



A Bibliometric Analysis of Articles on High-Stakes Testing Between 2000 and 2024

•Gönül Yazar ¹, •Vural Tünkler²

1 Independent Researcher, Isparta, Türkiye

2 Department of Turkish and Social Sciences Education, Faculty of Education, Süleyman Demirel University, Isparta, Türkiye

ABSTRACT

This study aims to examine articles on high-stakes testing using a bibliometric analysis method. High-stakes testing represents a critical tool through which schools and educators are evaluated based on student performance, thereby shaping educational policies. Within the scope of this research, 1,947 articles published between 2000 and 2024 were retrieved from the Web of Science Core Collection (WoS) database using the "Topic" field with the keywords "education AND high stakes testing OR accountability" and subsequently analyzed. The analysis employed the VOSviewer software. Findings indicate a notable increase in publications in this field, particularly after 2010. Among collaborating organizations, Stanford University emerged as prominent, underscoring the United States' leadership in the domain. Co-citation analysis identified the American Educational Research Journal as the most influential journal. Furthermore, "accountability" was the most frequently used keyword. The trends revealed in the study highlight the role of high-stakes testing within educational systems, the increase in the number of studies over the years, and its multidimensional nature. The results demonstrate how high-stakes testing shape education and indicate that more research could be conducted in this area in the future.

Keywords: High-stakes testing, assessment systems, education policies, standardized tests, bibliometric analysis.

Corresponding Author: Vural Tünkler

Department of Turkish and Social Sciences Education, Faculty of Education, Süleyman Demirel University, Isparta, Türkiye

E-mail: vuraltunkler@sdu.edu.tr **ORCID ID:** <https://orcid.org/0000-0002-3536-968X> **ROR ID:** <https://ror.org/04fjte88>

Received Date: 06.21.2025 **Accepted Date:** 02.16.2026 **Publication Date:** 04.15.2026

Citation: Yazar, G., & Tünkler, V. (2026). Bibliometric Analysis of Articles on High-Stakes Testing (2000-2024)

Ankara University Journal of Faculty of Educational Sciences, 59(1), 425-473. <https://doi.org/10.30964/auebfd.1724404>

All ethical declarations related to this article are provided on the final page of the manuscript (Page: 473)

Measurement is an important element of education systems. Burdett (2016) clearly emphasized the importance of feedback, namely measurement results, at every level of education for students to succeed and develop. Measurements encompass various methods aimed at determining students' knowledge and skills. In light of the data obtained, objectives such as student development, guidance, and the review of teaching practices are targeted (Rhoten et al., 2003). The type of measurement focused on in this article is large-scale standardized tests administered in any region, state, country, or across countries, prepared by various experts. This aforementioned large-scale measurement approach is not a new factor for education systems (Verger et al., 2019). Policy makers have turned their attention to education reforms, particularly as concerns grow about the current performance and effectiveness of education systems in providing fair opportunities for every student (Camphuijsen et al., 2021). Politicians prefer test-based standardized systems in order to enhance educational quality and establish a transparent and accountable structure in the public eye (Supovitz, 2009). In his study, Supovitz (2009) stated that there are four main theories promoting these systems and listed them as follows: (i) motivational theory (test-based accountability will motivate all educational stakeholders to improve their performance), (ii) alignment theory (test-based systems can help ensure alignment among all educational elements such as curriculum, teaching methods, and assessment), (iii) informational theory (such systems provide data to guide educators in their improvement efforts), (iv) idea of symbolism (the accountability system signals important values to stakeholders). These theories have laid the groundwork for the adoption and widespread use of test-based assessment systems since the 1990s, leading to the desire to create a common set of standards encompassing all students through high-standard-based education initiatives in all 50 U.S. states (Goertz & Duffy, 2003). Among the components of this standards-based system are assessments that measure student performance and accountability focused on the results of these assessments (Goertz & Duffy, 2003; Levatino et al., 2024; Ozga, 2009). Xie (2022) defined accountability as holding someone responsible for the task assigned to them, that is, a contract between parties to achieve predetermined goals. On the other hand, many researchers describe accountability as the process of interpreting schools' performance based on the results obtained from standardized tests administered to students and holding school administrators and educators responsible for these results (Bae, 2018; Verger et al., 2019). Erdağ (2019), however, stated that the perspective claiming that educational problems, especially since the 1980s, stemmed from school administrators and educators, has brought school accountability to the forefront as a means to overcome these issues.

Over the past quarter century, test-based accountability practices have secured a significant place within education systems around the world (Klinger & Rogers, 2011; Levatino et al., 2024). With the enactment of the No Child Left Behind (NCLB) Act in 2001, a performance-based accountability system (Kirby & Stecher, 2004), there has been an observed increase in the use of standardized tests to make decisions about students, teachers, school administrators, and schools (von der Embse & Witmer,

2014). Parkison (2009) has described NCLB as an ideology that determines and institutionalizes the priorities of education. This Act aimed to raise existing educational standards and close achievement gaps between schools through accountability (Tan, 2010). In 2015, the NCLB Act was updated with Every Student Succeeds Act (ESSA) (Finn, 2022), and ESSA sought to define student learning more broadly (Bae, 2018). According to the literature, when results from standardized tests administered to students are used to make critical decisions about their future educational pathways (e.g., graduation, grade retention, or promotion), affect school funding as well as the performance of administrators and teachers, or are publicly disclosed to impact the reputation of individuals and schools, such tests are defined as “high-stakes test” (McNeil, 2000; Orfield & Wald, 2000). In some countries, when established educational goals and standards are not met, sanctions such as the dismissal of school administrators and staff, salary reductions, or the closure of educational institutions, coupled with student attrition following institutional failure, have led to large-scale standardized tests being characterized as “high-stakes test” in certain regions (Hargreaves, 2020). Additionally, based on test results, some students are excluded from the education system, while others qualify to advance to the next educational level (Hursh, 2005; Jacobs et al., 2024). Examples of such countries and high-stakes tests include the Scholastic Aptitude Test (SAT) and American College Testing (ACT) in the United States, the International English Language Testing System (IELTS) and General Certificate of Secondary Education (GCSE) in the United Kingdom, and the Gaokao in China. The outcomes of these tests trigger the aforementioned sanctions (Hargreaves, 2020; Orfield & Wald, 2000; Wang et al., 2022). Similarly, in Australia, results from the National Assessment Program – Literacy and Numeracy (NAPLAN) are publicly shared on the MySchool website, enabling comparisons between schools (Thompson & Lašič, 2011). When low results are obtained, teachers and schools are required to design performance-based supplementary education programs, and underperforming students are expected to take additional classes (Thompson, 2013; Thompson & Lašič, 2011). In Türkiye, the High School Entrance Exam and the Higher Education Institutions Exam are considered high-stakes tests because they determine not only students’ progression to the next educational level but also their admission to prestigious institutions (Özaslan & Beyhan, 2023). All these consequences play a significant role in classifying standardized tests as high-stakes. Due to the reasons mentioned above, all stakeholders in the education system (students, teachers, administrators, and even parents) have shifted their focus from the educational process to the outcomes of standardized tests (Hanushek & Raymond, 2004).

The dominant theory behind accountability and high-stakes testing posits that educators will exert greater effort, schools will adopt effective teaching methods, and consequently, students will learn more and achieve inevitable success; however, the impacts of high-stakes tests remain a subject of debate (Heilig & Darling-Hammond, 2008). Falabella (2021) has noted that high-stakes tests and the accountability approach extend beyond school education policies to influence salary policies, teacher

training, and various fields from early childhood and special education to higher education, all of which occur despite insufficient research data on the effectiveness of standards-based education reforms. Sloane and Kelly (2003) categorized the debates surrounding high-stakes testing under four main headings and stated that these discussions provide sufficient grounds for reevaluating the results and implementation methods of high-stakes tests. These issues are: (i) the types of tests used, (ii) the impact of tests on students' morale and motivation, (iii) the degree of alignment between the test and the curriculum, (iv) the distinction between assessment of learning and assessment for learning. One of the most common concerns regarding high-stakes tests is the narrowing of instruction provided to improve test scores (Hargreaves, 2020; Hatfield & Soløst, 2025; Zakharov & Carnoy, 2021). The intention behind narrowing the scope of education is to prioritize subjects and topics that are tested and exclude those that are not from the curriculum (Au, 2011; Gulek, 2003). However, this is not always feasible. Some researchers have indicated that high-stakes tests measure certain topics not included in the school curricula, which students have thus never studied (Arviv Elyashiv & Avidov-Ungar, 2024; Thomas, 2005). Another problem arising from high-stakes tests is the intense demand from students and parents for private tutoring, also referred to in the literature as "shadow education" (Yung, 2021). The reason for this demand is that private tutors provide students with entirely exam-focused content aimed at improving outcomes (McCoy & Byrne, 2024). This situation creates a disadvantage for low-income students, exacerbating educational inequality by widening the achievement gap between disadvantaged and advantaged students rather than eliminating it (Hursh, 2005). Yet another issue caused by high-stakes tests is the pressure it imposes on students and educators. Numerous students experience anxiety due to the emphasis placed on tests, which negatively affects their academic performance (Hatfield & Soløst, 2025; von der Embse & Witmer, 2014) and mental health (Jerrim, 2023). Evaluating educators based on student performance outcomes has imposed significant pressure on school administrators and teachers (Arviv Elyashiv & Avidov-Ungar, 2024). Kenney et al. (2024), taking into account the pressure policymakers place on educators, have argued that policymakers must acknowledge the negative effects of this pressure on the educational process. All these findings demonstrate high-stakes tests' control over content, knowledge form, and pedagogy at the grade level (Au, 2007). Additionally, another criticism is that comparisons between schools in accountability-based education systems should be conducted with consideration of each school's unique context (Munoz-Chereau et al., 2022). Klinger and Rogers (2011), despite all these negative research findings, have argued that changes in the reporting and use of high-stakes test results would positively influence the teaching and learning process.

Upon review of the relevant literature, numerous studies conducted in various countries and on diverse topics regarding high-stakes testing in education and, in this context, accountability practices have been encountered (Au, 2016; Burgess et al., 2021; Elo & Nygren-Landgårds, 2021; Erduran et al., 2020; Escamilla et al., 2005; Fidjeland, 2023; Fischer et al., 2020; Fitchett & Heafner, 2010; Gorgodze &

Chakhaia, 2021; Hammack & Wilson, 2019; Hilton & Saunders, 2024; Hofflinger & von Hippel, 2020; Kavanagh & Fisher-Ari, 2020; Knoester & Meshulam, 2022; McElroy, 2023; Mintz & Kelly, 2021; Moore, 2025; Ober et al., 2022; Pace, 2011; Trujillo et al., 2021; Yin & Mu, 2022). Given the vast amount of data in the literature that cannot be analyzed manually, bibliometric analysis has been employed to reveal the overall trends and intellectual structure of articles on high-stakes testing. This is because bibliometric analysis serves as a crucial tool in demonstrating the processability of large-scale scientific data and its valuable contributions to research (Passas, 2024). This study aims to contribute to the literature and relevant researchers by examining articles on high-stakes testing through bibliometric analysis methods. In this context, the questions addressed throughout the research are as follows:

1. What is the distribution of articles related to high-stakes testing by years?
2. What kind of structure emerges regarding cooperation between authors, countries, and organizations?
3. What kind of structure emerged regarding co-citation analysis (reference, journal, and author)?
4. What kind of structure emerges in terms of keyword co-occurrence analysis?

Method

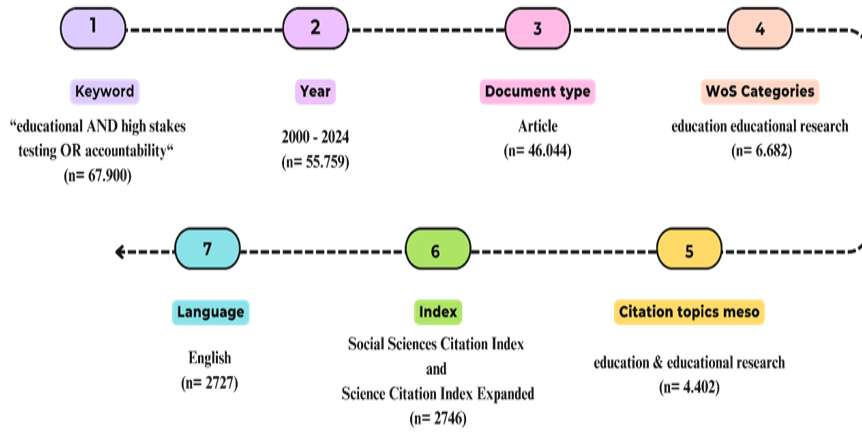
The bibliometric analysis method can be defined as a quantitative analysis of written materials to obtain an overview of any research field by classifying authors, articles, journals, citations, and keywords (Merigó & Yang, 2017). Such an analysis relies on identifying the literature, i.e., scientific publications, on a specific topic (Ellegaard & Wallin, 2015). Bibliometric analysis is highly suitable for making sense of large amounts of data that cannot be examined individually using rigorous techniques, thereby revealing and mapping the scientific knowledge and evolutionary process of established research fields from the past to the present (Donthu et al., 2021; Passas, 2024). Therefore, a meticulously conducted and well-prepared bibliometric study can provide a strong foundation for the field under investigation (Passas, 2024). This analysis method is increasingly preferred because it enables performance evaluation of comprehensive literature (Wallin, 2005). For this reason, the study utilized the bibliometric analysis method to reveal the trends and map of articles related to high-stakes testing.

Data Source and Analysis

The research process began with the determination of the research topic, followed by the clarification of the research questions. To select the keywords for data collection, the relevant literature was reviewed, and the most frequently used terms and concepts in the field were identified and tested. The results were compared to determine the keywords that would yield the most comprehensive and reliable data. Additionally, the literature review on high-stakes testing revealed that “high-stakes

testing” and “accountability” were the most commonly used terms. Considering the frequency and inclusivity of the selected keywords, this study is considered valid and reliable in line with its purpose. To access the articles for analysis, the WoS database, which has the oldest and most comprehensive records and allows direct data transfer to the analysis program, was chosen (Ellegaard & Wallin, 2015). On October 16, 2024, using the “advanced search” option in the WoS database with the “topic” tag, the keywords “education AND high stakes testing OR accountability” were searched, yielding 67.900 studies. To ensure that the analyses and findings provide more robust and reliable data, certain filtering steps were deemed necessary. The filtering steps applied are shown in Figure 1.

Figure 1
Filtering Steps Performed



After all these filtering steps, the abstracts of the articles were reviewed to determine whether the topic of high-stakes testing was central, and duplicate articles were identified. As a result, 1.947 articles were included in the study. Subsequently, a graph of the publication years of the articles related to high-stakes testing was created using the Microsoft Word program. To pragmatically conduct the bibliographic analysis of the available data (Donthu et al., 2021), the latest version of the VOSviewer program, version 1.6.20, was used. The primary reason for choosing the VOSviewer software is its ability to directly import data sets from the WoS database. VOSviewer utilizes two standard weights, namely the number and total strength of the links, in nodal network visualization, thereby presenting the relevance and strength of the links through node sizes and interlinking lines connecting the nodes denote (Donthu et al., 2020). To conduct bibliometric analysis or interpret the results of such analyses, it is essential to understand the meaning of certain fundamental terms. According to the explanations provided by Van Eck and Waltman

(2023), some of these terms are as follows:

- **Link:** A link is a connection or a relation between two items. The higher this value, the stronger the relation between two items.
- **Total link strength (TLS):** It indicates the total strength of the co-authorship links of a given researcher with other researchers.
- **Items:** These are objects of interest. Items can be authors, organizations, countries, or keywords.
- **Cluster:** A cluster is a set of items included in a map. Clusters are non-overlap, meaning an item can belong to only one cluster.

In this program, which enables various analyses, co-authorship, co-citation, and co-occurrence analyses were used to create network maps and tables based on the data provided by the VOSviewer software. The reason for selecting these three types of analyses is that the primary objective of the study is to reveal the collaboration structure, citation relationships, and conceptual trends within the relevant literature. The purpose of the co-authorship analysis is to examine documents written by multiple authors, organizations, and countries to uncover their collaboration network (Merigó et al., 2018). Co-citation refers to the frequency with which two articles of earlier literature are cited together by the later article (Small, 1973), while co-occurrence analysis identifies the keywords that appear more frequently in the documents (Merigó et al., 2018). During all these analyses, a threshold value is set to determine the minimum number of items to be included in the analysis (Karakus et al., 2019). In brief, the threshold value refers to the minimum condition that an item must meet to be included in the analysis. For example, in co-authorship analysis, if the threshold value is set at 5, it means that only authors with at least 5 documents are included in the analysis. In this study, threshold values were determined through trial, considering the comprehensibility of the network visualizations generated by the VOSviewer software. Initially, low values were tested, and to ensure clarity in overly complex network maps, the threshold value gradually increased. The threshold value was fixed at the point where meaningful clusters and networks became interpretable. The threshold value used for each analysis is specified under each finding. After determining the threshold value, prominent elements of the topic (e.g., author, journal, reference) were tabulated based on the total link strength provided by the software. The generated network maps and tables are presented and interpreted in the next section, under the heading of findings.

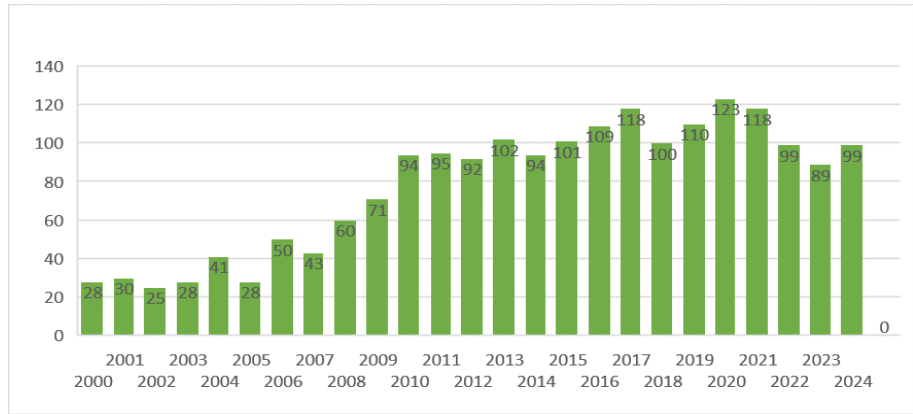
Results

This section includes the bibliometric analyses of 1.947 articles on "high-stakes testing" obtained from the WoS database, along with interpretations of these analyses.

Distribution of the Articles by Years

The distribution of articles on high-stakes testing in education by year is presented in Figure 2.

Figure 2
Distribution of Articles by Year



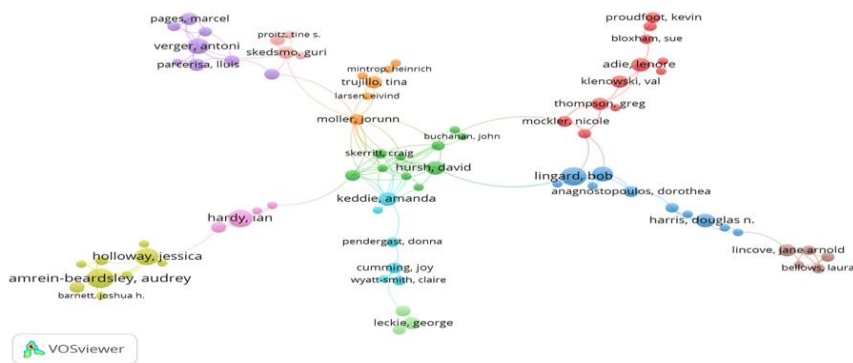
As can be observed from the figure, although there are slight fluctuations in the number of articles over the years, there has been a significant increase in the number of articles related to high-stakes testing, particularly after 2010, compared to other years. The years with the highest number of published articles are 2020 (n=123), followed by 2017 and 2021, which have an equal number of publications (n=118). The years with the fewest articles published are 2002 (n=25), followed by 2000, 2003, and 2005, which have an equal number of publications (n=28), and 2001 (n=30).

Co-authorship Analysis (Authors)

Figure 3 presents the authors engaged in collaboration and the resulting network map. According to the co-authorship analysis of the 1.947 articles examined, the total number of authors with at least one document is 3.396. When the threshold value, i.e., the number of documents, is set to 2, only 488 authors exceed this threshold. It was observed that 76 of these authors have established connections. This finding indicates that, among the 488 authors meeting the threshold, only 76 have collaborated with each other to form a collaboration network. Accordingly, the network map shown in Figure 3 consists of 76 nodes across 11 clusters, with a total link strength of 188. Antoni Verger emerges as the author with the highest number of co-publications, with a total link strength of 16. Within Verger's cluster, it was found that the link strength with authors Lluís Parcerisa, Marcel Pagès, and Gerard Ferrer-Esteban is higher compared to other authors in the same cluster. Additionally, according to the network map, Jorunn Møller is identified as the author with the highest link strength, indicating

the broadest network structure. However, this has not been sufficient to increase Møller's total link strength. Consequently, due to the low intensity of collaboration, Jorunn Møller did not rank among the top 10 authors with the highest collaboration and is not shown in the table.

Figure 3
Co-authorship Network Map of Authors



When examining Table 1, it is observed that authors such as Antony Verger and Meg Maguire rank higher on the list due to their higher total link strength, despite having fewer documents than Audrey Amrein-Beardsley and fewer citations than Barry Fishman. This indicates that Verger and Maguire hold a more central position within the research network and engage in more intensive collaborations. On the other hand, some authors with relatively high document and citation counts but lower total link strength can be said to work more individually or have limited collaborations, as they do not rank among the top positions.

Table 1
Most Collaborating Authors

Authors	Document	Citation	TLS
Verger, A.	8	219	16
Maguire, M.	5	184	14
Amrein-Beardsley, A.	12	85	13
Saultz, A.	7	104	13
Eisenkraft, A.	2	89	12
Fischer, C.	2	89	12
Fishman, B.	3	295	12
Frumin, K.	2	89	12
Gewirtz, S.	4	77	12
Lawrenz, F.	2	89	12

Co-authorship Analysis (Organizations)

Figure 4 presents the organizations involved in collaboration and the resulting network map. According to the co-authorship analysis of the 1.947 articles examined, it was found that there are 1.079 organizations with at least one document. When the threshold value is set to 10, only 68 organizations exceeded this threshold. It was determined that 67 of these organizations are interconnected, forming a collaboration network with 9 clusters and a total link strength of 244. According to the network map, Stanford University is identified as the organization with the most collaborations. However, the highest level of collaboration between organizations, with a link strength of 4, occurs between the University of Florida and Pennsylvania State University.

Figure 4
Co-authorship Network Map of Organizations

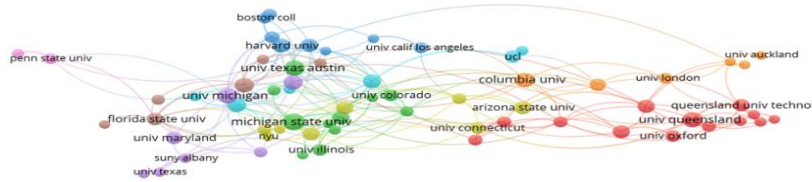


Table 2 lists the top 10 organizations with the highest collaboration based on total link strength. Stanford University, ranking first with a total link strength of 16, has collaborated most frequently with Columbia University and Vanderbilt University, each with a link strength of 2. It can be noted that other organizations on the list, despite having high citation or document counts, have lower total link strength due to fewer collaborations, placing them in the middle or lower ranks of the list.

Table 2
Most Collaborating Organizations

Organizations	Document	Citation	TLS
Stanford University	28	1247	16
University of Michigan	31	1222	15
The University of North Carolina at Chapel Hill	30	771	14
Florida State University	21	886	13
Michigan State University	38	541	13
University of Virginia	26	977	13
University of Rochester	17	976	12
University of Southern California	14	173	12
University of Wisconsin	28	1082	12
Deakin University	23	677	11

Co-authorship Analysis (Countries)

Figure 5 presents the countries involved in collaboration and the resulting network map. According to the co-authorship analysis of the 1,947 articles examined, it was found that 77 countries have at least one document. When the threshold value was set to 2, only 45 countries exceeded this threshold. It was observed that 41 of these countries are interconnected, forming 10 clusters with a total link strength of 350. According to the network map, the two countries that collaborated the most with each other were the United States and Australia.

Figure 5
Co-authorship Network Map of Countries

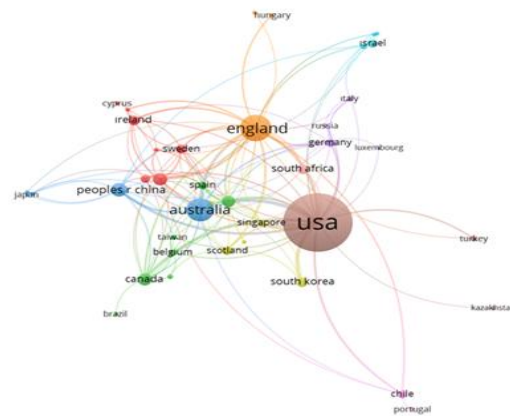


Table 3 lists the top 10 countries with the highest collaboration based on total link strength. The United States ranks first with a total link strength of 133. Considering the link strengths between them, the United States has collaborated most frequently with Australia, the United Kingdom, South Korea, and Canada, respectively. However, South Korea did not make it into the top 10 countries due to its low total link strength.

Table 3
Most Collaborating Countries

Countries	Document	Citation	TLS
USA	1124	31404	133
United Kingdom	227	4817	100
Australia	184	6236	60
Canada	50	880	34
People's Republic of China	61	1274	34
Norway	47	743	30
Germany	22	299	27
Netherlands	43	1527	24
Sweden	29	485	24
Scotland	27	1380	19

Co-citation Analysis (Journals)

Figure 6 presents the network map of scientific journals resulting from the co-citation analysis. Of the 37.053 scientific journals that received at least one citation in the analysis, it was observed that only 93 journals exceeded the citation threshold of 100 and were included in the map. These 93 journals formed 5 clusters with a total link strength of 327.273. According to the resulting network map, the strongest co-citation connection was found to be between the journals Educational Evaluation and Policy Analysis and American Educational Research Journal.

Figure 6
Co-citation Network Map of Journals

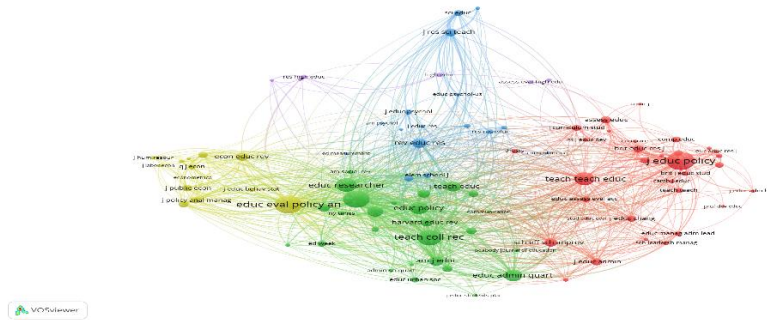


Table 4 lists the top 10 most influential scientific journals based on total link strength. At the top of the list is the American Educational Research Journal with a total link strength of 35.465. It was found that this journal is connected to all other journals included in the network map.

Table 4
The Top 10 Most Influential Journals

Journal	Citation	TLS
American Educational Research Journal	1396	35465
Educational Evaluation and Policy Analysis	1488	34233
Educational Researcher	996	24297
Journal of Education Policy	1524	21830
Teaching and Teacher Education	920	21676
Educational Administration Quarterly	800	20669
Educational Policy	864	19855
American Journal of Education	596	15280
Review of Educational Research	606	15053
Journal of Teacher Education	637	13639

Co-citation Analysis (References)

Figure 7 presents the network map of references resulting from the co-citation analysis. The threshold value for references to be included in the map was set at 30. A total of 60 references exceeded this threshold, forming 4 clusters with a total link strength of 4.358. According to the resulting network map, the reference with the strongest co-citation relationship was identified as Booher-Jennings (2005), with a total link strength of 458. Upon examining the other references with which this reference has co-citation relationships, it was found to be most strongly connected to Diamond and Spillane (2004), with a link strength of 27.

Figure 7
Co-citation Network Map of References

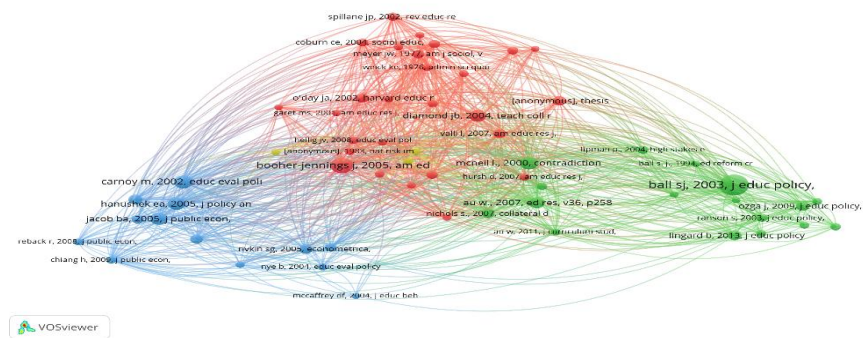


Table 5 lists the top 10 most influential references based on total link strength. The references Carnoy and Loeb (2002) and Hanushek and Raymond (2005) were identified as having the highest co-citation relationship, with a link strength of 48. However, Booher-Jennings (2005), with connections to 55 references and a total link strength of 458, surpassed Carnoy and Loeb (2002) and Hanushek and Raymond (2005) to rank first on the list.

Table 5
The Top 10 Most Influential References

References	Citation	TLS
Booher-Jennings (2005)	114	458
Carnoy & Loeb (2002)	90	359
Hanushek & Raymond (2005)	80	324
Diamond & Spillane (2004)	76	305
Jacob (2005)	82	303
Ball (2003)	171	301
Dee & Jacob (2011)	58	274
McNeil (2000)	83	226
Au (2007)	71	224
Neal & Schanzenbach (2010)	47	220

Co-citation Analysis (Authors)

Figure 8 presents the network map of authors resulting from the co-citation analysis. When the threshold value, i.e., the number of citations an author must have, was set at 80, 82 authors exceeded this threshold and were included in the map. These authors formed 5 clusters with a total link strength of 51.289. According to the network map, James P. Spillane was identified as the author with the strongest co-citation relationship, with a total link strength of 4.497. James P. Spillane and Cynthia E. Coburn were found to be the two authors with the highest link strength between them, with a link strength of 523.

Discussion, Conclusion and Suggestions

The aim of this study is to examine articles related to high-stakes testing through bibliometric analysis to reveal the general trends in the relevant literature. Based on 1,947 articles obtained from the WoS database, a significant increase in the number of studies on high-stakes testing has been observed between 2000 and 2024, particularly after the 2010s. The low number of studies in the early 2000s may be attributed to the fact that high-stakes testing had not yet been fully adopted in education systems during those years and had not found a place for themselves. The mandatory use of standardized tests introduced by the NCLB Act in the United States in 2001 (Sloane & Kelly, 2003), and its update with the ESSA in 2015 (Finn, 2022), directly influenced the increase in the number of studies in this field over the years. Considering the role of high-stakes testing and accountability systems within education systems, as well as the innovations brought to the field of assessment by technological advancements (such as digital assessment tools and artificial intelligence-supported assessment systems), it is likely that the number of studies in this area will continue to increase in the coming years.

According to the co-authorship analysis, although Audrey Amrein-Beardsley is the author most frequently mentioned in the studies, Antony Verger was identified as the author with the highest collaboration due to his high total link strength, indicating a strong intensity of collaboration. This finding suggests that Verger has a scientific production network with a broad impact. In general, Verger has conducted studies in areas such as education policies, privatization in education, and test-based accountability. This may be one of the important reasons for Verger's collaboration with different researchers. Furthermore, this finding highlights the importance of the impact achieved through the diversity of collaborations rather than the number of publications. Following the work of these authors can greatly contribute to researchers working in the field. Another finding from the co-authorship analysis is that Stanford University is the organization with the highest level of collaboration. Although Michigan State University, the University of Michigan, and the University of North Carolina at Chapel Hill have more publications than Stanford University, Stanford University has successfully expanded its network by collaborating with 14 different organizations. The final finding from the co-authorship analysis is that the United States ranks first in international collaboration. It leads the field by a wide margin in terms of document count, citation count, and total link strength, outperforming other countries. Considering the pioneering steps taken by the United States in standardization and accountability policies in education (NCLB and ESSA), this result is not surprising. Given Stanford University's leadership in this field and the prominence of the United States, researchers working on high-stakes testing could enhance the visibility of their studies by collaborating with American researchers and organizations.

In the study, the intellectual structure of the field was aimed to be revealed

through co-citation analysis of journals, articles, and authors rather than direct citation analysis of the collected data. In this context, according to the co-citation analysis, the scientific journal with the strongest co-citation relationship, considering total link strength, is the American Educational Research Journal. This journal has co-citations with all other journals in the network map. This indicates that the journal holds a central position in knowledge production related to high-stakes testing. Additionally, it was determined that the journal with the strongest co-citation relationship is Educational Evaluation and Policy Analysis, and this relationship represents the strongest co-citation link across the entire network map. The frequent joint referencing of these two journals by researchers demonstrates that they form a guiding structure for studies on high-stakes testing. Furthermore, it was found that the American Educational Research Journal publishes in areas such as education policies and practices, learning and teaching, research methods and innovative approaches in education, technology and education, and similar topics. The other journals listed in Table 4 were also identified as publishing in areas such as educational assessment, education policies, teacher education, and educational administration. Regularly following these journals can help researchers working in the field meet their need for up-to-date information. Another key finding from the analysis is that the reference with the strongest co-citation relationship is Booher-Jennings (2005). This reference is co-cited with 55 distinct references in total and has successfully expanded its network to become the most influential reference. In this study, Booher-Jennings (2005) examines how high-stakes test scores alter teachers' identities and interpersonal relationships, aiming to understand the impact of educational accountability systems on teacher behavior. Another significant reference, Diamond and Spillane (2004), investigates how four different elementary schools respond to high-stakes accountability policies. Consequently, Diamond and Spillane concludes that schools under supervision focus on policy requirements by adopting a narrowed curriculum and targeting the performance of specific students, whereas higher-performing schools strive to enhance the success of all students, emphasizing the need to address equity issues in education. According to the network map, the reference most frequently co-cited with Booher-Jennings (2005) is Diamond and Spillane (2004). The co-citation relationship between these two references indicates that studies in the field concentrate on teacher behaviors, educational equity, and schools' responses to assessment policies. However, the network map reveals that the two references with the highest link strength are Carnoy and Loeb (2002) and Hanushek and Raymond (2005). These studies demonstrate that high-stakes testing and accountability systems are effective in improving student achievement while also highlighting complex issues such as systemic inequalities. These prominent references in the literature indicate that high-stakes testing have social and political dimensions in addition to student achievement. In this context, future research may benefit from considering this multidimensional structure. The final finding from the co-citation analysis is that James P. Spillane is the author with the strongest co-citation relationship, with a total link strength of 4.497. This author is most frequently co-cited with Cynthia E. Coburn, and together they have formed a link strength of 523, making

them the two authors with the highest link strength across the entire network map. An examination of the authors' research interests reveals active publications in themes such as educational leadership, implementation variations, and teacher perceptions. Following James P. Spillane, the author with the highest total link strength is Stephen J. Ball. According to the network map, the author most frequently co-cited with Ball is Bob Lingard, with a link strength of 456 between them. These two authors were found to focus on themes in education policy. This thematic diversity demonstrates that the literature encompasses a structure addressing multiple dimensions, ranging from specific aspects (e.g., schools, educators) to broader ones (e.g., education systems, policy). For researchers in the field, reading these references and following the mentioned authors may prove highly beneficial in gaining mastery over the fundamental concepts and trends in the high-stakes testing literature.

Finally, when the collected data were examined in terms of keyword relationship networks, it was observed that the keywords clustered under 7 groups. According to the network map, the most frequently used keyword was determined to be "accountability." This keyword is extensively used with numerous keywords from other clusters beyond its own, resulting in a highly developed network structure. When considering total link strengths, other significant keywords linked to "accountability" were found to be, in order, "educational policy," "assessment," "educational reform," "policy," and similar terms. This finding indicates that studies on high-stakes testing are predominantly addressed in conjunction with education policies, educational reforms, and assessment methods. For researchers conducting searches in various databases, utilizing these keywords may facilitate access to the most accurate data and enable rapid familiarity with the field.

The trends emerging from the analysis reveal the position of high-stakes testing within education systems, the increase in the number of studies over the years, and its multidimensional nature. Based on the overall evaluation of the findings, co-authorship analysis identified Antony Verger as the author with the highest collaboration strength and Stanford University as the organization with the most collaborations. These findings indicate that studies on high-stakes testing are predominantly conducted within international networks. Therefore, project and publication collaborations that enhance cooperation between Turkish researchers and international individuals and organizations should be encouraged. Given the leadership of the United States in this field, bilateral or multilateral collaborations between researchers and organizations, particularly with U.S.-based scholars, can increase the visibility of both researchers and their studies.

Another finding from the results is that the number of publications alone is not sufficient for academic impact. The diversity of collaborations and the breadth of the network structure were found to be more decisive in generating academic influence. Consequently, authors should prioritize establishing collaborations with scholars from diverse disciplines rather than focusing solely on individual publications.

The American Educational Research Journal and Educational Evaluation and Policy Analysis stand out as leading journals in the field. Researchers may find it beneficial to follow these journals. An examination of prominent references reveals that they emphasize the social and political dimensions of high-stakes testing. Future studies focusing on these dimensions could make significant contributions to the field.

The keywords demonstrate that research on high-stakes testing is closely intertwined with themes of education policy, reform, and assessment. Future investigations could explore the interconnections among these themes. Moreover, the clustering of keywords into seven groups once again underscores the multidimensional nature of high-stakes testing. Future research may examine the relationship between the themes found in these clusters.



2000-2024 Arası Yüksek Riskli Testlerle İlgili Makalelerin Bibliyometrik Analizi

•Gönül Yazar¹, •Vural Tünkler²

1 Bağımsız Araştırmacı, Isparta, Türkiye
2 Türkçe ve Sosyal Bilimler Eğitimi Bölümü, Eğitim Fakültesi, Süleyman Demirel Üniversitesi, Isparta, Türkiye

ÖZ

Bu çalışmada, yüksek riskli testlerle ilgili yapılan makalelerin bibliyometrik analiz yöntemiyle incelenmesi amaçlanmıştır. Yüksek riskli testler, öğrenci performansı üzerinden okul ve eğitimcilerin değerlendirildiği ve eğitim politikalarının şekillendiği önemli bir araçtır. Çalışma kapsamında, Web of Science Core Collection (WoS) veri tabanından “Topic” etiketi ile “education AND high stakes testing OR accountability” anahtar kelimeleri kullanılarak 2000-2024 yılları arasında konuyla ilgili yapılan 1.947 makale analiz edilmiştir. Analizde VOSviewer programı kullanılmıştır. Araştırma sonucunda özellikle 2010'dan sonra bu alandaki yayınlarda artış gözlenmiştir. En çok iş birliği yapılan kuruluşlar arasında Stanford Üniversitesi öne çıkarken, ABD'nin bu alandaki liderliği vurgulanmıştır. Ortak atf analizinde, American Educational Research Journal en etkili dergi olarak belirlenmiştir. Ayrıca, anahtar kelimeler arasında “accountability” en çok kullanılan kavram olmuştur. Araştırmada ortaya çıkan eğilimler, yüksek riskli testlerin eğitim sistemleri içindeki yerini, yıllar içinde çalışma sayısında yaşanan artışı ve çok boyutlu yapısını ortaya koymaktadır. Sonuçlar, yüksek riskli testlerin eğitimi nasıl şekillendirdiğini ve gelecekte bu alanda daha fazla araştırma yapılabileceğini göstermektedir.

Anahtar sözcükler: Yüksek riskli testler, değerlendirme sistemleri, eğitim politikaları, standart testler, bibliyometrik analiz.

Sorumlu yazar: Vural Tünkler

Türkçe ve Sosyal Bilimler Eğitimi Bölümü, Eğitim Fakültesi, Süleyman Demirel Üniversitesi, Isparta, Türkiye

E-posta: vuraltunkler@sdu.edu.tr **ORCID ID:** <https://orcid.org/0000-0002-3536-968X> **ROR ID:**

<https://ror.org/04fjtte88>

Geliş tarihi: 21.06.2025 **Kabul tarihi:** 16.02.2026 **Yayın Tarihi:** 15.04.2026

Atf Bilgisi: Yazar, G., & Tünkler, V. (2026). 2000-2024 arası yüksek riskli testlerle ilgili makalelerin bibliyometrik analizi. *Ankara University Journal of Faculty of Educational Sciences*, 59(1), 425-473. <https://doi.org/10.30964/auebfd.1724404>

Makaleye ilişkin tüm etik beyanlar, çalışmanın son sayfasında yer almaktadır (Sayfa:473)

Ölçme, eğitim sistemlerinin önemli bir ögesidir. Burdett (2016), öğrencilerin başarılı olmaları ve gelişebilmeleri için eğitimin her kademesindeki geri bildirimlerin yani ölçme sonuçlarının önemini açıkça vurgulamıştır. Ölçme, öğrencilerin bilgi ve becerilerini belirlemeye yönelik çeşitli yöntemleri kapsar. Elde edilen veriler ışığında öğrenci gelişimi, yönlendirmesi ve öğretim uygulamalarının gözden geçirilmesi vb. amaçlar hedeflenir (Rhoten vd., 2003). Bu makalede odaklanılan ölçme türü, çeşitli uzmanlar tarafından hazırlanmış herhangi bir bölge, eyalet, ülke ya da ülkeler kapsamında uygulanan geniş ölçekli standart testlerdir. Bahsi geçen bu geniş ölçekli ölçme yaklaşımı eğitim sistemleri için yeni bir faktör değildir (Verger vd., 2019). Özellikle eğitim sistemlerinin her öğrenci için adil fırsat sunması, mevcut eğitim performansı ve etkililiği hakkında artan endişeler neticesinde politika yapıcılar dikkatlerini eğitim reformlarına yöneltmişlerdir (Camphuijsen vd., 2021). Politikacılar, eğitim kalitesini artırmak ve kamuoyu karşısında şeffaflık ve hesap verebilir bir yapı oluşturmak amacıyla test temelli standart sistemleri tercih etmektedir (Supovitz, 2009). Supovitz (2009) çalışmasında bu sistemleri teşvik eden dört temel teori olduğunu dile getirmiş ve bu teorileri şu şekilde sıralamıştır: (i) motivasyonel teori (test tabanlı değerlendirmenin tüm eğitim paydaşlarını performanslarını iyileştirmeleri noktasında motive edeceği), (ii) uyum teorisi (test tabanlı sistemlerin müfredat, öğretim yöntemleri, değerlendirme vb. tüm eğitim unsurlarının uyum içinde olmasına yardımcı olabileceği), (iii) bilgi teorisi (bu tür sistemlerin eğitimcilere iyileştirme çalışmalarına rehberlik etmelerinde veri sağlaması), (iiii) sembolizm teorisi (hesap verilebilirlik sisteminin paydaşlara önemli değerleri iletmesi). Bu teoriler 1990'lı yıllardan beri test tabanlı değerlendirme sistemlerinin benimsenip yaygınlaşmasına zemin hazırlamış ve bu sebeple 50 ABD eyaletinde yüksek standartlara dayalı eğitim girişimleriyle tüm öğrencileri kapsayan ortak bir standartlar seti oluşturulmak istenmiştir (Goertz ve Duffy, 2003). Standartlara dayalı bu sistemin bileşenleri arasında öğrenci performansını ölçen değerlendirmeler ve bu değerlendirme sonuçlarına odaklanan hesap verilebilirlik yer almaktadır (Goertz ve Duffy, 2003; Levatino, vd., 2024; Ozga, 2009). Xie (2022), hesap verilebilirliği, bir kimseyi kendisine verilen görevden sorumlu tutmak yani önceden belirlenen amaçlara ulaşmak için taraflar arasında yapılan bir sözleşme olarak tanımlamıştır. Öte yandan birçok araştırmacı tarafından hesap verilebilirlik, öğrencilere uygulanan standart testlerden elde edilen sonuçlara göre okulların performansının yorumlanıp bu sonuçlardan okul yöneticileri ve eğitimcileri sorumlu tutma sürecidir (Bae, 2018; Verger vd., 2019). Erdağ (2019) ise özellikle 1980'li yıllardan sonra eğitimde yaşanan sorunların okul yöneticileri ve eğitimcilerden kaynaklandığını ileri süren anlayışın bu sorunları aşmak için okul hesap verilebilirliğini öne çıkardığını dile getirmiştir.

Geçtiğimiz çeyrek asırda test tabanlı hesap verilebilirlik uygulamaları tüm dünyadaki eğitim sistemleri içinde önemli bir yer tutmuştur (Klinger ve Rogers, 2011; Levatino vd., 2024). Performansa dayalı bir hesap verilebilirlik sistemi olan (Kirby ve Stecher, 2004) No Child Left Behind (Hiçbir Çocuk Geride Kalmasın [NCLB]) yasasının 2001 yılında kabul edilmesiyle birlikte öğrenciler, öğretmenler, okul

idarecileri ve okullar hakkında karar alabilmek için standart testlerin kullanımında artış gözlemlenmiştir (von der Embse ve Witmer, 2014). Parkison (2009), NCLB'yi eğitimin önceliklerini belirleyen ve kurumsallaştıran bir ideoloji olarak belirtmiştir. Bu yasa ile mevcut eğitim standartlarının yükseltilmesi ve okullar arasındaki başarı farklarını hesap verilebilirliği kullanarak kapatmak amaçlanmıştır (Tan, 2010). 2015 yılında NCLB yasası Every Student Succeeds Act (Her Öğrenci Başarır [ESSA]) yasasıyla yeniden güncellenmiş (Finn, 2022) ve ESSA yasası ile öğrenci öğrenmelerinin daha geniş şekilde tanımlanması hedeflenmiştir (Bae, 2018). Alanyazına bakıldığında, öğrencilere uygulanan standart testlerden elde edilen sonuçların, öğrencilerin sonraki eğitim süreçlerine dair önemli kararlar (örneğin mezuniyet, sınıfta kalma veya sınıf geçme) almak için kullanılması, okullara sağlanan finansmanın yanı sıra yöneticiler ve öğretmenlerin performansını etkilemesi veya sınav sonuçlarının kamuya açıklanarak kişilerin ve okulların itibarını etkilemesi durumunda bu testler “yüksek riskli test (high-stakes test)” olarak tanımlanmaktadır (McNeil, 2000; Orfield ve Wald, 2000). Bazı ülkelerde belirlenen eğitim hedefleri ve ölçütlerinin gerçekleşmemesi durumunda okul yöneticileri ve personelinin işten çıkarılması, maaş kesintileri veya eğitim kurumlarının kapatılması gibi yaptırımlar ve kurum başarısızlığının ardından gelen öğrenci kayıpları neticesinde bazı bölgelerde büyük ölçekli standart testler karakter olarak “yüksek riskli testler” olarak adlandırılmıştır (Hargreaves, 2020). Ek olarak, uygulanan test sonuçlarına göre bazı öğrenciler eğitim sisteminden dışlanırken, bazıları bir sonraki eğitim kademesine geçmeye hak kazanmaktadır (Hursh, 2005; Jacobs vd., 2024). Bu ülkelere ve yüksek riskli testlere örnek verecek olursak, ABD’de Scholastic Aptitude Test (SAT) ve American College Testing (ACT), İngiltere’de International English Language Testing System (IELTS) ve General Certificate of Secondary Education (GCSE), Çin’de ise Gaokao adlı sınavlar uygulanmaktadır. Bu sınavların sonuçlarına göre yukarıda ifade edilen yaptırımlar söz konusudur (Hargreaves, 2020; Orfield ve Wald, 2000; Wang vd., 2022). Benzer şekilde, Avustralya’da gerçekleştirilen National Assessment Program – Literacy and Numeracy (NAPLAN) sınavının sonuçları, kamuya açık olarak MySchool adlı sitede paylaşılmakta ve bu durum okullar arası karşılaştırmaların yapılmasına imkân tanımaktadır (Thompson ve Lašič, 2011). Düşük sonuçlar elde edildiğinde ise öğretmenler ve okulların performans temelli ek eğitim uygulamaları tasarlamaları istenmekte ve başarısız öğrencilerin ek ders alması beklenmektedir (Thompson, 2013; Thompson ve Lašič, 2011). Türkiye’de uygulanan Liselere Geçiş Sistemi (LGS) ve Yükseköğretim Kurumları Sınavı (YKS) ise, öğrencilerin bir sonraki eğitim kademesine geçmelerinin yanı sıra nitelikli kurumlara kabul edilmelerini sağladıkları için yüksek riskli testler arasında yer almaktadır (Özaslan ve Beyhan, 2023). Tüm bu yaptırımlar, standart testlerin yüksek riskli testler olarak adlandırılmasında önemli bir etkidir. Bahsi geçen sebepler yüzünden eğitim sisteminin tüm paydaşları (öğrenci, öğretmen, yönetici hatta veliler) eğitim sürecinden ziyade standart testlerden elde edilen sonuçlara odaklanmışlardır (Hanushek ve Raymond, 2004).

Hesap verilebilirlik ve yüksek riskli testlerin ardındaki hâkim teori; eğitimcilerin daha fazla çaba gösterip, okulların etkili öğretim yöntemleri tercih etmeleri ve dolayısıyla öğrencilerin daha çok şey öğrenip kaçınılmaz başarıya ulaşmaları olsa da yüksek riskli testlerin etkileri tartışma konusudur (Heilig ve Darling-Hammond, 2008). Falabella (2021), yüksek riskli testler ve hesap verilebilirlik yaklaşımının sadece okul eğitim politikalarıyla kalmayıp maaş politikalarını, öğretmen eğitimini dahası erken çocukluk ve özel eğitimden yükseköğretime kadar birçok alanı etkilediğini belirtmiş ve tüm bunların standarda dayalı eğitim reformlarının etkililiği hakkında yeterli araştırma verisi olmamasına rağmen gerçekleştiğini dile getirmiştir. Sloane ve Kelly (2003), yüksek riskli testlerle ilgili tartışmaları dört ana konu başlığıyla sıralamış ve tartışmaların yüksek riskli testlerin sonuçları ve uygulama yöntemlerinin yeniden değerlendirilmesi için yeterli dayanağı oluşturduğunu belirtmiştir. Bu konular: (i) tercih edilen sınav türleri, (ii) testlerin öğrencilerin moral ve motivasyonuna etkisi, (iii) uygulanan testlerin müfredatla olan uyumu, (iv) öğrenmenin değerlendirilmesi ve öğrenme için değerlendirme arasındaki farktır. Yüksek riskli testler hakkındaki en yaygın endişelerden birisi, test sonuçlarını iyileştirmek için verilen eğitimin teste göre daraltılmasıdır (Hargreaves, 2020; Hatfield ve Soløst, 2025; Zakharov ve Carnoy, 2021). Eğitimin daraltılmasındaki kasıt teste tabi tutulan ders ve konuları önemseyip tabi olmayanları müfredattan dışlamaktır (Au, 2011; Gulek, 2003). Ancak bu her zaman mümkün değildir. Bazı araştırmacılar yüksek riskli testlerin resmi müfredatta yer almayan dolayısıyla öğrencilerin hiç çalışmadıkları bazı konuları ölçtüğünü belirtmişlerdir (Arviv Elyashiv ve Avidov-Ungar, 2024; Thomas, 2005). Yüksek riskli testlerin doğurduğu bir başka problem ise öğrenci ve velilerin alanyazında “gölge eğitim” olarak da adlandırılan özel derslere olan yoğun talebidir (Yung, 2021). Özel derse olan talebin sebebi ise özel ders veren öğretmenlerin öğrencilere tamamen sınav odaklı ve sonuçları iyileştirmeye yönelik eğitim içerikleri sunmasıdır (McCoy ve Byrne, 2024). Bu durum düşük gelirli öğrenciler için bir dezavantaj oluşturarak eğitimdeki eşitsizliği ortadan kaldırmak yerine dezavantajlı ve avantajlı öğrenciler arasındaki başarı farkını daha da açmaktadır (Hursh, 2005). Yüksek riskli testlerin yol açtığı bir diğer problem ise öğrenciler ve eğitimciler üzerinde kurduğu baskıdır. Çok sayıda öğrenci testlere verilen önemden dolayı kaygı duymakta, bunun sonucunda akademik performansları (Hatfield ve Soløst, 2025; von der Embse ve Witmer, 2014) ve ruh sağlıkları olumsuz etkilenmektedir (Jerrim, 2023). Öğrenci performans sonuçlarına göre eğitimcilerin değerlendirilmesi ise okul yöneticileri ve öğretmenler üzerinde ciddi bir baskı kurmuştur (Arviv Elyashiv ve Avidov-Ungar, 2024). Kenney ve arkadaşları (2024), politika yapımcıların eğitimciler üzerindeki baskılarını göz önüne alarak, bu baskıların eğitim sürecine olan olumsuz etkilerini kabul etmeleri gerektiğini dile getirmiştir. Tüm bu bulgular bizlere yüksek riskli testlerin sınıf düzeyinde içerik, bilgi formları ve pedagojiler üzerinde kontrolünü göstermektedir (Au, 2007). Ayrıca hesap verilebilirliğe dayalı eğitim sistemlerinde okullar arası karşılaştırmaların her okulun kendi bağlamı gözetilerek yapılması gerektiği de bir başka eleştiridir (Munoz-Chereau vd., 2022). Klinger ve Rogers (2011) ise tüm bu olumsuz araştırma sonuçlarına

rağmen yüksek riskli test sonuçlarının raporlanması ve kullanımında gerçekleştirilecek değişimlerin öğretim ve öğrenme sürecini olumlu yönde etkileyeceğini savunmuştur.

İlgili alanyazın incelendiğinde eğitimde yüksek riskli testler ve bu bağlamda hesap verilebilirlik uygulamaları ile ilgili çeşitli ülkelerde ve konularda yapılmış birçok araştırmayla karşılaşmıştır (Au, 2016; Burgess vd., 2021; Elo ve Nygren-Landgärds, 2021; Erduran vd., 2020; Escamilla vd., 2005; Fidjeland, 2023; Fischer vd., 2020; Fitchett ve Heafner, 2010; Gorgodze ve Chakhaia, 2021; Hammack ve Wilson, 2019; Hilton ve Saunders, 2024; Hofflinger ve von Hippel, 2020; Kavanagh ve Fisher-Ari, 2020; Knoester ve Meshulam, 2022; McElroy, 2023; Mintz ve Kelly, 2021; Moore, 2025; Ober vd., 2022; Pace, 2011; Trujillo vd., 2021; Yin ve Mu, 2022). Alanyazında manuel olarak analiz edilemeyecek kadar çok veriyle karşılaşıldığı için bibliyometrik analiz, yüksek riskli testlerle ilgili yapılmış makalelerin genel eğilimi ve entelektüel yapısını ortaya çıkarmada kullanılmıştır. Çünkü bibliyometrik analiz, geniş ölçekli bilimsel verilerin işlenebilirliğini ve araştırmalara yaptığı değerli katkıları ortaya koymada önemli bir araçtır (Passas, 2024). Bu çalışma yüksek riskli testler ile ilgili yapılmış makaleleri bibliyometrik analiz yöntemiyle inceleyerek alan yazına ve ilgili araştırmacılara katkı sunmayı hedeflemiştir.

Çalışmanın amacı eğitim alanında yüksek riskli testler ile ilgili yapılmış olan makaleleri bibliyometrik analiz yöntemiyle incelemektir. Bu bağlamda araştırma boyunca cevapları aranan sorular aşağıdaki gibidir:

1. Yüksek riskli testlere yönelik makalelerin yayın yılına ait dağılımları nasıldır?
2. Yüksek riskli testlere yönelik makalelerin yazar, kuruluş ve ülkelere ait iş birliği konusunda nasıl bir yapı ortaya çıkmaktadır?
3. Yüksek riskli testlere yönelik makalelerin dergi, referans ve yazarlara ait ortak atıf ağı konusunda nasıl bir yapı ortaya çıkmaktadır?
4. Yüksek riskli testlere yönelik makalelerin ortak kelime ağı konusunda nasıl bir yapı ortaya çıkmaktadır?

Yöntem

Bibliyometrik analiz yöntemi, yazarlar, makaleler, dergiler, atıflar ve anahtar kelimelere göre sınıflandırmalar yapabildiğimiz herhangi bir araştırma alanının genel görüntüsüne ulaşmak için yazılı materyallerin nicel analizi olarak tanımlanabilir (Merigó ve Yang, 2017). Böyle bir analiz belirli bir konudaki alanyazının yani bilimsel yayınların tanımlanmasına dayanır (Ellegaard ve Wallin, 2015). Bibliyometrik analiz, teker teker incelenemeyecek kadar çok verinin titiz tekniklerle anlamlandırılması ve bu sayede köklü araştırma alanlarının bilimsel bilgisini ve geçmişten bugüne evrimsel sürecini ortaya çıkarmak ve haritalandırmak için oldukça elverişlidir (Donthu vd., 2021; Passas, 2024). Bu yüzden tüm süreçleriyle titiz çalışılmış ve iyi hazırlanmış bir bibliyometrik araştırma ilgili alan için güçlü bir dayanak sağlayabilir (Passas, 2024). Bu analiz yöntemi kapsamlı alanyazının

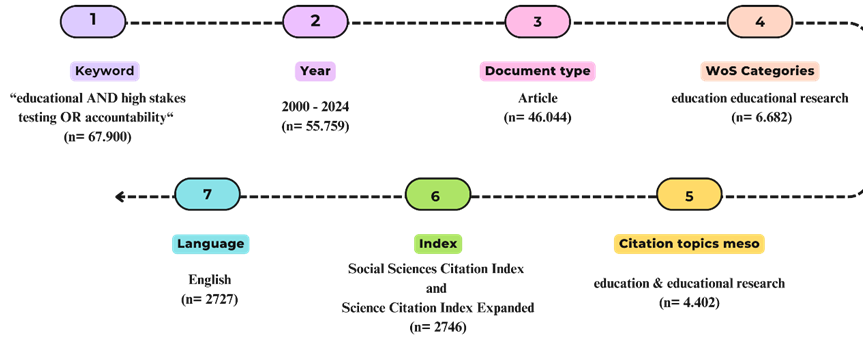
performans değerlendirmesine olanak sağladığı için giderek daha fazla tercih edilmektedir (Wallin, 2005). Bu sebeple çalışmada yüksek riskli testlerle ilgili makalelerin eğilimleri ve haritasını ortaya çıkarabilmek için bibliyometrik analiz yönteminden faydalanılmıştır.

Verilerin Toplanması ve Analizi

Araştırma sürecine ilk olarak araştırma konusunun belirlenmesiyle başlanmış, ardından araştırma soruları netleştirilmiştir. Veri toplamak için kullanılacak anahtar kelimelerin seçiminde ilgili alanyazın incelenip alanda en sık kullanılan terimler ve kavramlar tespit edilerek denenmiştir. Elde edilen sonuçlar karşılaştırılarak en kapsamlı ve en sağlıklı verilerin elde edilmesini sağlayacak anahtar kelimeler belirlenmiştir. Ayrıca, yüksek riskli testlerle ilgili alanyazın incelemesi sonucunda, “high-stakes testing” ve “accountability” terimlerinin en sık kullanılan terimler olduğu görülmüştür. Seçilen anahtar kelimelerin kullanım sıklığı ve kapsayıcılığı göz önüne alındığında bu çalışmanın amacı doğrultusunda geçerli ve güvenilir olduğu düşünülmektedir. Analizi yapılacak makalelere ulaşmak için, atıf dizinleri arasında en eski ve en kapsamlı kayıtlara sahip olan ve kullanılacak analiz programına doğrudan veri aktarımı sağlayan WoS veri tabanı tercih edilmiştir (Ellegaard ve Wallin, 2015). 16 Ekim 2024 tarihinde WoS veri tabanında “gelişmiş arama” seçeneği kullanılarak “topic” etiketi ile “education AND high stakes testing OR accountability” anahtar kelimeleri aratılmış ve 67.900 çalışmaya ulaşılmıştır. Analizlerin ve bulguların daha sağlıklı ve güvenilir veriler sağlaması için bazı filtrelemelere ihtiyaç duyulmuştur. Yapılan filtreleme adımları Şekil 1’de gösterilmiştir.

Şekil 1

Alanyazın Taraması Aşamasında Yapılan Filtreleme Adımlarının Gösterimi



Tüm bu filtreleme aşamalarından sonra makalelerin özet kısımları incelenerek yüksek riskli test konusunun merkezde olup olmadığı belirlenmiş ve tekrar eden makaleler ayıklanmıştır. Sonuç olarak 1.947 makale çalışmaya dâhil edilmiştir. Sonrasında ilk olarak Microsoft Word programı kullanılarak yüksek riskli testlerle

ilgili yapılmış makalelerin yayın yıllarına ait bir grafik oluşturulmuştur. Eldeki verilerin bibliyografik analizini pragmatik şekilde yapabilmek (Donthu vd., 2021) için VOSviewer programının son sürümü olan 1.6.20 kullanılmıştır. VOSviewer yazılımının tercih edilmesindeki en büyük etken WoS veri tabanından doğrudan veri setlerinin bu yazılıma aktarılabilmesidir. VOSviewer, düğüm ağı (nodal network) görselleştirmesinde bağlantıların sayısı ve toplam gücü gibi iki standart ağırlık kullanarak düğüm boyutları ve bağlantı çizgileri ile bağlantıların önemini ve gücünü bizlere sunmaktadır (Donthu vd., 2020). Bibliyometrik analiz yapabilmek veya yapılan analizleri yorumlayabilmek için bazı temel terimlerin ne anlama geldiğini bilmek gereklidir. Van Eck ve Waltman (2023) tarafından yapılan açıklamalara göre temel terimlerden bazıları şöyledir:

- Bağlantı: İki düğüm (öge) arasındaki ilişkiyi ifade etmektedir. Bu değer ne kadar yüksekse, iki öge arasındaki ilişki o kadar güçlüdür.
- Toplam bağlantı gücü (TBG): Bir düğümün yani ögenin diğer tüm ögelerle oluşturduğu tüm bağlantıların gücü demektir.
- Düğüm (öge): Analizde incelenen yazar, kuruluş, ülke veya anahtar kelimelerin ağ haritasında gösterilme şeklidir.
- Küme: Ağ haritasını oluşturan benzer özelliklere sahip ögelerin yer aldığı alt gruplardır. Her öge yalnızca bir küme içinde yer alabilir.

Çeşitli analizlerin yapılabildiği bu programda ortak yazarlık, ortak atıf ve birlikte bulunma analizleri kullanılarak veri setinden ağ haritaları ve VOSviewer yazılımının sunduğu veriler doğrultusunda tablolar oluşturulmuştur. Bu üç analiz türünün tercih edilme sebebi, çalışmanın temel amacının ilgili alanyazının iş birliği yapısını, atıf ilişkilerini ve kavramsal eğilimlerini açığa çıkarmaktır. Yapılan ortak yazarlık analizinin amacı birden fazla yazar, kuruluş ve ülkenin yazmış olduğu dokümanları inceleyerek aralarındaki iş birliği ağını ortaya çıkarmaktır (Merigó vd., 2018). Ortak atıf analizinin amacı, farklı dergilerde yayımlanan iki dokümanın, başka bir dergideki üçüncü bir dokümandan atıf alması durumunu ortaya çıkarmaktır (Small, 1973), birlikte bulunma analizi dokümanlarda en sık kullanılan anahtar kelimeleri belirlemeyi hedeflemektedir (Merigó vd., 2018). Tüm bu analizler yapılırken, analize dahil edilecek ögelerin minimum sayısını belirleyen bir eşik değer bulunmaktadır (Karakus vd., 2019). Kısaca eşik değer bir ögenin analize dahil edilebilmesi için sağlaması gereken minimum koşulu ifade eder. Örneğin ortak yazarlık analizinde bir yazarın analize dahil olabilmesi için eşik değer 5 olarak belirlenmişse bu en az 5 dokümana sahip olan yazarların analize dahil edildiği anlamına gelir. Bu çalışmada eşik değerler VOSviewer yazılımının oluşturduğu ağ görsellerinin anlaşılabilirliği dikkate alınarak deneme yoluyla belirlenmiştir. Başlangıçta düşük değerler denenmiş, aşırı karmaşık ağ haritalarında anlaşılabilirliği sağlamak amacıyla eşik değer kademeli olarak yükseltilmiştir. Anlamli kümelerin ve ağların yorumlanabilir halde geldiği noktada eşik değer sabitlenmiştir. Analizlerin hangi eşik değer dikkate alınarak yapıldığına her bulgunun altında yer verilmiştir. Eşik değer belirlenmesinin ardından yazılımın sunduğu veriler doğrultusunda konunun öne çıkanları (yazar,

dergi, referans vb.) toplam bağlantı gücü dikkate alınarak tablolaştırılmıştır. Oluşturulan ağ haritaları ve tablolara bir sonraki bölüm olan bulgular başlığı altında yer verilerek her biri yorumlanmıştır.

Bulgular

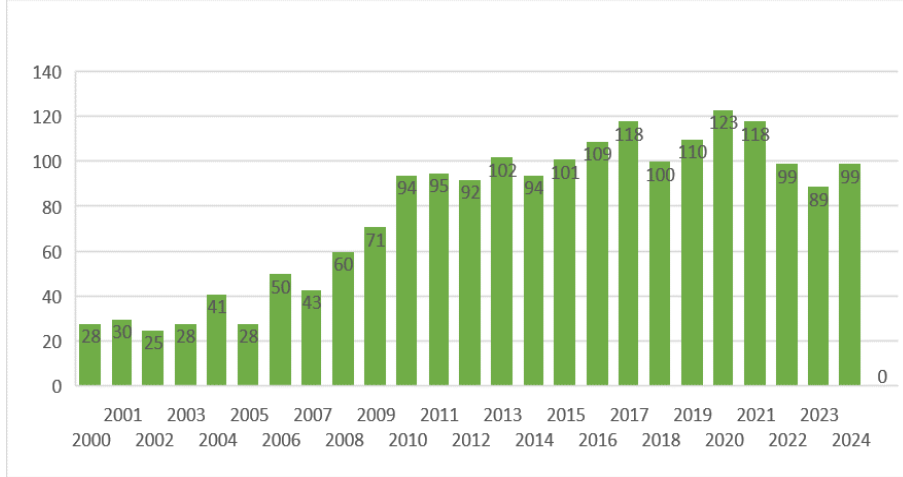
Bu bölümde WoS veri tabanından elde edilen “yüksek riskli testler” konulu 1.947 makalenin bibliyometrik analizleri ve bu analizlerin yorumları yer almaktadır.

Yapılan Çalışmaların Yıllara Göre Dağılımı

Eğitimde yüksek riskli testler üzerine yapılan makalelerin yıllara göre dağılımı Şekil 2’de verilmiştir.

Şekil 2

Eğitimde Yüksek Riskli Testlere Yönelik Çalışmaların Yayın Yıllarına Göre Dağılımı



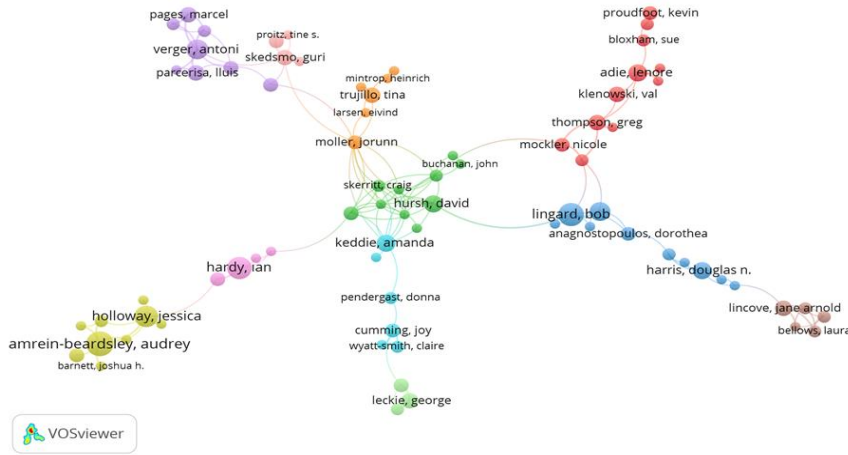
Şekilden anlaşılacağı üzere yıllara göre makale sayısında hafif dalgalanmalar olsa da genel olarak yüksek riskli testlerle ilgili yapılan makalelerin sayısında diğer yıllara oranla 2010 yılından sonra oldukça yüksek bir artış gözlenmektedir. En fazla makalenin yayınladığı yılların başında sırasıyla 2020 (n=123), eşit yayın sayısına sahip 2017 ve 2021 (n=118) yılları gelmektedir. En az makalenin yapılmış olduğu yıllar ise sırasıyla 2002 (n=25), eşit yayın sayısına sahip 2000–2003–2005 (n=28) ve 2001 (n=30) yıllarıdır.

Ortak Yazarlık Analizi (Yazarlar)

Şekil 3'te iş birliği içinde bulunan yazarlara ve bu iş birliği neticesinde ortaya çıkan ağ haritasına yer verilmiştir. Ortak yazarlık analize tabi tutulan 1.947 makaleden elde edilen sonuca göre en az bir dokümana sahip yazar sayısı toplamda 3.396'dır. Doküman sayısı yani eşik değer 2 olarak belirlendiği takdirde sadece 488 yazar bu eşiği aşmaktadır. Yazarlardan 76 tanesinin bağlantı kurduğu görülmüştür. Bu bulgu, eşiği karşılayan 488 yazar içerisinde yalnızca 76 yazarın birbirleriyle çalışarak iş birliği ağı oluşturdukları anlamına gelmektedir. Bu doğrultuda Şekil 3'te gösterilen ağ haritası 11 küme altında toplamda 76 düğümden oluşmaktadır ve 188 toplam bağlantı gücüne sahiptir. Antoni Verger adlı yazar 16 toplam bağlantı gücü ile en çok ortak yayın yapan yazar olarak karşımıza çıkmaktadır. Yazarın bulunduğu küme içerisinde Lluís Parcerisa, Marcel Pagès ve Gerard Ferrer-Esteban adlı yazarlarla arasındaki bağlantı gücünün aynı kümedeki diğer yazarlara göre daha fazla olduğu sonucuna ulaşılmıştır. Ayrıca ağ haritasına göre Jorunn Møller adlı yazarın en fazla bağlantı gücüne yani en geniş ağ yapısına sahip yazar olduğu görülmüş fakat bu durum yazarın toplam bağlantı gücünü yükseltmede yetersiz kalmıştır. Bu sebeple Jorunn Møller'in iş birliği yoğunluğu düşük olduğu için en çok iş birliği yapan ilk 10 yazar arasına girememiş ve tabloda gösterilmemiştir.

Şekil 3

En Çok İş Birliği İçinde Bulunan Yazarlar Ağ Haritası



Tablo 1'e bakıldığında Antony Verger ve Meg Maguire gibi yazarların doküman sayısının Audrey Amrein-Beardsley ve atıf sayısının Barry Fishman adlı yazarlardan az olmasına rağmen toplam bağlantı güçleri fazla olduğu için listenin ilk sıralarında yer almıştır. Bu durum Verger ve Maguire'in araştırma ağı kapsamında daha merkezi bir yere sahip oldukları ve daha yoğun iş birlikleri yürüttükleri anlamına gelmektedir.

Doküman ve atıf sayısı görece yüksek olan ama toplam bağlantı gücü düşük olduğu için ilk sıralarda yer almayan bazı yazarların ise daha çok bireysel çalıştıklarını veya sınırlı sayıda iş birlikleri gerçekleştirdiklerini söylemek mümkündür.

Tablo 1

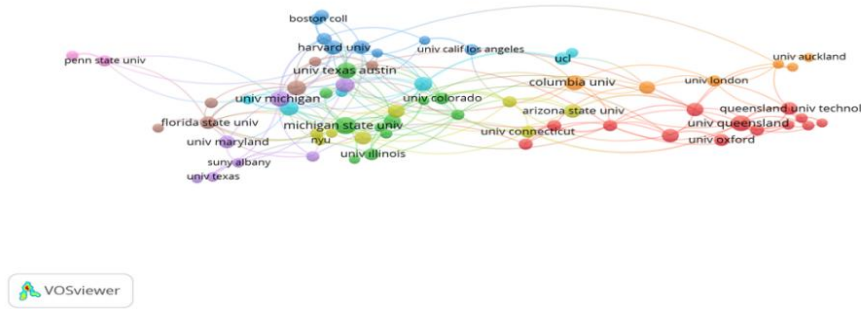
Yüksek Riskli Test Alanında En Çok İş Birliği Yapan Yazarlar (Toplam Bağlantı Gücüne Göre Sıralanmıştır)

Yazarlar	Doküman	Atıf	TBG
Verger, A.	8	219	16
Maguire, M.	5	184	14
Amrein-Beardsley, A.	12	85	13
Saultz, A.	7	104	13
Eisenkraft, A.	2	89	12
Fischer, C.	2	89	12
Fishman, B.	3	295	12
Frumin, K.	2	89	12
Gewirtz, S.	4	77	12
Lawrenz, F.	2	89	12

Ortak Yazarlık Analizi (Kuruluşlar)

Şekil 4'te iş birliği içinde bulunan kuruluşlara ve bu iş birliği sonucunda ortaya çıkan ağ haritasına yer verilmiştir. Ortak yazarlık analizine tabi tutulan 1.947 makaleden elde edilen sonuca göre en az bir dokümana sahip 1.079 kuruluş olduğu görülmüştür. Eşik değeri 10 olarak belirlendiğinde ise sadece 68 kuruluş bu eşiği aşabilmiştir. Bu kuruluşların 67 tanesinin birbiriyle bağlantılı olduğu yani iş birliği ağı sağladığı, 9 küme ve 244 toplam bağlantı gücü oluşturdukları belirlenmiştir. Ağ haritasına göre en fazla iş birliği yapan kuruluş Stanford Üniversitesi'dir. Fakat kuruluşlar arasındaki en çok iş birliği 4 bağlantı gücü ile Florida Üniversitesi ve Pennsylvania Devlet Üniversitesi arasında gerçekleşmiştir.

Şekil 4
En Çok İş Birliği İçinde Bulunan Kuruluşlar Ağ Haritası



Tablo 2’de toplam bağlantı gücüne dayalı olarak en çok iş birliği sağlayan ilk 10 kuruluşa yer verilmiştir. 16 toplam bağlantı gücüyle listenin ilk sırasında yer alan Stanford Üniversitesi, 2 bağlantı gücü ile en çok Columbia Üniversitesi ve Vanderbilt Üniversitesi ile iş birliği yapmıştır. Listedeki diğer kuruluşların atıf sayıları veya doküman sayıları fazla olsa da daha az iş birliğinde buldukları için toplam bağlantı güçlerinin az olduğunu ve listenin ortalarında veya sonlarında kendilerine yer bulduklarını söylemek mümkündür.

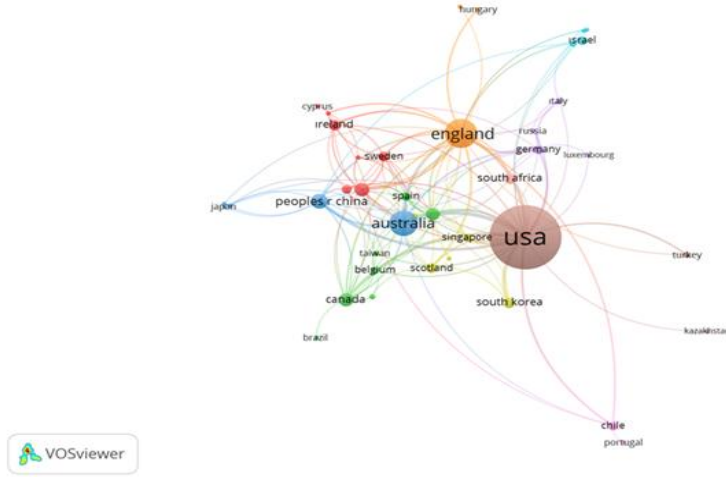
Tablo 2
Yüksek Riskli Test Alanında En Çok İş Birliği Yapan Kuruluşlar (Toplam Bağlantı Gücüne Göre Sıralanmıştır)

Kuruluşlar	Doküman	Atıf	TBG
Stanford Üniversitesi	28	1247	16
Michigan Üniversitesi	31	1222	15
Chapel Hill Kuzey Karolina Üniversitesi	30	771	14
Florida Devlet Üniversitesi	21	886	13
Michigan Devlet Üniversitesi	38	541	13
Virginia Üniversitesi	26	977	13
Rochester Üniversitesi	17	976	12
Güney Karolina Üniversitesi	14	173	12
Wisconsin Üniversitesi	28	1082	12
Deakin Üniversitesi	23	677	11

Ortak Yazarlık Analizi (Ülkeler)

Şekil 5’te iş birliği içinde bulunan ülkeler ve bu iş birliği sonucunda ortaya çıkan ağ haritasına yer verilmiştir. Ortak yazarlık analizine tabi tutulan 1.947 makaleden elde edilen sonuca göre en az bir dokümana sahip 77 ülke olduğu görülmüştür. Eşik değer 2 olarak belirlendiğinde ise sadece 45 ülke bu eşik değeri aşabilmiştir. Bu ülkelerden 41’inin birbiri ile bağlantılı olduğu, 10 küme altında ve 350 toplam bağlantı gücü oluşturdukları görülmüştür. Ağ haritasına göre birbiri arasında en çok iş birliği yapan iki ülkenin ABD ve Avustralya olduğu sonucuna ulaşılmıştır.

Şekil 5
En Çok İş Birliği İçinde Bulunan Ülkeler Ağ Haritası



Tablo 3’te toplam bağlantı gücüne dayalı olarak en çok iş birliği sağlayan ilk 10 ülkeye yer verilmiştir. Tablonun ilk sırasında 133 toplam bağlantı gücüyle ABD bulunmaktadır. Aralarındaki bağlantı güçleri dikkate alındığında, ABD sırasıyla en çok Avustralya, İngiltere, Güney Kore ve Kanada ülkeleriyle iş birliğinde bulunmuştur. Ancak Güney Kore toplam bağlantı gücü düşük olduğu için ilk 10 ülke arasına girememiştir.

Tablo 3

Yüksek Riskli Test Alanında En Çok İş Birliği Yapan Ülkeler (Toplam Bağlantı Gücüne Göre Sıralanmıştır)

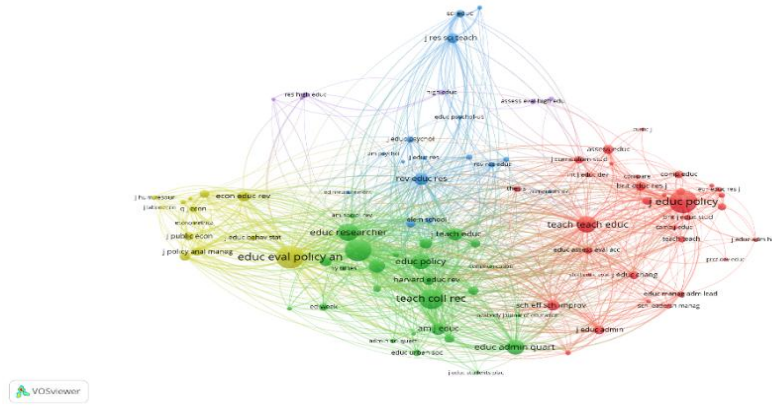
Ülkeler	Doküman	Atıf	TBG
ABD	1124	31404	133
İngiltere	227	4817	100
Avustralya	184	6236	60
Kanada	50	880	34
Çin Halk Cumhuriyeti	61	1274	34
Norveç	47	743	30
Almanya	22	299	27
Hollanda	43	1527	24
İsveç	29	485	24
İskoçya	27	1380	19

Ortak Atıf Analizi (Dergiler)

Şekil 6'da ortak atıf analizi sonucunda ortaya çıkan bilimsel dergilere ait ağ haritası görseline yer verilmiştir. Analize tabi tutulan ve en az bir kez atıf almış 37.053 bilimsel dergiden, atıf eşiği 100 olarak belirlendiğinde sadece 93 derginin bu eşiği aştığı ve haritada kendine yer edindiği gözlemlenmiştir. Bu 93 dergi 5 küme ve 327.273 toplam bağlantı gücü oluşturmuştur. Oluşan ağ haritasına göre ortak atıf bağlantısının en çok Educational Evaluation and Policy Analysis ve American Educational Research Journal dergileri arasında olduğu tespit edilmiştir.

Şekil 6

Ortak Atıf Analizine Göre Dergi Ağ Haritası



Tablo 4'te toplam bağlantı gücüne dayalı olarak en etkin ilk 10 bilimsel dergiye yer verilmiştir. Tablonun ilk sırasında 35.465 bağlantı gücü ile American Educational

Research Journal dergisi yer almaktadır. Bu derginin ağ haritasında yer alan diğer tüm dergilerle bağlantılı olduğu bulunmuştur.

Tablo 4
En Etkili 10 Dergi (Toplam Bağlantı Gücüne Göre Sıralanmıştır)

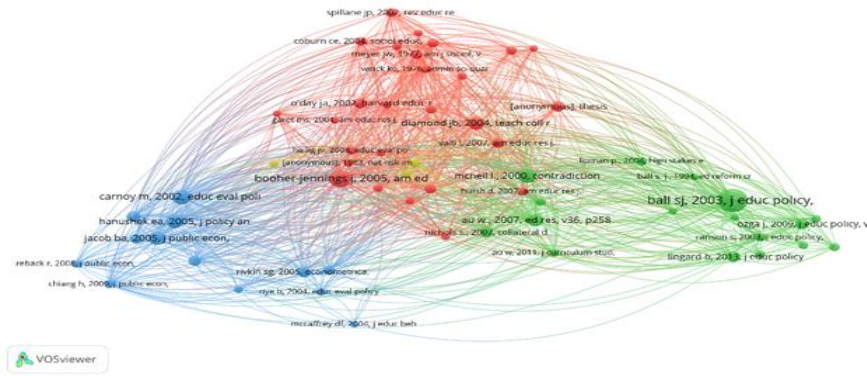
Alıntılanan Dergi	Atıf	TBG
American Educational Research Journal	1396	35465
Educational Evaluation and Policy Analysis	1488	34233
Educational Researcher	996	24297
Journal of Education Policy	1524	21830
Teaching and Teacher Education	920	21676
Educational Administration Quarterly	800	20669
Educational Policy	864	19855
American Journal of Education	596	15280
Review of Educational Research	606	15053
Journal of Teacher Education	637	13639

Ortak Atıf Analizi (Referanslar)

Şekil 7’de ortak atıf analizi sonucunda ortaya çıkan referanslara ait ağ haritası görseline yer verilmiştir. Referansların haritada olabilmesi için eşik değeri 30 olarak belirlenmiştir. Toplamda 60 referans bu eşiği aşmış ve 4 küme altında toplanıp 4.358 toplam bağlantı gücü oluşturmuşlardır. Oluşan ağ haritasına göre ortak atıf ilişkisi en güçlü olan referans 458 toplam bağlantı gücüyle Booher-Jennings (2005) olarak belirlenmiştir. Bu referansın ortak atıf ilişkisi içinde olduğu diğer referanslar incelendiğinde, 27 bağlantı gücüyle en fazla Diamond ve Spillane (2004) ile bağlantılı olduğu sonucuna ulaşılmıştır.

Şekil 7

Ortak Atıf Analizine Göre Referanslar Ağ Haritası



Tablo 5'te toplam bağlantı gücüne dayalı olarak en etkili ilk 10 referansa yer verilmiştir. Carnoy ve Loeb (2002) ve Hanushek ve Raymond (2005) referansları, aralarındaki 48 bağlantı gücüyle ortak atıf ilişkisi en yüksek referanslar olarak belirlenmişlerdir. Ancak Booher-Jennings (2005) 55 referansla ilişki kurarak 458 toplam bağlantı gücüyle Carnoy ve Loeb (2002) ve Hanushek ve Raymond (2005) referanslarının önüne geçerek listenin ilk sırasında yer almıştır.

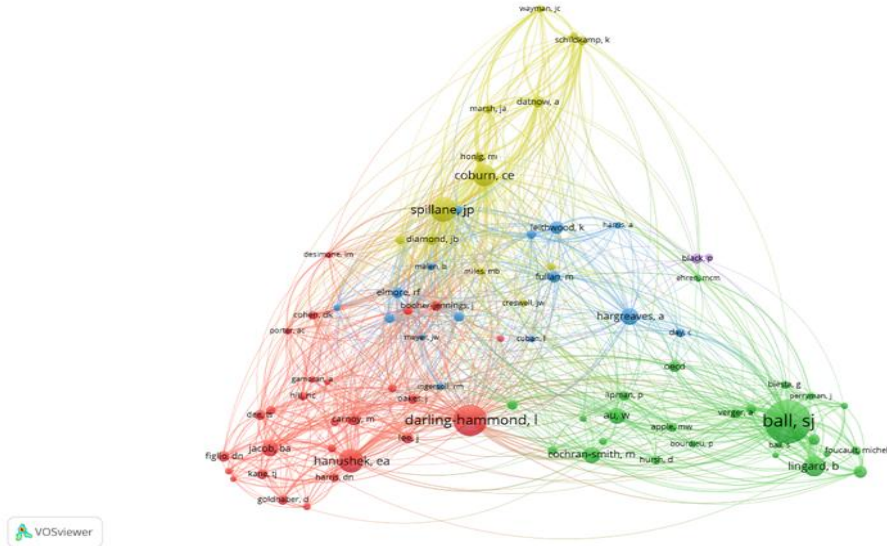
Tablo 5
En Etkili 10 Referans (Toplam Bağlantı Gücüne Göre Sıralanmıştır)

Alıntılanan Referans	Atıf	TBG
Booher-Jennings (2005)	114	458
Carnoy & Loeb (2002)	90	359
Hanushek & Raymond (2005)	80	324
Diamond & Spillane (2004)	76	305
Jacob (2005)	82	303
Ball (2003)	171	301
Dee & Jacob (2011)	58	274
McNeil (2000)	83	226
Au (2007)	71	224
Neal & Schanzenbach (2010)	47	220

Ortak Atıf Analizi (Yazarlar)

Şekil 8'de ortak atıf analizi sonucunda ortaya çıkan yazarlar ağ haritası görseline yer verilmiştir. Eşik değer yani yazarların sahip olması gereken atıf sayısı 80 olarak belirlendiğinde 82 yazar bu eşiği aşmış haritada yer almıştır. Bu yazarlar 51.289 toplam bağlantı gücüyle 5 küme altında toplanmıştır. Ağ haritasına göre James P. Spillane adlı yazarın 4.497 toplam bağlantı gücüyle ortak atıf ilişkisi en güçlü yazar olduğu sonucuna ulaşılmıştır. James P. Spillane ve Cynthia E. Coburn, 523 bağlantı gücüyle aralarında en fazla bağlantı gücü oluşan iki yazar olarak tespit edilmiştir.

Şekil 8
Ortak Atıf Analizine Göre Yazarlar Ağı Haritası



Tablo 6’da toplam bağlantı gücüne dayalı olarak en etkili ilk 10 yazara yer verilmiştir. James P. Spillane adlı yazarın atıf sayısı Linda Darling-Hammond ve Stephen J. Ball adlı yazarlardan az olmasına rağmen en yüksek toplam bağlantı gücüne sahip olan yazar olduğu için tablonun ilk sırasında yer almıştır.

Tablo 6
En Etkili 10 Yazar (Toplam Bağlantı Gücüne Göre Sıralanmıştır)

Yazarlar	Atıf	TBG
Spillane, J.P	361	4497
Ball, S.J.	610	4427
Coburn, C.E.	301	3808
Darling-Hammond, I.	462	3800
Hanushek, E.A.	340	2904
Datnow, A.	149	2224
Lingard, B.	279	2186
Diamond, J.B.	149	1978
Au, W.	226	1874
Jacob, B.A.	216	1753

Tablo 7
En Fazla Kullanılan İlk 10 Anahtar Kelime (Toplam Bağlantı Gücüne Göre Sıralanmıştır)

Anahtar Kelime	Birlikte Bulunma	TBG
Accountability	542	848
Educational Policy	125	297
Assessment	98	165
Educational Reform	43	119
Policy	53	105
Leadership	43	90
Equity	46	87
Education Policy	59	83
Teacher Evaluation	40	78
Teacher Education	52	77

Tartışma, Sonuç ve Öneriler

Bu çalışmanın amacı, yüksek riskli testlerle ilgili yapılmış makaleleri bibliyometrik analiz yöntemiyle inceleyerek ilgili alanyazının genel eğilimini ortaya çıkarmaktır. WoS veri tabanından elde edilen 1.947 makaleye göre, 2000-2024 yılları arasında yüksek riskli testlerle ilgili yapılan çalışmaların sayısında özellikle 2010'lu yıllardan sonra ciddi bir artış olduğu gözlemlenmiştir. Özellikle 2000'li yılların başındaki çalışma sayılarının az olmasının nedeni, yüksek riskli testlerin o yıllarda henüz eğitim sistemlerinde tam olarak benimsenmemiş veya kendine bir uygulama alanı bulamamış olmasından kaynaklanıyor olabilir. ABD'de 2001 yılında kabul edilen NCLB yasasının standart testlerin kullanımını zorunlu kılması (Sloane ve Kelly, 2003) ve bu yasanın 2015 yılında ESSA yasasıyla güncellenmesi (Finn, 2022) alandaki araştırma sayısının yıllar içindeki artışını doğrudan etkilemiştir. Yüksek riskli test ve hesap verilebilirlik sistemlerinin eğitim sistemleri içindeki rolü ve yaşanan teknolojik gelişmelerin (dijital ölçme araçları, yapay zekâ destekli ölçme sistemleri) ölçme değerlendirme alanına getirdiği yenilikler göz önüne alındığında bu alanda gelecek yıllarda yapılacak olan çalışma sayılarında artış yaşanabilir.

Yapılan ortak yazarlık analizine göre en çok çalışmada adı geçen yazar Audrey Amrein-Beardsley olmasına rağmen, toplam bağlantı gücünün yani iş birliği yoğunluğunun yüksek olması sebebiyle Antony Verger adlı yazarın en çok iş birliği sağlayan yazar olduğu belirlenmiştir. Bu bulgu Verger'in geniş etki alanı olan bilimsel bir üretim ağına sahip olduğunu göstermektedir. Genel olarak Verger, eğitim politikaları, eğitimde özelleştirme ve test tabanlı hesap verilebilirlik gibi konu alanlarında çalışmalar yürütmüştür. Bu durum Verger'in farklı araştırmacılarla iş birliği sağlamasının önemli nedenlerinden biri olabilir. Öte yandan bu bulgu yayın sayısından ziyade iş birliği çeşitliliği ile elde edilen etkinin önemini ortaya koymaktadır. Bu yazarların çalışmalarını takip etmek ilgili alanda çalışacak olan araştırmacılara büyük oranda katkı sağlayabilir. Ortak yazarlık analizi sonucunda ulaşılan diğer bulgu ise en çok iş birliği sağlayan kuruluşun Stanford Üniversitesi

olmasıdır. Michigan Devlet Üniversitesi, Michigan Üniversitesi ve Chapel Hill Kuzey Karolina Üniversiteleri Stanford Üniversitesine göre daha fazla çalışmaya sahip olsalar da Stanford Üniversitesi 14 farklı kuruluşla iş birliği yaparak ağ yapısını genişletmeyi başarmıştır. Ortak yazarlık analizi sonuçlarına göre son bulgu ise ülkeler arası iş birliğinde ilk sırada yer alan ABD'dir. Doküman sayısı, atıf sayısı ve toplam bağlantı gücü sıralamasında diğer ülkelere fark atarak bu alanda liderliği elinde tutmaktadır. ABD'nin eğitimde standartlaşma ve hesap verilebilirlik politikaları (NCLB ve ESSA) yönünde attığı öncü adımlar göz önüne alındığında bu sonuca şaşırılmaması gerekir. Yüksek riskli testlerle ilgili çalışma yürütecek yazarların Stanford Üniversitesi veya -ABD'nin bu alandaki liderliği dikkate alındığında- ABD'li araştırmacılar ve kuruluşlarla iş birliği yapması araştırmaların görünürlüğünü artırabilir.

Araştırmada toplanan veriler doğrudan atıf analizi yerine dergi, makale ve yazarların ortak atıf analizi yapılarak alanın entelektüel yapısı ortaya çıkarılmaya çalışılmıştır. Bu bağlamda ortak atıf analizine göre toplam bağlantı gücü dikkate alındığında ortak atıf ilişkisi en güçlü olan bilimsel dergi American Educational Research Journal adlı dergidir. Bu dergi ağ haritasındaki diğer tüm dergilerle ortak atıfa sahiptir. Bu durum derginin yüksek riskli testlerle ilgili bilgi üretiminde merkezi bir konumda olduğunu göstermektedir. Ayrıca bu derginin aynı anda ortak atıf ilişkisinin en güçlü olduğu derginin Educational Evaluation and Policy Analysis dergisi olduğu ve bu ilişkinin ağ haritası genelinde de en kuvvetli ortak atıf ilişkisi taşıdığı tespit edilmiştir. Bu iki derginin araştırmacılar tarafından sıklıkla birlikte referans gösterilmesi yüksek riskli testlerle ilgili çalışmalara yön veren bir yapı ortaya koyduklarını göstermektedir. Ayrıca American Educational Research Journal dergisinin eğitim politikaları ve uygulamaları, öğrenme ve öğretim, eğitimde araştırma yöntemleri ve yenilikçi yaklaşımlar, teknoloji ve eğitim vb. konu alanlarında yayımlar yapan bir dergi olduğu saptanmıştır. Tablo 4'te yer alan diğer dergilerin de aynı şekilde eğitimin değerlendirilmesi, eğitim politikaları, öğretmen eğitimi, eğitim yönetimi gibi konu alanlarında yayınlar yaptıkları tespit edilmiştir. Alanda çalışacak olan araştırmacıların bu dergileri düzenli olarak takip etmesi araştırmacıların güncel bilgiye olan ihtiyacını karşılayabilir. Analiz sonucunda ulaşılan bir diğer bulgu ise ortak atıf ilişkisi en güçlü olan referansın Booher-Jennings (2005) olduğudur. Bu referans toplamda 55 farklı referansla ortak atıf ilişkisi içinde bulunmuştur ve ağ yapısını genişleterek en etkili referans olmayı başarmıştır. Booher-Jennings (2005) bu çalışmada yüksek riskli test puanlarının öğretmenlerin kimliklerini ve birbirleriyle olan ilişkilerini nasıl değiştirdiğini inceleyerek, eğitimde hesap verebilirlik sisteminin öğretmen davranışlarını nasıl etkilediğini anlamayı amaçlamıştır. Diğer bir önemli referans olan Diamond ve Spillane (2004), dört farklı ilkokulda, okulların yüksek riskli hesap verebilirlik politikalarına nasıl tepki verdiğini incelemiştir. Sonuç olarak Diamond ve Spillane, denetimli okulların dar bir müfredat kullanarak politika gereksinimlerine odaklanıp belirli öğrencilerin performansını artırmaya çalıştığını, daha yüksek performanslı okulların ise tüm öğrencilerin başarısını artırmaya çalıştığını belirlemiş ve eğitimde eşitlik sorunlarının dikkate

alınması gerektiğini vurgulamıştır. Ağ haritasına göre Booher-Jennings (2005) referansına en çok ortak atfın verildiği referans ise Diamond ve Spillane (2004) olarak gösterilmiştir. Bu iki referansın ortak atfı ilişkisi, alandaki çalışmaların öğretmen davranışları, eğitimde eşitlik ve okulların değerlendirme politikalarına yönelik tepkileri çerçevesinde yoğunlaştığını göstermektedir. Fakat ağ haritasına göre bağlantı gücü en yüksek iki referansın Carnoy ve Loeb (2002) ve Hanushek ve Raymond (2005) olduğu saptanmıştır. Bu iki çalışma yüksek riskli testlerin ve hesap verilebilirlik sistemlerinin öğrenci başarısını artırmada etkili olduğunu ve sistemdeki eşitsizlik gibi karmaşık sorunları ortaya koymuştur. Alanyazında öne çıkan bu referanslar yüksek riskli testlerin öğrenci başarısı yanında sosyal ve politik boyutlarının da olduğunu göstermektedir. Bu bağlamda gelecekteki çalışmaların bu çok boyutlu yapının dikkate alınarak yürütülmesi yararlı olabilir. Ortak atfı analizine ait son bulgu James P. Spillane adlı yazarın 4.497 toplam bağlantı gücüyle ortak atfı ilişkisi en güçlü yazar olduğudur. Bu yazara en çok Cynthia E. Coburn adlı yazarla ortak atfı verilmiş ve bu yazarlar aralarında 523 bağlantı gücü oluşturarak aynı zamanda tüm ağ haritasında aralarında en fazla bağlantı gücü oluşan iki yazar olmayı başarmışlardır. Yazarların ilgi alanları incelediğinde, eğitimde liderlik, uygulama farkları ve öğretmen algısı gibi temalarda aktif yayınlar yaptıkları anlaşılmıştır. James P. Spillane'dan sonra toplam bağlantı gücü en yüksek olan yazar Stephen J. Ball çıkmıştır. Ağ haritasına göre bu yazarla birlikte en çok atfı verilen yazar ise aralarındaki 456 bağlantı gücüyle Bob Lingard olarak karşımıza çıkmaktadır. Bu iki yazarın ise çalışma alanlarının eğitim politikaları temaları üzerine olduğu tespit edilmiştir. Bu tematik çeşitlilik alanyazının özelden (okul, eğitimciler vb.) genele (eğitim sistemi, politika, vb.) birçok farklı boyutu içeren bir yapı sergilediğini göstermektedir. İlgili araştırmacıların bu referansları okumaları ve bahsi geçen yazarları takip etmeleri, yüksek riskli testler konusunda literatürdeki temel kavram ve eğilimlere hâkimiyet sağlaması açısından oldukça faydalı olabilir.

Son olarak ulaşılan veriler, kullanılan anahtar kelimelerin ilişki ağları açısından incelendiğinde, anahtar kelimelerin 7 küme altında toplandığı gözlemlenmiştir. Ağ haritasına göre anahtar kelimeler içinde en çok kullanılan anahtar kelimenin “accountability” olduğu sonucuna varılmıştır. Bu anahtar kelime kendi kümesi haricinde diğer kümelerdeki birçok anahtar kelimeyle birlikte kullanılarak ağ yapısı bakımından oldukça gelişmiştir. Toplam bağlantı güçleri dikkate alındığında önem taşıyan ve “accountability” anahtar kelimesiyle bağlantı oluşturan diğer anahtar kelimelerin ise sırasıyla “educational policy”, “assessment”, “educational reform”, “policy” anahtar kelimeleri olduğu görülmüştür. Bu bulgu yüksek riskli testlerle ilgili yapılan çalışmaların daha çok eğitim politikaları, eğitim reformları ve değerlendirme yöntemleriyle birlikte ele alındığını göstermektedir. Çeşitli veri tabanlarında tarama yapacak olan araştırmacıların bu anahtar kelimeleri kullanması, en doğru verilere ulaşılması ve alana hızlı bir şekilde hakim olunması açısından kolaylık sağlayabilir.

Yapılan analiz sonucunda ortaya çıkan eğilimler, yüksek riskli testlerin eğitim sistemleri içindeki yerini, yıllar içinde çalışma sayısında yaşanan artışı ve çok boyutlu

yapısını ortaya koymaktadır. Bu çalışmanın bulguları çerçevesinde, ortak yazarlık analizi en yüksek işbirliği gücüne sahip yazarın Antony Verger, en fazla işbirliğine sahip kurumun ise Stanford Üniversitesi olduğunu göstermektedir. Bu bulgular yüksek riskli testlerle ilgili çalışmaların daha çok uluslararası ağlarda gerçekleştiğini ortaya koymaktadır. Bu sebeple Türkiye'deki araştırmacıların uluslararası kişi ve kuruluşlarla iş birliğini artıracak proje ve yayın iş birlikleri teşvik edilmelidir. ABD'nin alandaki liderliği dikkate alındığında, araştırmacı ve kuruluşların özellikle ABD merkezli araştırmacılarla ikili veya çok taraflı iş birlikleri sağlamaları araştırmacı ve araştırmaların görünürlüğünü artırabilir.

Bulgulardan elde edilen diğer bir sonuç ise akademik etki için yayın sayısının tek başına yeterli olmadığıdır. İş birliği çeşitliliği ve ağ yapısı genişliğinin akademik etki için daha belirleyici olduğu görülmüştür. Bu sebeple yazarlar bireysel yayımlar yapmaktan ziyade farklı disiplinlerden çeşitli akademisyenlerle iş birliği kurmalıdırlar.

American Educational Research Journal ve Educational Evaluation and Policy Analysis dergileri alanda öne çıkmaktadır. Araştırmacıların bu dergileri takip etmeleri yararlı olabilir. Öne çıkan referanslar incelendiğinde bu referanslar yüksek riskli testlerin sosyal ve politik boyutlarını vurgulamaktadır. Gelecek araştırmaların bu boyutlara odaklanmaları alana katkı sağlayabilir.

Anahtar kelimeler yüksek riskli testlerle ilgili çalışmaların eğitim politikaları, reform ve değerlendirme temalarıyla yoğun ilişkili olduğunu göstermiştir. Gelecekteki araştırmalar bu temalar arasındaki ilişkileri inceleyebilir. Ayrıca anahtar kelimelerin yedi küme etrafında şekillenmesi bir kez daha yüksek riskli testlerin çok boyutlu yapısına işaret etmektedir. Gelecek araştırmalarda bu kümelerde yer alan temaların ilişkisini inceleyen çalışmalar yapılabilir.

References

- Arviv Elyashiv, R., & Avidov-Ungar, O. (2024). Teachers' perceptions of national large-scale assessment: the pedagogical dimension. *Educational Review*, 76(6), 1691–1707. <https://doi.org/10.1080/00131911.2023.2256996>
- Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, 36(5), 258–267. <https://doi.org/10.3102/0013189X07306523>
- Au, W. (2011). Teaching under the new Taylorism: High-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies*, 43(1), 25–45. <https://doi.org/10.1080/00220272.2010.521261>
- Au, W. (2016). Meritocracy 2.0: High-stakes, standardized testing as a racial project of neoliberal multiculturalism. *Educational Policy*, 30(1), 39–62. <https://doi.org/10.1177/0895904815614916>
- Ball, S. J. (2003). The teacher's soul and the terrors of performativity. *Journal of education policy*, 18(2), 215–228. <https://doi.org/10.1080/0268093022000043065>
- Bae, S. (2018). Redesigning systems of school accountability: A multiple measures approach to accountability and support. *Education Policy Analysis Archives*, 26(8). <http://dx.doi.org/10.14507/epaa.26.2920>
- Booher-Jennings, J. (2005). Below the bubble: “Educational triage” and the Texas accountability system. *American Educational Research Journal*, 42(2), 231–268. <https://doi.org/10.3102/00028312042002231>
- Burdett, N. (2016). *The good, the bad, and the ugly—testing as a part of the education ecosystem* (RISE-WP-16/010). Working paper series, research on improving systems of education (RISE).
- Burgess, S., Metcalfe, R., & Sadoff, S. (2021). Understanding the response to financial and non-financial incentives in education: Field experimental evidence using high-stakes assessments. *Economics of Education Review*, 85, 102195. <https://doi.org/10.1016/j.econedurev.2021.102195>
- Camphuijsen, M. K., Møller, J., & Skedsmo, G. (2021). Test-based accountability in the Norwegian context: Exploring drivers, expectations and strategies. *Journal of Education Policy*, 36(5), 624–642. <https://doi.org/10.1080/02680939.2020.1739337>
- Carnoy, M., & Loeb, S. (2002). Does external accountability affect student outcomes? A cross-state analysis. *Educational evaluation and policy analysis*, 24(4), 305–331. <https://doi.org/10.3102/01623737024004305>
- Dee, T. S., & Jacob, B. (2011). The impact of No Child Left Behind on student achievement. *Journal of Policy Analysis and Management*, 30(3), 418–446. <https://doi.org/10.1002%2Fpam.20586>

- Diamond, J. B., & Spillane, J. P. (2004). High-stakes accountability in urban elementary schools: Challenging or reproducing inequality?. *Teachers College Record*, 106(6), 1145-1176. <https://doi.org/10.1111/j.1467-9620.2004.00375.x>
- Donthu, N., Kumar, S., & Pattnaik, D. (2020). Forty-five years of Journal of Business Research: A bibliometric analysis. *Journal of Business Research*, 109. <https://doi.org/10.1016/j.jbusres.2019.10.039>
- 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>
- Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, 105, 1809-1831. <https://doi.org/10.1007/s11192-015-1645-z>
- Elo, J., & Nygren-Landgärds, C. (2021). Teachers' perceptions of autonomy in the tensions between a subject focus and a cross-curricular school profile: A case study of a Finnish upper secondary school. *Journal of Educational Change*, 22(3), 423-445. <https://doi.org/10.1007/s10833-020-09412-0>
- Erdağ, C. (2019). Bir hesap verebilirlik teknolojisi olarak merkezi sınavlar: Finlandiya, Estonya ve Yeni Zelanda örnekleri. *OPUS International Journal of Society Researches*, 14(20), 1528-1563. <https://doi.org/10.26466/opus.592335>
- Erduran, S., El Masri, Y., Cullinane, A., & Ng, Y. (2020). Assessment of practical science in high stakes examinations: A qualitative analysis of high performing English-speaking countries. *International Journal of Science Education*, 42(9), 1544-1567. <https://doi.org/10.1080/09500693.2020.1769876>
- Escamilla, K., Chávez, L., & Vigil, P. (2005). Rethinking the “gap” high-stakes testing and Spanish-speaking students in Colorado. *Journal of Teacher Education*, 56(2), 132-144. <https://doi.org/10.1177/0022487104273791>
- Falabella, A. (2021). The seduction of hyper-surveillance: Standards, testing, and accountability. *Educational Administration Quarterly*, 57(1), 113-142. <https://doi.org/10.1177/0013161X20912299>
- Fidjeland, A. (2023). Using high-stakes grades to incentivize learning. *Economics of Education Review*, 94. <https://doi.org/10.1016/j.econedurev.2023.102377>
- Finn, C. E. (2022). School accountability: Yesterday, today, and tomorrow. *Phi Delta Kappan*, 104(3), 12-17. <https://doi.org/10.1177/00317217221136588>
- Fischer, C., Fishman, B., Levy, A. J., Eisenkraft, A., Dede, C., Lawrenz, F., Jia, Y., Kook, J. F., Frumin, K., & McCoy, A. (2020). When do students in Low-SES schools perform better-than-expected on a high-stakes test? Analyzing school, teacher, teaching, and professional development characteristics. *Urban Education*, 55(8-9), 1280-1314. <https://doi.org/10.1177/0042085916668953>

- Fitchett, P. G., & Heafner, T. L. (2010). A national perspective on the effects of high-stakes testing and standardization on elementary social studies marginalization. *Theory & Research in Social Education*, 38(1), 114-130. <https://doi.org/10.1080/00933104.2010.10473418>
- Goertz, M., & Duffy, M. (2003). Mapping the landscape of high-stakes testing and accountability programs. *Theory Into Practice*, 42(1), 4-11. https://doi.org/10.1207/s15430421tip4201_2
- Gorgodze, S., & Chakhaia, L. (2021). The uses and misuses of centralised high stakes examinations-Assessment Policy and Practice in Georgia. *Assessment in Education: Principles, Policy & Practice*, 28(3), 322-342. <https://doi.org/10.1080/0969594X.2021.1900775>
- Gulek, C. (2003). Preparing for high-stakes testing. *Theory into practice*, 42(1), 42-50. https://doi.org/10.1207/s15430421tip4201_6
- Hammack, R., & Wilson, E. (2019). The dangers of high stakes testing in social studies. *Journal of Social Studies and History Education*, 2-26. Retrieved from <https://www.uhd.edu/documents/academics/public-service/urban-education/jsshe/archives/2019/the-dangers-high-stakes-rhammack2.pdf>
- Hanushek, E. A., & Raymond, M. E. (2004). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association*, 2(2-3), 406-415. Retrieved from <https://academic.oup.com/jeea/article-abstract/2/2-3/406/2194942>
- Hanushek, E. A., & Raymond, M. E. (2005). Does school accountability lead to improved student performance?. *Journal of Policy Analysis and Management*, 24(2), 297-327. <https://doi.org/10.1002/pam.20091>
- Hargreaves, A. (2020). Large-scale assessments and their effects: The case of mid-stakes tests in Ontario. *Journal of Educational Change*, 21(3), 393-420. <https://doi.org/10.1007/s10833-020-09380-5>
- Hatfield, J. L., & Soløst, T. E. T. (2025). Assessing the assessment: Exploring Norwegian primary education teachers' perceptions of national accountability testing. *Scandinavian Journal of Educational Research*, 69(2), 391-408. <https://doi.org/10.1080/00313831.2024.2308883>
- Heilig, J. V., & Darling-Hammond, L. (2008). Accountability Texas-Style: The Progress and learning of urban minority students in a high-stakes testing context. *Educational Evaluation and Policy Analysis*, 30(2), 75-110. <https://doi.org/10.3102/0162373708317689>
- Hilton, A. L., & Saunders, R. (2024). LANTITE's impact on teacher diversity: Unintended consequences of testing pre-service teachers. *The Australian Educational Researcher*, 51(4), 1063-1082. <https://doi.org/10.1007/s13384-023-00628-6>

- Hofflinger, A., & von Hippel, P. T. (2020). Missing children: How Chilean schools evaded accountability by having low-performing students miss high-stakes tests. *Educational Assessment, Evaluation and Accountability*, 32(2), 127-152. <https://doi.org/10.1007/s11092-020-09318-8>
- Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline in educational equality. *British Educational Research Journal*, 31(5), 605–622. <https://doi.org/10.1080/01411920500240767>
- Jacob, B. A. (2005). Accountability, incentives and behavior: The impact of high-stakes testing in the Chicago public schools. *Journal of Public Economics*, 89(5-6), 761-796. <https://doi.org/10.1016/j.jpubeco.2004.08.004>
- Jacobs, M., van der Velden, R., & van Vugt, L. (2024). High-stakes testing and educational careers: Exploiting the differences in cutoffs between test recommendations in the Netherlands. *Journal of Research on Educational Effectiveness*, 17(4), 711–745. <https://doi.org/10.1080/19345747.2023.2242342>
- Jerrim, J. (2023). Test anxiety: Is it associated with performance in high-stakes examinations?. *Oxford Review of Education*, 49(3), 321–341. <https://doi.org/10.1080/03054985.2022.2079616>
- Karakus, M., Ersozlu, A., & Clark, A. C. (2019). Augmented reality research in education: A bibliometric study. *Eurasia Journal of Mathematics, Science and Technology Education*, 15(10). <https://doi.org/10.29333/ejmste/103904>
- Kavanagh, K. M., & Fisher-Ari, T. R. (2020). Curricular and pedagogical oppression: Contradictions within the juggernaut accountability trap. *Educational Policy*, 34(2), 283-311. <https://doi.org/10.1177/0895904818755471>
- Kenney, A. W., Dulong Langley, S., Hemmler, V., Callahan, C. M., Gubbins, E. J., & Siegle, D. (2024). Different or differentiated? Recoupling Policy and practice in an era of accountability. *Educational Policy*, 38(1), 134-160. <https://doi.org/10.1177/08959048231153612>
- Kirby, S. N., & Stecher, B. M. (2004). Accountability in education. In B.M. Stecher, & S. N. Kirby (Eds.), *Organizational improvement and accountability: Lessons for education from other sectors* (pp. 1-10). Rand Corporation.
- Klinger, D. A., & Rogers, W. T. (2011). Teachers' perceptions of large-scale assessment programs within low-stakes accountability frameworks. *International Journal of Testing*, 11(2), 122–143. <https://doi.org/10.1080/15305058.2011.552748>
- Knoester, M., & Meshulam, A. (2022). Beyond deficit assessment in bilingual primary schools. *International Journal of Bilingual Education and Bilingualism*, 25(3), 1151–1164. <https://doi.org/10.1080/13670050.2020.1742652>
- Levatino, A., Parcerisa, L., & Verger, A. (2024). Understanding the stakes: The influence of accountability policy options on teachers' responses. *Educational Policy*, 38(1), 31-

60. <https://doi.org/10.1177/08959048221142048>
- McCoy, S., & Byrne, D. (2024). Shadow education uptake in Ireland: Inequalities and wellbeing in a high-stakes context. *British Journal of Educational Studies*, 72(6), 693–719 <https://doi.org/10.1080/00071005.2024.2331476>
- McElroy, K. (2023). Does test-based accountability improve more than just test scores?. *Economics of Education Review*, 94, 102381. <https://doi.org/10.1016/j.econedurev.2023.102381>
- McNeil, L. M. (2000). *Contradictions of school reform: Educational costs of standardized testing*. New York: Routledge.
- Merigó, J. M., & Yang, J. B. (2017). A bibliometric analysis of operations research and management science. *Omega*, 73, 37-48. <https://doi.org/10.1016/j.omega.2016.12.004>
- Merigó, J. M., Pedrycz, W., Weber, R., & de la Sotta, C. (2018). Fifty years of Information Sciences: A bibliometric overview. *Information Sciences*, 432, 245-268. <https://doi.org/10.1016/j.ins.2017.11.054>
- Mintz, J. A., & Kelly, A. M. (2021). Science teacher motivation and evaluation policy in a high-stakes testing state. *Educational Policy*, 35(1), 3-40. <https://doi.org/10.1177/0895904818810520>
- Moore, D. P. (2025). Culturally sustaining catch-up? How English language arts teachers “compensate” with culturally sustaining pedagogy after the state exam. *Urban Education*, 60(7), 2016-2050. <https://doi.org/10.1177/00420859241266605>
- Munoz-Chereau, B., González, Á., & Meyers, C. V. (2022). How are the ‘losers’ of the school accountability system constructed in Chile, the USA and England? *Compare: A Journal of Comparative and International Education*, 52(7), 1125–1144. <https://doi.org/10.1080/03057925.2020.1851593>
- Neal, D., & Schanzenbach, D. W. (2010). Left behind by design: Proficiency counts and test-based accountability. *The Review of Economics and Statistics*, 92(2), 263-283. <https://doi.org/10.1162/rest.2010.12318>
- Ober, T. M., Hong, M. R., Carter, M. F., Brodersen, A. S., Rebouças-Ju, D., Liu, C., & Cheng, Y. (2022). Are high school students accurate in predicting their AP exam scores?: Examining inaccuracy and overconfidence of students’ predictions. *Assessment in Education: Principles, Policy & Practice*, 29(1), 27–50. <https://doi.org/10.1080/0969594X.2022.2037508>
- Orfield, G., & Wald, J. (2000). Testing, testing: The high-stakes testing mania hurts poor and minority students the most. *The Nation*, 270(22), 38–40. <https://thenation.s3.amazonaws.com/pdf/testingtesting2000.pdf>
- Ozga, J. (2009). Governing education through data in England: from regulation to self-evaluation. *Journal of Education Policy*, 24(2), 149–162.

<https://doi.org/10.1080/02680930902733121>

- Özaslan, A. & Beyhan, Ö. (2023). Students' understandings related to their upcoming high-stakes tests: A phenomenographic study. *Journal of Teacher Education and Lifelong Learning*, 5(2), 789-803. <https://doi.org/10.51535/tell.1288890>
- Pace, J. L. (2011). The complex and unequal impact of high stakes accountability on untested social studies. *Theory & Research in Social Education*, 39(1), 32-60. <https://doi.org/10.1080/00933104.2011.10473446>
- Parkison, P. (2009). Political economy and the NCLB regime: Accountability, standards, and high-stakes testing. *The Educational Forum*, 73(1), 44-57. <https://doi.org/10.1080/00131720802539606>
- Passas, I. (2024). Bibliometric analysis: the main steps. *Encyclopedia*, 4(2), 1014-1025. <https://doi.org/10.3390/encyclopedia4020065>
- Rhoten, D., Carnoy, M., Elmore, R., & Chabran, M. (2003). The conditions and characteristics of assessment and accountability: The case of four states. In M. Carnoy, R. Elmore, & L. S. Siskin (Eds.), *The new accountability: High schools and high-stakes testing* (pp. 13-53). Routledge.
- Sloane, F. C., & Kelly, A. E. (2003). Issues in high-stakes testing programs. *Theory Into Practice*, 42(1), 12-17. https://doi.org/10.1207/s15430421tip4201_3
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265-269. <https://doi.org/10.1002/asi.4630240406>
- Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal of Educational Change*, 10, 211-227. <https://doi.org/10.1007/s10833-009-9105-2>
- Tan, S. H. (2010). Singapore's educational reforms: The case for un-standardizing curriculum and reducing testing. *AASA Journal of Scholarship and Practice*, 6(4), 50-58.
- Thomas, R. M. (2005). *High-stakes testing: Coping with collateral damage*. Routledge.
- Thompson, G. (2013). NAPLAN, MySchool and accountability: Teacher perceptions of the effects of testing. *International Education Journal: Comparative Perspectives*, 12(2), 62-84. <https://openjournals.library.sydney.edu.au/IEJ/article/view/7456>
- Thompson, G., & Lašič, T. (2011). *Doing something about it: Representations of NAPLAN in the public domain* [Paper presentation]. AARE Annual Conference, Hobart, TAS, Australia. <https://www.aare.edu.au/data/publications/2011/aarefinal00577.pdf>
- Trujillo, T., Møller, J., Jensen, R., Kissell, R. E., & Larsen, E. (2021). Images of educational leadership: How principals make sense of democracy and social justice in two distinct policy contexts. *Educational Administration Quarterly*, 57(4), 536-569. <https://doi.org/10.1177/0013161X20981148>

- Van Eck, N. J., & Waltman, L. (2023, Ekim 31). *VOSviewer manual (Sürüm 1.6.20)*. Centre for Science and Technology Studies (CWTS), Leiden University. <https://www.vosviewer.com>
- Verger, A., Parcerisa, L., & Fontdevila, C. (2019). The growth and spread of large-scale assessments and test-based accountabilities: A political sociology of global education reforms. *Educational Review*, 71(1), 5–30. <https://doi.org/10.1080/00131911.2019.1522045>
- Von der Embse, N. P., & Witmer, S. E. (2014). High-stakes accountability: Student anxiety and large-scale testing. *Journal of Applied School Psychology*, 30(2), 132–156. <https://doi.org/10.1080/15377903.2014.888529>
- Wallin, J. A. (2005). Bibliometric methods: Pitfalls and possibilities. *Basic & Clinical Pharmacology & Toxicology*, 97(5), 261-275. https://doi.org/10.1111/j.1742-7843.2005.pto_139.x
- Wang, J., Li, Q., & Luo, Y. (2022). Physics identity of Chinese students before and after Gaokao: The effect of high-stake testing. *Research in Science Education*, 52(2), 675-689. <https://doi.org/10.1007/s11165-020-09978-y>
- Xie, Q. (2022). The role of parents in the school accountability system: Insights from a Hong Kong case. *Studies in Educational Evaluation*, 72. <https://doi.org/10.1016/j.stueduc.2021.101114>
- Yin, Y. M., & Mu, G. M. (2022). Examination-oriented or quality-oriented? A question for fellows of an alternative teacher preparation program in China. *The Australian Educational Researcher*, 49(4), 727-742. <https://doi.org/10.1007/s13384-021-00462-8>
- Yung, K. W. H. (2021). Shadow education as a form of oppression: conceptualizing experiences and reflections of secondary students in Hong Kong. *Asia Pacific Journal of Education*, 41(1), 115–129. <https://doi.org/10.1080/02188791.2020.1727855>
- Zakharov, A., & Carnoy, M. (2021). Does teaching to the test improve student learning?. *International Journal of Educational Development*, 84, 102422. <https://doi.org/10.1016/j.ijedudev.2021.102422>

Ethical and Author Declarations | Etik ve Yazar Beyanları

Authors' Contributions

Both authors contributed equally to the conceptualization of the study, the development of the methodology, the data collection and analysis processes, and the writing and revision of the manuscript. Both authors approved the final version of the manuscript.

Conflict of Interest

The authors declare that they have no competing interests relevant to the content of this article.

Ethical Approve

This study was conducted within the scope of a research design that does not require ethics committee approval, as it is based on a bibliometric analysis of the existing literature on high-stakes testing. Throughout the research process, the principles of scientific research and publication ethics were strictly followed.

Use of Artificial Intelligence

No artificial intelligence tools were used during the writing, analysis, or data processing stages of the study.

Yazarların Katkı Düzeyleri

Bu çalışmada her iki yazar da araştırmanın kavramsallaştırılması, yöntemin oluşturulması, veri toplama ve analiz süreçleri ile makalenin yazımı ve gözden geçirilmesi aşamalarına eşit düzeyde katkı sağlamıştır. Her iki yazar da makalenin son hâlini onaylamıştır.

Çıkar Çatışması Beyanı

Yazarların, bu makalenin içeriğiyle ilgili beyan edilecek herhangi bir çıkar çatışması bulunmamaktadır.

Etik Onay

Bu çalışma, yüksek riskli testlere ilişkin mevcut alanyazının bibliyometrik analizine dayandığı için etik kurul onayı gerektirmeyen bir araştırma türü kapsamında yürütülmüştür. Araştırma sürecinde bilimsel araştırma ve yayın etiği ilkelerine tam olarak uyulmuştur.

Yapay Zeka Kullanımı

Araştırmanın yazım, analiz ve veri işleme aşamalarında herhangi bir yapay zekâ aracı kullanılmamıştır.

This study has been evaluated under double-blind peer review and verified to be free of plagiarism using iThenticate software.

Bu çalışma, çift taraflı kör hakemlik kapsamında değerlendirilmiş ve iThenticate yazılımı kullanılarak intihal içermediği teyit edilmiştir.

The studies published in our journal are published as open access under a CC-BY-NC-ND license.

Dergimizde yayımlanan çalışmalar CC-BY-NC-ND lisansı altında açık erişim olarak yayımlanmaktadır.

Ethical disclosure | Etik bildirim: ebfd@ankara.edu.tr