seen in the pandemic process including students participating in distance education. In addition, there became sharp decreases in the general participation and particularly face-to-face participation of the children in poverty (known as the families have an annual income under 25.000 dollars). It was found that only 13% of the children in the families in this economic level were able to participate in preschool education.

Carrying the office home and turning home into school in terms of students is one of the important reasons making blended learning more difficult to conduct. Teachers expressed that they found it hard to motivate both themselves and students in the courses while practicing blended learning. It was found that teachers had challenges in classroom management, adapting students to the course and producing contents. This result is parallel to some studies in the literature. As an example, Rasheed et al. (2020) and Wut and Xu (2021) indicated in their studies there were some challenges regarding using technology true and effectively in teaching processes, while Burke and Dempsey (2020) with Sarıdaş and Topdağı (2021) emphasized that teachers must be supported to overcome the insufficiencies in digital literacy. Similarly, according to Xie et al. (2020), students could easily be distracted by social chats, news, or games in online education. Some students cannot complete their school activities because there is no teacher to encourage and remind them their tasks and overcome their lack of motivation. The case of home and expectation cannot be taken into consideration during blended learning. It is because students can be in trouble since they have to look after family member (siblings, elderly, or patients).

It is likely to say that teachers agreed on the fact that blended learning is not suitable in the distance education part, particularly at the first and second grades of primary education with the preschool stage. Similar findings were found in a study by Cao et al. (2021) in a study carried out into the effects of teachers in China on the mathematic teaching in online teaching process during Covid-19 pandemic process. In this study, it was determined that distance education made mathematics teaching difficult particularly for the lower grades. In online teaching, as each student participated in the courses with their own devices, teachers had challenges in grouping students and arranging classroom activities, and they pointed out that distance education was not much suitable for young age group. The authors of the well-known book "Understanding Mathematics for Young Children", Haylock & Cockburn (2014) indicate that some hard to teach subjects could be taught easily to the students in this age group by using classroom activities, discussions, and some methods comprising peer cooperation.

It is likely to say that parents who cannot take necessary care of the education of their children is another factor making blended learning harder. In their study, Barnett and Jung (2021) found that parents from USA were less contented from distance education in preschool period, teachers were bored of the demands of supporting distance education, and childcare and education process affected work life negatively. This limitation in young age groups comes to the forefront as insufficiencies of technological literacy for the students in older age groups. According to Lake and Makori (2020), even though most students regard themselves as "digital natives" today, a great many of them, particularly young students, still need supports to overcome the problems of becoming online, and of technical troubles as well.

Effects regarding theory and application

The study provides comprehensive insights into the theory and application of blended learning. Initially, it explores the blended learning model, which is perceived as restructuring the fundamental operations of the educational system within schools. It emphasizes the importance of considering students' developmental levels, particularly in early education settings such as preschool and primary education, when implementing the blended learning approach. Moreover, the study highlights the crucial role of school administrators in effectively managing various aspects of education management, including planning, coordination, organization, communication, and supervision, within the context of blended learning.

During the Covid pandemic, the adoption of the blended learning model led to significant challenges for teachers as physical classrooms transitioned to virtual settings. Teachers' technical competencies developed in traditional classrooms may not be as effective in virtual environments. Consequently, there is a growing necessity for teachers to enhance their technical skills, including planning virtual courses, creating activities and content, utilizing technology, managing virtual classrooms, and facilitating communication.

While blended learning offers numerous advantages for educators, students, and parents, it also presents certain drawbacks. The study aims to identify and evaluate both the benefits and limitations associated with blended learning.

Additionally, the study highlights that while blended learning offers flexibility, models prioritizing face-to-face instruction may be more suitable for younger learners, particularly in early education settings. The findings emphasize that the success of blended learning relies not only on the structure of the model but also on the teacher's ability to effectively integrate technology and adapt to both virtual and physical classroom environments. This adaptability is crucial for addressing the unique developmental needs of preschool and primary school students, ensuring that educational methods remain effective and engaging across different learning contexts.

Limitations and research recommendations

The current study is a qualitative study carried out with a small participant group made up of teachers working in preschool and primary education in Ankara in Türkiye. For that reason, the results cannot be generalized with larger samplings. It is recommended that future studies should be carried out with the support of quantitative research methods having larger samplings. Samplings from the cities in different sizes could be chosen. The results of the studies investigating how blended learning is carried out in the student groups with older ages as in secondary education and university could be compared with the results of the current study.

The contents of the courses given to prospective teachers in the faculties of education could be rearranged in a way to increase the competencies in distance and blended learning applications. Policy makers could be recommended to carry out studies regarding how to apply blended learning at schools by taking the results of the current study into consideration. It is also recommended that teachers teaching in preschool and first and second grades of primary education should teach in distance education for one day while applying blended learning, and those teaching in the second and third grades should teach in distance education for two days.

Acknowledgements

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Ethics

The ethics application for the study was made on 07/05/2021 and the research was carried out with the approval of Haci Bektas Veli University Ethics Commission dated 14/06/2021 and numbered 2021.07.221.

References

Alhan, S. S. (2019). Pre-Service science teachers' views towards blended learning environment. *Erzincan University Faculty of Education Journal, 22*(2), 397-414. https://doi.org/10.17556/erziefd.633189

Anadolu Agency-AA (2022a). *Tüm eğitim öğretim kurumlarında uzaktan eğitime geçiş* Republic of Türkiye Anadolu Agency. https://www.aa.com.tr/tr/egitim/tum-egitim-ogretim-kurumlarinda-uzaktan-egitime-gecis-29-nisanda-baslayacak/2221506

Anadolu Agency-AA (2022b). *Tüm eğitim öğretim kurumlarında uzaktan eğitime geçiş 29 Nisan'da başlayacak*. Republic of Türkiye Anadolu Agency. https://www.aa.com.tr/tr/egitim/tum-egitim-ogretim-kurumlarinda-uzaktan-egitime-gecis-29-nisanda-baslayacak/2221506

Barnett, W. S., & Jung, K. (2021). Seven impacts of the pandemic on young children and their parents: initial findings from NIEER's December 2020. Preschool Learning Activities Survey. Research Report. National Institute for Early Education Research.

Burke, J., & Dempsey, M. (2020). *Covid-19 practice in primary schools in Ireland report*. Maynooth.

Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2018). *Eğitimde bilimsel araştırma yöntemleri.* Pegem.

Çakıt, İ. & Karadeniz, Ş. (2020). Harmanlanmış öğrenme ortamlarının, hentbolda temel becerilerin gelişimine etkisi. *Canakkale 18 March University Sport Sciences Journal, 3*(3), 34-52. https://dergipark.org.tr/tr/pub/comusbd/issue/59247/819808

Cao, Y., Zhang, S., Chan, M. C. E., & Kang, Y. (2021). Post-pandemic reflections: lessons from Chinese mathematics teachers about online mathematics instruction. *Asia Pacific Education Review, 22*(2), 157-168. https://doi.org/10.1007/s12564-021-09694-w

Cho, H. (2021). Carving out a hybrid space: a self-study of contextualizing teaching for social justice in South Korea. *Asia Pacific Education Review, 22*(3), 495-513. https://doi.org/10.1007/s12564-021-09683-z

Cresswell, J. W. (2014). *Research design: qualitative, quantitive and mixed method approaches*. Sage.

Dağ, F. (2011). Harmanlanmış karma öğrenme ortamları ve tasarımına ilişkin öneriler. *Journal of Ahi Evran University Kırsehir Education Faculty, 12*(2), 73-97. https://dergipark.org.tr/en/pub/kefad/issue/59495/855161

Diana, E., Rahmah, N., & Rofiki, M. (2022). Blended learning management: The efforts to develop students' soft skills in the new normal era. *Jurnal Basicedu, 6*(3), 4272-4281. https://doi.org/10.31004/basicedu.v6i3.2835

Dikmen, C. H. (2021). *Investigation of the social, cognitive and teaching presence of teachers participating in programming education in the blended learning environment* [Doctoral dissertation, Gazi University]. YÖK Tez Merkezi.

Dziuban, C., Graham, C. R., Moskal, P. D., Norberg, A., & Sicilia, N. (2018). Blended learning: The new normal and emerging technologies. *International Journal of Educational Technology in Higher Education, 15*(1), 3. https://doi.org/10.1186/s41239-017-0087-5

Erdoğan, İ. (2021). Yeniden yeni eğitim sistemi. *Journal of Insan ve Insan, 8*(27), 13-25. https://doi.org/10.29224/insanveinsan.814514

Finlay, M. J., Tinnion, D. J., & Simpson, T. (2022). A virtual versus blended learning approach to higher education during the COVID-19 pandemic: The experiences of a sport and exercise science student cohort. *Journal of Hospitality, Leisure, Sport & Tourism Education, 30,* 100363. https://doi.org/10.1016/j.jhlste.2021.100363

Gilroy, K. (2001). Collaborative e-learning: The right approach. *Ars Digita Systems Journal, 1*-10. http://www.eveandersson.com/arsdigita/asj/elearning/

Grupe, F. H. (2002). An internet-based expert system for selecting an academic major. *Internet and Higher Education, 5*(4), 333-344. https://doi.org/10.1016/S1096-7516(02)00129-X

Gürdoğan, M. & Bağ, H. (2019). Harmanlanmış uygulamaların akademik başarı, motivasyon ve öğrenci görüşlerine etkisi. *Akdeniz University Faculty of Education Journal, 2*(1), 36-61. https://dergipark.org.tr/tr/pub/akuned/issue/48807/570309

Hastie, P. A., & Curtner-Smith, M., (2006). Influence of a hybrid sport education–teaching games for understanding unit on one teacher and his students. *Physical Education and Sport Pedagogy 11*(1), 1-27. https://doi.org/10.1080/17408980500466813

Haylock, D., & Cockburn, A. (2014). *Understanding mathematics for young children: A guide for teachers of children 3-8*. Sage.

Hiğde, E. & Aktamış, H. (2021). Probleme dayalı harmanlanmış öğrenme ortamının etkililiğinin ve öğrencilerin tutumlarının incelenmesi. *Journal of Manisa Celal Bayar University Education Faculty, 9*(1), 81-103. https://doi.org/10.52826/mcbuefd.884752

Jumabaeva, C., Sait Kyzy, A., Baryktabasov, K., & Ismailova, R. (2020). The hybrid learning implementation in Kyrgyzstan. *Science, Education, Art and Technology Journal (SEAT Journal), 4*(1), 23-30. https://dergipark.org.tr/en/pub/bestdergi/issue/50074/600354

Junco, R., Heiberger, G. & Loken, E. (2011). The effect of Twitter on college student engagement and grades. *Journal of Computer Assisted Learning, 27*(2), 119-132. https://doi.org/10.1111/j.1365-2729.2010.00387.x

Korkmaz, Ö. & Kadirhan, M. (2020). EBA içerikleriyle harmanlanmış öğretim uygulamasının öğrencilerin fen bilimleri dersindeki akademik başarılarına ve tutumlarına etkisi. *Trakya Journal of Education, 10*(1), 64-75. https://doi.org/10.24315/tred.529721

Korkmaz, O., & Karakus, U. (2009). The impact of blended learning model on student attitudes towards geography course and their critical thinking dispositions and levels. *Turkish Online Journal of Educational Technology-TOJET, 8*(4), 51-63.

Lago M. E. (2000), The hybrid experience: How sweet it is!. *Converge Magazine, 3*(9), 5-7.

Lake, R. & Makori, A. (2020). The digital divide among students during COVID-19: Who has access? Who doesn't? The Lens. https://www.crpe.org/thelens/digital-divide-among-students-during-covid-19-who-has-access-who-doesnt

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: an expanded source book* (2nd Edition). Sage.

Ministry of National Education-MoNE (2022a). *Koronavirüs'e karşı eğitim alanında alınan tedbirler*. Republic of Türkiye Ministry of National Education. https://www.meb.gov.tr/bakan-selcuk-koronaviruse-karsi-egitim-alaninda-alinan-tedbirleri-acikladi/haber/20497/tr

Ministry of National Education-MoNE (2022b). *Türkiye, koronavirüs salgınında ulusal çapta uzaktan eğitim veren iki ülkeden biri*. Republic of Türkiye Ministry of National Education. https://www.meb.gov.tr/turkiye-koronavirus-salgininda-ulusal-capta-uzaktan-egitim-veren-2-ulkeden-biri/haber/20618/tr

Ministry of National Education-MoNE (2022c). *Yüz yüze eğitime ara verilmesi*. Republic of Türkiye Ministry of National Education. http://covid19.meb.gov.tr/meb_iys_dosyalar/2021_05/04130426_yuz-yuze-egitime-araverilmesi.pdf

Ministry of National Education-MoNE (2022d). *Uzaktan eğitim kararı*. Republic of Türkiye Ministry of National Education. http://covid19.meb.gov.tr/covid19.html?-cat=rapor

Ministry of National Education-MoNE (2022e). 2020-2021 eğitim öğretim yılının ikinci dönemi uzaktan ve yüz yüze eğitimle başlıyor. Republic of Türkiye Ministry of National Education, Anadolu Agency. https://www.aa.com.tr/tr/egitim/okullarda-2020-2021-egitim-ogretim-donemi-yarin-sona-eriyor/2290792

Ministry of National Health, (2022). Dünya Sağlık Örgütü'nün pandemi ilanı. Republic of Türkiye Ministry of National Health. https://covid19.saglik.gov.tr/TR66494/

National Council of Teachers of Mathematics-NCTM (2000). *Principles and standards for school mathematics*. Reston.

Osguthorpe, R. T. & Graham, C. R. (2003). Blended learning environments. *Quarterly Review of Distance Education, 4*(3), 227-233. https://www.learntechlib.org/p/97576/

Porter, W. W., Graham, C. R., Spring, K. A., & Welch, K. R. (2014). Blended learning in higher education: Institutional adoption and implementation. *Computers & Education, 75*, 185-195. https://doi.org/10.1016/j.compedu.2014.02.011

Prifti, R. (2022). Self–efficacy and student satisfaction in the context of blended learning courses. *Open Learning: The Journal of Open, Distance and e-Learning, 37*(2), 111-125. https://doi.org/10.1080/02680513.2020.1755642

Rad, F. A., Otaki, F., Baqain, Z., Zary, N., & Al-Halabi, M. (2021). Rapid transition to distance learning due to COVID-19: Perceptions of postgraduate dental learners and instructors. *PLoS ONE, 16*(2). https://doi.org/10.1371/journal.pone.0246584

Rasheed, R. A., Kamsin, A., & Abdullah, N. A. (2020). Challenges in the online component of blended learning: A systematic review. *Computers & Education, 144*. https://doi.org/10.1016/j.compedu.2019.103701

Rovai, A.P. & Jordan, H.M. (2004). Blended learning and sense of community: a comparative analysis with traditional and fully online graduate courses. *The International Review of Research in Open and Distance Learning, 5*(2), 1-13. https://doi.org/10.19173/irrodl.v5i2.192

Sarıdaş, G. & Topdağı, M. (2021). Öğretmenlerin mesleki gelişim ihtiyaçlarının pandemi sonrası harmanlanmış öğrenme bağlamında incelenmesi. *Journal of Educational Administration and Policies, 2*(2), 25-38. https://dergipark.org.tr/tr/pub/eypd/issue/67243/949603

Singh, H., & Reed, C. (2001). A white paper: achieving success with blended learning. *Centra Software, 1*, 1-11. http://www.leerbeleving.nl/wbts/wbt2014/blend-ce.pdf

Singh, J., Steele, K., & Singh, L. (2021). Combining the best of online and face-to-face learning: hybrid and blended learning approach for COVID-19, Post Vaccine, & Post-Pandemic World. *Journal of Educational Technology Systems, 50*(2), 140–171. https://doi.org/10.1177/00472395211047865

Sloman, M. (2003). *Training in the age of the learner. Chartered ınstitute of personnel and development.* CIPD.

Stein, J., & Graham, C. R. (2014). *Essentials for blended learning a standards-based guide.* New York: Routledge.

Türker, M. S. (2021). An investigation of the opinions of teachers who teach Turkish as a second/foreign language about blended learning environments. *Uluslararası Türkçe Edebiyat Kültür Eğitim Dergisi, 10*(3), 1049-1069. https://doi.org/10.52826/teke.1000807

Turkish Statistical Institute-TURKSTAT (2022). *Öğretim yılına göre okul öncesinde net okullaşma oranı, okul, öğrenci, öğretmen ve derslik sayısı.* Republic of Türkiye Statistical Institute. https://data.tuik.gov.tr/Kategori/GetKategori?p=egitim-kul-tur-spor-ve-turizm-105&dil=1

Uluyol, A. G. Ç., & Karadeniz, Ş. (2009). Bir harmanlanmış öğrenme ortamı örneği, öğrenci başarısı ve görüşleri. *Van Yüzüncü Yıl University Journal of the Faculty of Education, 6*(1), 60-84. https://dergipark.org.tr/en/pub/yyuefd/issue/13711/165995

United Nations Educational Scientific and Cultural Organization-UNESCO (2022). *Covid-19 impact on education.* UNESCO. https://en.unesco.org/covid19/educationresponse

Ünsal, H. (2010). Yeni bir öğrenme yaklaşımı: harmanlanmış öğrenme [A new learning approach: blended learning]. *Journal of National Education, 40*(185), 130-137. https://dergipark.org.tr/tr/pub/milliegitim/issue/36199/407092

Usta, E. & Mahiroğlu, A. (2008). Harmanlanmış öğrenme ve çevrimiçi öğrenme ortamlarının akademik başarı ve doyuma etkisi. J*ournal of Ahi Evran University Kirsehir Education Faculty, 9*(2), 1-15. https://dergipark.org.tr/en/pub/kefad/issue/59525/856022

Van de Walle, J., Karp, K. S., & Bay-Williams, J. M. (2010). *Elementary and middle school mathematics.* Pearson.

Wut, T. M., & Xu, J. (2021). Person-to-person interactions in online classroom settings under the impact of Covid-19: a social presence theory perspective. *Asia Pacific Education Review, 22*(3), 371-383. https://doi.org/10.1007/s12564-021-09673-1

Xie, X., Siau, K., & Nah, F. F. H. (2020). Covid-19 pandemic-online education in the new normal and the next normal. Journal of Information *Technology Case and Application Research 22*(3), 175–187. https://doi.org/10.1080/15228053.2020.1824884

Yıldırım, A., & Şimşek, H. (2016). *Sosyal bilimlerde nitel araştırma yöntemleri*. Seçkin.

Yıldırım, İ., & Vural, Ö. F. (2016). Matematik öğretimine entegre edilmiş harmanlanmış öğrenme süreci hakkındaki öğrenci görüşleri. *Journal of Ahi Evran University Kırsehir Education Faculty (KEFAD), 17*(2), 1-15. https://dergipark.org.tr/en/pub/kefad/issue/59426/853554

Yılmaz, Ö. (2018). Fen öğretiminde harmanlanmış öğrenme: avantajlar ve alışkanlıklar. *Hitit University Journal of Social Sciences Institute, 11*(3), 2111-2121. https://doi.org/10.17218/hititsosbil.439414

International Journal of Educational Studies and Policy (IJESP)

Volume: 5, Issue: 2, November 2024

Investigation of Pre-Service Teachers' Statistical Literacy Levels

Muhammed Celal Uras[1], Mehmet Şata[2], Yasin Soylu[3]

ABSTRACT

Statistical literacy is one of the most important actors in the information age. There has been an increase in the importance of education. Therefore, teachers' mastery of statistical concepts is necessary for quality education. The present study aims to examine the statistical literacy levels of pre-service elementary mathematics teachers. The study is a quantitative research with ex post facto design. The sample consists of 530 pre-service teachers in two universities. "Statistical Literacy Scale" was used for data collection. For the validity analysis of the scale, a first level confirmatory factor analysis was carried out and the Cronbach alpha coefficient was calculated for reliability. The data were analyzed using independent samples t-test and one-way ANOVA. The statistical literacy levels of pre-service teachers increase with the statistics and probability courses they take during their undergraduate education. It was also concluded that there was a significant difference between the students who took the courses and those who did not. It is recommended to conduct new studies for curriculum development and educational policies by revealing the factors affecting the development of statistical literacy.

Keywords: Statistics literacy, pre-service teachers, statistics and probability, mathematics

DOI: https://doi.org/10.5281/zenodo.14016307

Article Info:

Received: 29.05.2024 **Accepted: 27.10.2024** **Article Type: Research Article**

Cite as: Uras, M. C., Şata. M. & Soylu. Y. (2024). Investigation of pre-service teachers' statistical literacy levels. *International Journal of Educational Studies and Policy, 5*(2), 174-185.

[1]Corresponding Author: Dr. Muhammed Celal Uras, Ağrı İbrahim Çeçen University, mcuras@agri.edu.tr, ORCID: 0000-0003-3994-8723

[2] Assoc. Prof. Mehmet Şata, Van Yüzüncü Yıl University, mehmetsata@yyu.edu.tr, ORCID: 0000-0003-2683-4997

[3] Prof. Dr. Yasin Soylu, Atatürk University, yasinsoylu@atauni.edu.tr, ORCID: 0000-0003-0906-4994.

Introduction

Education has a vital role in the development of societies (Balı & Dönmez, 2018). Therefore, teachers' competencies and skills are a determining factor in the effectiveness of education (Wu et al., 2022). Statistical literacy (SL) is gaining more importance in education. In addition, SL has become an indispensable requirement in today's information age. It is also important that teachers will use both in their professional development and teaching (Sharma, 2017). Therefore, determining and improving the SL levels of pre-service teachers (PST) is important in improving the quality of education. The studies conducted in this direction have a significant place in the literature. The present study examines the SL levels of PST and the gap in this field.

Literature Review

In today's information age, individuals need literacy skills for personal development and social participation. Literacy requires mastery in various areas of knowledge. Within the education system, teachers must have literacy in various fields to ensure that future generations grow up as individuals who can make informed decisions (Mhlongo et al., 2023). In this context, SL is becoming increasingly important at all levels of education.

Literacy is one of the most basic needs of modern societies. Beyond the ability to read and write texts, literacy now refers to the ability to critically evaluate and use information effectively. Today, the ability to access and interpret information accurately plays a critical role in determining individuals' personal and professional development (Pilgrim & Martinez, 2013). Literate individuals can solve problems more effectively, make more informed decisions and contribute to society. Literacy supports individuals' lifelong learning and is an essential requirement for general well-being. Therefore, individuals need to acquire literacy skills. This can be realized through education.

Literacy is needed in many different areas of society. For example, it manifests in various fields such as finance, media and digital technologies, information management, and mathematics. Financial literacy is concerned with the ability to manage finances (Goyal & Kumar, 2021). Digital literacy, which is indispensable today, includes the ability to interact with digital tools and understand media content (Kaya et al., 2024; Nicholson, 2017). Information literacy is related to the skills of accessing, sharing, evaluating, and using information (Walsh, 2011). Mathematical literacy requires performing basic mathematical operations and recognizing symbols and terms (Holenstein et al., 2021). Different types of literacy promote individuals' involvement in social life and enhance their overall well-being. One of the types of literacy is SL. It is essential in today's data-driven world. It basically involves understanding and interpreting statistical data (Sharma, 2017). Many situations and events encountered both in professional life and in the public domain are based on data. For example, statistical methods are used to evaluate the effectiveness of constructive policies, measure the performance of employees, and evaluate education. Therefore, to make rational and efficient decisions, it is necessary to make inferences by mastering the data obtained. SL is necessary and essential for everyone involved in social life. Moreover, SL is needed to calculate probabilities when shopping individually or when planning a career. SL is also becoming increasingly important today due to its use in many areas of social life. It involves understanding and using statistics' basic concepts and language (Garfield, 2011). SL also includes being able to recognize and interpret different representations of data. Moreover, its scope is not limited to these but goes beyond the basic skills. SL includes the ability to interpret and critique

real-world data, such as graphs presented in the media on climate change. It also develops the ability to understand and evaluate, and cultural contexts of data's social, environmental, and cultural contexts. Thus, it strengthens individuals' statistical thinking skills, enabling them to solve real-life problems more effectively. There are many different models related to SL in the literature as it concerns society in general. Gal (2002) proposed a SL model that includes knowledge and attitude or disposition components and emphasized the importance of making inferences after understanding and interpretation. For this, mathematics, statistics, and literacy skills are employed, and dispositions are effective in this process. Watson and Callingham (2003) constructed the structure of SL in a hierarchical way. This structure is from personal and informal views (the lowest) to the stage that has grounds and requires critical thinking (the highest). Sharma et al. (2011) presented a four-stage framework for diagnosing students' thinking. The aim of this framework, which is similar to the Watson and Callingham (2003) model, is to provide teachers with a tool that can be used to construct and assess students' SL constructs. Finally, the GAISE framework defined three statistical stages (levels A, B, and C) through which students' progress to develop their statistical understanding in order to determine the level of development of K-12 students (Carver et al., 2016). In these models and frameworks, SL levels were determined according to individuals' knowledge. In addition, the importance of the components of SL was emphasized. SL has an essential place in education, especially in the case of Sharma et al. (2011) and Carver et al. (2016), who focus on students' statistical skills.

SL aims to develop individuals' statistical thinking skills. Therefore, teachers and PST must help students develop their SL skills. In this way, educational practices become more effective, students' statistical thinking skills increase and their academic achievement and social contributions are strengthened (Darling-Hammond et al., 2020). Therefore, for teachers, SL is of critical importance in education. Teachers need to have SL enabling students to think and make inferences based on data. SL empowers teachers to make informed and effective decisions not only in their own field but also in education. It also allows teachers to monitor student achievement, evaluate educational programs, and develop strategies that are appropriate to student needs (Ridgway et al., 2011). Moreover, SL is becoming increasingly important in education because today's decision-making processes require making inferences from data. Teachers are responsible for equipping their students with the skills that they will need in society. Therefore, teachers must continuously improve their SL competencies and apply these skills effectively in and out of the classroom (Ben-Zvi & Makar, 2013). In order to strengthen teachers' knowledge and skills in SL, it is important that they successfully complete the teacher training program and follow the best practices in this field. Also, teacher training programs must emphasize the importance of SL (Guven et al., 2021). In this way, future generations can be better educated and more successful in a data-driven world. Moreover, school administrators and educational institutions must also provide appropriate environments to support teachers' SL skills and offer professional development opportunities in this area.

In summary, developing and promoting SL should become a key priority of education systems. This increases success in individuals' personal and professional lives and contributes to society's general well-being. Furthermore, statistical skills are needed in daily life. Therefore, statistics is an essential integral part of education and training. Statistics also goes beyond the transfer of mathematical knowledge. It encompasses an understanding of statistical concepts and their applications. Mathematics teachers play an important role in developing individuals' SL (Callingham & Watson, 2017). However, teachers experience difficulties because they struggle with statistical content or lack confidence in teaching statistics (Marshman & Dunn, 2024).

Therefore, knowledge and experience in undergraduate education are required for PST of mathematics. Within the scope of probability and statistics courses, basic concepts of probability, probability types, probability function, sampling, data organization and analysis, sampling distribution and estimation, and confidence intervals are included. Topics related to probability and statistics are included in all levels of education from primary school to undergraduate level. PSTs are expected to be statistically literate individuals to teach in their future professional lives.

Research conducted to evaluate the SL levels of PST is considered an important step toward increasing the effectiveness of education. Statistics is developing and, therefore, the focus of numerous research. Many studies are focusing on SL (Badenes-Ribera et al., 2018; De Vetten et al., 2023; Zieffler et al., 2018). In the literature, various tools and methods for measuring PST's SL skills have been examined, and research on the effect of these skills on PST's professional performance has been discussed. In addition, studies on the importance of SL training of PST and the effectiveness of these trainings have a prominent place in the literature. Ozmen and Baki (2019) examined the secondary school mathematics curriculum in the context of SL and revealed that SL is shaped around statistical process components and indicators. Cakmak and Durmus (2015) stated that secondary school students had difficulty learning probability concepts. Regarding this result, it is known that PST has difficulties in providing appropriate feedback based on student mistakes. They also have difficulty solving problems that require going beyond procedures and comparing measures of central tendency (Kazunga et al., 2023). Guven et al. (2021) stated that the SL levels of PST are generally low, which affects the competencies. PST who had and had not taken mathematics courses in their undergraduate education responded similarly to the SL questions. There was no significant difference between them (Forgasz et al., 2024). (Schreiter et al., 2024) reported that PST had low conceptual knowledge in the basic areas of statistics. PST understand statistical data and present them in different representations, but they are inadequate in interpreting and making inferences (De Vetten et al., 2023; Nahdi et al., 2021). In addition, there is a greater focus on the procedural aspects of statistics with significant differences in their knowledge levels. Aydin et al. (2019), who examined the self-efficacy and attitudes of PST toward the statistics course, reported that their self-efficacy beliefs and attitudes toward the statistics course were high and moderate, respectively.

The theoretical framework and empirical studies reveal the importance of SL. Recent studies on PST's SL have focused on variables such as attitude and competence. PST's SL levels are expected to increase with the statistics and probability courses they take during their undergraduate education. The present study aims to investigate the relationship between statistics and probability courses on PST's SL level. The research questions are:

1. Is there a relationship between statistics and probability courses and the SL levels of pre-service elementary mathematics teachers?
2. Is there a significant difference in the SL levels of pre-service elementary school mathematics teachers according to grade level?

Method

Research Design

The present study utilized a descriptive survey, which is one of the quantitative research approaches (Şata, 2020). The model is preferred to explore the relationship between changing conditions and subsequent behaviors after the pre-existing conditions and situations are defined.

Sample

The study's population consists of PSTs studying in an elementary mathematics teaching program in Turkey. The sample consists of 530 PSTs (67.7% female) selected by convenient sampling method. Table 1 shows the demographic variables of the participants.

Table 1. Distribution of participants according to demographic variables

Variables	Categories	N	%
Gender	Female	359	67.7
	Male	171	32.3
Grade	1st grade	122	23.0
	2nd grade	161	30.4
	3rd grade	130	24.5
	4th grade	117	22.1
Take a statistics and probability course*	1	253	47.7
	2	277	52.3
	Total	530	100.0

**1: PST who have taken statistics and probability courses*
2: PST who have not taken statistics and probability courses

Of the sample who continue their education in 2 different universities, 23.0% of them are at the 1st-grade level, 30.4% are at the 2nd-grade level, 24.5% are at the 3rd-grade level, and 22.1% are at the 4th grade level. In addition, the percentage of participants who took and did not take statistics and probability courses is almost equal to each other. The fact that the participants were selected from different universities in Turkey and from different grade levels contributes to the generalizability of the study.

Data Collection

The data for this study was collected through an online survey using Google Forms. All participants took part in the study voluntarily and were informed of the purpose and procedures. Responses were assured of confidentiality.

Data Collection Tools

SL Scale (SLS)

SLS, developed by Sahin (2012), consisting of 17 items, was designed to measure the SL levels of PST. Items (e.g., "There are 30 numbers. The standard deviation of these numbers is found to be zero. Which of the following can you be sure of?") are scored one-point for questions with four options and 0.5 points for questions with two options. The score obtained from the SLS varies between 0 and 16. High scores mean that PST's levels of SL have increased, while low scores mean that their levels of SL have decreased. In the present study, the Cronbach α coefficient calculated for the whole scale was. 601, while the McDonald ω coefficient was determined to be .933. CFA was performed for the validity of the measurements obtained from the measurement

tool, and the fit values were χ^2 /df = 297.82/118 = 2.523, CFI = .981, NNFI = .980, NFI = .965, RMSEA (%90 CI) = .054 (.040 - .050), SRMR = .060.

Data Analysis

Descriptive statistics of the measurements, Cronbach alpha coefficients for the reliability of the measurements, and confirmatory factor analysis were performed to provide evidence for the validity of the measurements. One-way ANOVA was conducted to determine the change in SL levels of PST according to their taking statistics and probability courses. In addition, Levene's test was used to determine whether the variance was equally distributed before the variance analysis. SPSS and LISREL software packages were used for data analysis. In data analysis, the .05 level was taken into consideration for statistical significance.

Ethic

The study was conducted with the permission of Ağrı İbrahim Çeçen University Scientific Research Ethics Committee dated 24.03.2022 and numbered 84.

Results

Descriptive statistics calculated for the measurements obtained from the measurement tools are presented in Table 2.

Table 2. Descriptive statistics of measurements

Variable	Variable level	Skewness		Kurtosis	
		Value	Std. Error	Value	Std. Error
Take a statistics and probability course	1	-0.217	0.153	-0.243	0.305
	2	-0.223	0.146	-0.095	0.292
Grade	1	-0.145	0.219	0.040	0.435
	2	-0.331	0.191	-0.124	0.380
	3	-0.240	0.212	-0.191	0.422
	4	0.147	0.224	-0.382	0.444

Table 2 shows that the skewness and kurtosis values of the measurements according to both the status of taking statistics and probability course and the grade level are within the range of ± 2.00. Accordingly, it was determined that the measurements had a normal distribution (Shiel & Cartwright, 2015).

The confirmatory factor analysis of the SLS was made and the results are presented in Figure 1.

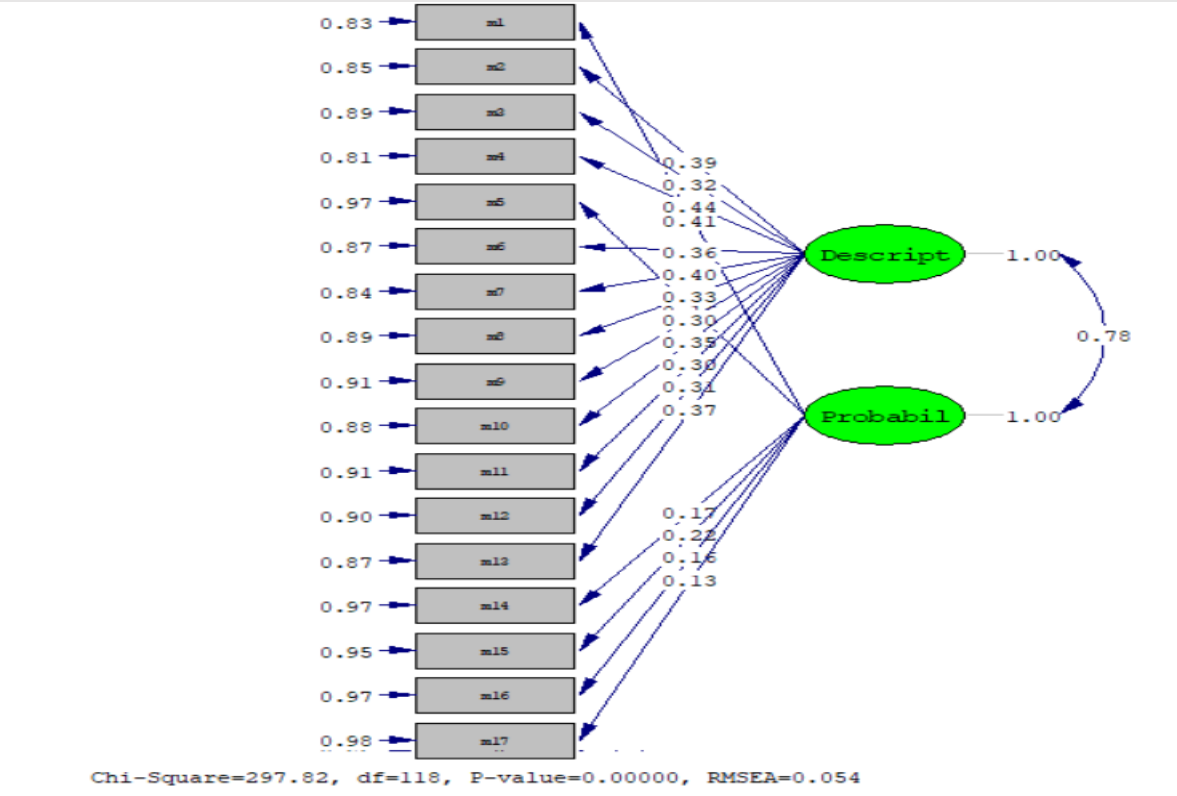


Figure 1. Measurement model defined for SLS first-level CFA results

Figure 1 shows that the factor structure of the scale was confirmed by confirmatory factor analysis. When Figure 1 is examined, it is seen that the factor loadings of some items are low. The analysis included these items because the model-data fit had a high fit value. No items were removed because there was sufficient evidence for construct validity. The results of the analysis conducted to determine the SL levels of PST according to their course-taking status are presented in Table 3.

Table 3. SL levels of PST

Variables	Categories	n	$\bar{x}$	SD
Take a statistics and probability course	1	253	9.84	3.082
	2	277	7.96	2.537
	Total	530	8.86	2.961

Table 3 shows that PSTs who did not take the statistics and probability course ($\bar{x}$ =7.96) had lower SL scores than those who did ($\bar{x}$ =9.84). The results of the independent samples t-test analysis to test the significance of this difference are presented in Table 4.

Table 4. Independent sample t-test analysis results related to taking statistics and probability course

Variables	Categories	n	$\bar{x}$	SD	df	t	*p*
Take a statistics and probability course	1	253	9.84	3.082	489.328	7.699	.000
	2	277	7.96	2.537			
	Total	530	8.86	2.961			

Levene test (F): 10.332; p= .001

Table 4 shows that the variances of the scores obtained by the groups who took and did not take statistics and probability courses were not homogeneous [F=10.332; p = .001, (p > .05)]. The results of the independent samples t-test when the variances are not homogeneous show that the mean scores of the groups who took and did not take statistics and probability courses differed significantly [t (489.328)= 7.699; p=0.000, (p < .05)]. According to the calculated eta-square value (η^2 = .099), it may propose that taking statistics and probability course has a moderate effect on the differentiation between PST's SL levels. The results of the analysis conducted to determine the SL levels of PST according to their grade levels are presented in Table 5.

Table 5. SL levels of PST at grade levels

Variables	Categories	n	$\bar{x}$	SD
Grade	1st grade	122	7.86	2.642
	2nd grade	161	7.95	2.507
	3rd grade	130	10.25	3.051
	4th grade	117	9.60	2.960
	Total	530	8.86	2.961

Table 5 shows that PST at the 1st ($\bar{x}$ =7.86) and 2nd ($\bar{x}$ =7.95) levels had lower SL scores than PST at the 3rd ($\bar{x}$ =10.25) and 4th ($\bar{x}$ =9.60) levels. One-way ANOVA results to test the significance of the difference are presented in Table 6.

Table 6. One-way ANOVA results according to the grade levels of PST

Variables	Categories	n	$\bar{x}$	Source of Variance	Sum of Squares	df	Sum of Squares	F	p	Difference (scheffe)
Grade	1st grade	122	7.86	between groups	571.41	3	190.470	24.63	0.000	1-3
	2nd grade	161	7.95							1-4
	3rd grade	130	10.25	within groups	4066.97	526	7.732			2-3
	4th grade	117	9.60							2-4
	Total	530	8.86		4638.38	529				

Levene test (F): 2.548; p= .055

Table 6 shows that the variances of SLS scores of PST at all grade levels are homogeneous (F=2.548; p > .05). One-way ANOVA results show that the mean scores of PST at all grade levels differed significantly ($F_{(3-526)}$=24.63; p < .05). Scheffe test showed that the significant difference was between 1-3, 1-4, 2-3, 2-4 grade levels. According to the calculated eta-square value (η^2= .118), it can be said that the grade level has a moderate effect on the differentiation between PST's SL levels.

Discussion

This study examined the relationship between statistics and probability courses and the SL level of PST. The findings provide insights into the SL levels of PST about their course-taking status and grade level. The results are interpreted in terms of both theoretical and empirical support. The results show that SL differs according to grade level and course-taking status. In other words, statistics and probability courses are related to SL. This result is consistent with the findings in the literature that emphasize the importance of statistics and probability courses in SL and skill development (Guven et al., 2021). The statistics and probability knowledge levels of PST who have taken statistics and probability courses are higher than those who have not. Therefore, integrating these courses into teacher education programs can be an effective strategy to strengthen the basic skills of pre-service teachers. Basic concepts and skills in statistics and probability courses affect the SL levels of PST. This training requires practice, a goal-oriented approach, and learning. This coincides with the conceptual understanding emphasized by GAISE (Carver et al., 2016). Therefore, PST can more effectively cope with the challenges they will encounter in their future professional lives. The results are similar across different grade levels. For example, the statistics and probability knowledge levels of 3rd and 4th grade PST are higher than the 1st and 2nd grade teacher candidates. This progress can be attributed to exposure to and interest in more statistical concepts over time. Upper-grade levels provide students with the opportunity to encounter and apply statistical concepts in a variety of contexts.

The importance of SL today is increasing as data-based decision-making and solving complex problems become basic skills. Teachers can promote SL by developing students' conceptual understanding and statistical thinking skills in this context. As previous research emphasizes, knowledge and practice are important in developing SL (Badenes-Ribera et al., 2018; Darling-Hammond et al., 2020). As pre-service teachers master statistical concepts and apply them in different contexts, they may increase their SL levels. Therefore, to improve the SL of PSTs, besides theoretical knowledge, opportunities to encounter concrete examples and apply them in real life should be provided. This approach can help them to increase their literacy levels. In this context, it should be remembered that this hands-on approach in education can more effectively develop PSTs' statistical thinking skills and provide effective teaching.

Conclusion and Implication

This research shows that statistics and probability courses taken at the undergraduate level are related to SL levels of PST. Findings reveal that these courses play a critical role in developing PST's statistical thinking skills and that these skills further increase at higher grade levels. The importance of this study emphasizes that statistics and probability courses should be included more in the context of teacher education programs. Statistics and probability education can enable PST to make data-based mathematics teaching decisions and transfer statistical thinking skills to students. More focus on these areas by teacher education programs can increase the professional competencies of future teachers. Additionally, this finding demonstrates how statistics and probability courses not only improve the SL levels of PST, but also how teacher education programs can contribute to developing these skills. The study results emphasize the importance of reconsidering teacher education programs to increase the potential of PST to increase their SL levels and be more successful in teaching mathematics. In this context, relevant stakeholders need to shape education policies by considering these results. Changes made in this direction may contribute to future teachers' ability to provide mathematics education more effectively.

Limitations and Suggestions

This study has several limitations. The data collected in this study are based on the participants' statements and are limited by the scope of the measurement tool. The gender distribution in the sample is unbalanced, with a higher representation of women (67.7%) compared to men (32.3%). This skewed distribution could potentially reveal gender-related biases and limit the generalizability of the results to a wider population. The results of the analyses show that statistics and probability courses are related to PST's SL. Therefore, longitudinal and experimental studies are needed to analyze the relationship in more detail. In addition, researchers can conduct studies to examine PST's attitudes toward statistics and probability courses. Teacher education programs can include more activities to increase PST's conceptual understanding of statistics and probability courses. It is also suggested to conduct new studies for curriculum development and educational policies by revealing the factors affecting the development of SL skills. Comparative studies related to statistical literacy can be conducted by including university students from different undergraduate programs in the sample.

Conflicts of Interest

The authors declare no competing interests.

Funding

The authors did not receive any financial support for this article.

Ethics

All procedures were conducted in accordance with the ethical standards of the institutional and/or national research committee. The study was conducted with the permission of Ağrı İbrahim Çeçen University Scientific Research Ethics Committee dated 24.03.2022 and numbered 84.

References

Aydin, E., Sevimli, E., & Saja, A. (2019). Self-efficacy beliefs among Palestinian pre-service mathematics teachers' in statistics. *Journal of Theoretical Educational Science, 12*(4), 1209-1222. https://doi.org/10.30831/akukeg.451395

Badenes-Ribera, L., Frias-Navarro, D., Iotti, N. O., Bonilla-Campos, A., & Longobardi, C. (2018). Perceived statistical knowledge level and self-reported statistical practice among academic psychologists. *Frontiers in Psychology, 9*, 996. https://doi.org/10.3389/fpsyg.2018.00996

Balı, O., & Dönmez, B. (2018). Eğitim bilimleri anabilim dalı doktora öğrencilerinin karşılaştıkları problemler ve çözüm önerileri. *İnönü Üniversitesi Eğitim Fakültesi Dergisi, 19*(3), 284-309. https://doi.org/10.17679/inuefd.399079

Ben-Zvi, D., & Makar, K. (2013). International perspectives on the teaching and learning of statistics. In D. Ben-Zvi & K. Makar (Eds.), *Teaching and Learning of Statistics: International Perspectives* (pp. 1-10). Springer Cham.

Cakmak, Z. T., & Durmus, S. (2015). İlköğretim 6-8. sınıf öğrencilerinin istatistik ve olasılık öğrenme alanında zorlandıkları kavram ve konuların belirlenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 15*(2), 27-58. https://doi.org/10.17240/aibuefd.2015.15.2-5000161312

Callingham, R., & Watson, J. M. (2017). The development of statistical literacy at school. *Statistics Education Research Journal, 16*(1), 181-201. https://doi.org/10.52041/serj.v16i1.223

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Roswell, G. H., Velleman, P., & Witmer, J. (2016). *Guidelines for assessment and instruction in statistics education (GAISE) college report 2016*. American Statistical Association. http://www.amstat.org/education/gaise/GaiseCollege_full.pdf

Darling-Hammond, L., Flook, L., Cook-Harvey, C., Barron, B., & Osher, D. (2020). Implications for educational practice of the science of learning and development. *Applied Developmental Science, 24*(2), 97-140. https://doi.org/10.1080/10888691.2018.1537791

De Vetten, A., Keijzer, R., & Schoonenboom, J. (2023). Pre-service primary school teachers' knowledge during teaching informal statistical inference. *Statistics Education Research Journal, 22*(2), 12-12. https://doi.org/10.52041/serj.v22i2.424

Forgasz, H., Hall, J., & Robinson, T. (2024). Evaluating pre-service teachers' statistical literacy capabilities. *Mathematics Education Research Journal, 36*(1), 231-258. https://doi.org/10.1007/s13394-022-00438-6

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review, 70*(1), 1-25. https://doi.org/10.1111/j.1751-5823.2002.tb00336.x.

Garfield, J. (2011). Statistical literacy, reasoning, and thinking. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 1439-1442). Springer Berlin Heidelberg.

Goyal, K., & Kumar, S. (2021). Financial literacy: A systematic review and bibliometric analysis. *International Journal of Consumer Studies, 45*(1), 80-105. https://doi.org/10.1111/ijcs.12605

Guven, B., Baki, A., Uzun, N., Ozmen, Z. M., & Arslan, Z. (2021). Evaluating the statistics courses in terms of the statistical literacy: Didactic pathways of pre-service mathematics teachers. *International Electronic Journal of Mathematics Education, 16*(2), em0627. https://doi.org/10.29333/iejme/9769

Holenstein, M., Bruckmaier, G., & Grob, A. (2021). Transfer effects of mathematical literacy: An integrative longitudinal study. *European Journal of Psychology of Education, 36*(3), 799-825. https://doi.org/10.1007/s10212-020-00491-4

Kaya, A., Şata, M., Türk, N., Özok, H. I., & Yıldırım, M. (2024). Reliability and validity of short-form generic scale of being phubbed and phubbing among Turkish adolescents and young adults. *Journal of Technology in Behavioral Science*. https://doi.org/10.1007/s41347-024-00428-4

Kazunga, C., Bansilal, S., & Chiromo, L. (2023). Primary pre-service teachers' knowledge of the concepts of mean and median. *African Journal of Research in Mathematics, Science and Technology Education, 27*(3), 367-382. https://doi.org/10.1080/18117295.2023.2277984

Marshman, M., & Dunn, P. K. (2024). Improving statistical thinking. *Mathematics Education Research Journal, 36*(1), 1-5. https://doi.org/10.1007/s13394-023-00477-7

Mhlongo, S., Mbatha, K., Ramatsetse, B., & Dlamini, R. (2023). Challenges, opportunities, and prospects of adopting and using smart digital technologies in learning environments: An iterative review. *Heliyon, 9*(6), e16348. https://doi.org/10.1016/j.heliyon.2023.e16348

Nahdi, D., Jatisunda, M., Cahyaningsih, U., Kurino, Y., Juliar, E., & Bilda, W. (2021). Statistical literacy analysis of pre-service elementary teachers education. *Journal of Physics: Conference Series, 1764*(1), 012126. https://doi.org/10.1088/1742-6596/1764/1/012126

Nicholson, K. (2017). Objectives and actions. In K. Nicholson (Ed.), *Innovation in Public Libraries* (pp. 127-135). Chandos Publishing.

Ozmen, Z. M., & Baki, A. (2019). 5-8. sınıf matematik öğretim programının istatistik okuryazarlığı bağlamında incelenmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi, 13*(2), 1063-1082. https://doi.org/10.17522/balikesirnef.603569

Pilgrim, J., & Martinez, E. E. (2013). Defining literacy in the 21st century: A guide to terminology and skills. *Texas Journal of Literacy Education, 1*(1), 60-69.

Ridgway, J., Nicholson, J., & McCusker, S. (2011). Developing statistical literacy in students and teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: A Joint ICMI/IASE Study: The 18th ICMI Study* (pp. 311-322). Springer Netherlands.

Sahin, F. (2012). *Lisans öğrencileri için istatistiksel okuryazarlık ölçeği geliştirilmesi çalışması* [Yayımlanmamış yüksek lisans tezi]. Boğaziçi Üniversitesi.

Schreiter, S., Friedrich, A., Fuhr, H., Malone, S., Brünken, R., Kuhn, J., & Vogel, M. (2024). Teaching for statistical and data literacy in K-12 STEM education: A systematic review on teacher variables, teacher education, and impacts on classroom practice. *ZDM – Mathematics Education, 56*(1), 31-45. https://doi.org/10.1007/s11858-023-01531-1

Sharma, S. (2017). Definitions and models of statistical literacy: A literature review. *Open Review of Educational Research, 4*(1), 118-133. https://doi.org/10.1080/23265507.2017.1354313

Sharma, S., Doyle, P., Shandil, V., & Talakia'atu, S. (2011). Developing statistical literacy with Year 9 students. *Set: Research Information for Teachers*(1), 43-50. https://doi.org/10.18296/set.0398

Shiel, G., & Cartwright, F. (2015). *Analyzing data from a national assessment of educational achievement*. World Bank Publications.

Şata, M. (2020). Nicel araştırma yaklaşımları. E. Oğuz (Ed.), *Eğitimde Araştırma Yöntemleri* (ss. 77-98). Eğiten Kitap.

Walsh, J. (2011). Methods of instruction. In J. Walsh (Ed.), *Information Literacy Instruction* (pp. 3-55). Chandos Publishing.

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2*(2), 3-46. https://doi.org/10.52041/serj.v2i2.553

Wu, D., Zhou, C., Li, Y., & Chen, M. (2022). Factors associated with teachers' competence to develop students' information literacy: A multilevel approach. *Computers & Education, 176*, 104360. https://doi.org/10.1016/j.compedu.2021.104360

Zieffler, A., Garfield, J., & Fry, E. (2018). What is statistics education? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 37-70). Springer.