A deep dive into statistical, data, and digital literacies: A bibliometric analysis (2000–2024)

R. Fernández-Pascual, M. Pinto, D. Caballero and A. E. Marín Jiménez
University of Granada, Spain
rpascual@ugr.es

This study aims to explore the field of statistical literacy in relation to quantitative, data, and digital literacy, highlighting their interconnected development and significance, to better understand how STEAM educators and researchers contribute to advancing these essential skills. Scientific literature indexed in Web of Science from 2000 to 2024 (n = 674 articles and books) is analyzed using co-occurrence analysis in VOSviewer. Five interrelated key trends are discovered: Data Literacy, Statistical Applications for Modeling and Prediction, Statistics Education, Mathematics, and Quantitative skills, Statistical and Quantitative Literacies and Digital Literacy. The findings highlight the growing relevance of statistical, data, and digital literacy, as well as the need of an effective integration of these literacies for education, research, and professional practice in STEAM disciplines.

MOTIVATION

In a data-driven world, data, statistical, and digital literacies are increasingly interconnected and essential. For STEAM students and graduates, mastering this multi-literacy is crucial for advancing the information and knowledge society, driving innovation, and contributing to societal progress. This study underscores the need for further research of the intersections between statistical, data, and digital literacy, particularly in the STEAM disciplines.

Bibliometric studies shed light on the main concerns and topics of interest for the educational and research community.

The overall objective of this study is to explore the thematic field of statistical literacy examining its development alongside related literacies, specifically quantitative, data and digital literacy, to gain a deeper understanding of how educators and researchers of STEAM area are engaging in the advancement of these critical skills. In order to better understand the development of academic production, the following specific research objectives are proposed:

- RO1: To identify the most significant terms used in titles and author-selected keywords, in order to uncover the main research trends in the field.
- RO2: To analyze the temporal evolution in the usage of key terms, highlighting how the focus of research has shifted over the study period.

LITERATURE REVIEW

Although it is difficult to establish clear boundaries between statistical literacy and data literacy (Schield, 2005; Weiland, 2017), statistical literacy is understood as the ability to think critically about statistical information, involving the analysis and interpretation and the comprehension of key statistical concepts (Garfield and Ben-Zvi, 2004; Schield, 2004), while data literacy refers to the ability to access, assess, manipulate, summarize, present data and comprehending data (Koltay, 2015). Data literacy is an essential component of both information literacy and statistical literacy which involves a series of skills, critically evaluating data and its sources; knowing how to select and synthesize data; using data ethically (Gould, 2017; Šorgo, 2018).

With regard to the evolving definition of Digital literacy, there is a lack of agreement about its scope (Heitin, 2016; Biezā, 2020; Dimitrakopoulou, 2022). Osterman (2013), Feerrar (2019) and the American College Research Libraries (2015), frame Digital literacy as the ability to effectively use information and communication technologies to locate, evaluate, create, and communicate information, combining both cognitive and technical skills.

We also identified in the literature a significant presence of works focused on numeracy and quantitative literacy, which incorporate concepts and practices from statistics (Steen, 2001; Lackie, 2004; Karaali et al., 2016; Wilkins, 2016; Brock et al., 2021), thereby blurring the boundaries among these related literacies.

In: Birk, L., Loth, G., Jotzo, L., Binder, K., & Frischemeier, D. (Eds.) (2026). *Statistics and Data Science Education in STEAM. Proceedings of the Satellite Conference of the International Association for Statistical Education (IASE), September/October 2025, Münster, Germany*.

There is a clear overlap among these literacy frameworks, as a result of the interconnected nature of digital literacy, information literacy, data literacy, quantitative literacy and statistical literacy. Schield (2005, p. 6) stated that it "is difficult to promote information literacy or data literacy without promoting statistical literacy".

We found several recent bibliometric analyses focused on data literacy, statistical literacy, and related fields. Naseema and Sevukan (2022) conducted a scientometric analysis of publications indexed in Scopus on data literacy, using CiteSpace and VOSviewer. Their study identified key areas of specialization (education, medicine, and technology). They conclude that enhancing data literacy requires a multidisciplinary approach and a stronger focus on foundational skills. Marchy and Juandi (2023) investigated the research landscape on statistical literacy, finding out that most of the studies were conducted at the secondary school level and the preferred learning method was project-based learning. The findings underscore the importance of conducting further investigations into statistical literacy in Asian and African contexts, especially at the primary school level. Ghodoosi et al. (2023) conducted a systematic review of data literacy education, excluding other related literacies such as digital and statistical literacy but this analysis drew from Google Scholar, Science Direct, ResearchGate, and Scopus, excluding Web of Science. Results show that data literacy education is largely grounded in established learning theories and pedagogical approaches, but lacks enough empirical evidence and a clear consensus on the core competencies that define it. These gaps point to an urgent need for empirical research to better define both the content and the most effective methods for teaching data literacy. Finally, Fernández-Pascual et al. (2024) conducted an analysis of the scientific production indexed in Web of Science regarding the thematic field of data literacy and its related literacies, combining co-occurrence (using VOSviewer) and co-citation (using CiteSpace) techniques. A progression in the data literacy terminology was discovered. Data literacy has evolved from early focuses on quantitative and statistical literacy to applied uses across disciplines, and more recently, to societal themes like citizenship and big data. Results revealed the most significant trends and the evolution of data literacy, highlighting its expansion across multiple disciplines.

METHODS: DATA COLLECTION

The selection criteria were established to include only journal articles and books written in English and published between 2000 and 2024. Consequently, conference proceedings, doctoral dissertations, and review articles were excluded from the analysis. The search was conducted using the following query:

> TS=("data literacy" OR "data literate" OR "data competence" OR "data competency" OR "data skill" OR "data literacies" OR "data literates" OR "data competences" OR "data competencies" OR "data skills" OR "data competent" OR "data skillful" OR "data skilful" OR "quantitative literacy" OR "quantitative literate" OR "quantitative competence" OR "quantitative competency" OR "quantitative skill" OR "quantitative literacies" OR "quantitative literates" OR "quantitative competences" OR "quantitative competencies" OR "quantitative skills" OR "quantitative competent" OR "quantitatively skillful" OR "quantitatively competent" OR "quantitatively skilful" OR "statistical literacy" OR "statistical literate" OR "statistical competence" OR "statistical competency" OR "statistical skill" OR "statistical literacies" OR "statistical literates" OR "statistical competences" OR "statistical competencies" OR "statistical skills" OR "statistical competent" OR "statistically skillful" OR "statistically competent" OR "statistically skilful" AND ("digital literacy" OR "digital literate" OR "digital competence" OR "digital competency" OR "digital skill" OR "digital literacies" OR "digital literates" OR "digital competences" OR "digital competencies" OR "digital skills"))

After verifying the relevance of the publications to the fields involved, the final dataset included a total of 674 references.

RESULTS

The descriptive overview of the quantitative scientific production throughout the analyzed period (2000–2024) reveals a steady growth in data literacy research over the past decade, with a significant increase that reaches its peak in 2024, with a maximum of 78 publications (Figure 1).

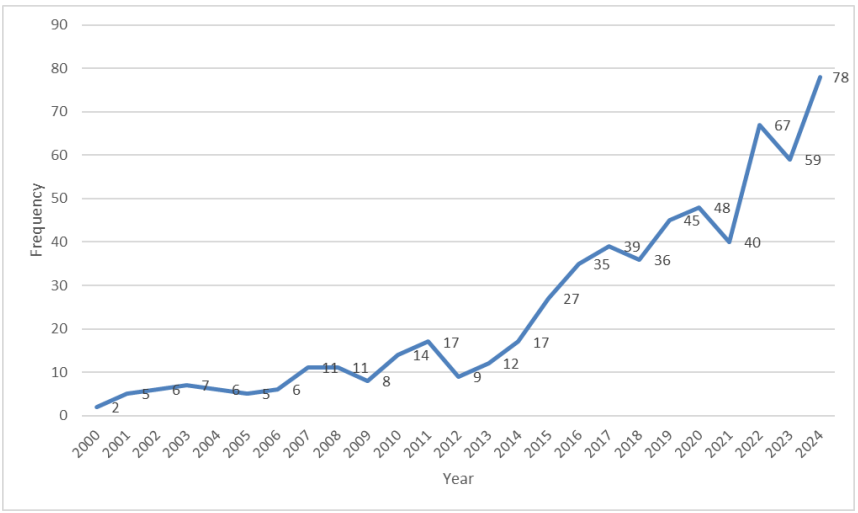

Figure 1. Evolution of scientific production.

With the aim of identifying recurring topics, major trends, and the evolution over time of the main terms used in titles and abstracts, we performed a co-occurrence analysis in VOSviewer (Van Eck and Waltman, 2013, 2017). Following the removal of stopwords (connectors, prepositions, conjunctions, and articles), a total of 15266 terms were identified for analysis and the 201 terms that appeared more than 20 times, were included in the maps. Finally, we generated a network visualization to examine the relationships among items, using the total link strength attribute to quantify the overall intensity of connections. In this representation, colors denote the cluster affiliation of each term, and label size indicates the relative weight or importance of each item within the network.

Regarding RO1, results show evidence of five main interrelated trending topics (Figure 2). First, data literacy and its societal impact, within their applications to the community and society. This trend focuses on data visualization, big data, artificial intelligence, and open data policies. Second, statistical applications for modeling and prediction, which focuses on the use of statistics to develop predictive models for assessment across disciplines. Third, there is growing concern among educators and researchers about the integration of statistical education, mathematics, and quantitative literacy in curricula within STEAM disciplines. The fourth trend relates to statistical and quantitative literacy from the perspective of statisticians, who explore the intersection of these literacies and their implications for education, research, and professional practice. Finally, the last trend addresses the role of digital literacy in content creation, communication and dissemination, and digital competencies, as well as the growing interest in attitudes and behaviors related to digital technologies, data communication, digital security, and the ethical sharing of information in an increasingly interconnected digital landscape.

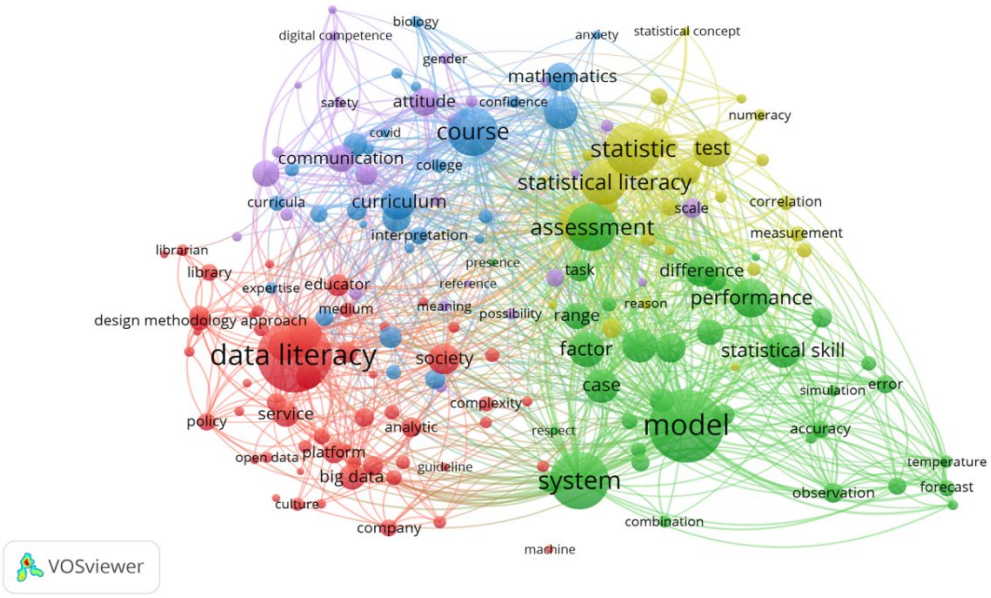

Figure 2. Main trending topics.

To address RO2, regarding the analysis of the temporal evolution in the usage of key terms, an examination of the chronological trends reveals a clear shift in both the terminology employed and the thematic focus of the research corpus over time (Figure 3). Starting in 2012, these publications were focused around quantitative skills, quantitative literacy, statistical skills and statistical concepts like probability, correlation, error, experiment, or forecast. Statistical literacy and related terms, along with the emerging concern for the assessment and performance, led to the incorporation of applied terms such as model, evaluation, assessment, scale, or score, which were published on average in 2018. In 2019-2020, studies covered both the application areas of statistical literacy in different disciplines, courses and curricula, and its application to citizen, society, community, government and data use. Finally, by 2021, the literature began to incorporate the terms data literacy, data literacy education, digital literacy, digital competence, big data, and data science.

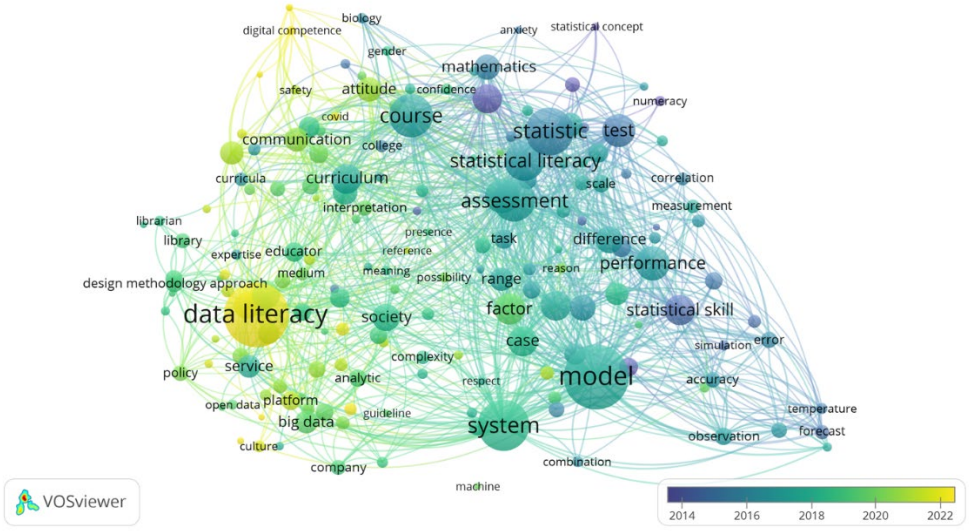

Figure 3. Overlay visualization of trending topics (colors indicate the average year of publication).

DISCUSSION

Findings of our bibliometric analysis of scientific production in WoS on the statistical, data and digital literacy fields in STEAM disciplines from 2000 to 2024 are consistent with the results obtained by Ghodoosi et al. (2023). Our results from the analysis of words in titles and abstracts, show that during the first decade of the 21st century, studies focused on the development and integration of quantitative and statistical skills in STEAM education. However, by 2012, there was a shift towards statistical literacy and applied methods for prediction and assessment, which became a focus of interest until 2018. Around 2018, publications covered both the application areas of statistical literacy and digital literacy to citizen and society. Although the term data literacy began emerging in the literature in 2011 (Ghodoosi et al., 2023; Fernández-Pascual et al., 2024), it was not until 2021 that the usage of terms such as data literacy, digital literacy, big data and data science was consolidated. This trend reflects the recent interest on these topics within the research community.

The integration of statistical, data and digital literacies, and their implications for education, research, and professional practice in STEAM disciplines is a topic of interest among educators and researchers. Meridha (2024) highlighted "the lack of integration of technology into teaching and learning practices is an obstacle for digital literacy in academic settings" (p. 1), and the importance of reducing infrastructure disparities between low- and high-ranking institutions to facilitate the adoption of digital technologies in the teaching–learning process. There is also a growing concern about teachers' digital literacy, as age and gender differences may hinder the development of fully integrated digital skills in education (Adeoye, 2023).

Encheva et al. (2024) examined the development of digital literacy tools in STEAM and concluded that a new generation of STEAM experts is needed, able of effectively integrating digital skills into teaching practices. Hafizzaturroyani et al. (2024) analyzed the attitudes of students and teachers regarding digital literacy and emphasized the need to shift teaching approaches toward more interactive methods that actively engage students in the learning process.

The literature suggests that, in order to promote the integration of statistical, data, and digital literacies, it is necessary to develop individuals capable of managing data, performing tasks, and solving statistical problems in digital environments. In this line, some authors recommend active learning as an effective method to foster digital literacy (Jesionkowska et al., 2020; Caratozzolo et al., 2021), while project-based learning is gaining support as a strategy to promote the integration of statistical, mathematical, and digital literacies, increasing positive attitudes and improving the skills levels (Koparan & Güven, 2015; Zhang & Fang, 2019; Vistara et al., 2022; Supianti et al., 2025).

The study reveals a shift in STEAM education research from basic statistical skills to integrated literacies (statistical, data, and digital) highlighting the need to update curricula accordingly. Emphasis should be placed on teaching students to manage and interpret data in digital contexts. The growing prominence of terms like *data literacy* and *data science* underscores the importance of embedding these literacies across disciplines. Active and project-based learning are recommended strategies to foster relevant skills and attitudes, calling for institutional support in teacher training and curriculum design.

CONCLUSION

The findings highlight the growing relevance of statistical, data, and digital literacy, particularly in STEAM field. It is crucial to conduct further research on the intersections among these literacies, drawing on diverse sources such as WoS and Scopus. Future research focused on the diagnosis of both attitudes and actual skill levels across different educational contexts is essential. It is not only important to understand how they are perceived by educators and students but also to have reliable tools that can accurately measure data and digital literacy skills in authentic learning environment. Developing and validating reliable instruments to assess data literacy levels in real-world educational settings is crucial.

Moreover, the growing presence and impact of artificial intelligence (AI) in both learning and research processes within STEAM field, as well as in broader society, must be considered. Consequently, understanding and identifying the connections and interrelationships among various literacies, such as data, digital, and AI literacy, has become an urgent necessity. However, this also presents a significant challenge, as it entails ongoing processes of diagnosis, reflection, and decision-making.

DISCLOSURE OF INTERESTS

The authors have no competing interests to declare.

ACKNOWLEDGMENT

This paper is part of the R+D+I "Knowledge Generation", PID2021-128808OB-I00, funded by MICIU/AEI/ 10.13039/501100011033 and, by the European Regional Development Fund (ERDF/EU).

REFERENCES

Adeoye, M. A. (2023). Gender differences in teachers' digital literacy skills in teaching STEAM. *Journal of Education Technology, 7*(3), 462–469.

Association of College and Research Libraries (2015). Framework for Information Literacy for Higher Education. Association of College and Research Libraries, Chicago, IL. http://www.ala.org/acrl/standards/ilframework

Biezā, K. E. (2020). Digital literacy: Concept and definition. *International Journal of Smart Education and Urban Society (IJSEUS), 11*(2), 1–15.

Brock, C.H, Wiest, L.R. & Thrailkill, L. (2021). Learning quantitative literacy: A sixth-grade disciplinary literacy unit. *The Reading Teacher, 74*(6), 733–746. https://doi.org/10.1002/trtr.2008

Caratozzolo, P., Alvarez-Delgado, A., & Sirkis, G. (2021). Fostering digital literacy through active learning in engineering education. In *2021 IEEE Frontiers in Education Conference* (*FIE*) (pp. 1-6). IEEE. https://doi.org/10.1109/fie49875.2021.9637304

Dimitrakopoulou, D. (2022). Digital literacy. In L. A. Schintler & C. L. McNeely (Eds.), *Encyclopedia of Big Data* (pp. 393–395). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-32010-6_72

Encheva, M., Tammaro, A. M., Yancheva, G., Zlatkova, P., Conti, G., & Maasilta, M. (2024). Towards a STEAM model for digital fluency skills: Perceptions of students and teachers. *IFLA Journal, 50*(1), 75–92. https://doi.org/10.1177/03400352231209491

Feerrar, J. (2019). Development of a framework for digital literacy. *Reference Services Review, 47*(2), 91–105.

Fernández-Pascual, R., Pinto, M., & García Marco, F. J. (2024). Emergence and evolution of data literacy: Insights from a bibliometric study. *Journal of Librarianship and Information Science*, *57*(4), 1079–1097. https://doi.org/10.1177/09610006241265102

Garfield, J., & Ben-Zvi, D. (2004). Research on statistical literacy, reasoning, and thinking: Issues, challenges, and implications. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 397–409). Dordrecht: Springer Netherlands.

Ghodoosi, B., West, T., Li, Q., Torrisi-Steele, G. & Dey, S. (2023). A systematic literature review of data literacy education. *Journal of Business & Finance Librarianship, 28*(2), 112–127. https://doi.org/10.1080/08963568.2023.2171552

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal, 16*(1), 22–25.

Hafizzaturroyani, H., Habibi, M., Fahruddin, F., Darmiany, D., & Subarinah, S. (2024). Development of digital literacy tools in the implementation of Science, Technology, Engineering, Art, and Mathematics (STEAM) Learning for the Implementation of the Independent Curriculum for Children Aged 5-6 Years. *International Journal of Education and Digital Learning (IJEDL), 2*(5), 63–72.

Heitin, L. (2016). What is Digital Literacy? Digital Literacy: an evolving definition. *Education Week, 36*(12).

Jesionkowska, J., Wild, F., & Deval, Y. (2020). Active learning augmented reality for STEAM education-A case study. *Education Sciences, 10*(8), 198. https://doi.org/10.3390/educsci10080198

Karaali, G., Villafane Hernandez, E. H., & Taylor, J. A. (2016). What's in a Name? A critical review of quantitative literacy, numeracy, and quantitative reasoning. *Numeracy*, *Journal of the National Numeracy Network, 9*(1), Article 6. http://dx.doi.org/10.5038/1936-4660.9.1.2

Koltay, T. (2015). Data literacy: in search of a name and identity. *Journal of Documentation, 71*(2), 401–415. https://doi.org/10.1108/JD-02-2014-0026

Koparan, T., & Güven, B. (2015). The effect of project-based learning on students' statistical literacy levels for data representation. *International Journal of Mathematical Education in Science and Technology, 46*(5), 658–686.

Lackie, P. (2004). Helping Students Understand & Use Data: A discussion of the jargon and trends in "Quantitative Literacy". Presented at *IASSIST 2004 Conference*, Madison.

Marchy, F. & Juandi, D. (2023). Student's statistical literacy skills (1980-2023): A systematic literature review with bibliometric analysis. *Journal of Education and Learning Mathematics Research, 4*(1), 31–45. https://doi.org/10.37303/jelmar.v4i1.105

Meridha, J. M. (2024). The causes of poor digital literacy in educational practice, and possible solutions among the stakeholders: a systematic literature review. *SN Soc Sci, 4,* 210. https://doi.org/10.1007/s43545-024-01010-8

Naseema, S. & Sevukan, R. (2022). Global research trends in research data management (RDM) – a scientometric view. *International Journal of Information Science and Management, 20*(4), 117–135.

Osterman, M. D. (2013). Digital literacy: Definition, theoretical framework, and competencies. In M. S. Plakhotnik, S. M. Nielsen, & D. M. Pane (Eds.), *Proceedings of the 11th Annual College of Education & GSN Research Conference* (pp. 135–141). Florida International University.

Schield, M. (2004). Statistical literacy curriculum design. *Curriculum Design Roundtable* sponsored by the International Association of Statistical Educators (IASE) in Lund, Sweden. https://www.statlit.org/pdf/2004SchieldIASE.pdf

Schield, M. (2005). Information Literacy, Statistical Literacy and Data Literacy. *IASSIST Quarterly*, *28*(2–3), 6. https://doi.org/10.29173/iq790

Šorgo, A. (2018). Information, data and statistical literacy as foundation stones of project-based education. In M. Rusek & K. Vojíř (Eds.), *Project-based education in science education empirical texts XV* (pp.12–20). Univerzita Karlova.

Steen, L. (Ed.). (2001). *Mathematics and democracy: The case for quantitative literacy*. United States: The National Council on Education and the Disciplines.

Supianti, I. I., Yaniawati, P., Bonyah, E., Hasbiah, A. W., & Rozalini, N. (2025). STEAM approach in project-based learning to develop mathematical literacy and students' character. *Infinity Journal, 14*(2), 283–302.

Van Eck, N. J. & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics, 111*, 1053–1070. https://doi.org/10.1007/s11192-017-2300-7

Van Eck, N. J., & Waltman, L. (2013). VOSviewer manual. Leiden: Universiteit Leiden. 1–53.

Vistara, M. F., Rochmad, R., & Wijayanti, K. (2022). Systematic literature review: STEM approach through engineering design process with project based learning model to improve mathematical creative thinking skills. *Mathematics Education Journal, 6*(2), 140–156.

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics, 96*(1), 33–47.

Wilkins, J. L. (2016). An assessment of the quantitative literacy of undergraduate students. *The Journal of Experimental Education, 84*(4), 639–665. https://doi.org/10.1080/00220973.2015.1111854

Zhang, H., Fang, L. (2019). Project-Based Learning for Statistical Literacy: A Gamification Approach. In T. Väljataga & M. Laanpere (Eds.), *Digital Turn in Schools—Research, Policy, Practice. Proceedings of ICEM 2018 Conference* (pp. 3–16). Singapore: Springer.