BIODIDAKTIKA: Jurnal Biologi dan Pembelajarannya, Vol.20, No.2, 2025, pp. 126-136
e-ISSN 2527-4562. DOI. 10.62870/biodidaktika.v20i2.31522

Assessing Biology Education Students' Self-Efficacy in Data Literacy: An Analysis of Their Confidence in Understanding Data

Submitted 5 March 2025, Revised 16 July 2025, Accepted 27 July 2025

Puti Siswandari*[1], Hardini Puspitaningrum[2], Rini Solihat[3]

[1,2,3]Biology Education Study Program, Faculty of Mathematics and Natural Sciences Education, Universitas Pendidikan Indonesia, Bandung, Indonesia

Corresponding Email: *puti.siswandari@upi.edu

Abstract

Data literacy is an essential aspect in science and education, yet its integration in biology education remains underexplored. Competence in data literacy is essential for biology education students to conduct learning evaluation analyses that increase student learning performance. Grounded in Bandura's social cognitive theory, this study investigates biology education undergraduates' self-efficacy in data literacy during a statistics course. A convenience sample of 70 students completed a questionnaire assessing confidence (7 items), experience (4 items), and attitude (4 items) of data on a 5-point Likert scale. Additionally, students completed a data interpretation test (with biology-based data scenarios), rated their confidence on a 4-point Likert scale, and provided reasoning. Result revealed high levels of self-efficacy across all domains, consistent confidence during data interpretation tasks, and positive attitudes towards learning with data. The study contributes conceptually by linking perceived confidence and performance-based evidence of data literacy, highlighting implications for improving quantitative reasoning in biology education.

Keywords: Data Literacy, Self-Efficacy, Statistics

INTRODUCTION

Data is increasingly disseminated through social media and educational platforms in the form of numbers, symbols, tables, and diagrams in the digital era. This rapid spread demands robust data literacy skills, including reading data, identifying patterns, analyzing data, making predictions, and making decisions from the data collection (Gittens, 2015; D'Ignazio & Bhargava, 2016; Giese et al., 2020). Data literacy skills are more than just the technical transformation of data from one form to another; they also include a literacy process, which is visualizing data such that it becomes information that is meaningful and easy to understand (D'Ignazio, 2022).

As defined by Gibson & Mourad (2018), data literacy comprises five key components: (1) conceptual framework: knowledge and understanding of the uses data; (2) data collection: perform data exploration to identify useful data; (3) data management: assess data organization methods and tools to organize data; (4) data interpretation: create meaningful table and graphical representation of data; (5) data evaluation: thinks critically to evaluate results of analysis and compares results of analysis with other findings (Bonikowska et al., 2019). Furthermore, data literacy competencies should be developed, as they are essential in all sectors. This includes not only the use of information technology but also the analysis and interpretation of large data sets (Coners et al., 2024). In biology education specifically, data literacy supports evidence-based teaching and scientific thinking.

Data literacy entails more than just collecting and transforming data; it also includes understanding word problems involving the use of statistics. Despite growing emphasis, students often face challenges in applying statistical concepts to new data scenarios (Hall & Vance, 2010), stemming from insufficient experience and low statistical self-efficacy (Kaufmann et al., 2022). Self-efficacy, as conceptualized in Bandura's social cognitive theory, refers to one's belief in their capacity to accomplish tasks and achieve goals (Hall & Vance, 2010; Abu Bakar & Ismail, 2020). Self-efficacy plays a central role in metacognitive processes, influencing persistence, motivation (engagement), and performance (Ouweneel et al., 2013). Students with high self-efficacy are more likely to persevere through analytical tasks and engage in data-driven problem solving. When students engage with data literacy and conduct research, they move beyond the threshold of being knowledge consumers to becoming knowledge creators (Burress, 2022).

In biology education, data literacy and self-efficacy are interdependent. Competence in data literacy enhances pedagogical content knowledge (Hoogland et al., 2016; Kickbusch et al., 2022), allowing future teachers to transform and communicate data in ways that support students' learning (Green et al., 2016; Mandinach & Gummer, 2016). However, Milton et al. (2007) found that some teachers believe they do not need strong data literacy skills for teaching. This is often due to a lack of formal training during their coursework.

Although research has examined either data literacy and self-efficacy separately, few studies have integrated both constructs to explore how biology education students engage with data-based tasks. This study addressed this gap by assessing students' self-efficacy, experiences, and attitudes toward data literacy during a statistics course. It also captures students' confidence and justification during a data interpretation test, providing insight into their metacognitive engagement. As future educators and researchers, biology education undergraduates must be equipped not only with technical proficiency but also with the confidence to interpret and communicate data accurately. By investigating both perceived self-efficacy and demonstrated confidence in data tasks, this study provides novel insight into preparing data-literate biology educators.

METHOD

The participants represent a convenience sample of 70 biology education undergraduate students during their Statistics course. Students take the course in their second semester. The course teaches descriptive statistics, producing data and sampling, frequency distribution, probability, confidence level, and tests of significance. Students participated in hands-on activities using both primary scientific literature and secondary datasets, with an emphasis on analyzing experimental designs and statistical outcomes to foster data literacy skills (Tong et al., 2022).

The data used in this study are students' self-efficacy in the data literacy questionnaire and the data interpretation test. All students were tested in class at the end of the term. The data literacy questionnaire was adapted from Öz & Özdemir (2022), consisted of 15 items divided into three subscales, including students' confidence (7 items), experience (4 items), and attitude (4 items). All items were rated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The level of students' confidence was interpreted using a guideline adapted from Almohtadi & Aldarabah (2021), as shown in Table 1. Cronbach's alpha was used to examine the internal consistency of each subscale.

Table 1. The Confidence Levels of The Mean Scores on a 5-point Likert Scale of Students' Self-Efficacy

Mean Range	Level
1.0 – 1.5	Very low
1.6 – 2.5	Low
2.6 – 3.5	Moderate
3.6 – 4.5	High
4.6 – 5.0	Very high

To provide further evidence about students' confidence in data literacy, a custom data interpretation test was developed consisting of four items, as shown in Table 3. Each item required students to write an explanation of their answer and indicate their confidence level on a 4-point Likert scale, from 1 (not confident at all) to 4 (completely confident) (see Table 2).

Table 2. The Confidence Levels of The Mean Scores on a 4-point Likert Scale of the Data Interpretation Test

Mean Range	Level
1.00 – 1.49	Not confident at all
1.50 – 2.49	Not confident
2.50 – 3.49	Confident
3.50 – 4.00	Completely confident

Descriptive statistics were used to summarize students' self-efficacy, experience, and attitude toward the data. In the relationship between perceived self-efficacy and performance, we use Spearman's rank correlation to analyze the association between confidence level and performance scores on the test. All data were analyzed using SPSS with a significance level of $\alpha = 0.05$.

BIODIDAKTIKA: Jurnal Biologi dan Pembelajarannya, Vol.20, No.2, 2025, pp. 126-136
e-ISSN 2527-4562. DOI. 10.62870/biodidaktika.v20i2.31522

Table 3. Rubrics and Scoring Guide for Data Interpretation Test

Topic	Items	Content	Rubrics	
			Score	Indicator
Case 1: Biology	4	Sugar's effect on media pH and pollen tube length during pollination	0 1 2 3	Incorrect answer Inaccurately answer Partially answer Correct answer
Case 2: Biology education	4	Interpretation of scoring rubrics on students' critical thinking skills (Reynders et al., 2020)	0 1 2 3	Incorrect answer Inaccurately answer Partially answer Correct answer
Case 3: Frequency distribution	2	Create a table of frequency distribution along with descriptive statistics using the data provided	0 1 2 3	Incorrect answer/blank Inaccurately answer Partially answer Correct answer
Case 4: Test of significance	3	Determine the hypothesis, data analysis method, and data interpretation based on the pretest and post-test findings	0 1 2 3	Incorrect answer/blank Inaccurately answer Partially answer Correct answer

RESULTS AND DISCUSSION

Students' confidence in data literacy reflected students' ability to recognize, access, interpret, and evaluate various forms of data. In addition to quantitative results from the self-efficacy questionnaire, examples of students responses are illustrated in Figure 1. The findings indicate that a majority of students (57% scored 4; 31,4% scored 5) were able to differentiate between qualitative and quantitative data. Furthermore, 60% of students demonstrated an understanding of how to retrieve relevant information and approximately 64,3% were capable of extracting data and drawing valid conclusions.

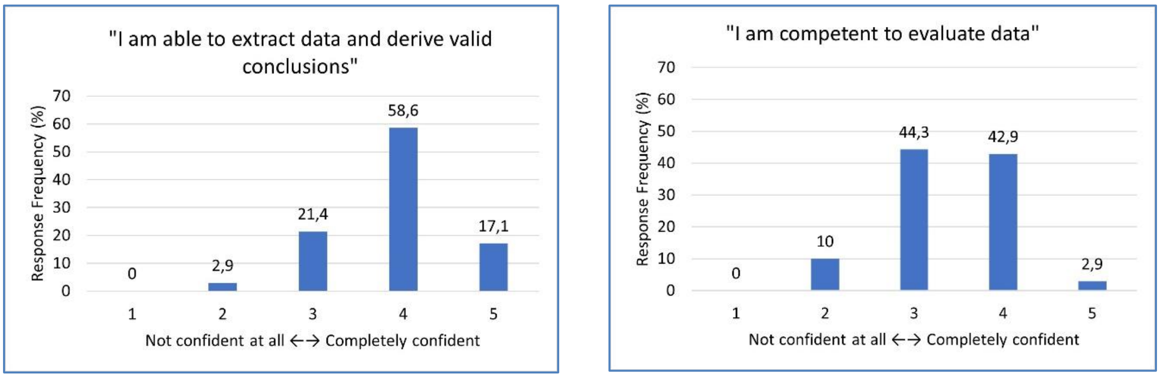

Figure 1. Students' Responses on Confidence in Data Literacy

Students' experience with data literacy was primarily demonstrated through their familiarity with data processing tools (Figure 2). Most students reported being comfortable using Microsoft Excel and SPSS. Additionally, they understood fundamental statistical concepts such as mean, median, and mode and were able to construct appropriate data visualizations including tables, graphs, and diagrams. Approximately 60% of students (score 4) indicated confidence in discussing and providing feedback on data analysis tasks.

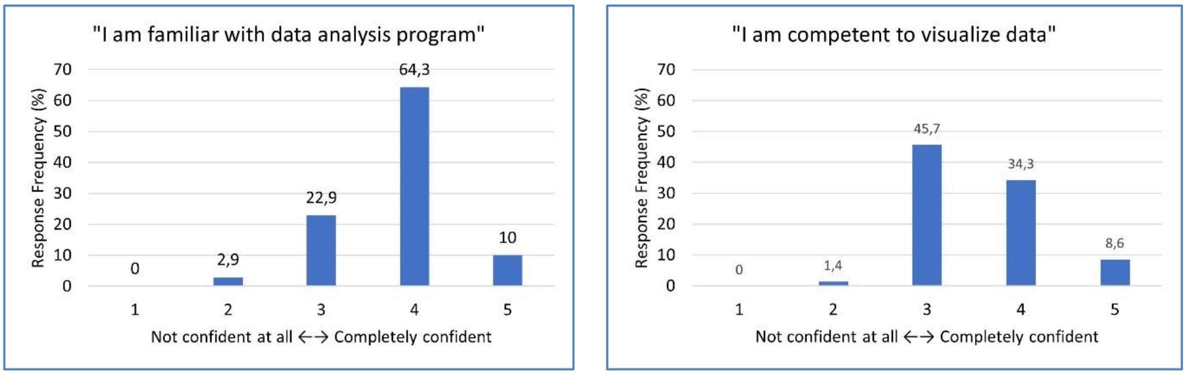

Figure 2. Students' Responses on Experience in Data Literacy

Students expressed positive attitudes toward data literacy, emphasizing its importance in organizing, documenting, and evaluating information. They recognized that data analysis skills enhance their pedagogical content knowledge as prospective biology teachers. Moreover, a significant proportion of students (55,7% scored 4; 32,9% scored 5) acknowledged that data literacy contributes to the development of critical thinking skills, enabling them to scrutinize information more carefully and avoid misinformation. Engagement with primary scientific literature and the use of data analysis tools provided contextualized learning experiences that supported the development of critical thinking (Ye & Jin, 2024).

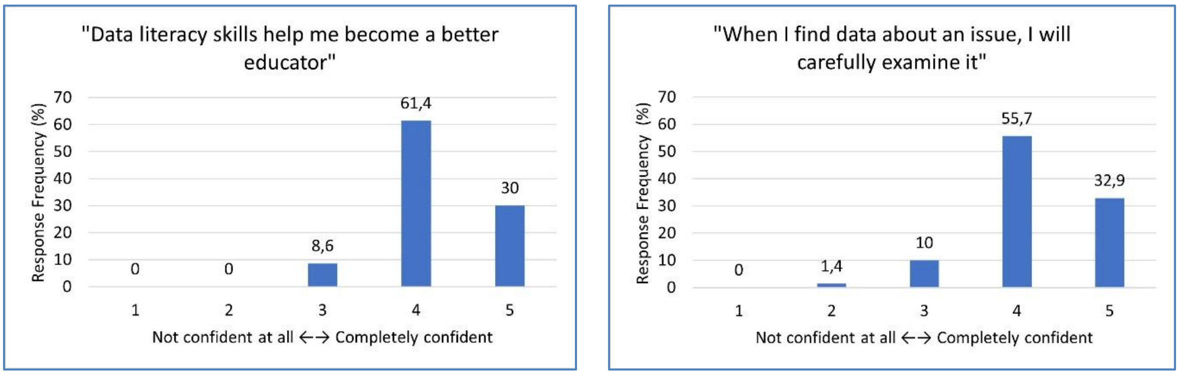

Figure 3. Students' Responses on Attitude in Data Literacy

Table 4 summarizes the questionnaire results, indicating that students had high levels of self-efficacy in all three aspects: self-efficacy (mean = 3,86), experience (mean = 3,78), and attitude (mean = 4,22). These findings are consistent with performance outcomes, where 70,88% of students expressed confidence in successfully solving data interpretation test. Previous studies highlighted that students' self-efficacy is enhanced when they engage with real-world data and contextual tasks (Kjelvik & Schultheis, 2019; Simon et al., 2022).

Table 4. Confidence Levels of Students' Self-Efficacy Results

Aspects	Items	Cronbach's Alpha	Mean ± SD	Level
Confidence	7	0,78	3,86 ± 0,42	High
Experience	4	0,73	3,78 ± 0,48	High
Attitude	4	0,76	4,22 ± 0,48	High

Table 5 shows that the students' performance on the data interpretation test further supports the questionnaire findings. In Cases 1 and 2 (biological and educational contexts) on the test, nearly 60% of students provided accurate and confident responses. These items assessed the interpretation of data presented in graphs and tables. Students who successfully interpreted these visualizations demonstrated an understanding of scientific content, even without direct or prior exposure to the specific concepts. Such performance reflects the positive impact of exposure to primary scientific data on students' data literacy (Brearley et al., 2023).

In Cases 3 and 4 (frequency distribution and test of significance), a larger proportion of students (approximately 80%) responded both accurately and confidently. However, nearly 30% of students who answered incorrectly reported being confident in their responses, suggesting that while confidence was generally high, some students overestimated their understanding, particularly in tasks involving graph interpretation. Nevertheless, they demonstrated high self-efficacy and a positive attitude toward learning. This weakness can be addressed through practice and exposure to data analysis (Stanford et al., 2015). In the learning process, the development of skills in understanding and analyzing is influences by both the input (content) and personal characteristics, such as motivation, creativity, and intellect (Dietrich et al., 2015).

Table 5. Confidence Level in Answering Data Interpretation Test

Aspect	Correct Answer		Incorrect Answer	
	% of students	Confidence Level	% of students	Confidence Level
Case 1	63,92	3,19 (Confident)	36,07	3,03 (Confident)
Case 2	55,72	3,43 (Confident)	44,28	3,24 (Confident)
Case 3	81,1	3,43 (Confident)	18,9	2,5 (Confident)
Case 4	82,8	3,00 (Confident)	17,2	3,00 (Confident)
Mean	70,88	3,26 (Confident)	29,11	2,94 (Confident)

The confidence levels demonstrated by students in answering data-related questions (Case 3) were consistent with the high proportion of correct responses they provided. For instance, approximately 53,6% of students achieved a perfect score (score = 3) on questions related to frequency distribution, while 4,4% of students answered incorrectly (score = 0). A similar pattern was observed in Case 4, which assessed students' understanding of significance testing. In this case, around 47,1% of students answered the questions correctly (score = 3), 35,7% responded partially correctly, and only 4,3% answered incorrectly or left the questions blank. These results suggest that, although students show strong performance in descriptive statistics, additional practice is needed to improve their understanding of significance testing (Setiawan & Sukoco, 2021). The findings indicate that students have a stronger grasp of descriptive statistical concepts compared to inferential statistical procedures.

Table 6. Spearman's Rank Correlation Results

Variable	*M*	*SD*	1	2
Data Interpretation Score	1,94	0,49	0,35**	—
Confidence Level	3,23	0,44	—	0,35**

N = 70. Spearman's rho coefficients are shown. *M* = Mean; *SD* = Standard Deviation. $p < 0,01$ (**)

Table 6 shows that the Spearman correlation coefficient (p) between the two variables was 0,35 and this relationship was statistically significant at the 0,01 level ($p < 0,01$). This result indicates a moderate positive correlation between students' confidence and their performance on the data interpretation test. This implies students who reported higher confidence tended to achieve higher scores on the test. Although the correlation is not strong. It is meaningful and suggest that self-perceived confidence is related to data interpretation ability. The results align with the research conducted in engineering students, which states that encouraging academic self-efficacy can enhance students' cognitive performance (Alias & Hafir, 2009).

The findings underscore the importance of integrating contextual data tasks into statistics and biology education courses to strengthen students' data interpretation skills (Gordon & Nicholas, 2010). Prior work has shown that students with greater confidence and familiarity with data tend to exhibit stronger self-efficacy in data-driven decision-making (Mendez-Carbajo, 2020). However, the study is limited by its small sample size and convenience sampling approach, which may reduce the generalizability of the findings. Additionally, self-efficacy may fluctuate throughout a course and should not be assumed to remain stable of all students (Ouweneel et al., 2013). Future studies should consider longitudinal approaches and incorporate additional performance measures to further validate the role of self-efficacy in developing data literacy among biology education students.

CONCLUSION

This study revealed that biology education students demonstrate high levels of self-efficacy in data literacy, as reflected in their confidence, experience, and attitudes toward working with data. Students were able to distinguish between qualitative and quantitative data, retrieve relevant information, and extract meaningful conclusions. They also reported familiarity with statistical tools such as Microsoft Excel and SPSS and showed competency in constructing data visualizations and engaging in analytical discussions. Performance in the data interpretation test aligned with self-reported confidence levels, particularly in descriptive statistics and frequency distribution. However, challenges remained in inferential statistics, where some students exhibited overconfidence despite incorrect responses. This highlights the importance of practice in more advanced statistical reasoning. The moderate, statistically significant correlation between students' confidence and performance further underscores the relationship between self-efficacy and data literacy competence.

The findings suggest that integrating contextual, data-drive tasks into the curriculum enhances both skill acquisition and students' belief in their ability to interpret data. Nonetheless, the study's limitations including a small convenience-based sample and the cross-sectional design, reestrict broader generalization. Future research should adopt longitudinal approaches and pre-post assessments to track the development of self-efficacy and performance over time, ensuring that instruction not only builds competence but also sustains students' confidence.

REFERENCES

Abu Bakar, M. A., & Ismail, N. (2020). Exploring Metacognitive Regulation and Students' Interaction in Mathematics Learning: an Analysis of Needs To Enhance Students' Mastery. *Humanities & Social Sciences Reviews*, *8*(2), 67–74. https://doi.org/10.18510/hssr.2020.82e07

Alias, M., & Hafir, N. A. H. M. (2009). The relationship between academic self-confidence and cognitive performance among engineering students. *2009 Research in Engineering Education Symposium, REES 2009*, 1–6.

Almohtadi, R. M., & Aldarabah, I. T. (2021). University Students' Attitudes toward the Formal Integration of Facebook in their Education: Investigation Guided by Rogers' Attributes of Innovation. *World Journal of Education*, *11*(1), 20. https://doi.org/10.5430/wje.v11n1p20

Bonikowska, A., Sanmartin, C., & Frenette, M. (2019). Data Literacy: What it is and how to Measure it in the Public Service. *European Journal of Communication*, *3*(11), 299. https://www.lrsd.net/Documents/Data Literacy What it is and How to Measure it in the Public Service.pdf

Brearley, A. M., Rott, K. W., & Le, L. J. (2023). A Biostatistical Literacy Course: Teaching Medical and Public Health Professionals to Read and Interpret Statistics in the Published Literature. *Journal of Statistics and Data Science Education*, *31*(3), 286–294. https://doi.org/10.1080/26939169.2023.2165987

Burress, T. (2022). Data Literacy Practices of Students Conducting Undergraduate Research. *College and Research Libraries*, *83*(3), 434–451. https://doi.org/10.5860/crl.83.3.434

Coners, A., Matthies, B., Vollenberg, C., & Koch, J. (2024). Data Skills for Everyone! (?)–An Approach to Assessing the Integration of Data Literacy and Data Science Competencies in Higher Education. *Journal of Statistics and Data Science Education*, *0*(0), 1–29. https://doi.org/10.1080/26939169.2024.2334408

D'Ignazio, C. (2022). Creative data literacy. *Information Design Journal*, *23*(1), 6–18. https://doi.org/10.1075/idj.23.1.03dig

D'Ignazio, C., & Bhargava, R. (2016). DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. *The Journal of Community Informatics*, *12*(3), 83–107. https://doi.org/10.15353/joci.v12i3.3280

Dietrich, H., Zhang, Y., Klopp, E., Brünken, R., Krause, U. M., Spinath, F. M., Stark, R., & Spinath, B. (2015). Scientific competencies in the social sciences. *Psychology Learning and Teaching*, *14*(2), 115–130. https://doi.org/10.1177/1475725715592287

Gibson, J. P., & Mourad, T. (2018). The growing importance of data literacy in life science education. *American Journal of Botany*, *105*(12), 1953–1956. https://doi.org/10.1002/ajb2.1195

Giese, T. G., Wende, M., Bulut, S., & Anderl, R. (2020). Introduction of data literacy in the undergraduate engineering curriculum. *IEEE Global Engineering Education Conference, EDUCON*, *2020-April*, 1237–1245. https://doi.org/10.1109/EDUCON45650.2020.9125212

Gittens, C. A. (2015). Assessing Numeracy in the Upper Elementary and Middle School Years. *Numeracy*, *8*(1). https://doi.org/10.5038/1936-4660.8.1.3

Gordon, S., & Nicholas, J. (2010). Teaching with examples and statistical literacy: Views from teachers in statistics service courses. *International Journal of Innovation in Science and Mathematics Education*, *18*(1), 14–25.

Green, J. L., Sarah Schmitt-Wilson, Tena Versland, Kelting-Gibson, L., & Gustave E. Nollmeyer. (2016). Teachers and Data Literacy: A Blueprint for Professional Development to Foster Data Driven Decision Making. *Journal of Continuing Education and Professional Development*, *3*(1), 14–32.

Hall, S., & Vance, E. A. (2010). Improving self-efficacy in statistics: Role of self-explanation & feedback. *Journal of Statistics Education*, *18*(3), 1–22. https://doi.org/10.1080/10691898.2010.11889583

Hoogland, I., Schildkamp, K., van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education*, *60*, 377–386. https://doi.org/10.1016/j.tate.2016.07.012

Kaufmann, L., Ninaus, M., Weiss, E. M., Gruber, W., & Wood, G. (n.d.). *29 Annals of the New York Academy of Sciences - 2022 - Kaufmann - Self-efficacy matters Influence of students perceived.pdf*.

Kickbusch, S., Dawes, L., Kelly, N., & Nickels, K. (2022). Developing Mathematics and

Science Teachers' Ability to Design for Active Learning: A Design-based Research Study. *Australian Journal of Teacher Education*, *47*(9), 80–99. https://doi.org/10.14221/ajte.2022v47n9.5

Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE Life Sciences Education*, *18*(2), 1–8. https://doi.org/10.1187/cbe.18-02-0023

Mandinach, E. B., & Gummer, E. S. (2016). Every teacher should succeed with data literacy. *Phi Delta Kappan*, *97*(8), 43–46. https://doi.org/10.1177/0031721716647018

Mendez-Carbajo, D. (2020). Baseline Competency and Student Self-efficacy in Data Literacy: Evidence from an Online Module. *Journal of Business and Finance Librarianship*, *25*(3–4), 230–243. https://doi.org/10.1080/08963568.2020.1847551

Milton, M., Rohl, M., & House, H. (2007). Secondary Beginning Teacher's Preparedness to Teach Literacy and Numeracy: A Survey. *Australian Journal of Teacher Education*, *32*(2). https://doi.org/10.14221/ajte.2007v32n2.4

Ouweneel, E., Schaufeli, W. B., & Le Blanc, P. M. (2013). Believe, and you will achieve: Changes over time in self-efficacy, engagement, and performance. *Applied Psychology: Health and Well-Being*, *5*(2), 225–247. https://doi.org/10.1111/aphw.12008

ÖZ, S., & ÖZDEMİR, A. (2022). Validity and Reliability Study on The Development of Data Literacy Scale for Educators. *International Journal of Contemporary Educational Research*, *9*(3), 649–661. https://doi.org/10.33200/ijcer.1079774

Reynders, G., Lantz, J., Ruder, S. M., Stanford, C. L., & Cole, R. S. (2020). Rubrics to assess critical thinking and information processing in undergraduate STEM courses. *International Journal of STEM Education*, *7*(1). https://doi.org/10.1186/s40594-020-00208-5

Setiawan, E. P., & Sukoco, H. (2021). Exploring first year university students' statistical literacy: A case on describing and visualizing data. *Journal on Mathematics Education*, *12*(3), 427–448. https://doi.org/10.22342/JME.12.3.13202.427-448

Simon, M., Prather, E., Rosenthal, I., Cassidy, M., Hammerman, J., & Trouille, L. (2022). New Curriculum Development Model for Improving Undergraduate Students' Data Literacy and Self-Efficacy in Online Astronomy Classrooms. *Astronomy Education Journal*, *02*(1), 1–13. https://doi.org/10.32374/aej.2022.2.1.043ra

Stanford, J. S., Rocheleau, S. E., Smith, K. P. W., & Mohan, J. (2015). Early undergraduate research experiences lead to similar learning gains for STEM and Non-STEM undergraduates. *Studies in Higher Education*, *42*(1), 115–129. https://doi.org/10.1080/03075079.2015.1035248

Tong, L., White, B. J. G., & Singh, J. (2022). Bridging statistics and life sciences undergraduate education. *Journal of Biological Education*, *00*(00), 1–13. https://doi.org/10.1080/00219266.2022.2118810

Ye, L., & Jin, Y. (2024). Teaching Students to Read COVID-19 Journal Articles in Statistics

Courses. *Journal of Statistics and Data Science Education*, *32*(2), 143–149. https://doi.org/10.1080/26939169.2024.2302185