ALEXANDRIA

Revista de Educação em Ciência e Tecnologia

Trends in research on the statistical education of future mathematics teachers: a bibliometric study

Tendências nas pesquisas sobre a formação estatística do futuro professor de matemática: um estudo bibliométrico

Sandra Cristina Martini Rostirola [1]
https://orcid.org/0000-0003-2482-6117
Elisa Henningi[2]
https://orcid.org/0000-0002-7754-9451
Ivanete Zuchi Siplei[2]
https://orcid.org/0000-0002-8640-1336

1. Departamento de Ensino, Pesquisa e Extensão, Instituto Federal Catarinense, Videira, Brasil. E-mail: sandra.rostirola@ifc.edu.br
2. Departamento de Matemática, Universidade do Estado de Santa Catarina, Joinville, Brasil. E-mail: elisa.henning@udesc.br; ivanete.siple@udesc.br

Abstract: The vast amount of data individuals currently handle makes knowledge of statistics essential for educators in their pedagogical practice. It is crucial to elucidate the statistical education offered in mathematics teacher education programs, as future teachers will address these topics in classrooms. This study aimed to outline the landscape in which mathematics teachers' statistical education takes place through a bibliometric analysis of 731 articles from the *Scopus* and *Web of Science* databases. A descriptive analysis and correspondence factor analysis were conducted to identify clusters related to emerging themes. The results revealed four main groupings guiding research on teacher education: teacher professional development; educators' statistical education; statistical literacy, reasoning, and thinking; and data literacy.

Keywords: mathematics teacher education, statistical education,teacher training, bibliometrics.

Resumo: A grande quantidade de dados que os indivíduos lidam atualmente torna o conhecimento da Estatística fundamental para os educadores em sua prática pedagógica. É essencial elucidar a formação estatística nos cursos de Licenciatura em Matemática, já que são os futuros professores que abordarão esses temas nas salas de aula. Este estudo teve como objetivo apresentar um esboço do cenário em que ocorre a formação estatística do professor de matemática por meio de uma análise bibliométrica de 731 artigos das bases *Scopus* e *Web of Science*. Foi realizada uma análise descritiva e análise fatorial por correspondência para

identificar clusters relacionados às temáticas emergentes. Os resultados revelaram quatro agrupamentos que orientaram as pesquisas relativas à formação docente: desenvolvimento profissional docente; formação estatística do educador, letramento, raciocínio e pensamento estatísticos e letramento em dados.

Palavras-chave: licenciatura em matemática, educação estatística, formação de professores, bibliometria.

Introduction

With the expansion of information and communication technologies in the 21st century, the need arose for the development of skills related to data analysis and interpretation. In this context, it is understood that statistical education in schools needs to be strengthened, which requires the training of teachers who teach mathematics. This discussion should start from initial teacher training, since it is there that professional teacher development begins.

Therefore, a bibliometric analysis was carried out with the aim of presenting an outline of the scenario in which statistical training of mathematics teachers takes place, looking for evidence of the approaches and trends that are being used in initial training programs and that impact teacher preparation for teaching statistics.

Bibliometrics is one of the most comprehensive types of systematic literature review. This technique consists in analyzing publications, whether books, reports or articles (Araújo, 2011), in order to quantify and evaluate the scientific production of different themes (Ramos-Rodríguez & Ruíz-Navarro, 2004). It summarizes large amounts of bibliometric data and presents the intellectual structure and emerging trends of a topic or field of research (Donthu et al. 2021). This possibility of summarization justifies the choice of this method in this study.

The methods used in this research relate to descriptive statistics and exploratory data analysis, initially proposing a synthesis of the portfolio of texts, considering the most relevant authors, works and sources, that is, those with the highest production and frequency of citations. Subsequently, a thematic analysis was performed considering titles, abstracts and keywords using clustering tools available in the *Biblioshiny* web-based tool (Aria & Cuccurullo, 2017). This article reports on two phases of the research: the first portraying bibliometric aspects and the second seeking to characterize the trends present in the various articles that make up the portfolio in question.

The results allowed for the clarification of theoretical concepts and their relationship with the development of ideas and actions regarding literacy, reasoning and statistical thinking in the initial training of mathematics teachers. It is noteworthy that, unlike applying a microscope and exhaustively detailing the portfolio, this study

outlines its characteristics, highlighting its trends, in a manner analogous to exploring previously unknown territory using a probe.

Method

The research has a mixed-methods approach (Creswell & Clark, 2013), as it uses descriptive statistics techniques and exploratory data analysis associated with textual analysis of a portfolio of scientific articles that refer to the approach of statistics education in initial teacher training programs in mathematics.

Data collection was carried out on September 2, 2022, based on a plan that included defining the descriptors, procedures and bibliometric tools that would lead to achieving the research objective, in the *Scopus* and *Web of Science* (WOS) databases, chosen for their broad coverage of scientific journals related to education, as well as for their reputation in the academic community for applying rigorous indexing and publishing criteria.

Thus, the descriptors “training teacher”, “pre-service teacher”, “preservice teacher” and “prospective teacher” were applied to these databases, using the item “search in all fields” associated with the logical connector OR. This procedure generated a set of 82,700 studies in the *Scopus* database and 4,224 in *WOS*. However, this set did not ensure the detection of works related to statistics education. Therefore, the “search within” tool was used, which associates the logical connector AND to the set from the first stage, now searching for the descriptors “statistics education” OR “statistical education”, since, while exploring the databases in the bibliometric planning, it was noticed that both terms are used in keywords of articles that deal with the subject.

It should be noted that, along with the connectors, a date filter was applied, considering the years 2000 to 2022 in both databases. The choice of the initial date refers to the year following the “Experiences and Expectations of Statistics Education - Challenges for the 21st Century” international conference, held in Florianópolis/Santa Catarina/Brazil in 1999, considered a milestone for statistics education (Cazorla, 2006), not to mention internationally impactful movements, such as the publication of the first GAISE report (Guidelines for Assessment and Instruction in Statistics Education) in 2005.

At this stage, the number of studies was reduced to 754 in *Scopus* and 10 in *WOS*, totaling 764. Of these, works identified in the platform data as books (18), editorials (4), notes and letters (2) were also removed, resulting in 740 documents

(730 in *Scopus* and 10 in *WOS*). The bibliometric data for this portfolio were then downloaded from their respective databases in *.bib* and *.txt* extensions.

These spreadsheets were processed in RStudio (R Core Team, 2022), andmerged into a single corpus using the merge function. Furthermore, in this step, duplicate texts were removed using the remove.duplicate tool, resulting in 732 documents. However, the software uses the document title to remove duplicates, which may differ in some contexts or be missing terms or letters. Therefore, filters were applied in Excel based on the DOI (Digital Object Identifier), an action recommended by Silva et al. (2022), resulting in a portfolio of 731 documents, including research articles, conference proceedings and book chapters, which were analyzed using the *bibliometrix* package in its web interface known as Biblioshiny (Aria & Cuccurullo, 2017). It is also noted that some studies from proceedings and book chapters were retained in the research because they were filtered as standalone articles during data collection by the software and dealt with the research topics.

A descriptive analysis of the portfolio and a bibliometric approach to the study topic was carried out, considering the three fundamental principles of bibliometrics used to analyze patterns in scientific production and the productivity of authors and research sources: Zipf's Law, Lotka's Law, and Bradford's Law. The intellectual structure of the texts was analyzed through connections among citations, co-citations, and similarity between authors based on grouping by bibliographic coupling. Regarding keywords, the analysis was conducted through trend topics, and the document abstracts were analyzed using thematic maps and conceptual structure based on groupings. The data analysis and results are presented below.

Data Analysis And Results

This section will be subdivided into a quantitative description of articles, book chapters and extended abstracts presented at events, contained in the portfolio in question, and a thematic analysis of the conceptual and intellectual structure of these studies, with the aim of presenting an outline of the scenario in which the statistical training of mathematics teachers takes place.

Descriptive Analysis of the Portfolio

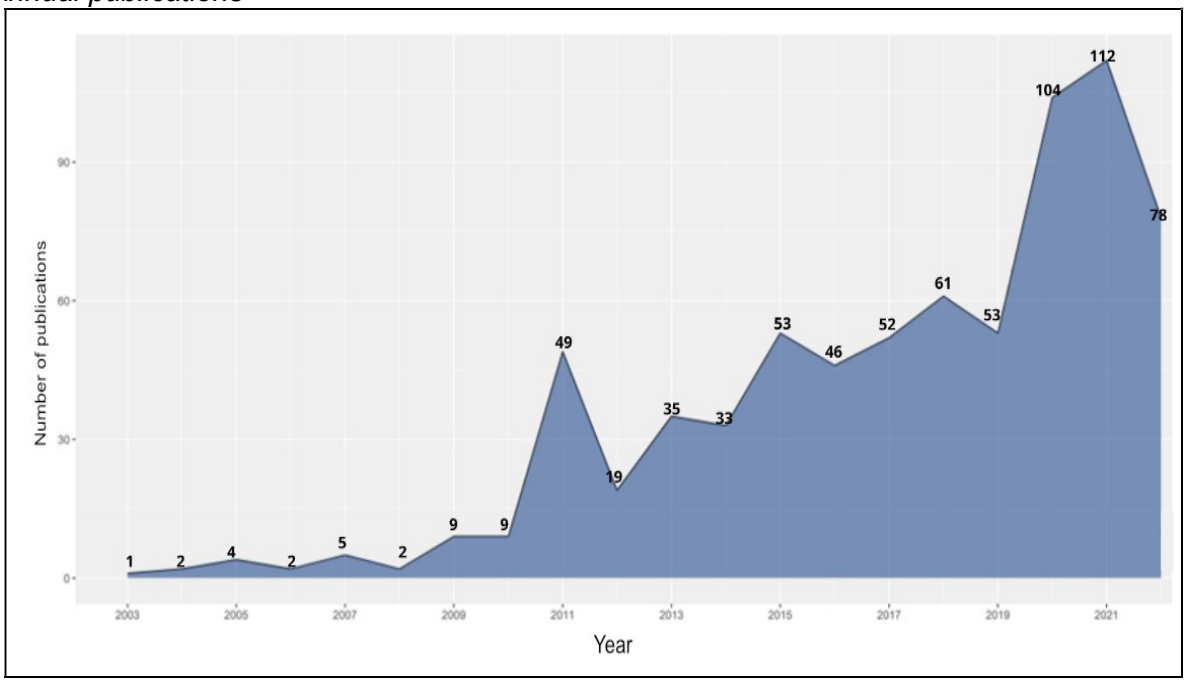
The portfolio, composed of 731 documents, is distributed across 316 research sources (including journals, conference proceedings and books). The most frequent

document type is articles (627), however, there are 63 book chapters and 41 documents that refer to events (conferences, congresses and similar) – the latter filtered by the software as individual articles and kept in the search due to their convergent themes.

The predominant language in the portfolio is English, with 656 documents. To this total, we must also add articles that have bilingual versions: 1 in English and Croatian; 8 in English and Spanish; and 4 in English and Turkish. Other languages found in the portfolio are: Chinese (1 article); French (1); German (3); Portuguese (7); Spanish (47); and Turkish (3).

The documents were published between 2003 and 2022, with an average of 36.55 publications per year. Figure 1 describes the annual quantity up to the date of data collection, considering that two articles were preprints that were not used in the graph. The average annual growth rate was 25.77%, with notable growth between 2008 and 2009, 2010 and 2011, and 2019 and 2022.

Figure 1
Annual publications



Note: [Image description] A rectangular área chart with a light gray background and white grid lines. It features a blue-filled silhouette forming an ascending polygonal line from left to right. The vertical axis representes the number of publications (with tick marksat 0, 30, 60 and 90), while the horizontal axis represents time in years (from 2003 to 2022). Above each peak and trough of the blue line, a black numerical label indicates the exact number of publications for that year. The line starts at a low level in 2003 with 1 publication, remains below 10 until 2010, makes na initial jump to 49 in 2011, drops to 19 in 2012, and then rises steadily to reach a historical peak of 112 publications in 2021, ending the chart with a decline to 78 publications in 2022. [End of description].

Source: research data (2026)

The documents, cited an average of 8 times, were produced by 1,488 authors, including 129 single-author publications and 1,359 articles in collaboration with other authors, resulting in 2.04 authors per document. In terms of content, the documents include 37,669 references (an average of 51.5 per document) and 1,767 keywords (an average of 2.4 per document).

Following the description of this general information about the portfolio, a bibliometric approach to the topic in question will be presented, considering the three fundamental principles of bibliometrics used to analyze patterns in scientific production and the productivity of authors and research sources: Zipf's Law, which analyzes the most frequent words; Lotka's Law, which contextualizes the production of authors; and Bradford's Law, whose object is the production of research sources (Cassettariet al., 2015; Pinheiro, 1983; Cândido et al., 2018).

Bradford's Law and the Emphasis on research Sources

Bradford's Law, in the field of bibliometrics, describes how scientific production on a given topic is distributed among journals. According to this principle, there is a small core of journals that concentrates most of the publications, while a progressively larger number of journals shows less productivity on the same topic. The number of journals (*n*), in the core and subsequent zones, will vary in the ratio 1: $n:n^2$ (Brookes, 1969).

The documents analyzed come from sources that can be identified as periodicals, books or events. Table 1 highlights the fourteen main sources, presented as the most frequent and belonging to Zone 1 of Bradford's Law. This law allows for the ordering of scientific periodicals in descending order of productivity, enabling the establishment of exponentially divided groupings (zones), with Zone 1 being the most productive (Pinheiro, 1983). It is noteworthy that the first and third sources correspond to periodicals specializing in statistics education, while the second presents texts on statistics education in teacher training.

Table 1
Zone 1 of Bradford's Law

Source	Articles
Statistics Education Research Journal	57
New ICMI Study Series[1]	25
Journal of Statistics Education	24
Bolema - Mathematics Education Bulletin	19
International Journal of Mathematical Education in Science and Technology	17
International Journal of Science and Mathematics Education	16
Journal of Physics: Conference Series	15
Educational Studies in Mathematics	13
Mathematical Thinking and Learning	12
Teaching Statistics	12
ZDM - Mathematics Education	11
Teaching And Teacher Education	12
Mathematics	13
Acta Scientiae	14

Source: research data (2026)

Productivity was also analyzed using the H-index (Table 2), which qualifies scientific production through the number of citations. This index is calculated by the ratio of the number of published works to their citations. For example: the H-index is 5 when a given journal has had at least five of its articles cited at least five times (Thomazet al., 2011).

Table 2
H-index

Source	H-index
New ICMI Study Series	12
Statistics Education Research Journal	10
Teaching And Teacher Education	8
Journal Of Statistics Education	7
ZDM - Mathematics Education	7
Educational Studies In Mathematics	5
International Journal for Lesson and Learning Studies	5
International Journal of Science and Mathematics Education	5
Bolema - Mathematics Education Bulletin	4
Journal Of Mathematics Teacher Education	4
Plos One	4
Sustainability (Switzerland)	4
Teaching Statistics	4
Canadian Journal if Science, Mathematics and Technology Education	3

Source: research data (2026)

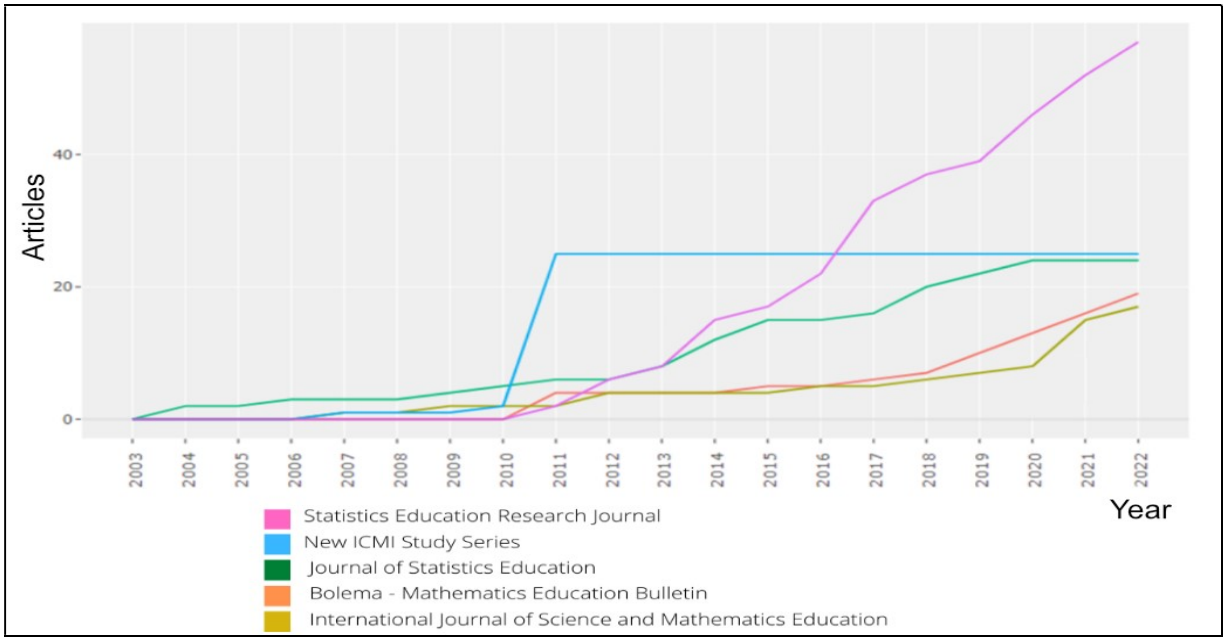
Comparing productivity in terms of frequency from Table 1 and the H-index in Table 2, we find some recurring sources such as: *New ICMI Study Series; Statistics Education Research Journal; Teaching and Teacher Education; Journal of Statistics Education; ZDM - Mathematics Education; Educational Studies in Mathematics; International Journal of Science and Mathematics Education; Bolema - Mathematics*

[1]This is a series of academic publications in the field of mathematics education from the International Commission on Mathematical Instruction.

Education Bulletin; and *Teaching Statistics*. This points to the paths taken by researchers in statistics education and alludes to the sources to which one should refer for the study of the subject.

Figure 2 shows the accumulated growth of articles in the various journals that are part of this study. It contains periodicals whose frequency of productivity on the topic is within Zone 1, according to Bradford's Law. The biggest growth observed was in the *Statistics Education Research Journal* with 57 articles; followed by the *New ICMI Study Series* of books with 25; *Journal of Statistics Education* with 24, *Bolema - Mathematics Education Bulletin* with 19; and *International Journal of Science and Mathematics Education* with 17 articles relating to the topic and the period in question. The increase in articles referring to the topic in sources not specialized in statistical education brings indications that curricular reforms introduced since the 1970s in several countries may be having effects in the field of research.

Figure 2
Growth of the topic in journals



Note. [Image description] Line graph on a light gray background, with the vertical axis indicating the number of articles and the horizontal axis representing the period from 2003 to 2022. Five lines cross the graph, extending from the left toward the upper right. The pink line represents the Statistics Education Research Journal, which stands out for its steepest growth and ends the period at nearly 60 articles. The blue line represents the New ICMI Study Series, which shows a sharp increase in 2011 and stabilizes at 25 articles through the end of the period. The remaining three lines show continuous growth: the green line represents the Journal of Statistics Education, the orange line represents Bolema, and the gold line represents the International Journal of Science and Mathematics Education. [End of description].

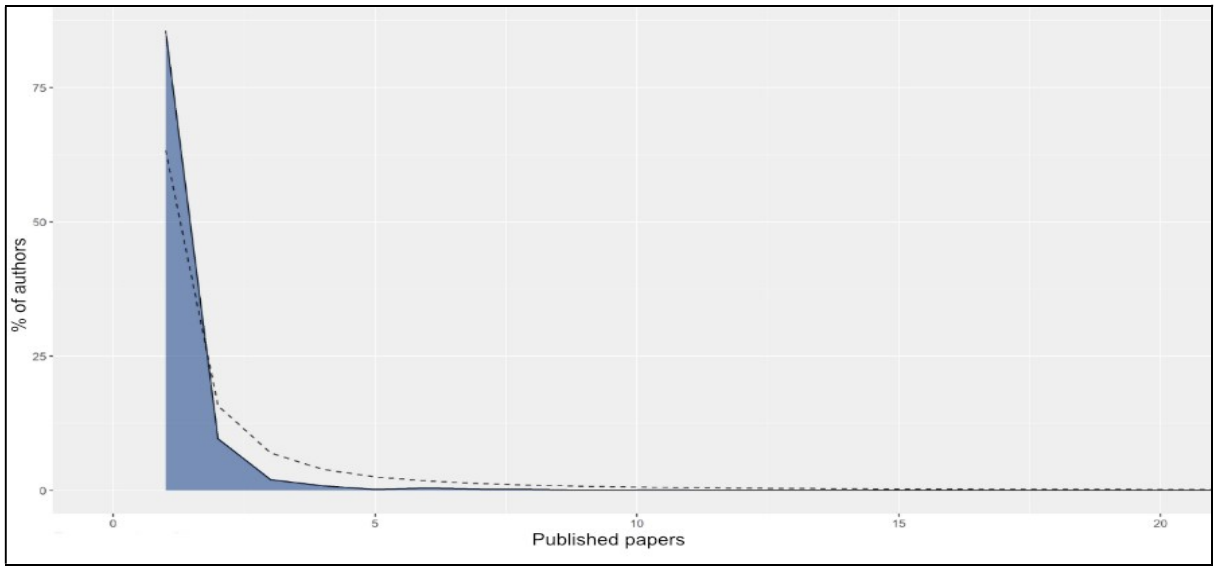
Source: research data (2026)

The next section will present productivity data for the authors who contributed to this set of documents through descriptive analyses and Lotka's Law.

Lotka's Law and Portfolio Authorship

Lotka's Law is a principle of bibliometrics that explains how the productivity of scientific authors is distributed. It is based on the premise that few authors produce a lot and many authors produce little. It allows us to study the frequency of publication of authors in a given research field, determining that the number of authors who make x contributions is approximately $\frac{1}{x^2}$ of those who make only one. Furthermore, the proportion of authors with a single contribution is about 60%; that is, the number of authors who publish a certain number of articles is a fixed proportion to the number of authors who publish a single article (Silva et al., 2022; Cândido et al., 2018). The graph in Figure 3 shows the frequency distribution of scientific production among authors, establishing a proportion of 85.6% for authors with a single publication and 0.1% for those who have 22 articles.

Figure 3
Frequency in the distribution of scientific production



Note. [Image description] Rectangular area chart with a light gray background and white grid lines. It contains a blue filled silhouette forming a curve with a sharp decline from left to right, overlaid by a black dashed line following the same downward movement. The vertical axis indicates the percentage of authors (from 0 to 75), and the horizontal axis indicates the number of documents written (from 0 to 20). [End of description].

Source: research data (2026)

With this in mind, the most productive authors in the intersection of statistics education and teacher training were analyzed – those who published the most on the

subject within the portfolio in question, considering the fractional indicator, which is an index whose mathematical model considers the co-authorship of articles (Table 3). For informational purposes, the number of articles and their H-index at the time of data collection are also presented.

Table 3
Authors with the highest production on the topic

Author	Fractional indicator	Article frequency	H-index
Leavy A	10.42	22	9
Groth R	7.50	8	4
Batanero C	6.95	21	4
Hourigan M	6.92	15	6
Sharma S	6.40	10	3
Lesser L	6.15	14	4
Frischemeier D	5.54	11	3
Biehler R	3.79	10	5
Da P J	3.50	6	3
Contreras J	3.28	12	4
Gea M	3.20	11	2
Ben-Zvi D	3.17	8	6
Lee H	3.13	8	5
Koparan T	3.00	3	2
Shin D	3.00	3	2
Casey S	2.75	6	3
Watson J	2.67	6	3
Arteaga P	2.53	9	2
Bansilal S	2.50	5	3
Estrella S	2.50	4	1

Source: research data (2026)

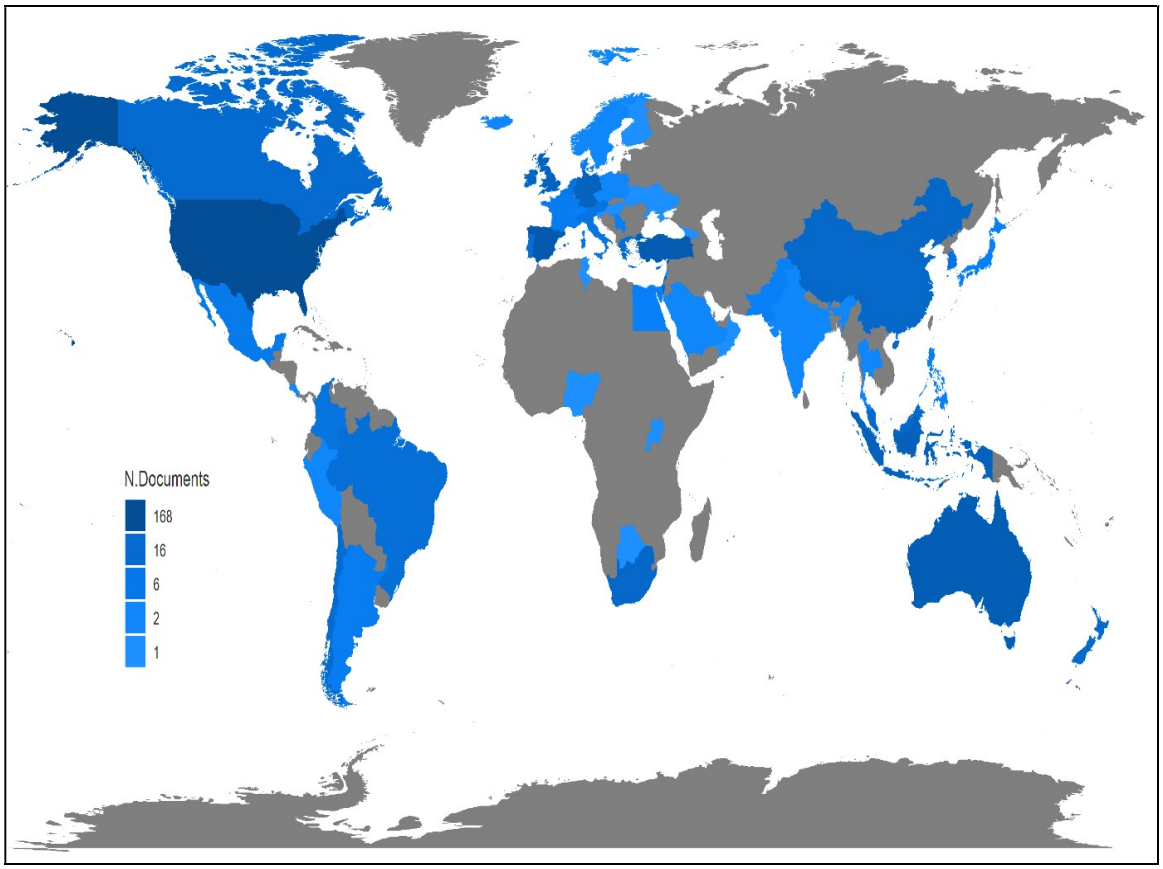
The H-index (Table 3) identifies authors with the highest number of citations relative to their articles. It is a metric that measures both the productivity and citation impact of a researcher's publications. The index is based on the set of the scientist's most cited works and the number of citations they have received in other publications (Aria & Cuccurullo, 2017). In this research project, it was calculated using the H-index function in *bibliometrix*, which refers to data obtained from the *Scopus* and *WOS* databases. In this respect, we can exemplify this with Aisling M. Leavy (Leavy A.[2]) who is in the most productive range of Lotka's Law with 22 articles and an H-index of 9. It is also worth noting that many of these authors are considered pioneers in the field, such as Carmen Batanero (Batanero C.) and Dani Ben-Zvi (Ben-zvi D.), who were already writing about the topic in the 2000s.

These researchers are distributed across various countries spanning all continents, as we can see in Figure 4, with highlights for the United States (168),

[2] Some abbreviations of names may appear differently in other studies. The format presented refers to the compilation of Biblioshiny for the portfolio under analysis.

Spain (62), Turkey (55) and Australia (47), with publications in the area to which this research converges.

Figure 4
Researchmap



Note. [Image description] Rectangular world map with a white background, where countries are shaded in blue tones to indicate the quantity of documents produced, while countries without records are shown in dark gray. In the lower-left corner, there is a vertical legend entitled “N.Documents” with five numerical markers: the darkest shade of blue represents the highest value, with 168 documents, and the tones progressively lighten for the values 16, 6, 2, and 1 document. Visually, the United States stands out with the darkest blue color on the map (168 documents). Countries such as Canada, Brazil, Spain, China, and Australia appear in an intermediate shade of blue. Several other countries in Latin America, Europe, Africa (such as South Africa and Egypt), and Asia (such as India and Japan) are shaded in light blue tones, indicating lower production, between 1 and 6 documents. [End of description].

Source: research data (2026)

The articles in the portfolio in question have an overall average of 84 citations per country; however, the median is 16 citations (with a standard deviation of 176.3). The data presented in Table 4 show the countries of origin of the publications, along with the total number of citations (TC) and the average number of citations per article, which is influenced by the number of articles published. This explains, for example, why a country that publishes more, such as the United States, has a lower average number of citations (9.40) than Belgium (72.20), which publishes less. This

means that, in bibliometric terms, productivity is not always convergent with the relevance of the studies.

Table 4

Citations per article

**: Total citations. **: Average number of citations per article*

Country	Tc*	Average**	Country	Tc*	Average**
United States	1009	9.40	Finland	16	16.00
United Kingdom	733	48.90	Indonesia	16	1.10
Australia	389	11.40	Mexico	16	4.00
Belgium	361	72.20	Norway	15	7.50
Ireland	258	14.30	Georgia	14	4.70
Germany	220	8.10	Chile	7	1.00
Spain	216	6.50	Slovenia	7	7.00
Turkey	181	5.00	France	5	1.70
China	144	8.50	Switzerland	5	2.50
New Zealand	123	20.50	Botswana	4	4.00
Israel	114	6.70	Japan	3	1.00
Denmark	97	32.30	Slovakia	3	1.50
Canada	87	6.70	Argentina	2	0.70
Italy	86	17.20	Egypt	2	0.70
Netherlands	81	7.40	Philippines	2	0.70
Singapore	74	37.00	Uganda	2	2.00
Cyprus	69	11.50	Czech Republic	1	1.00
South Africa	61	4.40	Hong Kong	1	0.30
Austria	46	9.20	Lebanon	1	1.00
Portugal	44	4.00	Malta	1	1.00
Saudi Arabia	35	11.70	Oman	1	1.00
Greece	32	4.00	Serbia	1	1.00
Malaysia	26	2.90	Sweden	1	0.50
Pakistan	26	13.00	Thailand	1	0.50
Iceland	18	9.00	Hungary	0	0.00
Colombia	17	2.80	India	0	0.00
Korea	17	2.10	Peru	0	0.00
Brazil	16	3.20	Tunisia	0	0.00

Source: research data (2026)

The analysis of author affiliations involved selecting the 20 most productive institutions, which were compared to their countries of origin. The results are presented in Table 5, which includes the country, university name and number of publications from each institution. The results are consistent with previous findings, with a predominance of universities from the United States, which contributed 65 articles distributed across eight affiliations. Next is Spain, represented by 34 articles originating exclusively from the University of Granada, and Australia, with 20 articles originating from Queensland and Tasmania. The predominance of universities from countries considered developed is noteworthy. Also highlighted are the universities

where pioneering researchers in the field are found, such as Batanero (University of Granada), Ben-Zvi (University of Haifa) and Leavy (University of Limerick). In Latin America, there is the Pontifical Catholic University of Valparaiso, with research by Estrella.

Table 5

Affiliations

Country	Affiliation	Articles
United States	University of Texas - El Paso	10
	North Carolina StateUniversity	8
	Salisbury University	8
	University of Delaware	8
	Cincinnati Children's Hospital Medical Center	7
	Eastern Michigan University	7
	University of California	7
Spain	University of Granada	34
Australia	University of Queensland	11
	University of Tasmania	9
Ireland	University of Limerick	19
Netherlands	University of Utrecht	11
Portugal	University of Lisbon	9
Taiwan	National Taiwan Normal University	8
Germany	University of Paderborn	8
Chile	Pontifical Catholic University of Valparaiso	7
Israel	University of Haifa	7
United Kingdom	University of Leicester	7
Austria	University of Vienna	7

Source: research data (2026)

The next section will use mixed methods to discuss aspects of the intellectual structure of the articles that make up the portfolio, as well as the themes that emerge from the analysis of keywords, titles and abstracts.

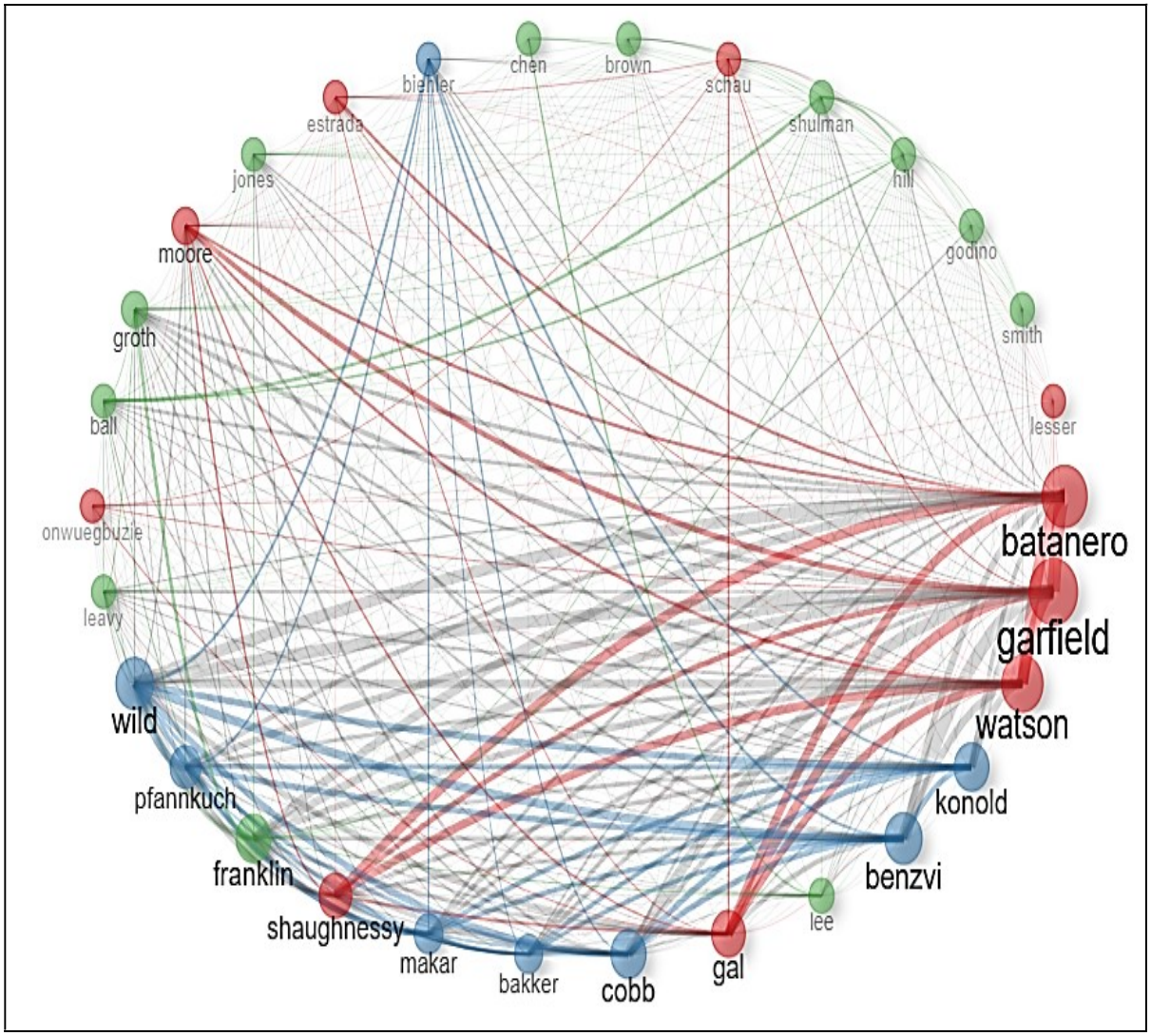
Intellectual and Conceptual Structure of the Articles

The intellectual structure of the texts was analyzed through the connections among citations – co-citations, that is, by the measure that reveals which authors are cited together most frequently, indicating the epistemological basis of an article (Grácio, 2016). According to Silva et al. (2022), the representation of co-citations occurs in the form of networks of authors, which show which researchers are cited together, as well as the intensity of their links with other groups, contributing to the understanding of the intellectual structure of the portfolio.

The colors (as well as the ramifications) of the networks refer to chains of texts that have a common intellectual structure. The size of the nodes indicates that a particular author was cited more frequently in the portfolio, determining theirrelevance to the area. In this way, it is observed which ideas (in terms of epistemological assumptions) are interrelated in the works in question.

So, the texts that make up the portfolio have three intellectual aspects (as shown in Figure 5). The first is represented by authors related to the foundations of statistics education (Batanero, Garfield, Watson, Gal, Shaughnessy, Onwuegbuzie, More, Estrada, Schau and Lesser); the second is inspired by methods of teaching statistics (Konold, Ben-zvi, Cobb, Makar, Pfannkuch, Wild, Biehler); and thethird directly represents the context of teacher training (Lee, Franklin, Leavy, Ball, Jones, Chen, Brown, Shulmann, Hill, Godino and Smith). In Figure 5, a predominant core of highly influential authors is observed in red, which shows more frequent connections, followed by the cluster in blue. Both clusters deal with themes directly linked to statistics education, whether for teaching or learning, focusing on statistical content, student learning and perceptions, and teacher training (initial and continuing). The green cluster, while encompassing authors who are fundamental to understanding teaching, is responsible for underlying themes and the theoretical framework that make up the articles in this corpus, such as Shulman, whose intellectual weight for studies focusing on teacher knowledge is undeniable.

Figure 5
Intellectual structure of the texts

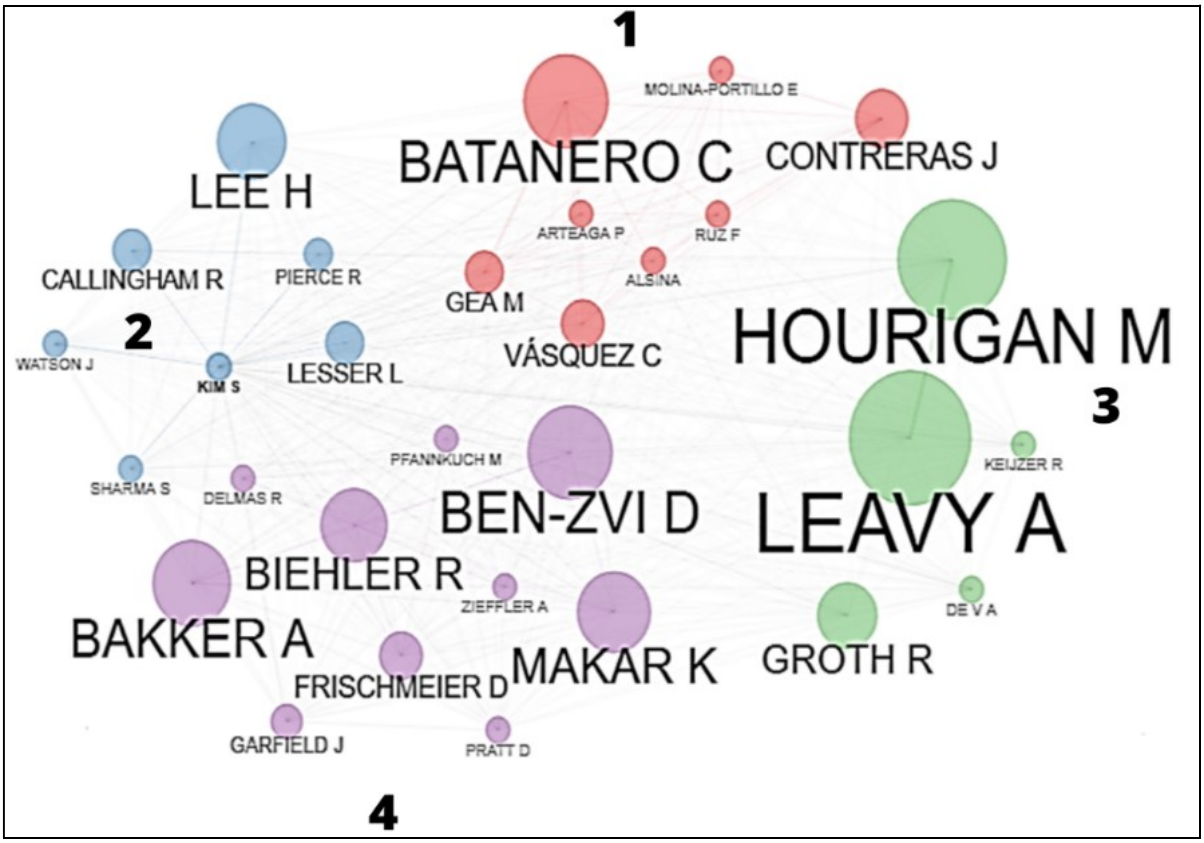


Note. [Image description] Circular network diagram composed of colored nodes connected by lines. The nodes represent authors and vary in size according to their relevance within the network. The authors “Batanero,” “Garfield,” “Watson,” and “Wild” stand out, represented by larger circles. The colors red, blue, and green indicate different clusters of authors, while the lines show the connections between them, with varying thicknesses indicating greater or lesser intensity of the relationships. [End of description].

Source: research data (2026)

With the intent to identify similarities between authors, the clustering by coupling tool was also used to couple authors with the theoretical references they used. Two works are coupled when they cite the same references (Grácio, 2016). To improve the visualization of the clusters, the tool’s parameters were limited to 30 authors, who were grouped into four clusters, as shown in Figure 6.

Figure 6
Coupling based on references and authors



Note. [Image description] Network diagram with authors represented by colored circles connected by thin lines. The names “Hourigan M,” “Leavy A,” “Batanero C,” “Ben-Zvi D,” and “Bakker A” are highlighted through larger labels and circles, indicating greater relevance within the network. The groups are organized by colors — green, red, blue, and purple — representing different clusters of authors and their connections. [End of description].

Source: research data (2026)

Through coupling, common themes among authors can also be identified. This reveals four clusters. The green cluster centers on Hourigan M. and Leavy A., who address teacher training, pedagogical practices and statistics in the classroom; the purple cluster centers on Ben-Zvi D., which focuses on modeling statistical thinking and statistical learning; the red cluster represents structuring themes of statistical education, centered on Batanero C. The blue cluster, highlighting Lee H., presents more general educational discussions. In summary, Table 6 shows four groupings numbered and identified by color in Figure 6, allowing for the characterization of the central themes of each network of authors.

Table 6
Themes resulting from coupling

Group/color	Themes	Frequency among authors
1 (red)	Assessment; Correlation and regression; Statistical literacy.	8
2 (blue)	Research in statistics education; Basic state standards for mathematics (c*ommon core standards*)	7
3 (green)	Initial teacher training; Mathematics education; Teacher training.	5
4 (purple)	Initial teacher training; Research in statistics education; Statistical reasoning.	10

Source: research data (2026)

By coupling the authors and their references, we can consider that Groups 3 and 4 are the closest to the initial training of teachers in relation to statistics, as shown in Table 6. The next section will address the words that stand out in this portfolio considering their titles, keywords and Zipf's Law, which is a mathematical-linguistic basis that analyzes the frequency and distribution of words contained in a text, making it possible to map and create rankings of occurrence (Zipf, 1949).

Although the analysis highlights the centrality of certain authors and themes, it is relevant to problematize the absences and silences in the field. There is a low incidence of perspectives related to educational technology contexts, critical approaches such as those that comprise critical statisticseducation (Perin, 2019), and movements for civic statistics, and a field limited to what refers to inclusion and research in the field of inclusive statistical education, such as that of Correia and Cazorla (2021).

Titles and Keywords

First, it is important to define Zipf's Law (Zipf, 1949), which is understood as the way words are distributed in a scientific text. It states that the frequency of a word is inversely proportional to its position in the frequency ranking: $f(r) = \frac{f}{r}$, where *f* is frequency and *r* is ranking (Guedes&Borschiver, 2005).

The 20 most frequent words in the titles are shown in Table 7 and highlight ideas that emerge from the documents. The first word denotes the theme guiding the research, "mathematics teachers", followed by the perspective of teacher training with the terms "preservice teacher","prospective teachers" and"teacher education". Words referring to methodologies are also frequent, such as "lesson study" and "flipped classroom". In addition, themes related to knowledge and professional development of teachers are present: "content knowledge" and"professional

development". The words "statistical literacy","statistical reasoning","statistics education" and "introductory statistics" are also highlighted in the titles.

Table 7
Most frequent expressions in document titles
n:*absolute frequency*

Expression	n	Expression	n
Mathematics teachers	42	Flipped classroom	14
Preservice teachers	42	School mathematics	14
Prospective teachers	26	Primary school	13
Teacher education	24	Professional development	13
Mathematics education	18	School teachers	13
Content knowledge	17	Statistics education	13
Lesson study	17	Secondary school	12
Statistical literacy	17	Introductory statistics	11
Statistical reasoning	16	Systematic review	11
School students	15	Teachers' understanding	11

Source: research data (2026)

Keywords highlight the themes that stand out in the documents and indicate the subject that the author will address (Martins Jr., 2017). As illustrated in Figure 7, the most frequent term is "statistics education research", which has 56 mentions in texts directly related to statistics education. The term "statistics education" has 39 mentions, "statistics"" 28, "probability" 23, "statistical literacy" 22, and "statistics reasoning" 18, and belong to texts that specify the theme. From the perspective of initial teacher training, the terms "teacher education" (38), "mathematics education" (25) and "professional development" (20) stand out.

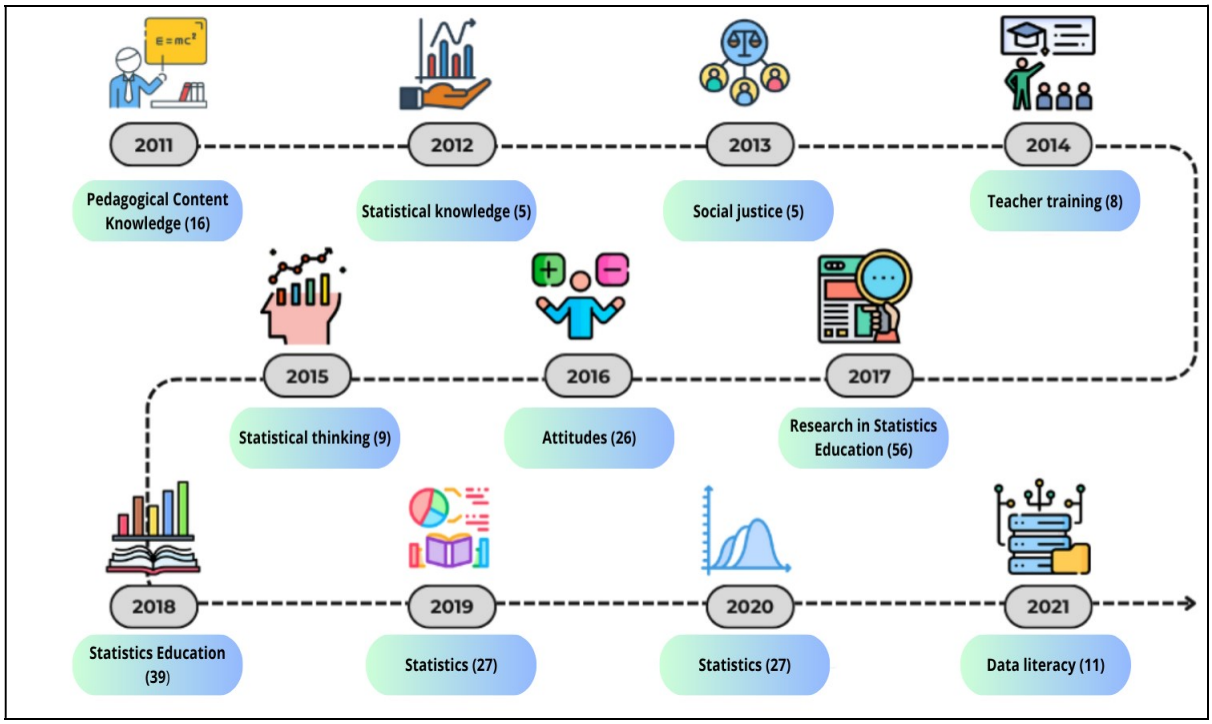
Figure 7
Word cloud ofkeywords

Note. [Image description] Irregularly shaped word cloud composed of terms related to Statistics Education. The expression “statistics education research” stands out in the center, displayed in larger red letters. Other prominent terms include “statistics education,” “teacher education,” “professional development,” “statistical literacy,” “probability,” and “statistics.” The words vary in size and appear in shades of red, brown, orange, and yellow, indicating higher or lower frequency of occurrence. [End of description].

Source: research data (2026)

Keywords were also analyzed using Trending Topics, understood as the themes that are most frequently repeated in research over specific years. For this, the standard configuration of the *bibliometrix* tool was adopted, considering words with a frequency greater than or equal to five and limiting it to a maximum of five terms per year. Figure 8 presents the most frequent topics from 2011 onwards, since, in previous years, the portfolio shows a large terminological dispersion, making it impossible to identify consistent trends. A centrality of research in statistics education is observed, structured around two main keyword cores: statistics education and teacher training. The latter core highlights the relevance of investigations focused on the training of teachers who teach mathematics, demonstrating its centrality in the field.

Figure 8
Trending topics



Note. [Image description] Horizontal timeline spanning the years from 2011 to 2021, illustrated with colorful icons and themes related to Statistics Education. Each year presents a topic and the number of occurrences in parentheses. The most prominent topics include “Statistics Education (39),” “Pedagogical Content Knowledge (27),” “Statistics (27),” “Attitudes (26),” and “Research in Statistics Education (26).” The elements are distributed in green and blue colored boxes, connected by a dotted line representing the temporal sequence. [End of description].

Source: research data (2026)

In 2011, the most prominent topic was “pedagogical content knowledge”, reflected, for example, in Ponte’s (2011) article, which discusses the challenges of statistical teacher training. Burgess’s (2011) article also stands out, addressing the knowledge teachers need to explore statistical investigations in the classroom. In 2012, the trending topic was “statistical knowledge”. The highlighted texts refer to the development of statistical thinking, statistical reasoning and attitudes towards statistics. Here, Jaafar et al. (2012) are noteworthy, as they seek to determine the influence of prior mathematical knowledge and statistical reasoning on student performance outcomes.

The term “social justice”, in 2013, brings the idea of ways to implement aspects related to social justice during mathematics education, as we can find in Garii and Appova (2013). For 2014, “teacher preparation” stands out; that is, the perspective of initial teacher training, evidenced by Groth (2014), who discusses statistics programs for future teachers, programs developed in response to the recent

K-12 curriculum recommendations, for which follow-up assessments must be designed.

The term “statistical thinking” is the most prominent topic of discussion in 2015. One example is the article by Sánchez and Gómez-Blancarte, which emerges from a professional development context, from a project emphasizing the learning and teaching of statistical content through the Lesson Study methodology, whose objective was to develop elements of statistical thinking in students.

For the year 2016, the trend is the term “attitude(s)”, referring to the positioning or reflection regarding statistics. An example is Aydin (2016), who seeks to investigate the attitudes of future teachers towards statistics with the use of Excel.

The years 2017, 2018, 2019 and 2020 have trending topics with a more generic meaning: “statistics education research”, “statistics education”, “statistics” and “probability”, reporting texts that address teaching and learning, teacher training, curriculum and teaching methodologies. For example, we can cite the article by Diamond and Stylianides (2017), which presents an epistemological evaluation of the teaching actions of a group of academics in the field of statistics, observing a preference for constructivism. Similarly, Sharma (2018) reports a study on the use of the mother tongue as a resource for students learning about statistical methods. Finally, Molina-Portillo et al. (2019) emphasize statistical literacy at the basic and secondary education levels. In 2020, Hourigan and Leavy (2020) report the probabilistic understandings of a sample of 104 pre-service teachers, focusing on one particular aspect – the understanding of probabilistic justice[3].

The trending topic of 2021 (“data literacy”) suggests new discussions considering what is commonly called the Information Society. An example is the article by Khan and Mason (2021), which illustrates post-pandemic scenarios that combine three distinct themes relevant to changes in the requirements of educational systems worldwide: STEM education, mathematical thinking and data literacy. The year 2022 was not evaluated, as the sample was collected before its end. The following section will address the themes that emerged in the analysis of the abstracts.

Abstracts and Emerging Themes

The document abstracts were analyzed using the thematic map tool, which uses a graph built on Cartesian coordinate axes to compare centrality with the density of themes, indicating their relevance. Density is understood as a measure of

[3]Concepts of probability in every day situations (such as games, for example) where the individual understands the concepts of uncertainty in the occurrence of a given phenomenon.

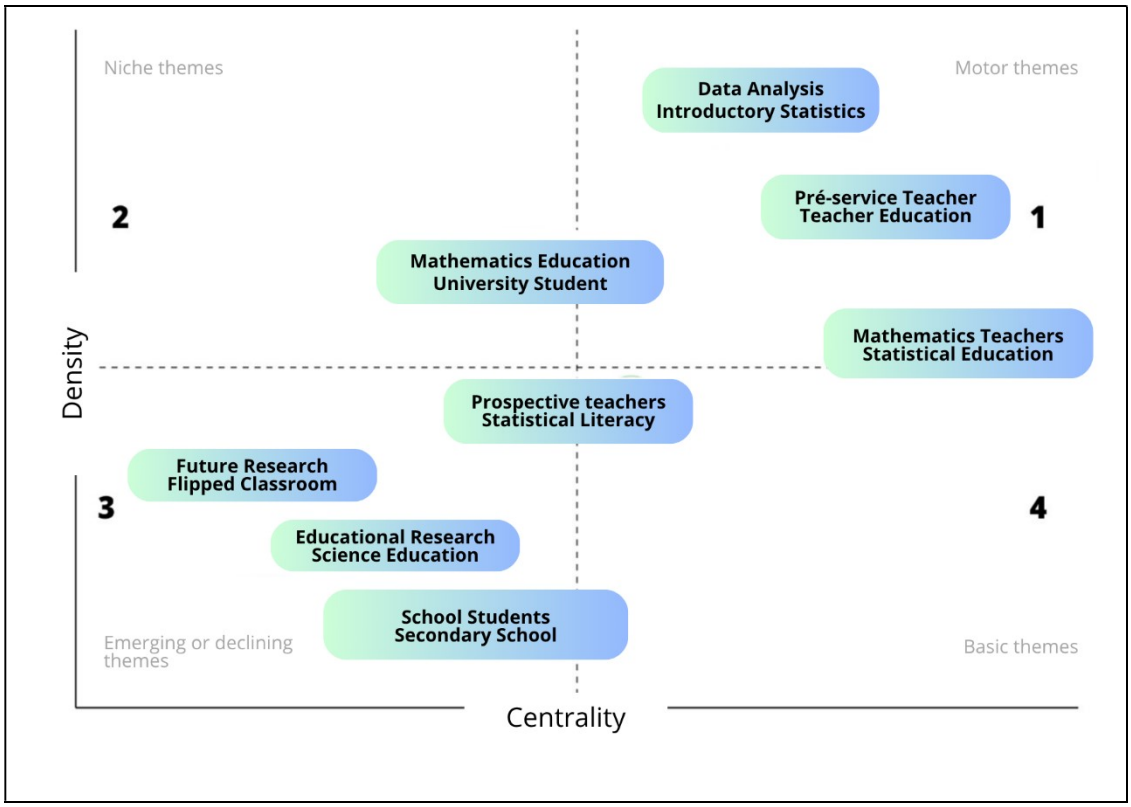
the network's internal strength, and centrality is the interaction with other networks (Cobo et al., 2011; Rojo; Lacruz, 2022). So, the four quadrants determine the degree of integration and the most developed themes within a given research portfolio.

Regarding the graph's ordering, the first quadrant indicates themes with high density and centrality, called driving themes, and points to well-developed topics. The second contains themes with high density and weak centrality; that is, well-developed themes, but with low integration withthe other networks. The third has low density and centrality, which means it integrates themes that are declining (or emerging). The fourth quadrant, with strong centrality but low density, contains the fundamental themes for the research field, but weakly developed.

Thus, Figure 9 shows words contained in the abstracts revealing well-developed themes related to "data analysis" and "introductory statistics"; "pre-service teachers" and "teacher education"; and "mathematics teachers" and "statistics education". These themes were recurrent in the analysis of the articles and allow us to conclude what the researchers'interests are in relation to statistics education and teacher training.

In the second quadrant, as a thematic niche, that is, well-developed but with few nodes (networks) among the research projects, are topics related to "mathematics education" and "university students", which can be understood as marginal topics within the research field. For the third quadrant, we highlight emerging or declining topics such as "future research", "flipped classroom", "educational research", "science education", "elementary school students" and "secondary school students". As basic themes, the fourth quadrant shows the general ideas of the research, such as "prospective teachers" and "statistical literacy".

Figure 9
Theme clusters



Note. [Image description] Rectangular diagram with a white background, divided into four quadrants forming a cross. The vertical axis represents “Density” and the horizontal axis represents “Centrality.” In the upper-right quadrant, entitled “Motor themes,” there are three rounded blocks: “Data Analysis and Basic Statistics” in yellow, “Preservice Teachers and Teacher Education” in purple, and “Statistics Education and Mathematics Teachers” in red. In the upper-left quadrant, entitled “Niche themes,” there is the block “Mathematics Education and University Students” in teal blue. In the lower-left quadrant, entitled “Emerging or declining themes,” there are the blocks “Future Research and Flipped Classroom” in gray, and “Educational Research and Science Education” in dark green. Crossing the lower vertical dividing line are the block “Prospective Teachers and Statistical Literacy” in light blue and, just below it, the block “Elementary and High School Students” in light green. [End of description].

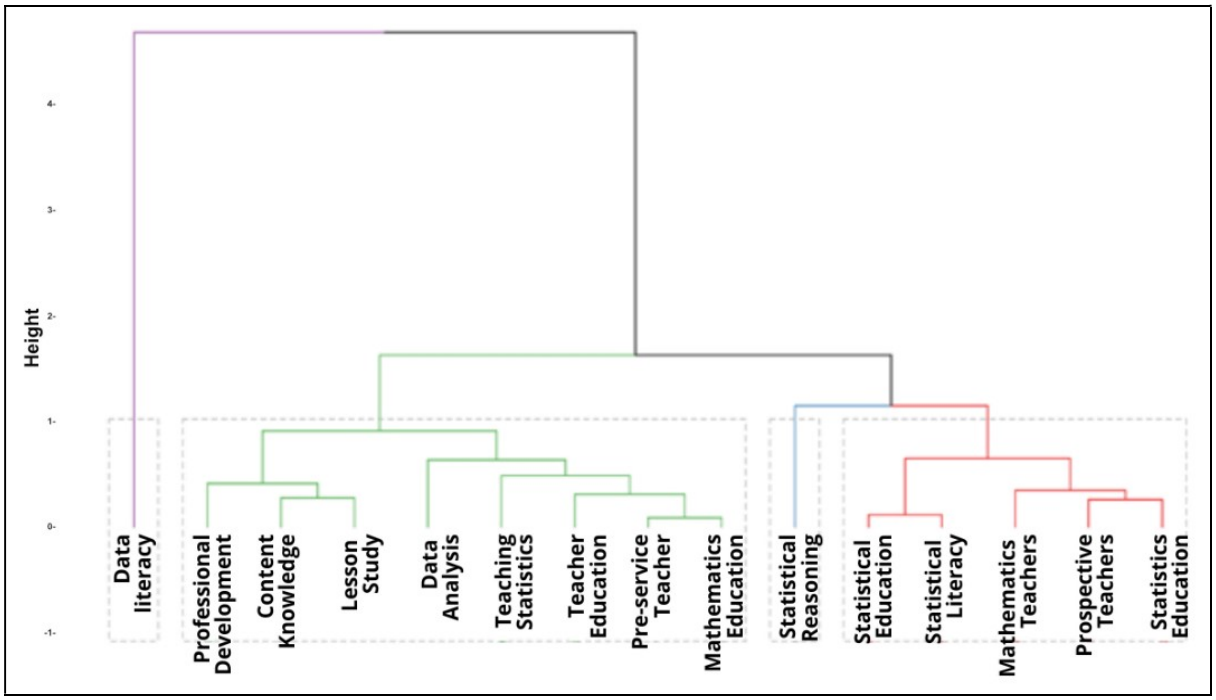
Source: research data (2026)

Considering the analysis illustrated by Figure 9, the abstracts of the articles show how well-developed and interrelated themes are found in articles that deal with data analysis, teacher training and statistics education. This shows the composition of the portfolio and indicates the paths and thematic options of the research field as a whole.

In order to detail these themes, looking for a research scenario for the initial training of mathematics and statistics education teachers, a detailed approach was carried out using the factorial approach tool, producing a dendrogram (Figure 10) to categorize the documents by main subject and choose those that best converge to the objectives of this study. Thus, the abstracts of the documents were considered

and a correspondence analysis was carried out, considering bigram parameters (two words) for the 20 main terms calculated for fourclusters[4].

Figure 10
Dendrogram



Note. [Image description] Vertically oriented dendrogram with a white background, structured as a hierarchical clustering of academic terms. From the top, branches in different colors descend and split until reaching the base, where sixteen terms are written vertically and organized into four blocks outlined by dashed rectangular boxes. On the far left, a purple line descends in isolation to the first box containing the term “Data Literacy.” In the second box, colored green, there is a large grouping of eight terms related to professional development and teacher education, such as “Professional Development,” “Data Analysis,” and “Mathematics Education.” In the third box, an isolated dark blue line connects only to the term “Statistical Reasoning.” In the fourth and final box, on the far right, a reddish-brown branch expands to group five terms written in red, including “Statistics in Education,” “Statistical Literacy,” “Mathematics Teachers,” “Prospective Teachers,” and “Statistics Education.” [End of description].

Source: research data (2026)

Using clustering tools, through correspondence analysis from Biblioshiny, each of the texts that make up the clusters was verified, understanding their relevance through the number of citations and highlighting the general theme of each grouping. Cluster 1 highlights the terms “pre-service teacher”, “teacher education”, “mathematics education”, “professional development”, “content knowledge”, “lesson study”, “data analysis” and “teaching statistics”. The first four terms are associated with teacher training itself, and the others refer to content knowledge and pedagogical content, such as the methodology of lesson studies, in addition to an

[4]The choice of four clusters was made after na exploratory analysis of the thematic map tool in *bibliometrix*, which identified that the themes were present in greater quantity and relevance in four groupings.

emphasis on data analysis and the teaching of statistics, highlighting ideas related to teacher professional development with regard to statistics. Cluster 2 deals more specifically with statistics education and initial teacher training, considering the terms "statistical education", "statistical literacy", "mathematics teachers", "prospective teachers" and "statistics education". In the case of Cluster 3, ideas related to "statistical reasoning" and its development predominate. Cluster 4 presents a new trend –"data literacy", which includes articles that address data literacy, one of the needs of today's society. For a more detailed analysis of the portfolio's themes, five articles from each cluster were selected, using the highest number of citations as a criterion. These are presented in Table 8, which indicates the quantity of citations (QC), cluster number (NC), some bibliometric characteristics (title, authors and year), and a description of the theme addressed.

Table 8
Featured articles in the clusters

QC	NC	Title, authors and year	Description
244	1	*Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research* - F. Depaepea; L. Verschaffelb; G. Kelchtermansa (2013)	Systematic literature review that deals with different meanings of pedagogical content knowledge.
61	1	*Scaffolding and dialogic teaching in mathematics education: introduction and review* - A. Bakker; J. Smit; R. Wegerif (2015)	This article aims to discuss the dialogic processes of assisting students in their learning. It deals with a teacher's active and sensitive involvement in student learning.
58	1	*Pre-Service Elementary Teachers' Perceptions of Factors in an Holistic Methods Course Influencing their Confidence in Teaching Science* – C. Howitt (2007)	This article deals with results from a survey that aims to measure the level of confidence of prospective teachers. Factors that influence confidence were identified, such as: knowledge of pedagogical content, content knowledge, learning environment, evaluation and reflection, scientific activities, children's view of science, knowledge of scientific content and scientific inquiry.
58	1	*Fundamental Statistical Ideas in the School Curriculum and in Training Teachers* – G. Burrill; R. Biehler (2011)	The article considers several approaches in the teaching and learning of statistics and their implications in the education of mathematics teachers.
52	1	*Structural Model of the Effects of Cognitive and Affective Factors on the Achievement of Arabic-Speaking Pre-service Teachers in Introductory Statistics* – F.M. Nasser (2004)	This study examines attitudes towards mathematics and statistics of 162 prospective mathematics teachers in Israel. Results also indicated that math aptitude, attitudes toward math and statistics, and motivation, together, accounted for 36% of the variation in performance on introductory statistics for the sample.
81	2	*The Importance of Attitudes in Statistics Education* – C. Ramirez; C. Schau; E. Emmioglu (2012)	The authors went through 15 surveys related to attitudes towards statistics, as well as, in the same sense, described the SATS-M (Model of

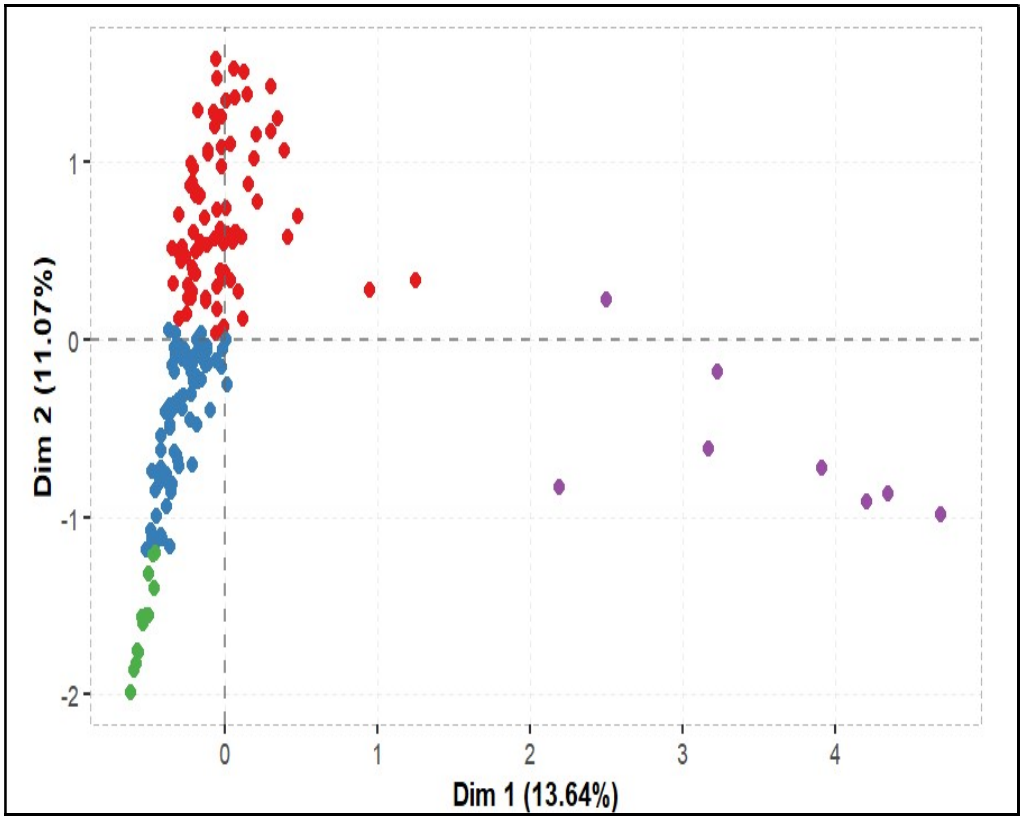
QC	NC	Title, authors and year	Description
			Students' Attitudes Toward Statistic). SATS-M includes three broad constructs that affect the statistics course outcomes: student characteristics, experiences related to prior achievement, and statistical attitudes.
59	2	*The Reasoning Behind Informal Statistical Inference* – K. Makar; A. Bakker; D. Ben-Zvi (2011)	This article uses educational literature on informal statistical inference to argue that, for students to make informal statistical inferences, there are a number of interrelated key elements that are needed to support their informal inferential reasoning.
47	2	*Training Teachers to Teach Probability* – C. Batanero; J. D. Godino; R. Roa (2004)	The authors express the reasons why teaching probability is difficult for mathematics teachers, describe the necessary contents in the didactic preparation of teachers to teach and analyze some examples of activities in an initial training course for mathematics teachers.
38	2	*An investigation of prospective secondary mathematics teachers' conceptual knowledge of and attitudes towards statistics* – A. Hannigan; O. Gill; A. M. Leavy (2013)	The goals of this study were to investigate the conceptual understanding of statistics by prospective high school mathematics teachers; the nature of their attitudes towards statistics and if there was a relationship between the attitude towards statistics and the conceptual understanding of statistics through an instrument – Survey of attitudes towards statistics.
29	2	*Technology for Enhancing Statistical Reasoning at the School Level* – R. Biehler; D. Ben-Zvi, A. Bakker; K. Makar (2013)	This study offers an up-to-date overview of digital technologies relevant to teaching statistics for the purpose of developing students' statistical reasoning at the school level. It also reports on trends in statistics education, the history of technologies in statistics and statistics education, and digital learning environments.
56	3	*Mathematical Knowledge and Practices Resulting from Access to Digital Technologies* – J. Olive; K. Makar; V. Hoyos; L. Kee Kor; O. Kosheleva; R. Sträßer. (2010)	This paper deals with a literature review that indicates that interactions between students, teachers, tasks and technologies can generate mathematical knowledge. Recent developments in dynamic technologies have the potential to promote new mathematical practices in different contexts: for example, dynamic geometry, statistics education, robotics and digital games.
46	3	*Attitudes and Achievement in Statistics: a meta-analysis study* – E. Emmioğlu; Y. Capa-Aydin (2012)	This article brings a meta-analysis addressing the relationships between learning statistics and four components of attitudes towards statistics (Cognitive Competence, Affection, Value and Difficulty) assessed through SATS (Survey of Attitudes Towards Statistics) in the United States, comparing the results with those from other countries.
38	3	*Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics* – M. Hood; P. A. Creed; D. L. Neumann (2012)	In this article, the authors test a model of the relationship between attitudes towards statistics and accomplishments based on Eccles' Expectancy Value Model (1983) in university students in Australia.
37	3	*Surveys Assessing Students' Attitudes Toward Statistics: A Systematic Review of Validity and*	This is a literature review that presents data on surveys aiming to assess attitudes and learning of statistics.

QC	NC	Title, authors and year	Description
		Reliability – M. M. Nolan; T. Beran; K. G. Hecker (2012)	
22	3	*How strongly does statistical reasoning influence knowledge and acceptance of evolution?* – D. Fiedler; G.C. Sbeglia; R. H. Nehm; U. Harms (2019)	This work deals with the application of surveys (RAPROMATH and RAPROEVO) to measure statistical reasoning and its evolution in a group of 564 university students in the United States. The data provide clues about the role of statistical reasoning in evolutionary knowledge and motivate future works to explore statistical literacy in educational processes.
35	4	*Exploring the challenge of developing student teacher data literacy* – B. Cowie; B. Cooper (2016)	The article deals with aspects related to the professional development of teachers regarding data literacy. The authors state that evaluative data on students provide much information about the educational processes and that they should be analyzed. However, the necessary knowledge in mathematics and statistics to carry out these analyses is still not enough, which suggests an opportunity for this literacy in data and evaluation.
8	4	*Preparing preservice teachers to be data literate: a Queensland case study* – M. Carey; P. Grainger; M. Christie (2017)	The authors' concern is whether teachers in training are being prepared for so-called data-driven pedagogy. Thus, a training course was held to expand the teaching potential for data analysis and allow the professional development of teachers in data literacy.
7	4	*Toward data-scientific thinking* – R. Gould (2021)	The author argues that the current data culture makes data science literacy necessary for students and teachers. So, it proposes a course that involves statistical, mathematical and computational reasoning.
3	4	*Visualizing the teaching of data visualizations in social studies: A study of teachers' data literacy practices, beliefs, and knowledge* – T.L. Shreiner; B.M. Dykes (2020)	This work values data literacy as essential knowledge for citizens of today's society. It questions how prepared teachers are to teach this subject. So, it uses a survey to seek understanding about attitudes, beliefs and knowledge of the teachers in this regard.
2	4	*Exploring variability during data preparation: a way to connect data, chance, and context when working with complex public datasets* – M. H. Wilkerson; K. Lanouette; R. L. Shareff (2021)	The work deals with the use of large socio-governmental databases for data literacy. Through analyses of task-based interviews with ten teenage participants, specific actions were found during data preparation, such as filtering data or calculating new measures, and opportunities were presented to engage students in preparing and analyzing various datasets. Some changes in statistics teaching that are needed for a focus on "big data literacy" are presented.

Source: research data (2026)

For the study of the conceptual structure of the documents, correspondence analysis was used, resulting in the factorial map shown in Figure 11, which displays four clusters and highlights the divergence between texts dealing with teacher training and statistics education and those whose theme is data literacy.

Figure 11
Dimensions of cluster coverage



Note. [Image description] Rectangular scatter plot with a white background and very light gray grid lines, divided into four quadrants by two central dashed lines intersecting at the zero point. The horizontal axis represents Dimension 1 (with 13.64% variance), and the vertical axis represents Dimension 2 (with 11.07% variance). The graph displays numerous colored points organized into four distinct groupings or clusters. On the left side, forming a dense and elongated vertical column crossing the center of the graph, there are three groups: at the top, the red cluster with points concentrated in the upper-left quadrant; in the middle, the blue cluster with points grouped around the central line; and at the bottom, the green cluster with points aligned along the lower-left edge. Separated from this main column and dispersed horizontally across the right side of the graph, mainly in the lower-right quadrant, are a few isolated points forming the fourth cluster in purple. [End of description].

Source: research data (2026)

The highlighted categories reveal some trends present in the portfolio in question. This includes texts that address theoretical conceptions related to the training of mathematics teachers and ideas about teaching, obtained from the opinions of the teachers themselves or future teachers, often through surveys. In this context, there is also research that measures attitudes towards statistics. This demonstrates a preference among researchers for studying the context of statistical training considering the opinions, knowledge and beliefs of the individuals involved in the teaching-learning processes.

Another point is that probability content may be taught separately from statistical content, not contributing to a stochastic training of future teachers. This uncertainty arises because, in this bibliometric analysis, 81 articles (11.1%) were

obtained whose descriptors or titles contain the word “probability”. The result is similar to that obtained by Franco and Alsina (2022), who report that probability was the most frequently found content in their research. Thus, one can question whether the emphasis on a given theme in research may indicate a weak integration of other curricular content in teacher training, which corroborates a fragmented view of the content originating from teacher education.

Furthermore, since these are reflections on the conception of statistical knowledge in teacher training, Franco and Alsina (2022) highlight that many studies consider aspects of teacher knowledge such as the Didactic-Mathematics Knowledge and Competencies Model (CCDM) of Godino (2009) and systems similar to the TechnologicalPedagogicalContent Knowledge (TPACK) of Shulman (1986), Koehler and Mishra (2008), and Ball, Thames, and Phelps (2008). This pattern is also reflected in this study, as evidenced in the intellectual structure section of the articles.

Bibliometric analysis points toward the methodologies used in the initial teacher training found in the portfolio, such as games, flipped classroom, lesson study, and the use of diverse technologies, which allow the development of competencies for the future professional practices of educators and highlight the associations that should occur in terms of pedagogical and content knowledge. Research on statistical reasoning and the factors that influence its development, as well as data literacy and discussions of the skills of students and teachers for living in the “big data” society, are also highlighted.

With regard to teaching practice, which presupposes dealing with complex situations arising from multiple factors and requiring different possible responses from the teacher (Broi; Gessinger; Lima, 2011), as well as the mobilization of different dimensions of professional knowledge (Shulman, 1986), the research analyzed points to a scenario still in the initial stages, in which statistical knowledge (Burgess, 2011) is incipient.

This is the panorama evidenced in studies situated at the intersection between statistics education and the initial training of mathematics teachers. These investigations indicate trends focused on understanding the development of statistical literacy, reasoning and thinking in different teaching and learning contexts, considering the use of active methodologies such as lesson studies, flipped classrooms and investigative practices. In this scenario, the teaching of statistics is consolidated as a privileged space for the critical reading of data and for awakening students’interest in research as a formative principle.

Final Considerations

This study provided evidence that reveals the themes emerging from the research field that influence the ways in which the process of educating statistics teachers takes place. Although it does not provide an exact picture of trends in Brazil, it offers robust indications of them.

The articles considered in this bibliometric analysis show a predominance of research related to statistics education; to statistical literacy, reasoning and thinking; and that emphasize methodologies such as flipped classroom, lesson study and the use of technologies in the professional practices of educators and future educators. It is also worth highlighting more recent texts that use the descriptor “data literacy”, a theme that differs from research in statistics education, but which is valued as the interpretation of large amounts of data becomes essential to living in the current social context. This shows that changes are needed in statistics education and, consequently, in teacher training.

One of the limitations of the portfolio composition process was the idiomatic differences regarding the organization of education systems in various countries, which led to an exploratory analysis prior to the research, opting for the most frequent terms. It is also worth noting that some differences in the spelling of words were accepted due to their widespread use, such as “pre-service” and “preservice”.

Furthermore, gaps in the corpus stand out, particularly those concerning the critical perspective of statistics and its uses, as well as research on inclusive statistics and its implications for teacher training. These gaps open possibilities for future research, notably investigating the potential of statistics classes as privileged spaces for developing the scientific spirit of future teachers, fostering critical analysis skills, data interpretation and the adoption of an investigative approach to the phenomena studied.

This study did not intend to focus on any particular nuance of the statistics education scenario, but rather to portray general aspects; that is, to outline the themes that stand out in the research and, based on them, observe patterns, trends in the training of mathematics educators, or paradigm shifts. The results here suggest a convergence between statistics education and teacher training. However, this relationship is still fragmented in most cases, requiring critical further development that integrates the fundamentals of statistics education into teacher training, especially regarding teachers’knowledge and experiences. Nevertheless, the study opens opportunities for questioning the knowledge valued in teacher

training and intrinsic factors in training, which require specific analysis in future studies.

References

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R. & Witmer, J. (2005). *Gaise college report. In American Statistical Association*. Alexandria, VA: American Statistical Association.

Araújo, R. F., & Alvarenga, L. (2011). A bibliometria na pesquisa científica da pós-graduação brasileira de 1987 a 2007*. Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação,* 16(31), 51-70.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of *Informetrics*, 11(4), 959-975.

Aydın, S. (2016). Using excel in teacher education for sustainability. *Journal of Teacher Education for Sustainability*, 18(2), 89-104.

Bakker, A., Smit, J., & Wegerif, R. (2015). Scaffolding and dialogic teaching in mathematics education: Introduction and review. *Zdm*, 47, 1047-1065.

Ball, D., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? Journal of Teacher Education, 59(5), 389-407.

Batanero, C., Godino, J. D., & Roa, R. (2004). Training teachers to teach probability. *Journal of Statistics Education*, 12(1).

Biehler, R., Ben-Zvi, D., Bakker, A., & Makar, K. (2013). *Technology for enhancing statistical reasoning at the school level*. Third international handbook of mathematics education, 643-689.

Broi, M. P., Gessinger, R. M., & do Rosário Lima, V. M. (2011). Trajetória docente do professor de Matemática. *Alexandria: Revista de Educação em Ciência e Tecnologia*, 4(1), 133-151.

Burgess, T. A. (2011). Teacher knowledge of and for statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), Teaching statistics in school mathematics: *Challenges for teaching and teacher education.* A joint ICMI/IASE study (pp. 259–270). Springer.

Burrill, G., & Biehler, R. (2011). Fundamental statistical ideas in the school curriculum and in training teachers. *Teaching statistics in school mathematics-challenges for teaching and teacher education*: A Joint ICMI/IASE study: The 18th ICMI study, 57-69.

Brookes, B. C. (1969). Bradford's law and the bibliography of science. *Nature*, 224(5223).

Cândido, R. B., Garcia, F. G., Campos, A. l. S., & Tambosi Filho, E. (2018). Lei de Lotka: um olhar sobre a produtividade dos autores na literatura brasileira de finanças. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 23(53), 1-15.

Carey, M., Grainger, P., & Christie, M. (2017). Preparing preservice teachers to be data literate: A Queensland case study. Asia-Pacific *Journal of Teacher Education*, 46(3), 267-278.

Cassettari, R.R.B.; Pinto, A.L.; Rodrigues, R.S.; Santos, L.S. dos. (2015). Comparação da Lei de Zipf em conteúdos textuais e discursos orais. *El profesional de la información*, v. 24, n. 2, p. 157-167.

Cobo, M. J., López-Herrera, A. G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the Fuzzy Sets Theory field. *Journal of Informetrics*, 5(1), 146-166.

Cazorla, I. M. A. (2006). O ensino de Estatística no Brasil. *Sociedade Brasileira de Educação Matemática*.http://www.sbem.com.br/gt_12/arquivos/cazorla.htm.

Cowie, B., & Cooper, B. (2016). Exploring the challenge of developing student teacher data literacy. *Assessment in Education: Principles, Policy & Practice*, 24(2), 147-163.

Correia, G. S., & Cazorla, I. M. (2021). Diálogos entre o ensino de probabilidade e estatística e a educação inclusiva nas pesquisas brasileiras. *Perspectivas da Educação Matemática*, 14(35), 1-20.

Creswell, J.W & Clarck, V.L.P. (2013). *Pesquisa de métodos mistos*. 2ª Ed. Porto Alegre, RS: Penso.

Depaepe, F., Verschaffel, L., & Kelchtermans, G. (2013). Pedagogical content knowledge: A systematic review of the way in which the concept has pervaded mathematics educational research. *Teaching and teacher education*, 34, 12-25.

Diamond, A. H., & Stylianides, A. (2017). Personal epistemologies of statisticians in academia: An exploratory study. *Statistics Education Research Journal*, 16(2), p. 335-361.

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.

Emmioğlu, E. S. M. A., & Capa-Aydin, Y. E. S. I. M. (2012). Attitudes and achievement in statistics: A meta-analysis study. *Statistics Education Research Journal*, 11(2), 95-102.

Fiedler, D., Sbeglia, G. C., Nehm, R. H., & Harms, U. (2019). How strongly does statistical reasoning influence knowledge and acceptance of evolution? *Journal of Research in Science Teaching*, 56(9), 1183-1206.

Franco, J.S., & Alsina, Á. (2022). Conhecimento especializado de professores do ensino fundamental para o ensino de estatística e probabilidade. *Educação Matemática*, 34 (3), 65-96. https://doi.org/10.24844/em3403.03

Garii, B., & Appova, A. (2013). Crossing the great divide: Teacher candidates, mathematics, and social justice. *Teaching and teacher education*, 34, 198-213.

Guedes, V. L da S.; Borschiver, S. (2005) *Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica*. Universidade Estadual Paulista.

Godino, J. D. (2009). Categorías de análisis de los conocimientos del profesor de matemáticas. *Unión-Revista Iberoamericana de Educación Matemática*, 5(20).

Gould, R. (2021). Toward data-scientific thinking. Teaching Statistics, 43, S11-S22.

Grácio, M. C. C. (2016). Acoplamento bibliográfico e análise de cocitação: revisão teórico-conceitual. *Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação*, 21(47), 82-99.

Groth, R. E. (2014). Using work samples from the National Assessment of Educational Progress (NAEP) to design tasks that assess statistical knowledge for teaching. *Journal of Statistics Education*, 22(3).

Hannigan, A., Gill, O., & Leavy, A. M. (2013). An investigation of prospective secondary mathematics teachers' conceptual knowledge of and attitudes towards statistics. *Journal of Mathematics Teacher Education*, 16, 427-449.

Hood, M., Creed, P. A., & Neumann, D. L. (2012). Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics*. Statistics Education Research Journal*, 11(2), 72-85.

Hourigan, M., & Leavy, A. M. (2020). Pre-service teachers' understanding of probabilistic fairness: Analysis of decisions around task design. International *Journal of Mathematical Education in Science and Technology*, 51(7), 997-1019.

Howitt, C. (2007). Pre-service elementary teachers' perceptions of factors in an holistic methods course influencing their confidence in teaching science. *Research in Science Education*, 37, 41-58.

Jaafar, Z., Kheng, F. K., Ab Malek, H., & Ali, R. (2012, September). Cognitive factors influencing statistical performance of diploma science students: A structural equation model approach. *In 2012 International Conference on Statistics in Science, Business and Engineering* (ICSSBE) (pp. 1-5). IEEE.

Khan, K., & Mason, J. (2021, November). The M in STEM and issues of data literacy. *In 29th International Conference on Computers in Education Conference, ICCE 2021-Proceedings* (Vol. 1, pp. 632-637).

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13(1-2), 152-173.

Martins Junior, J. (2017). *Como escrever trabalhos de conclusão de curso: instruções para planejar e montar, desenvolver, concluir, redigir e apresentar trabalhos monográficos e artigos*. Editora Vozes Limitada.

Mishra, P., & Koehler, M. J. (2008). Introducing technological pedagogical content knowledge. *In: annual meeting of the American Educational Research Association*, 1, 16.

Molina-Portillo, E., Contreras, J. M., Godino, J. D., & Ruz, F. (2019). Statistical literacy in the information society. *Boletín de Estadística e Investigación Operativa*, 35(2), 148-169.

Nasser, F. M. (2004). Structural model of the effects of cognitive and affective factors on the achievement of Arabic-speaking pre-service teachers in introductory statistics. *Journal of Statistics Education*, 12(1).

Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys assessing students'attitudes toward statistics: a systematic review of validity and reliability. *Statistics Education Research Journal*, 11(2), 103-123.

Olive, J., Makar, K., Hoyos, V., Kor, L. K., Kosheleva, O., & Strässer, R. (2010). Mathematical knowledge and practices resulting from access to digital technologies. *Mathematics education and technology-rethinking the terrain: the 17th ICMI study*, 133-177.

Perin, A. P. (2019). Educação estatística crítica: um estudo das práticas discentes em um curso de tecnologia [Universidade Estadual Paulista (Unesp)].

R Core Team. (2022). R: *A language and environment for statistical computing.*

Rojo, R. O., & Lacruz, A. (2022). Empreendedorismo Internacional: mapeamento temático e proposta de agenda de pesquisa. *Organizações & Sociedade*, 30, 301-328.

Pinheiro, L. V. R. (1983). Lei de Bradford: uma reformulação conceitual. *Ciência da Informação*, 12 (2), 59-80.

Da Ponte, J. P. (2011). Preparing teachers to meet the challenges of statistics education. *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: a Joint ICMI/IASE Study*: the 18th ICMI Study, 299-309.

Ramirez, C., Schau, C., & Emmioğlu, E. S. M. A. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal*, 11(2), 57-71.

Ramos-Rodríguez, A. R., & Ruíz-Navarro, J. (2004). Changes in the intellectual structure of strategic management research: A bibliometric study of the Strategic Management Journal, 1980–2000. *Strategic Management Journal*, 25(10), 981-1004.

Sánchez, E. A. S., & Gómez-Blancarte, A. L. (2015). Negotiation of meaning as a learning process: the case of a professional development program in statistics teaching. *Revista Latinoamericana de Investigación en Matemática Educativa*, 18(3), 387-420.

Sharma, S. (2018). Bridging Language Barriers in Statistics for Year-12 Pasifika Students: A Collaborative Study. *Waikato Journal of Education*, 23(1), 107-120.

Sharma, S., Sharma, S., Doyle, P., Marcelo, L., & Kumar, D. (2021). Teaching and learning probability using games: A systematic review of research from 2010-2020. *Waikato Journal of Education*, 26(2), 51-64.

Shreiner, T. L., & Dykes, B. M. (2020). Visualizing the teaching of data visualizations in social studies: A study of teachers' data literacy practices, beliefs, and knowledge. *Theory & Research in Social Education*, 49(2), 262-306.

Shulman, L. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4-14.

Silva, C.L., Sgarbossa, M., Grzybovski, D; Mozzato, R.A. (2022). *Manual prático para estudos bibliométricos com o uso do Biblioshiny*. EDIUPF.

Thomaz, P. G., Assad, R. S., & Moreira, L. F. P. (2011). Uso do fator de impacto e do índice H para avaliar pesquisadores e publicações*. Arquivos Brasileiros de Cardiologia*, 96, 90-93.

Wilkerson, M. H., Lanouette, K., & Shareff, R. L. (2021). Exploring variability during data preparation: a way to connect data, chance, and context when working with complex public datasets. *Mathematical Thinking and Learning*, 24(4), 312-330.

Zipf, G.K. (1949). *Human behavior and the principle of least effort.* Addison-Wesley.

Text received on 07/28/2025

Text approvad on 03/06/2026

Text published on 08/26/2026