Student’s Statistical Literacy Skills (1980-2023): A Systematic Literature Review with Bibliometric Analysis

Febrianna Marchy & Dadang Juandi


To link this article: https://doi.org/10.37303/jelmar.v4i1.105

Copyright (c) 2023 Journal of Education and Learning Mathematics Research (JELMaR) is licenced under CC-BY-SA

Publisher
Department of Mathematics Education,
Faculty of Teacher Training and Education,
Universitas Wisnuwardhana Malang
Student’s Statistical Literacy Skills (1980-2023): A Systematic Literature Review with Bibliometric Analysis

1 Febrinna Marchy, 2 Dadang Juandi
Mathematics Education Study Program, Universitas Pendidikan Indonesia, Indonesia
Email: febrinnamarchy@upi.edu

Abstract: There have been demands to enhance the statistical literacy in the mathematics curriculum due to the vital role statistics plays in both everyday life and professional environments. The main purpose of this research is to investigate the heterogeneity and research landscape pertaining to students’ statistical literacy skills. This research employs a systematic literature review approach combined with bibliometric analysis as the chosen methodology. Data obtained from the Scopus database which has been refined through the PRISM method. The results show that the publication trend has increased in recent years. Publications in 2017 had the most impact in this area. The United States of America is the most influential country in this study. Most of the research was conducted at the secondary school level and most were published in Q3 journals. The learning method most often used in research related to students’ statistical literacy skills is project based learning. Information on research developments related to students’ statistical literacy skills can serve as a valuable reference for future researchers. It is strongly advised for educators and researchers to conduct further investigations into the statistical literacy skills of students in Asia and Africa, particularly focusing on the primary school level.

Keyword: Statistical Literacy, bibliometric, vosviewer

INTRODUCTION

Statistical literacy refers to the skill of comprehending, critically analyzing, and effectively communicating statistical information and messages (Gal, 2002). Statistical literacy is typically defined as the ability to analyze and evaluate statistical information in a critical manner (Koga, 2022). Statistical literacy, like reading, writing, and speaking, is a skill, statistical literacy requires two types of reading abilities: comprehension and interpretation (Schield, 1999). Statistical literacy includes essential skills that can be utilized to understand statistical information or research findings. These abilities encompass organizing data, generating and presenting tables, and handling diverse forms of data. (Callingham & Watson, 2017). Teachers are expected to have difficulties in implementing statistical literacy. In order to enhance statistical literacy in the classroom, educators may necessitate practical concepts and interactive learning activities (Sharma, 2013). Statistical literacy is essential to educate students particularly students in primary to become parts of a today’s society that is constantly bombarded with arguments based on statistical decisions (Aziz & Rosli, 2021). There have been demands to enhance the incorporation of statistical literacy into the mathematics curriculum due to the significance of statistics in everyday life and professional environments (Sharma, 2017). Several agents, such as educational institutions, statistics agencies, statistical associations, and the media, can help improve statistical literacy, the more these many agents work together to increase statistical literacy, the better the outcomes (Ferligoj, 2015).
We currently live in an open data era, with data providers such as national statistics offices, Eurostat, the OECD (Organisation for Economic Co-operation and Development), and the United Nations aiming to make their data available to the public (Kokotsaki et al., 2014). Technological and communication advancements have expanded the number of statistical information available through popular media (Sharma, 2013). Statistics are widely disseminated in the media, yet anyone without a background in statistics may transmit information and make decisions (Budgett & Pfannkuch, 2010). Reading and analysing statistical data needs more than just basic literacy, it needs statistical literacy (Chick & Pierce, 2013). Statistical literacy is necessary for productive citizens in the 4.0 industrial revolution era (Yuniawatika, 2018). In order to actively participate in society, individuals must possess the ability to critically evaluate statistical data. To cultivate this skill, it is imperative to commence the enhancement of statistical literacy among young individuals at different educational levels (Gonda et al., 2022). If children are not exposed to different aspects of statistical literacy during their schooling experience, exposing statistics to the environment might be a worthless experience since pupils will not form ideas about the data's contents (Singer et al., 2015).

Many research on statistical literacy have been undertaken between 1980 and 2023, both in elementary school, middle school, and college, such as (Carel & Juandi, 2023; Weiland & Sundrani, 2022; Towse et al., 2022; Büscher, 2022; Gómez-Blancarte et al., 2021; Valentini, 2016) and other research. It is important to understand analyze how the study description connects to statistical literacy in order to collect comprehensive information and proper data. As a result, this study was carried out as a Systematic Literature Review (SLR) with bibliometric analysis (Farman, 2021). A systematic literature review provides a concise overview of the existing body of research conducted on a specific topic (Ariati & Juandi, 2022). It is carried out in order to find, select, assess, and synthesize all trustworthy research data relevant to that topic (Cronin et al., 2013). Based on the social, intellectual, and conceptual structure of disciplines, bibliometric analysis examines the evolution of a study domain, encompassing subjects and authors (Donthu et al., 2020). SLR was carried out using bibliometric analysis, which is one of the most extensive and intelligent data study methods in the literature (Farman, 2021).

The primary goal of this systematic literature review with bibliometric analysis are to find out how the heterogeneity of studies related to the students’ statistical literacy skills and how the research landscape related to this study from 1980 to 2023. The research questions used in this study are as follows: (1) What are the current trends in publications on students' statistical literacy skills? (2) At which education level is there a lot of research related to students' statistical literacy skills? (3) What is the teaching method that impacts the development of students' statistical literacy skills? (4) What is the trend of citation in students' statistical literacy skills? (5) What is the geographical distribution of publications and patterns of cooperation between countries in research involving students' statistical literacy skills? (6) What is the relationship between the distribution of journal rankings based on quartile values and students' statistical literacy skills? (7) What is the research focus on students' statistical literacy skills?

METHOD

The research protocol was PRISM (Preferred Reporting Items for Systematic Review and Meta-Analyses). The selection procedure consists of four steps: identification, screening, eligibility, and inclusion (Juandi & Tamur, 2020; Liberati et al., 2009; Moher et al., 2009). In this study, the researchers used the keywords (“statistical literacy” AND "students") in the Scopus database to do identification. According to the original search, 247 documents about statistical literacy and students were found. The next stage was Screening, in the screening process the researcher determined the inclusion criteria.
including, namely, publications must be in English, publications in the form of article and conference paper, subject area of publications is mathematics. Following the screening process, 93 publications that fit the predefined criteria were identified. The third process was looking at the manuals, titles and abstracts of the 93 publications. The criteria are whether or not statistical literacy is discussed with students across all levels of education (from primary until university level). Statistical literacy publications at the students will be further analyzed at a later stage. From the feasibility results, 23 publications do not fulfil the criteria. The number of publications that are included in the inclusion stage is 70 documents, 5 documents from conference paper and 65 documents from article. Descriptive analysis and bibliometric analysis are employed to examine publication trends concerning students’ statistical literacy using data extracted from the Scopus database.

Data that meets the inclusion criteria and passes the four prism stages will then be analyzed with several applications including Microsoft Excel, Publish or Perish and Vosviewer. Microsoft Excel is used to see the trend of publications related to statistical literacy from 1980 to 2023 and examine the dispersion of journal publications according to quartile values. Publish or Perish is used to calculate the researcher's annual citations, calculate the total citations from the annual publication and calculate the h-index and g-index values. In looking at patterns of relations between countries related to students’ statistical literacy and to see research focus and research novelty, the VOSviewer application is used. The selection of articles is depicted in Figure 1 as follows.
RESULT AND DISCUSSION
Publications related to student’s statistical literacy skills that have gone through the process of collecting data so that 70 publications are obtained in the range from 1980 to 2023 and are in accordance with the criteria then a systematic literature review with bibliometric analysis is carried out. Publication trends, education level, teaching materials used in research, citation trends, country and journal distribution and research focus will be discussed further.

What are the current trends in publications on students' statistical literacy skills?

Figure 2 depicts the progression of publications concerning students' statistical literacy skills from 1980 to 2023. A total of 70 publications are grouped by year of publication, which can be seen in Figure 2 below.

![Trends in publications on students' statistical literacy skills by year](image)

Figure 2. Study Trends by Year of Publication

Figure 2 indicates that publications relating to students' statistical literacy skills have increased over the previous 44 years, from 1980 to February 2023, although there was little change from 1980 to 1998. The greatest increase was 167% from 2015 to 2016, followed by decreases from 2017 to 2019 and 2021 to 2022. In 2016, there were 8 publications related to students' statistical literacy skills, up to 11 in 2017. According to Garfield et al. (2008) in the 1990s, there was a significant increase in the demand for statistical education to prioritize research on statistical literacy.

At which education level is there a lot of research related to students' statistical literacy skills?

Figure 3 shows the specifics of the primary dissemination study depending on education level, from Primary to university.
According to Figure 3, the majority of the research was conducted in secondary schools, with as many as 27 studies. With 23 studies, the following was largely done at the university level. Meanwhile, there was less study on students' statistical literacy skills at the primary level, with seven studies. Three studies were conducted concurrently in primary and secondary schools, and ten publications did not specify at what level of education they were conducted.

Developing statistical literacy is crucial for educating students, especially those in primary schools, so they can effectively navigate today’s society, which is consistently exposed to arguments based on statistical judgments (Aziz & Rosli, 2021). However, as seen in Figure 3, research on students' statistical literacy is still quite limited at the primary school level, implying that research at the primary school level should be increased.

What is the teaching method that impacts the development of students' statistical literacy skills?

Some methods were found to be applied to increase statistical literacy skills based on an analysis of 70 papers relating to students' statistical literacy skills. Some studies are shown in Table 1 below.

Table 1. Teaching Method that Impacts the Development of Students' Statistical Literacy Skills

<table>
<thead>
<tr>
<th>Authors</th>
<th>Teaching Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Lasater et al., 2023)</td>
<td>To explore the data, use a visualization program to produce tables and charts.</td>
</tr>
<tr>
<td>(Hsu et al., 2022)</td>
<td>Familiarize students with Excel, Python, or other tools to assist them in achieving statistical literacy through the analysis of real-world data.</td>
</tr>
<tr>
<td>(Delport, 2022)</td>
<td>Use media articles to provide students with opportunity to reflect on such (mis)representations and to develop statistics literacy.</td>
</tr>
<tr>
<td>(De Souza Oliveira &amp; De Faria Reis, 2021)</td>
<td>Nepso is an approach that promotes statistical literacy in students by combining abilities and skills required for project work.</td>
</tr>
<tr>
<td>(Souza et al., 2020)</td>
<td>Develop and carry out statistical inquiry that harnesses parts of creative dissent to enhance student learning experiences via engaging inquiry assignments that stimulate collaborative exploration of ideas, data analysis, and reporting.</td>
</tr>
<tr>
<td>(Batanero &amp; Gea, 2020)</td>
<td>Collaborate with statistical and investigative projects to further educate citizens about statistics.</td>
</tr>
<tr>
<td>(Odom &amp; Bell, 2017)</td>
<td>Planning and teaching statistical concepts in a PK-12 classroom using inquiry-based teaching and learning approaches</td>
</tr>
</tbody>
</table>
The PPDAC cycle was adopted as a framework for arranging students' statistical investigative activities.

On academic achievement in statistics, matching multiple memory methods with computer-assisted instruction strengthen students' abilities in the areas of arithmetic, interpretation and graphical display of data. The laboratory instrument is another tool offered to help students develop critical thinking skills. High school students can observe the complete data generation process, from microdata through synthesis procedures, in the laboratory.

A project-based learning approach at the statistical literacy level of grade 8 middle school students for data representation a collaborative project to develop students' statistical literacy

TBL is a teaching technique in which groups of students work together in teams to study subject matter. The primary goal of TBL is to provide students with opportunities to practice course content during class time, providing projects to let students understand how to collect and process the information offered in the media articles and scientific publications they read.

From table 1 above it is known that the learning method most often used to develop students' statistical literacy skills is project-based learning. One of the recommended learning trends in the 2013 Curriculum is Project Based Learning (PjBL), which aligns with the demand for 21st-century skills (Yunita et al. (2021)). This is in accordance with Koparan (2015) that the use of project-based learning will increase students' statistical literacy. There are other learning methods that are also used, namely use a visualization program, familiarize students with Excel, Python, or other tools, use media articles, team based learning, inquiry based learning, and others.

What is the trend of citation in students' statistical literacy skills?
The table provided as Table 2 illustrates the citation trends associated with student's statistical literacy skills from 1980 to 2023. As with publication trends, as many as 70 publications were collected based on the year of publication, which will then be seen based on total publications per year, NCP, TC, C/P scores, and others, which can be seen in table 2 below.

Table 2. Citation Analysis of Publications

<table>
<thead>
<tr>
<th>Year</th>
<th>T</th>
<th>NCP</th>
<th>TC</th>
<th>C/P</th>
<th>C/CP</th>
<th>h</th>
<th>g</th>
</tr>
</thead>
<tbody>
<tr>
<td>2023</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1,0</td>
<td>1,0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2022</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0,2</td>
<td>1,0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2021</td>
<td>7</td>
<td>6</td>
<td>14</td>
<td>2,0</td>
<td>2,3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2020</td>
<td>5</td>
<td>4</td>
<td>13</td>
<td>2,6</td>
<td>3,3</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2019</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>3,5</td>
<td>7,0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2018</td>
<td>5</td>
<td>4</td>
<td>17</td>
<td>3,4</td>
<td>4,3</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2017</td>
<td>11</td>
<td>10</td>
<td>133</td>
<td>12,1</td>
<td>13,3</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>2016</td>
<td>8</td>
<td>7</td>
<td>105</td>
<td>13,1</td>
<td>15,0</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td>2015</td>
<td>3</td>
<td>3</td>
<td>20</td>
<td>6,7</td>
<td>6,7</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2014</td>
<td>4</td>
<td>4</td>
<td>7</td>
<td>1,8</td>
<td>1,8</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2013</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1,0</td>
<td>1,0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2012</td>
<td>2</td>
<td>2</td>
<td>23</td>
<td>11,5</td>
<td>11,5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2011</td>
<td>2</td>
<td>2</td>
<td>16</td>
<td>8,0</td>
<td>8,0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>2010</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>17,0</td>
<td>17,0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2009</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2008</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>9,0</td>
<td>9,0</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Based on the data presented in Table 2, the value for the year 2017 is 10, indicating that it has the highest NCP among the preceding years. The number of publications in 2017 was 11, which was the most in comparison to the previous year. Then, based on the number of citations, articles in 2017 were cited 133 times, more than in any previous year. However, some years, particularly 1981–1997 and 2001–2002, do not have number of citations.

2017 had the greatest h-index and g-index values, with $h_{index} = 6$ and $g_{index} = 11$. As a result, 2017 had the highest influence on this research. 11 publications cited in 2017 received a total of 133 citations, implying that at least 11 of these papers were cited 12 times, as evident in the table presented below as Table 3.

### Table 3. Articles Published in 2017 (citation > 1)

<table>
<thead>
<tr>
<th>No</th>
<th>Author (year)</th>
<th>Title</th>
<th>Sources</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Gould, 2017)</td>
<td>Data literacy is statistical literacy</td>
<td>Statistics Education Research Journal</td>
<td>52</td>
</tr>
<tr>
<td>4</td>
<td>(Mclaughlin &amp;</td>
<td>A flipped classroom model for a</td>
<td>Statistics Education</td>
<td>10</td>
</tr>
</tbody>
</table>
Research conducted by Gould (2017) with the title "Data literacy is statistical literacy" gets the highest number of citations, namely 52 times, the article explores the need to revise previous definitions of statistical literacy in order to encompass the increasingly prominent role of data in our everyday lives. Through an innovative data science program conducted with high school students, it has been demonstrated that teaching statistical literacy, alongside data literacy, can commence at an early stage. The research with the second highest number of citations is the research conducted by Weiland (2017) which discusses the study of data, statistical literacy holds significant importance in mathematics education, as it is intricately integrated into the K-12 school-level mathematics curriculum. This study examines the essential concepts of critical literacy within formal classroom settings, specifically in the context of mathematics education, and explores the intersection of these key ideas with those of statistical literacy to establish a framework for critical statistical literacy. The journals in table 3 above can be used as a reference and purpose of publication related to this field such as the journal "Statistics Education Research Journal", “Educational Studies in Mathematics”, and others.

**What is the geographical distribution of publications and patterns of cooperation between countries in research involving students' statistical literacy skills?**

The geographical distribution of countries and patterns of relations between countries are seen based on the origin of the authors of published documents related to students' statistical literacy skills from 1980 to 2023. The results of the analysis can be seen in Figure 4.
Figure 4. The Geographical Distribution of Publication

Figure 4 shows that the country that has the most influence on students’ statistical literacy skills is the United States with 21 publications. Furthermore, Australia is the second most influential country with a total of 16 publications. Germany, Ireland, Italy, and United Kingdom each with 3 publications, Brazil, New Zealand, and Turkey with a total of 3 publications. China, Indonesia, and Spain with a total of 2 publications. Then Chile, Czech Republic, Israel, Japan, Serbia, Slovakia, South Africa, South Korea, Taiwan, and Thailand each with 1 publication related to the topic of students’ statistical literacy skills. In this case the United States have the most influence on this topic and this number is very far when compared to publications from other country. This is in accordance with what was conveyed by Aziz & Rosli (2021) that the United States is the country with the most authors related to students’ statistical literacy skills.

The pattern of collaboration between countries can be seen in Figure 5 below. In this stage, researchers do not establish a specific threshold. Consequently, even countries with only one document related to this field will be included, regardless of whether they have collaborative relationships with other countries. The 22 countries that are displayed can be observed in Figure 5 below, as indicated.

Figure 5. Patterns of State Collaboration
Based on the above Figure 5, it is evident that the United States exhibits the largest circle diameter in comparison to other countries, indicating the highest degree of collaborative relations. The VOSviewer visualization reveals that the United States has engaged in collaborations with 21 other countries, indicating comprehensive collaboration with all the displayed countries.

**What is the relationship between the distribution of journal rankings based on quartile values and students' statistical literacy skills?**

The distribution of journal rankings is seen from the quartile (Q) value of journals related to publications on students' statistical literacy skills. The website Scimagojr displays journal rankings. The distribution of journal rankings can be seen in Figure 6 below:

![Journal Ranking Distribution](image)

**Figure 6. Ranking based on Journal Quartile Values**

Figure 6 shows that 34% of publications in journals relevant to research on students' statistical literacy skills are at the Q3 level, with a total of 22 journals. The journal ranks in second place, with a Q2 ranking of 32% and a total of 21 journals. Journals with a Q1 ranking of 29%, totaling 19 journals, and journals with a Q4 ranking of 3%, totaling 2 journals. There is one journal that does not yet have a quartile value, which means that statistical literacy writing must be improved again before it can be published in a journal with a quartile value.

**What is the research focus on students' statistical literacy skills?**

The threshold is chosen at this step by the researcher, specifically shared keywords that are presented with a minimum of two publications, which means that keywords that have been used together in two or more publications will be displayed on the keyword co-appearance using VOSviewer. Figure 7 shows that when the threshold was established, 22 keywords were found.
Figure 7. Keyword Co-occurrence (Occurrence Threshold ≥ 2)

The focus of the research can be seen from the clusters shown; in Figure 7 above, it can be seen that there are two different colors. This color indicates a research cluster related to students’ statistical literacy skills. There are two clusters, namely red and green. This means that the research focus is divided into two parts: 1) The first cluster (in red) consists of 12 items; if seen from the size of the circle in this first cluster, the keywords that have the largest diameter are the keywords teaching, statistics education, and teaching statistics, meaning that these keywords are the first research focus together with statistical literacy; 2) The second cluster (in green) consists of 10 items, with the keywords statistics and students being the largest circle in the cluster, meaning that these keywords are the first research focus together with statistics education research.

CONCLUSION

Research on students’ statistical literacy skills has increased over the last 4 decades. Publications in 2017 have been cited more than in any other year and are also the most published. The United States of America is the most influential country in this study. Most of the research was conducted at the secondary school level and most were published in Q3 journals. The learning method most often used in research related to students’ statistical literacy skills is project based learning. Information on research developments related to students' statistical literacy skills can serve as a valuable reference for future researchers. It is strongly advised for educators and researchers to conduct further investigations into the statistical literacy skills of students in Asia and Africa, particularly focusing on the primary school level. Furthermore, it is proposed that educators and researchers explore innovative approaches to enhance students’ statistical literacy skills through the development of new instructional methods. The limitation of this study is that the data was taken on February 4, 2023, meaning that articles published after that date are not part of the analysis of this research. The database used only comes from the Scopus database, this is still not enough to describe students' statistical literacy skills, because there are many other databases that can be used.
REFERENCES


