

Article

Evolution and Future Prospects of Education Evaluation Research in China over the Last Decade

Shanshan Wang ¹, Junping Qiu ^{1,*}, Jia Zhou ²  and Yunlong Yu ^{1,*}
¹ Zhejiang Academy of Higher Education, Hangzhou Dianzi University, Hangzhou 310018, China

² Teaching Research Office of Vocational Education and Adult Education, Wuhan Academy of Educational Science, Wuhan 430030, China

* Correspondence: jpqiu@hdu.edu.cn (J.Q.); yuyi@hdu.edu.cn (Y.Y.)

Abstract: With the popularity of educational evaluation studies, researchers have begun to mine text data to provide helpful information for educational stakeholders. Despite the frequent adoption of textual analysis, few studies have been found to search and analyze data from research projects. As the highest-level and most authoritative research projects in China, the Chinese National Education Science Planning projects could fully reflect the research fronts and hotspots in the field. This study examines the educational evaluation projects in the Chinese National Education Science Planning from 2011 to 2020, adopting bibliometric, textual analysis, and visualization methods to present an overview of the educational evaluation research in China. The results have revealed a steady growth of projects on educational evaluation in China as there were 219 projects approved during the 13th Five-Year Plan period, 68 more than that of the 12th Five-Year Plan period, and accounted for 9.00% of the total number of projects launched in the same period. The educational evaluation project organizations are diversified, including universities, scientific research institutions, primary and secondary schools, administrative departments, educational societies, and publishing units. The research trends focus on the close integration of educational evaluation with informatization, needs in educational practice, and national strategic development. This study offers an applicable methodology and framework for the future bibliometric study of educational evaluation.

Keywords: educational evaluation in China; funded research projects; evolution; trends; visualization analysis



Citation: Wang, S.; Qiu, J.; Zhou, J.; Yu, Y. Evolution and Future Prospects of Education Evaluation Research in China over the Last Decade. *Sustainability* **2022**, *14*, 14340. <https://doi.org/10.3390/su142114340>

Academic Editor: Alexander Mikroyannidis

Received: 11 October 2022

Accepted: 28 October 2022

Published: 2 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the rapid growth in the number of published papers on educational evaluation research, researchers have begun to use text mining methodology to get an overview of the evolution of educational evaluation research [1,2]. The vast amount of data in text mining can provide critical information and insight for outlining published research outcomes and research trends that require attention [3,4]. Pereira et al. [5] reviewed 30 papers published in the *Journal Assessment and Evaluation in Higher Education* from 2006 to 2013 and summarized the cutting-edge themes in European higher education assessment research following the Bologna Processes based on the papers' purpose, content, and methodology. Marx et al. [6] performed a review of the products from the first decade of the student standards health education assessment project. Chugh et al. [7] collected 43 papers on postgraduate evaluation feedback from six online databases such as Scopus (Elsevier) from 2010 to 2020 and outlined four critical issues: the content of feedback, the process, personnel, and expectations through a narrative literature review. By using the Citespace software, Wang et al. [8] conducted a bibliometric and visualization analysis of research papers published over the last decade (2010–2019) in six international authoritative SSCI journals in educational evaluation and, based on highly cited literature, summarised the hotspots and trends of international educational evaluation research over the last decade.

Similar text-mining-based approaches have been applied to various areas of educational evaluation, such as research on health literacy assessment [9], educational learning outcomes evaluation [10–12], peer evaluation [13–15], teaching evaluation [16–18], feedback on outcomes [19,20], self-assessment [21–23], teacher evaluation [24–26], student evaluation [27–31], formative assessment [32–34], educational technology evaluation [35,36], the impact of mid-term assessment [37], and assessment of prior learning [38]. These studies reviewed and outlined the current development of educational evaluation research but are limited to bibliometric analyses or literature reviews of published research papers. To the best of our knowledge, most researchers have adopted a monotonous perspective as they mainly examine published research papers on educational evaluation and discuss from dimensions such as the number of papers published, country distribution, research institutions, authors' characteristics, research themes, and development trends. In contrast, few have analyzed the research projects on educational evaluation.

The Chinese National Education Science Planning Project is the highest-level and most authoritative funding project in the education science research field, macroscopically reflecting the hotspots and research fronts in the field and with high academic value and reference value for decision-making. To gain a more comprehensive view, we set the projects on educational evaluation in Chinese National Education Science Planning from 2011 to 2020 as the sample data and adopted bibliometric, textual analysis, and visualization methods to analyze in depth the number of projects on educational evaluation, research units, research themes and the evolution of the themes, revealing the main structure and trends of approved projects on educational evaluation in China.

The study's primary contribution is to provide an in-depth content analysis of the Chinese National Education Science Planning Project over the last decade, helping global education researchers to gain a comprehensive overview of the stage characteristics and overall development trends of educational evaluation research in China. It also facilitates Chinese education researchers to monitor the growth of the research projects and content of research over the last decade. The study also provides a plausible methodology and framework for the bibliometric study of educational evaluation in the future.

2. Materials and Methods

2.1. Data Collection and Processing

To help global education researchers gain a comprehensive understanding of the development of educational evaluation research in China, we obtained data from education science projects annually published on the official website of the Office of the Leading Group of China Education Science Planning. The approved dates of the selected projects range from 2011 to 2020. In total, 370 original records of projects containing the words "evaluation" or "assessment" in their titles were found. The ten years were divided into two periods, namely, 2011–2015 and 2016–2020, for comparative analysis. This time division is based on China's strategy to formulate national economic and social development plans in five-year cycles, mainly to plan major national construction projects, productivity distribution, and important proportional relations in the national economy and to set goals and directions for national economic development. China began to formulate its first Five-Year Plan in 1953, and the period 2011–2015 is for the 12th Five-Year Plan and 2016–2020 is for the 13th Five-Year Plan. Bibliometric analysis was conducted based on the projects' data to explore the development and distribution of institutions of these projects.

As an essential part of the literature, keywords are the core terms that the researcher has thoughtfully refined to reflect essential concepts [39,40]. The title of a research project is generally made up of a number of words (more than one). Following a glossary of education terms, we manually extracted title keywords from the 370 project titles. Keyword pre-processing is essential to obtain reliable analysis results because there are often redundant or synonymous words in the project titles, potentially reducing the frequency of keywords or leading to inaccurate answers if not appropriately handled [41]. First, the researchers deleted Chinese words such as "of", "and" and "with", and redundant words such as

“research”, “based on” and “analysis”. A total of 432 keywords were obtained first. Next, synonyms or near-synonyms were classified into the same category. For example, “moral” and “character”, “health” and “physical and mental health”, “university” and “higher institutions” were classified under one term, respectively. A total of 202 keywords were obtained. The keywords of project titles over the 10 years were then cleaned and sorted using the R language text mining package (e.g., *bibliometrix*, *tidyverse*, *tidytext*).

We applied a combined method of bibliometric statistics, word cloud, keyword co-occurrence, strategic diagram, and Sankey diagram to analyze evolution of the projects. All analyses were conducted using R 4.1.3 software (R Core Team, 2020). The bibliometric, text analysis and visualization were carried out mainly using five open-source R packages, namely, *tidytext*, *bibliometrix*, *tidyverse*, *ggwordcloud* and *networkD3*. The package *tidytext* was mainly used to clean the keywords; the package *bibliometrix* was mainly used to create the network of keywords co-occurrence; the package *tidyverse* was mainly used to tidy the data and visualize the results of analysis; the package *ggwordcloud* was used to build a word cloud; the package *networkD3* was used to draw a Sankey diagram to explore the evolution of the themes in projects.

2.2. Analytical Methods

2.2.1. Descriptive Statistics

The bibliometric analyses of this study include the development of research project establishments, the distribution of research institutions, and the frequency of keywords. The development of research project establishment was mainly reflected by the changes in the number of approved projects on education evaluation in China from 2011 to 2020 and the percentage of these projects in the overall approved projects of education science planning for that year in China. The distribution of research institutions was mainly analyzed through exploring the distribution of the higher education institutions, primary and secondary schools, research institutions, administrative departments, and educational societies. Word cloud made by the R package named *ggwordcloud* was used to visualize the frequency of keywords. The sizes of keywords represented their frequencies, and the larger the word size, the higher the frequency of occurrence [42].

2.2.2. Keywords Co-Occurrence Analysis

The R package named *bibliometrix* was used to analyze the co-occurrences of 202 keywords in the titles of the approved projects and to construct a keyword co-occurrence network to reveal the relations between keywords in the titles of the projects [43]. The keyword co-occurrence represents the co-existence of several keywords in the same title, and the higher the frequency of the co-occurrence is, the more frequently these keywords are used by researchers. Moreover, the keyword co-occurrence analysis is a powerful method to discover and describe the interactions between different research topics. The keyword co-occurrence network was analyzed in clusters to form a number of distinct research themes [44]. In the map of research themes, the size of the nodes is proportional to the degree of centrality of the keywords in the network. The colors of the nodes represented different clusters, namely, themes. The thickness of the lines between nodes is proportional to the frequency of co-occurrence between keywords. The thicker the line was, the greater the number of keyword co-occurrence [42,45].

2.2.3. Thematic Analysis

A strategic diagram and a Sankey diagram of the research themes for approved projects on education evaluation of the 2011–2020 Chinese National Education Science Planning were drawn using the R packages named *ggplot2* in *tidyverse* and *networkD3*, respectively [46,47]. The number of keywords is an indicator of the richness of a theme, while the frequency of keywords is an indicator of the focus of a theme. Then, a strategic diagram was applied to show the themes with a different number of keywords (*x*-axis) and the frequency of keywords (*y*-axis). In the strategy diagram, research themes in the

first quadrant have a high degree of focus and richness, those in the second quadrant have a high degree of focus but a low degree of richness, those in the third quadrant have a low degree of focus and richness, and those in the fourth quadrant have a low degree of focus but a high degree of richness. The Sankey diagram presents the development process of the research themes of the educational evaluation projects, including the continuation, integration, differentiation, and extinction, from Period I to Period II, reflecting the change of the research hotspots over time [47,48].

3. Results

3.1. Scale of Education Evaluation Projects

Figure 1 shows the number of approved education evaluation projects from 2011 to 2020 and the percentage of the total number of approved projects in that year. As can be seen from the figure, the number of National Education Science Planning projects focusing on education evaluation increased from 21 in 2011 to 44 in 2020, and the percentage increased from 5.22% to 10.09%, showing an overall steady growth trend. Seen from a five-year time cycle, the number of education evaluation projects during China's 12th Five-Year Plan period (2011–2015) was 151, accounting for 7.16% of the total number of projects during that period. By contrast, the number increased to 219 during the 13th Five-Year Plan period (2016–2020), accounting for 9.00% of the total number of projects approved during that period. It indicates that China is attaching more and more importance to education evaluation research, and it has gradually become a hot spot for scientific research.

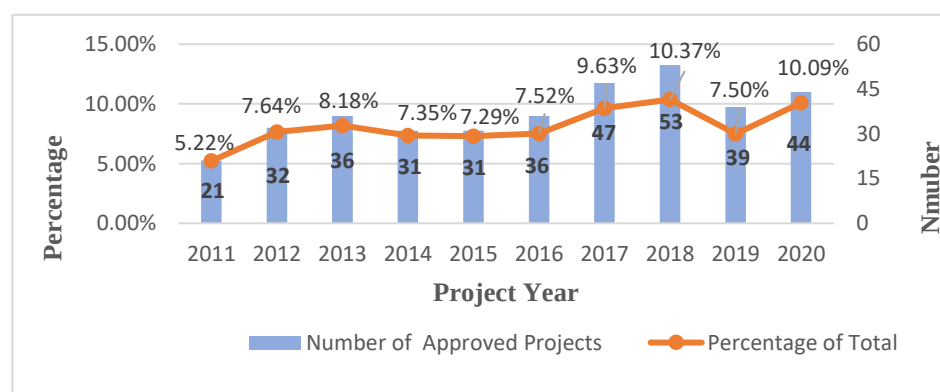


Figure 1. Distribution of Chinese National Education Science Planning projects between 2011 and 2020.

3.2. Distribution of Research Units

Figure 2 shows the distribution of various units undertaking education evaluation projects during the 12th Five-Year Plan period (2011–2015) and the 13th Five-Year Plan period (2016–2020). During the 12th Five-Year Plan period, higher education institutions took up the most significant number of education evaluation research projects, amounting to 129. Research institutions took up ten projects, administrative departments took up nine projects, primary and secondary schools took up two projects, and educational societies took up one project. During the 13th Five-Year Plan period, institutions of higher education still ranked first, with 179 educational evaluation projects, followed by research institutions with 25 projects, primary and secondary schools with eight projects, administrative departments with five projects, and educational societies and publishers with two projects. Compared with the 12th Five-Year Plan period, although the number of educational evaluation research projects undertaken by research institutions and primary and secondary schools has increased during the 13th Five-Year Plan period, a large gap remains between higher education institutions and them.

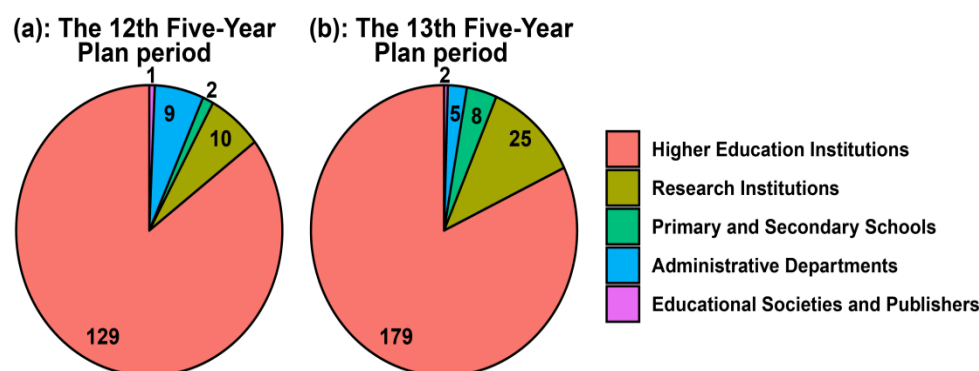


Figure 2. Distribution of research units between 2011 and 2020. This figure shows the distribution of various units undertaking education evaluation projects during the 12th Five-Year Plan period (2011–2015) and the 13th Five-Year Plan period (2016–2020).

3.3. Keyword Analysis of Project Themes

Figure 3 shows the word clouds formed by the keywords in the titles of the projects during the 12th and 13th Five-Year Plans period, respectively. Table 1 presents top 20 keywords of the project titles during the 12th and 13th Five-Year Plans period.

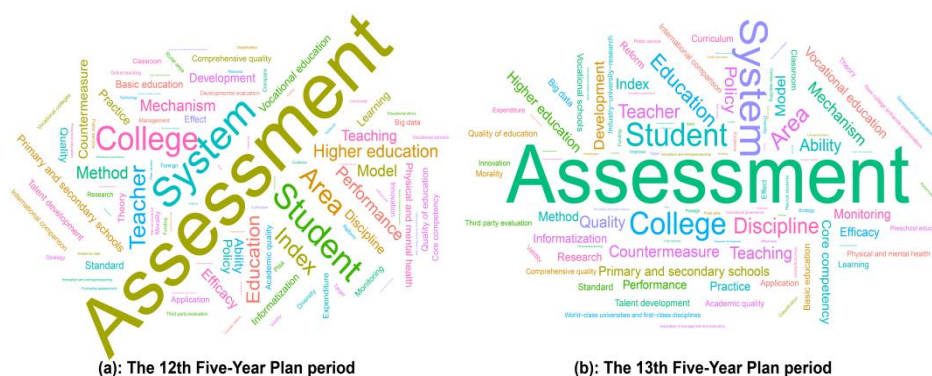


Figure 3. Keyword analysis of project themes between 2011 and 2020. This figure shows the most frequent words in the titles of the projects during the 12th Five-Year Plan period (2011–2015) and the 13th Five-Year Plan period (2016–2020). The centrality and size of the words indicate the frequency of the theme. The larger the word size is, the more important it is.

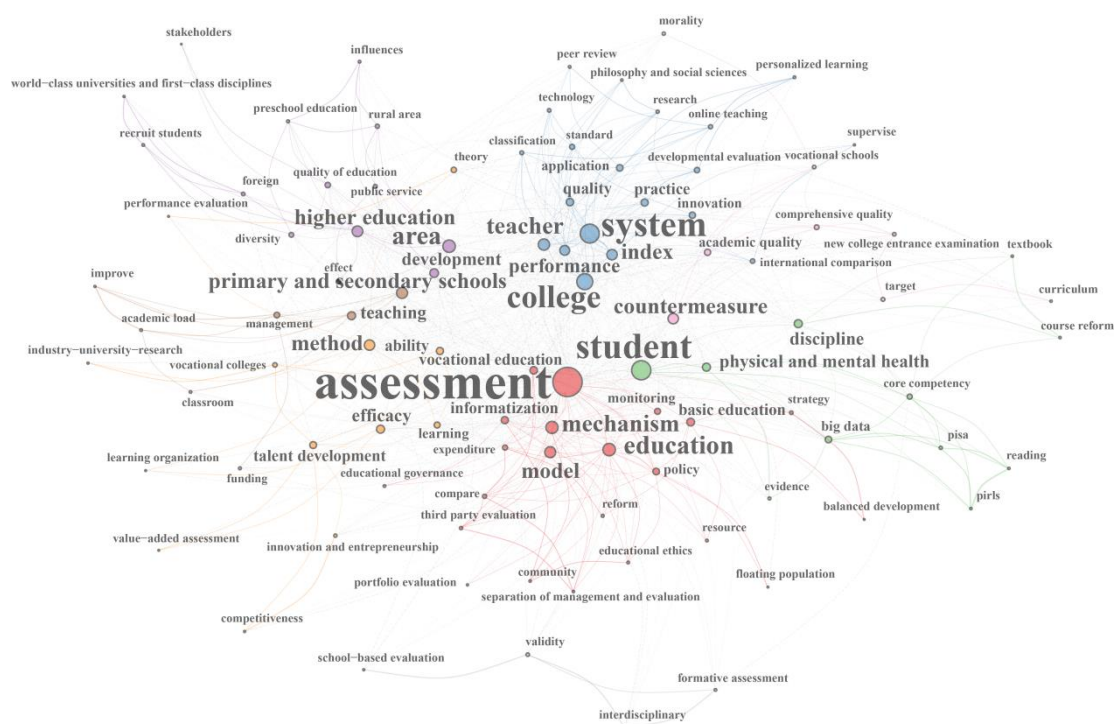
Figure 3a and Table 1 show that during the 12th Five-Year Plan period, “assessment” (143), “system” (48), “student” (38), “college” (36), “regional” (22), “teacher” (22), “higher education” (19), “index” (18), “performance” (16), “mechanism” (15) and “method” (14) have become high-frequency buzzwords. Figure 3b and Table 1 show that during the 13th Five-Year Plan period, “assessment” (199), “systems” (67), “student” (48), “college” (45), “subject” (29), “regional” (26), “teacher” (22), “teaching” (21), “primary and secondary schools” (20), “quality” (20) and “policy” (20) have become high-frequency buzzwords. During the 12th and 13th Five-Year Plan periods, evaluation systems, student evaluation, teacher evaluation, regional evaluation, higher education evaluation, and evaluation mechanisms have become common hot topics for scholar.

3.4. Theme Distribution of Projects

Figure 4 shows the distribution of research themes of projects on educational evaluation in the 12th Five-Year Plan period (2011–2015). The 99 keywords in the titles of the approved projects during the 12th Five-Year Plan period were divided into eight clusters with different colors.

Table 1. Top 20 keywords in the project titles during the 12th and 13th Five-Year Plans period.

Keyword	Frequency	Subperiod	Keyword	Frequency	Subperiod
assessment	143	2011–2015	assessment	199	2016–2020
system	48	2011–2015	system	67	2016–2020
student	38	2011–2015	student	48	2016–2020
college	36	2011–2015	college	45	2016–2020
regional	22	2011–2015	subject	29	2016–2020
teacher	22	2011–2015	regional	26	2016–2020
higher education	19	2011–2015	teacher	22	2016–2020
index	18	2011–2015	teaching	21	2016–2020
performance	16	2011–2015	primary and secondary schools	20	2016–2020
mechanism	15	2011–2015	quality	20	2016–2020
method	14	2011–2015	policy	19	2016–2020
primary and secondary schools	13	2011–2015	higher education	19	2016–2020
model	12	2011–2015	mechanism	19	2016–2020
teaching	12	2011–2015	ability	18	2016–2020
subject	11	2011–2015	development	18	2016–2020
efficacy	11	2011–2015	vocational education	18	2016–2020
vocational education	11	2011–2015	core competency	17	2016–2020
physical and mental health	10	2011–2015	informatization	15	2016–2020
development	10	2011–2015	model	15	2016–2020
ability	10	2011–2015	monitoring	15	2016–2020

**Figure 4.** Themes distribution of the eight clusters of projects between 2011 and 2015. This figure presents the distribution of research themes of projects on educational evaluation during the 12th Five-Year Plan period.

Cluster 1—Educational evaluation mechanism (22 keywords, a total word frequency of 263). This cluster focuses on evaluation, education, mechanisms, modes, and standards.

Cluster 2—Teacher performance evaluation (20 keywords, a total word frequency of 202). This cluster's most frequent research themes cover higher education institutions, teachers, evaluation indicators, performance, and application.

regions, policy, projects and vocational education, and industry–university–research cooperation and its impact.

Cluster 4—Evaluation of the high-quality development of first-class universities and first-class disciplines (abbreviated as “Double First-class”) initiative (19 keywords, a total word frequency of 154). The research themes in this cluster focus on universities, world-class universities, and first-class disciplines, effectiveness, and third-party evaluation.

Cluster 5—Monitoring the quality of basic education (12 keywords, a total word frequency of 87). The research topics of this cluster mainly include quality, monitoring, basic education, and balanced development.

Cluster 6—Evaluation of innovation and entrepreneurship (7 keywords, a total word frequency of 49). This cluster’s most frequent research themes include innovation, innovation and entrepreneurship, educational governance, and strategy.

Cluster 7—Evaluation reform of the college entrance examinations (8 keywords, a total word frequency of 42). The research themes in this group involve reform, new college entrance examinations, examinations, college enrollment, and practices.

3.5. Evolution of Research Themes

Figure 6 presents a strategic diagram of the research themes of projects on educational evaluation for the 12th (2011–2015) and 13th (2016–2020) Five-Year Plan periods (Figure 6a,b), as well as a Sankey diagram of the evolution of the themes (Figure 6c).

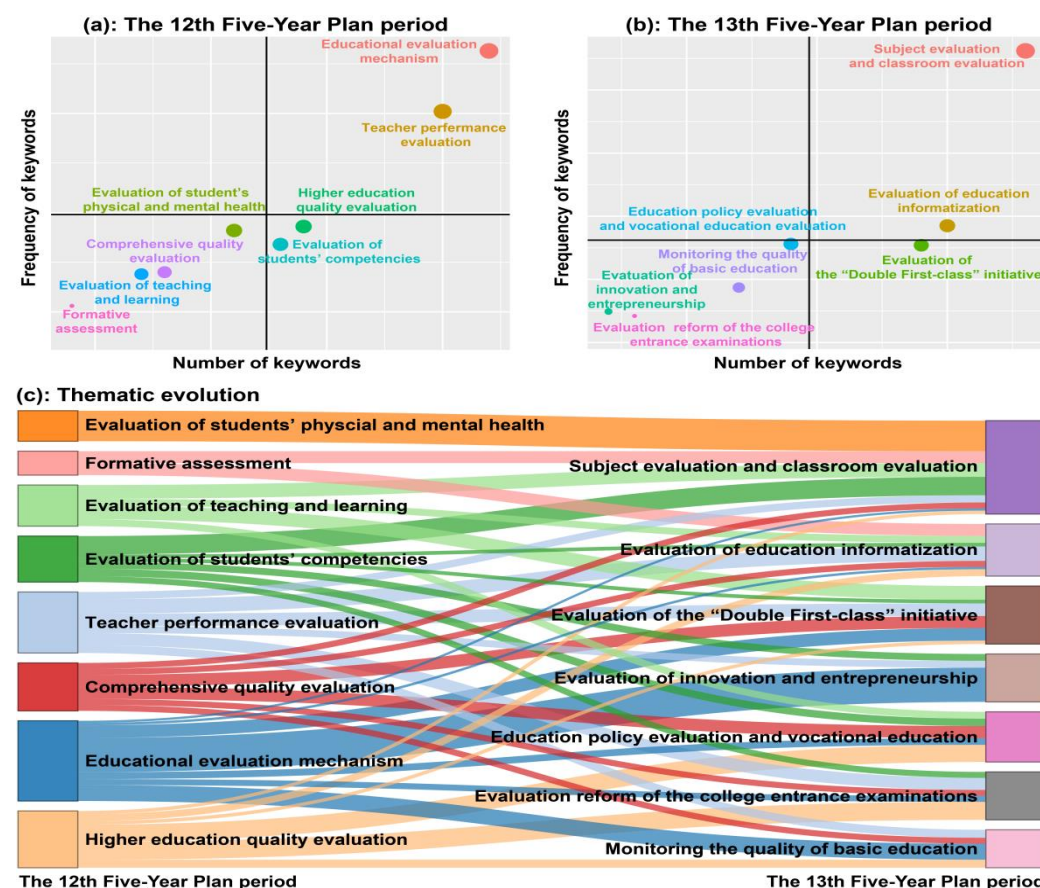


Figure 6. Thematic maps (strategic diagrams: (a,b)) and evolution (Sankey graph: (c)) of research themes. This figure presents the research themes of projects on educational evaluation during the 12th Five-Year Plan period and the 13th Five-Year Plan period (Figure 6a,b), as well as a Sankey diagram of the evolution of the themes (Figure 6c).

Figure 6a shows that the eight research themes for the 12th Five-Year Plan period are located in the first, second, and fourth quadrants, respectively. Research on educa-

tional evaluation mechanisms and teacher performance evaluation are located in the first quadrant, indicating that these two research themes have received significant attention and have been thoroughly explored. Students' physical and mental health evaluation, comprehensive quality evaluation, teaching evaluation, and formative assessment are in the third quadrant. Among these themes, except for students' physical and mental health evaluation, which is close to the average level, the other three research themes have received less attention and are characterized by a low degree of richness in the research contents. Higher education quality evaluation and evaluation of students' competencies are located in the fourth quadrant, reflecting a low degree of focus on the themes but a high degree of richness in research contents.

Figure 6b,c shows that the seven research themes for the 13th Five-Year Plan period are located in quadrants 1, 2, and 4, respectively, and have evolved from the eight research themes in the 12th Five-Year Plan period. Located in the first quadrant, subject evaluation, classroom evaluation, and evaluation of education informatization have been the focus of educational researchers, especially subject evaluation and classroom evaluation. With a high degree of focus and richness in research content, these two research themes have evolved from the teaching evaluation and evaluation of students' physical and mental health in the 12th Five-Year Plan period.

Located in the third quadrant are the research themes of education policy evaluation and vocational education evaluation, basic education quality monitoring, innovation and entrepreneurship evaluation, and evaluation reform of the college entrance examination. As shown in Figure 6c, the research themes, including basic education quality monitoring, innovation and entrepreneurship evaluation, and evaluation reform of the college entrance examination, are developed from the comprehensive quality evaluation, education mechanism evaluation, higher education evaluation, teaching evaluation, and teacher performance evaluation in the 12th Five-Year Plan period. Although the four research themes in the third quadrant have received less attention and are less richly researched, they are emerging themes in response to the development of educational reforms. In the fourth quadrant, the evaluation of the "Double First-class" initiative stands out in terms of the richness of research contents and the wide range of research scopes. The evaluation of the "Double First-class" initiative has evolved from themes such as the educational evaluation mechanisms, teacher performance evaluation, teaching evaluation, comprehensive quality evaluation, higher education evaluation, and student competency evaluation during the 12th Five-Year Plan period.

4. Discussion

4.1. Development in the 12th Five-Year Plan Period

During the 12th Five-Year Plan period, the number of approved projects on educational evaluation in Chinese National Education Science Planning was 151, accounting for 7.16% of the total number of approved projects. Regarding the units responsible for the educational evaluation projects, the 12th Five-Year Plan period witnessed diversity in the research units, which mainly included higher education institutions, scientific research institutions, primary and secondary schools, administrative departments, educational societies, and publishing units. Regarding the number of projects undertaken by each kind of research unit, higher education institutions occupied the highest place, accounting for 85.43% of all projects, and became the primary force of educational evaluation research. In comparison, the scientific research institutions only accounted for 6.62%, while the primary and secondary schools merely accounted for 1.32%. The scarcity of the research institutions' and schools' participation reflects the structural imbalance of the educational research team and the inadequacy of the research force from the staff of scientific research institutions and front-line teachers. According to this study, the research on educational evaluation during this period covers topics such as student evaluation, teacher evaluation, and evaluation methods. The research themes in this period focus on student comprehensive quality evaluation, teacher performance evaluation, and evaluation standards. Through a thor-

ough examination of international research papers on educational evaluation from 2010 to 2019, Wang et al. [9] found that student achievement evaluation, teacher effectiveness, professional development evaluation, and the validity of evaluation criteria were hot topics in international educational evaluation research. A similarity in topics of educational evaluation research is observed between Chinese educational science researchers and global researchers.

In terms of student evaluation, the research scope is no longer limited to the evaluation of knowledge levels, but extends to students' outcomes, with a particular emphasis on students' general competencies [49–51]. The themes of student evaluation in China's research projects mainly include the all-around performance of primary and secondary school students from character development, academic achievement, physical and mental health levels, artistic literacy, and social practice. For example, the city of Shanghai conducted a comprehensive quality evaluation of high school students from the aspects of students' social responsibility, professional aspirations and talents, personality traits, personal hobbies, character development and citizenship, courses of study and academic achievement, physical and mental health and artistic literacy, and innovation and practical skills. In terms of teacher evaluation research, teachers' professional development plays a pivotal role in educational improvement [52,53], and previous studies have found that performance evaluations can effectively promote teachers' professional development [54,55]. The themes of teacher evaluation in China's research projects focus on changes in teachers' instructional practices, students' achievement, and research achievements. Regarding the research of educational evaluation methods, evaluation standard is a popular keyword. Evaluation standards can enhance the validity of evaluation and promote the quality of education [56,57]. China's "Twelfth Five-Year Plan for National Educational Development" sets out the requirement to establish a sound national system of educational standards. In educational research, the approved projects primarily explore themes such as higher education quality evaluation standards, student evaluation standards, and subject evaluation standards.

4.2. Development in the 13th Five-Year Plan Period

During the 13th Five-Year Plan period, the number of approved projects on educational evaluation in Chinese National Education Science Planning was 219, with an increase of 68 compared with that of the 12th Five-Year Plan period. The rapid development of project establishment in educational evaluation reflects the intense attention and support from the Chinese education authorities, highlighting the importance and value of educational evaluation in educational reforms and development. Regarding the number of approved projects among various responsible units, higher education institutions still ranked first in the 13th Five-Year Plan period, accounting for 81.74% of the total. Nevertheless, it showed a decrease of 3.73%. On the other hand, the number of projects undertaken by scientific research institutions and primary and secondary schools had been steadily increasing, with their proportions being 11.42% and 3.65%, respectively. The proportion increase compared with the 12th Five-Year Plan period reflects the increased attention from research institutions and primary and secondary schools to education evaluation research and their enhanced participation, probably due to the support and promotion of education authorities. The study results have shown that during the 13th Five-Year Plan period, research on educational evaluation in China shares similar themes with international educational evaluation research. They both focus on classroom evaluation, program evaluation, and big-data-based educational evaluation, but research topics with Chinese characteristics have also emerged.

As a critical area of educational evaluation, the classroom is considered a primary source of effective feedback and diagnostic information for teachers [58]. The focus of educational evaluation research in China has shifted from teacher performance evaluation in the 12th Five-Year Plan period to the evaluation of teaching effectiveness. Program evaluation has been receiving increasing attention as educational accountability was estab-

lished [59]. The themes in China's projects focus on the implementation and effects of the National Basic Education Quality Monitoring Program, the STEM Program in schools, and the School Collaborative Education Program. As for big-data-based educational evaluation, teachers can use evaluation information to adjust teaching strategies, while students can improve their learning methods through feedback data [60]. The research themes in China's projects relate to big data in educational evaluation, including the areas of evaluation of classroom teaching, student learning evaluation, and evaluation of school education quality.

In addition, this research has shown that some educational evaluation research themes with Chinese characteristics have emerged for the first time in the 13th Five-Year Plan period. Examples include the evaluation of the "Double First-class" initiative; the evaluation of cooperation between industry, universities, and research institutions; the evaluation of innovation and entrepreneurship; and the evaluation reform of the college entrance examination. Two factors have mainly brought about these new themes. One of them is the influence of national policies. Education policies guide people's educational behavior and the direction of educational activities [61]. In recent years, China has formulated a series of policy documents to accelerate the development of the "Double First-class" initiative. The initiative refers to the development of first-class universities and disciplines of the world. It is China's strategic plan for the development of higher education, aiming to enhance the overall strength and international competitiveness of Chinese higher education, facilitate a number of prestigious universities and disciplines to become world-class and promote the level of talent cultivation, scientific research, social services and cultural inheritance and innovation in higher education. The research in China's research projects focuses on the evaluation of the effectiveness of the "Double First-class" initiative, the evaluation of the university disciplines included in "Double First-class" projects, and the evaluation of the scientific research performance of "Double First-class" universities. The other influencing factor is the demand for educational practices. Educational practices promote the development of scientific research [62]. With the rapid development of the economy and society, China is also facing many new challenges in education, such as the lack of close cooperation between enterprises, universities, and research institutions; the inadequacy of innovation and entrepreneurship among university students; and the requirement for improving the form and content of the college entrance examination. Responding to the needs in education practice, research themes such as evaluating the effectiveness of industry–university–research cooperation, evaluating university students' innovation and entrepreneurship, and reforming the college entrance examination have received more attention. These explorations reflect the importance of localized studies in educational science research in China.

5. Conclusions

The study has uncovered basic facts, research themes, and the evolution of educational evaluation research in China. It will facilitate China's high-quality development of educational science research in the 14th Five-Year Plan period and the educational evaluation research in other countries. To further promote China's educational research, the education administrative departments need to attach great importance to constructing the scientific research team of primary and secondary school teachers. They should encourage and support the teachers to strengthen cooperation with universities and research institutes to carry out scientific research, enhance their scientific research awareness, and improve their scientific research level through cooperation. Secondly, researchers should actively conduct local education theory research drawing from foreign research achievements. They should focus on themes such as the "Double First-class" initiative evaluation, industry-university-research evaluation, and the evaluation reform of the college entrance examinations to establish evaluation systems, indicators, standards, and mechanisms that are suitable for the national context and strive to build an education evaluation system that is consistent with China's reality and of world standard. Thirdly, the researchers need to keep pace with the times, actively respond to educational practice needs, address significant issues in

educational reform and development and hot issues of social concern, and seek effective ways to solve educational problems. At the same time, international researchers could draw upon the research method and framework in this study to further promote educational evaluation research. They could conduct the research with a combination of bibliometrics, textual analysis, and visualization methods, which will enable them to gain a more comprehensive and objective view, discover the development patterns, and reasonably identify the research trends.

This study has some limitations. On the one hand, only the approved research projects were analyzed, while relevant research papers were not included, potentially limiting the sample adequacy. On the other hand, only projects in Chinese National Education Science Planning were selected, with other types of research projects being excluded. The sample is relatively homogeneous. Nevertheless, the study can still shed light on the direction of educational evaluation research in China because the research projects examined are of the highest level and most authoritative in China's education field. This in-depth exploration of research projects is hoped to provide a new perspective on educational evaluation research and offer a practical methodology and framework for bibliometric research on educational evaluation.

Author Contributions: Conceptualization, S.W. and J.Z.; methodology, S.W. and Y.Y.; software, Y.Y.; validation, S.W., Y.Y. and J.Z.; formal analysis, S.W. and Y.Y.; writing-original draft preparation, S.W.; writing-review and editing, S.W. and J.Z.; project administration: S.W., J.Q., J.Z. and Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Zhejiang Provincial Philosophy and Social Sciences Planning Project (Grant No. 22NDJC016Z), the Educational Science Planning Project of Zhejiang Province (Grant No. 2021SCG003), and the Research Start-up Fund, Hangzhou Dianzi University (Grant No. KYS265621030).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: According to the data access policies, the data used to support the findings of this study are available from the Web of the National Office for Education Sciences Planning official website. Reasonable requests for data can be made by email: 42465@hdu.edu.cn.

Acknowledgments: We would like to thank all the participants of the research.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ferreira-Mello, R.; Andre, M.; Pinheiro, A.; Costa, E.; Romero, C. Text mining in education. *Wires Data Min. Knowl.* **2019**, *9*, e1332. [\[CrossRef\]](#)
2. Foll, P.; Thiesse, F. Exploring Information Systems Curricula A Text Mining Approach. *Bus. Inf. Syst. Eng.* **2021**, *63*, 711–732. [\[CrossRef\]](#)
3. Jalali, S.M.J.; Park, H.W.; Vanani, I.R.; Pho, K.H. Research trends on big data domain using text mining algorithms. *Digit. Scholarsh. Humanit.* **2022**, *36*, 361–370. [\[CrossRef\]](#)
4. He, W. Examining students' online interaction in a live video streaming environment using data mining and text mining. *Comput. Hum. Behav.* **2013**, *29*, 90–102. [\[CrossRef\]](#)
5. Pereira, D.; Flores, M.A.; Niklasson, L. Assessment revisited: A review of research in Assessment and Evaluation in Higher Education. *Assess. Eval. High. Educ.* **2016**, *41*, 1008–1032. [\[CrossRef\]](#)
6. Marx, E.; Hudson, N.; Deal, T.B.; Pateman, B.; Middleton, K. Promoting health literacy through the Health Education Assessment Project. *J. Sch. Health* **2007**, *77*, 157–163. [\[CrossRef\]](#)
7. Chugh, R.; Macht, S.; Harreveld, B. Supervisory feedback to postgraduate research students: A literature review. *Assess. Eval. High. Educ.* **2021**, *47*, 683–697. [\[CrossRef\]](#)
8. Wang, S.; Li, A.Y.; Su, J.Y. A Visual Analysis of the research Hot-spots and Trends of International Education Evaluation in recent Ten Years. *Heilongjiang Res. High. Educ.* **2021**, *321*, 14–22. (In Chinese) [\[CrossRef\]](#)
9. Altin, S.V.; Finke, I.; Kautz-Freimuth, S.; Stock, S. The evolution of health literacy assessment tools: A systematic review. *BMC Public Health* **2014**, *14*, 1207. [\[CrossRef\]](#)

10. Rogaten, J.; Rienties, B.; Sharpe, R.; Cross, S.; Whitelock, D.; Lygo-Baker, S.; Littlejohn, A. Reviewing affective, behavioural and cognitive learning gains in higher education. *Assess. Eval. High. Educ.* **2019**, *44*, 321–337. [[CrossRef](#)]
11. Wong, H.Y.H.; Chans, C.K.Y. A systematic review on the learning outcomes in entrepreneurship education within higher education settings. *Assess. Eval. High. Educ.* **2021**, 1–18. [[CrossRef](#)]
12. Nachbauer, M.; Kyriakides, L. A review and evaluation of approaches to measure equity in educational outcomes. *Sch. Eff. Sch. Improv.* **2019**, *31*, 306–331. [[CrossRef](#)]
13. Meek, S.E.M.; Blakemore, L.; Marks, L. Is peer review an appropriate form of assessment in a MOOC? Student participation and performance in formative peer review. *Assess. Eval. High. Educ.* **2017**, *42*, 1000–1013. [[CrossRef](#)]
14. Hopfenbeck, T.N.; Lenkeit, J.; El Masri, Y.; Cantrell, K.; Ryan, J.; Baird, J.A. Lessons Learned from PISA: A Systematic Review of Peer-Reviewed Articles on the Programme for International Student Assessment. *Scand. J. Educ. Res.* **2018**, *62*, 333–353. [[CrossRef](#)]
15. Panadero, E.; Alqassab, M. An empirical review of anonymity effects in peer assessment, peer feedback, peer review, peer evaluation and peer grading. *Assess. Eval. High. Educ.* **2019**, *44*, 1253–1278. [[CrossRef](#)]
16. Blackhart, G.C.; Peruche, B.M.; DeWall, C.N. Factors influencing teaching evaluations in higher education. *Teach. Psychol.* **2006**, *33*, 37–39. [[CrossRef](#)]
17. Brady, D.R.; Asselin, M.E. Exploring outcomes and evaluation in narrative pedagogy: An integrative review. *Nurse Educ. Today* **2016**, *45*, 1–8. [[CrossRef](#)]
18. Staff, A.I.; Oosterlaan, J.; van der Oord, S.; Hoekstra, P.J.; Vertessen, K.; de Vries, R.; van den Hoofdakker, B.J.; Luman, M. The Validity of Teacher Rating Scales for the Assessment of ADHD Symptoms in the Classroom: A Systematic Review and Meta-Analysis. *J. Atten. Disord.* **2020**, *25*, 1578–1593. [[CrossRef](#)]
19. Ryan, T.; Henderson, M.; Ryan, K.; Kennedy, G. Designing learner-centred text-based feedback: A rapid review and qualitative synthesis. *Assess. Eval. High. Educ.* **2020**, *46*, 894–912. [[CrossRef](#)]
20. Winstone, N.E.; Nash, R.A.; Parker, M.; Rowntree, J. Supporting Learners' Agentic Engagement With Feedback: A Systematic Review and a Taxonomy of Recipience Processes. *Educ. Psychol.* **2018**, *52*, 17–37. [[CrossRef](#)]
21. Panadero, E.; Jonsson, A.; Botella, J. Effects of self-assessment on self-regulated learning and self efficacy: Four meta-analyses. *Educ. Res. Rev.* **2018**, *22*, 74–98. [[CrossRef](#)]
22. Zell, E.; Strickhouser, J.E.; Sedikides, C.; Alicke, M.D. The Better-Than-Average Effect in Comparative Self-Evaluation: A Comprehensive Review and Meta-Analysis. *Psychol. Bull.* **2020**, *146*, 118–149. [[CrossRef](#)] [[PubMed](#)]
23. Panadero, E.; Brown, G.T.L.; Strijbos, J.W. The Future of Student Self-Assessment: A Review of Known Unknowns and Potential Directions. *Educ. Psychol. Rev.* **2016**, *28*, 803–830. [[CrossRef](#)]
24. Ford, T.G.; Van Sickle, M.E.; Clark, L.V.; Fazio-Brunson, M.; Schween, D.C. Teacher Self-Efficacy, Professional Commitment, and High-Stakes Teacher Evaluation Policy in Louisiana. *Educ. Policy* **2017**, *31*, 202–248. [[CrossRef](#)]
25. Derrington, M.L.; Kirk, J. Linking job-embedded professional development and mandated teacher evaluation: Teacher as learner. *Prof. Dev. Educ.* **2017**, *43*, 630–644. [[CrossRef](#)]
26. Donaldson, M.L.; Woulfin, S.; LeChasseur, K.; Cobb, C.D. The structure and substance of teachers' opportunities to learn about teacher evaluation reform: Promise or pitfall for equity? *Equity Excell. Educ.* **2016**, *49*, 183–201. [[CrossRef](#)]
27. Bartimote-Aufflick, K.; Bridgeman, A.; Walker, R.; Sharma, M.; Smith, L. The study, evaluation, and improvement of university student self-efficacy. *Stud. High. Educ.* **2016**, *41*, 1918–1942. [[CrossRef](#)]
28. Gotlib, D.; Saragoza, P.; Segal, S.; Goodman, L.; Schwartz, V. Evaluation and Management of Mental Health Disability in Post-secondary Students. *Curr. Psychiatr. Rep.* **2019**, *21*, 43. [[CrossRef](#)]
29. McKenna, J.W.; Kim, M.K.; Shin, M.; Pfannenstiel, K. An Evaluation of Single-Case Reading Intervention Study Quality for Students with and At Risk for Emotional and Behavioral Disorders. *Behav. Modif.* **2017**, *41*, 868–906. [[CrossRef](#)]
30. Tang, X.; Yin, Y.; Lin, Q.; Hadad, R.; Zhai, X. Assessing computational thinking: A systematic review of empirical studies. *Comput. Educ.* **2020**, *148*, 103798. [[CrossRef](#)]
31. Siddiq, F.; Hatlevik, O.E.; Olsen, R.V.; Throndsen, I.; Scherer, R. Taking a future perspective by learning from the past—A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educ. Res. Rev.* **2016**, *19*, 58–84. [[CrossRef](#)]
32. Panadero, E.; Jonsson, A. The use of scoring rubrics for formative assessment purposes revisited: A review. *Educ. Res. Rev.* **2013**, *9*, 129–144. [[CrossRef](#)]
33. Hartmeyer, R.; Stevenson, M.P. A systematic review of concept mapping-based formative assessment processes in primary and secondary science education. *Assess. Educ.* **2018**, *25*, 598–619. [[CrossRef](#)]
34. Schildkamp, K.; van der Kleij, F.M.; Heitink, M.C.; Kippers, W.B.; Veldkamp, B.P. Formative assessment: A systematic review of critical teacher prerequisites for classroom practice. *Int. J. Educ. Res.* **2020**, *103*, 101602. [[CrossRef](#)]
35. Warren, S.J.; Churchill, C. Strategic, operations, and evaluation planning for higher education distance education. *Distance Educ.* **2022**, *43*, 239–270. [[CrossRef](#)]
36. Harris, L.; Wyatt-Smith, C.; Adie, L. Using data walls to display assessment results: A review of their affective impacts on teachers and students. *Teach. Teach.* **2020**, *26*, 50–66. [[CrossRef](#)]
37. Day, I.N.Z.; van Blankenstein, F.M.; Westenberg, M.; Admiraal, W. A review of the characteristics of intermediate assessment and their relationship with student grades. *Assess. Eval. High. Educ.* **2018**, *43*, 908–929. [[CrossRef](#)]

38. Stenlund, T. Assessment of prior learning in higher education: A review from a validity perspective. *Assess. Eval. High. Educ.* **2010**, *35*, 783–797. [[CrossRef](#)]
39. Padilla, F.M.; Gallardo, M.; Manzano-Agugliaro, F. Global trends in nitrate leaching research in the 1960–2017 period. *Sci. Total Environ.* **2018**, *643*, 400–413. [[CrossRef](#)]
40. Ho, Y.S.; Satoh, H.; Lin, S.Y. Japanese lung cancer research trends and performance in Science Citation Index. *Intern. Med.* **2010**, *49*, 2219–2228. [[CrossRef](#)]
41. Zhu, J.J.; Dressel, W.; Pacion, K.; Ren, Z.Y.J. ES&T in the 21st Century: A Data-Driven Analysis of Research Topics, Interconnections, and Trends in the Past 20 Year. *Environ. Sci. Technol.* **2021**, *55*, 3453–3464. [[CrossRef](#)] [[PubMed](#)]
42. Baker, H.K.; Kumar, S.; Pattnaik, D. Twenty-five years of the Journal of Corporate Finance: A scientometric analysis. *J. Corp. Finance* **2021**, *66*, 101572. [[CrossRef](#)]
43. Aria, M.; Cuccurullo, C. bibliometrix: An R-tool for comprehensive science mapping analysis. *J. Informetr.* **2017**, *11*, 959–975. [[CrossRef](#)]
44. Lozano, S.; Calzada-Infante, L.; Adenso-Diaz, B.; Garcia, S. Complex network analysis of keywords co-occurrence in the recent efficiency analysis literature. *Scientometrics* **2019**, *120*, 609–629. [[CrossRef](#)]
45. Cobo, M.J.; Lopez-Herrera, A.G.; Herrera, F. An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *J. Informetr.* **2011**, *5*, 146–166. [[CrossRef](#)]
46. Stegmann, J.; Grohmann, G. Hypothesis generation guided by co-word clustering. *Scientometrics* **2003**, *56*, 111–135. [[CrossRef](#)]
47. Yu, Y.L.; Jin, Z.; Qiu, J.P. Global Isotopic Hydrograph Separation Research History and Trends: A Text Mining and Bibliometric Analysis Study. *Water* **2021**, *13*, 2529. [[CrossRef](#)]
48. Seabra, D.; Caldeira-pires, A. The thermodynamic rarity concept: A systematic review. *Ecol. Indic.* **2020**, *108*, 105689. [[CrossRef](#)]
49. Sadler, D.R. Three in-course assessment reforms to improve higher education learning outcomes. *Assess. Eval. High. Educ.* **2016**, *41*, 1081–1099. [[CrossRef](#)]
50. Kember, D. Nurturing generic capabilities through a teaching and learning environment which provides practise in their use. *High. Educ.* **2009**, *57*, 37–55. [[CrossRef](#)]
51. Barrie, S. Understanding What We Mean by the Generic Attributes of Graduates. *High. Educ.* **2006**, *51*, 215–241. [[CrossRef](#)]
52. McChesney, K.; Aldridge, J.M. A review of practitioner-led evaluation of teacher professional development. *Prof. Dev. Educ.* **2019**, *45*, 307–324. [[CrossRef](#)]
53. Ritter, G.W.; Barnett, J.H. Learning on the job: Teacher evaluation can foster real growth. *Phi Delta Kappan* **2016**, *97*, 48–52. [[CrossRef](#)]
54. Tuytens, M.; Devos, G.; Vanblaere, B. An integral perspective on teacher evaluation: A review of empirical studies. *Educ. Assess. Eval. Acc.* **2020**, *32*, 153–183. [[CrossRef](#)]
55. Sun, M.; Penuel, W.R.; Frank, K.A.; Gallagher, H.A.; Youngs, P. Shaping Professional Development to Promote the Diffusion of Instructional Expertise among Teachers. *Educ. Eval. Policy Anal.* **2013**, *35*, 344–369. [[CrossRef](#)]
56. Boleslavsky, R.; Cotton, C. Grading Standards and Education Quality. *Am. Econ. J. Microecon.* **2015**, *7*, 248–279. [[CrossRef](#)]
57. Noaman, A.Y.; Ragab, A.H.M.; Madbouly, A.I.; Khedra, A.M.; Fayoumi, A.G. Higher education quality assessment model: Towards achieving educational quality standard. *Stud. High. Educ.* **2017**, *42*, 23–46. [[CrossRef](#)]
58. Cohen, J.; Goldhaber, D. Building a More Complete Understanding of Teacher Evaluation Using Classroom Observations. *Educ. Res.* **2016**, *45*, 378–387. [[CrossRef](#)]
59. Ryan, K. Shaping educational accountability systems. *Am. J. Eval.* **2002**, *23*, 453–468. [[CrossRef](#)]
60. Prashanti, E.; Ramnarayan, K. Ten maxims of formative assessment. *Adv. Physiol. Educ.* **2019**, *43*, 99–102. [[CrossRef](#)]
61. Santos, M.; Fetzner, A.R. Procampo: Rural Education Policy. *Rev. Tempos Espac. Educ.* **2019**, *12*, 189–206. [[CrossRef](#)]
62. Gutierrez, E.J.D. Intercultural education practice in Secondary Education. *Rev. Educ.* **2014**, *363*, 12–34. [[CrossRef](#)]