Evaluating the Effectiveness of Spreadsheet-Based Instruction on High School Students' Data Literacy: A Pre-test–Post-test Study

Tubagus Robbi Megantara[1*], Rizki Apriva Hidayana[2], Nenden Siti Nurkholipah[3], Rika Amelia[4], Abdul Gazir Syarifudin[5]

[1,2,3,4,5] *Department of Mathematics, Faculty Mathematics and Natural Sciences, Universitas Kebangsaan Republik Indonesia, Bandung, Indonesia*

Corresponding author email: tubagusrobbimegantara@mipa.ukri.ac.id

Abstract

Data literacy has become an essential competency for students in the digital era, particularly at the secondary education level where foundational numeracy skills are developed. Spreadsheet software such as Microsoft Excel offers an accessible platform for integrating mathematical concepts with practical data analysis and visualization. This study aims to quantitatively evaluate the effectiveness of a spreadsheet-based instructional intervention in improving high school students' data literacy skills. A quasi-experimental one-group pre-test–post-test design was employed involving 12 senior high school students from SMAS Sebelas Maret, Bandung. The instructional intervention consisted of three integrated modules: Introduction to Excel and Data Literacy, Data Analysis and Logical Functions, and Data Visualization. Students' competencies were assessed before and after the intervention using a structured test aligned with these modules. Data were analyzed using descriptive statistics, paired-sample t-tests, and effect size estimation (Cohen's d). The results indicate a statistically significant improvement in overall performance from pre-test to post-test ($p < 0.05$), accompanied by a large effect size ($d = -0.96$), demonstrating substantial practical impact. The strongest learning gains were observed in data visualization skills, while data analysis and logical functions showed more moderate improvement. These findings suggest that spreadsheet-based instruction is an effective approach for enhancing data literacy and applied numeracy skills among high school students.

Keywords: data literacy, spreadsheet-based learning, numeracy education, pre-test–post-test, Microsoft Excel

1. Introduction

In recent years, data literacy has emerged as a critical component of numeracy education, particularly in response to the increasing dominance of data-driven decision-making in academic, professional, and societal contexts. Data literacy encompasses the ability to collect, analyze, interpret, and communicate data effectively, enabling individuals to make informed decisions based on empirical evidence (Bhargava, 2019; Griffin & Holcomb, 2023). Consequently, secondary education plays a crucial role in equipping students with foundational data competencies that support both further study and workforce readiness (Jeong & Lee, 2024).

The importance of data literacy is not limited to academic success but extends directly to employability and participation in a modern digital economy. Many industries increasingly expect workers to possess basic data handling and analytical skills, regardless of their specific roles (Formby et al., 2017; Miloradov et al., 2022). As a result, educational institutions are encouraged to integrate practical data-related skills into mathematics and technology-enhanced learning environments (Leon-Urrutia et al., 2022).

One instructional approach that has gained considerable attention is the integration of spreadsheet software into mathematics and data-oriented learning. Spreadsheet applications such as Microsoft Excel provide a flexible and accessible platform for numerical computation, logical analysis, and data visualization. Excel has been widely recognized as an industry-standard tool in business and management contexts (Liengme, 2000) and has demonstrated effectiveness in supporting students' understanding of statistical and numerical concepts in educational settings (Simaremare & Siregar, 2024). Furthermore, the use of spreadsheet functions allows learners to explore mathematical relationships while engaging directly with real-world data (Mustafy & Rahman, 2024; Talib & Mohamad, 2020).

Despite these advantages, prior studies on spreadsheet-based learning often emphasize program descriptions or qualitative outcomes, with limited emphasis on quantitative evaluation of learning effectiveness. In the context of community-based educational initiatives, empirical evidence demonstrating measurable learning gains remains

relatively scarce. This limitation highlights the need for instructional studies that apply inferential statistical methods to assess the actual impact of spreadsheet-assisted instruction on students' data literacy and numeracy skills.

This study addresses this gap by conducting a quantitative evaluation of a spreadsheet-based instructional intervention implemented for high school students within a school community setting. Using a one-group pre-test–post-test design, the study examines changes in students' competencies related to data processing, logical analysis, and data visualization. By employing descriptive statistics, paired-sample *t*-tests, and effect size estimation, this research aims to provide robust evidence regarding the effectiveness of spreadsheet-assisted learning as an instructional strategy.

The findings of this study are expected to contribute to the literature on community service–oriented educational research by demonstrating how technology-supported learning interventions can meaningfully enhance data literacy and applied numeracy skills at the secondary school level.

2. Materials and Methods

2.1. Research Design

This study employed a quasi-experimental one-group pre-test–post-test design to evaluate the effectiveness of a spreadsheet-based instructional intervention on students' data literacy and numeracy skills. This design enables the measurement of learning gains by comparing participants' performance before and after the intervention within the same group. The approach is commonly used in educational and community-based research where random assignment and control groups are not feasible.

2.2. Participants

The participants were students from SMAS Sebelas Maret, Bandung, who voluntarily took part in the program. An initial orientation session was attended by 31 students. From this group, 13 students registered for the technical instructional phase. One participant withdrew due to personal reasons, resulting in a final sample of 12 students who completed the entire instructional sequence and both assessment stages. These 12 participants constituted the dataset used for quantitative analysis.

2.3. Instructional Materials

The instructional materials were designed to integrate mathematical concepts with practical spreadsheet applications using Microsoft Excel. The content was organized into three main instructional modules:
1. Introduction to Excel & Data Literacy, focusing on spreadsheet navigation, data entry, basic arithmetic operations, and fundamental concepts of data representation.
2. Data Analysis & Logical Functions, emphasizing conditional logic and basic analytical reasoning through functions such as *IF*, *SUMIF*, *COUNTIF*, and *AVERAGEIF*.
3. Data Visualization, covering the construction and interpretation of charts (e.g., bar charts, line charts, and pie charts) and the principles of selecting appropriate visual representations for different data types.

All materials were delivered through hands-on activities to encourage active learning and direct engagement with real data.

2.4. Instruments

Students' learning outcomes were assessed using a structured test administered both before (pre-test) and after (post-test) the instructional intervention. The test consisted of items aligned with the three instructional modules and measured students' ability to perform basic data processing, apply logical functions, and interpret data visualizations. Scores were calculated for each category as well as for the total test score.

2.5. Data Collection Procedure

The pre-test was administered immediately prior to the instructional intervention to establish baseline competency levels. After the completion of all instructional modules, the post-test was administered using the same assessment structure. This procedure ensured consistency and allowed for direct comparison of performance across the two testing occasions.

2.6. Data Analysis

Descriptive statistics, including mean scores, standard deviations, minimum values, and maximum values, were calculated to summarize participants' performance on the pre-test and post-test. To evaluate the statistical significance

of observed learning gains, paired-sample t-tests were conducted for the total score and for each instructional category. Statistical significance was assessed at the $\alpha = 0.05$ level.

In addition to significance testing, effect sizes were calculated using Cohen's *d* to assess the practical magnitude of the observed differences. Effect size interpretation followed conventional benchmarks, where values around 0.2 are considered small, around 0.5 medium, and 0.8 or above large.

3. Results and Discussion

3.1. Training Program

The learning intervention was implemented through a structured sequence of instructional activities targeting high school students from SMAS Sebelas Maret, Bandung. The program began with an initial orientation session attended by 31 students, which served to introduce the relevance of mathematical literacy and its applications. Following this phase, a subset of students voluntarily continued to a focused spreadsheet-based instructional program. After one participant withdrew due to personal reasons, a total of 12 students completed the full learning sequence and were included in the quantitative analysis.

The instructional component was delivered through three integrated modules. The first module introduced fundamental spreadsheet operations, emphasizing data entry, basic arithmetic computations, and navigation of the spreadsheet environment. The second module focused on elementary data analysis using logical functions, enabling students to perform conditional calculations and data categorization. The final module addressed data visualization, guiding students to represent numerical information using appropriate graphical formats and to interpret data patterns effectively.

The impact of this instructional sequence was evaluated using a pretest–posttest design. The following subsection presents a statistical analysis of students' performance before and after the intervention, providing quantitative evidence of learning gains attributable to the spreadsheet-assisted instruction.

### 3.2. Analysis of Students' Competency Improvement Based on Pre-test and Post-test Results

The improvement in students' competency in using Microsoft Excel for basic data processing, analysis, and visualization is reflected in the comparison between pre-test and post-test results.

Table 1: Summary of Descriptive Statistics (Pre-test and Post-test)

| Data Category	Test	Mean	Standard Deviation	Min	Max
Total Correct Answers	Pre-test	6.83	3.49	1	12
	Post-test	9.75	2.49	5	13
Introduction to Excel & Data Literacy	Pre-test	2.17	1.53	1	5
	Post-test	3.42	1.16	2	5
Data Analysis & Logical Functions	Pre-test	2.25	1.22	0	4
	Post-test	2.75	1.29	1	5
Data Visualization	Pre-test	2.42	1.98	0	5
	Post-test	3.58	1.08	2	5

The descriptive statistics consistently indicate improved performance from the pre-test to the post-test. The mean total score increased by 2.92 points, from 6.83 to 9.75, suggesting a substantial improvement in students' understanding following the instructional intervention. Performance gains were observed across all content categories, with the largest mean increases found in Data Visualization (1.16 points) and Introduction to Excel & Data Literacy (1.25 points).

In addition, the post-test results exhibit lower score variability, as reflected by the reduced standard deviations. This pattern suggests that the intervention not only enhanced overall performance but also contributed to a more homogeneous level of understanding among participants.
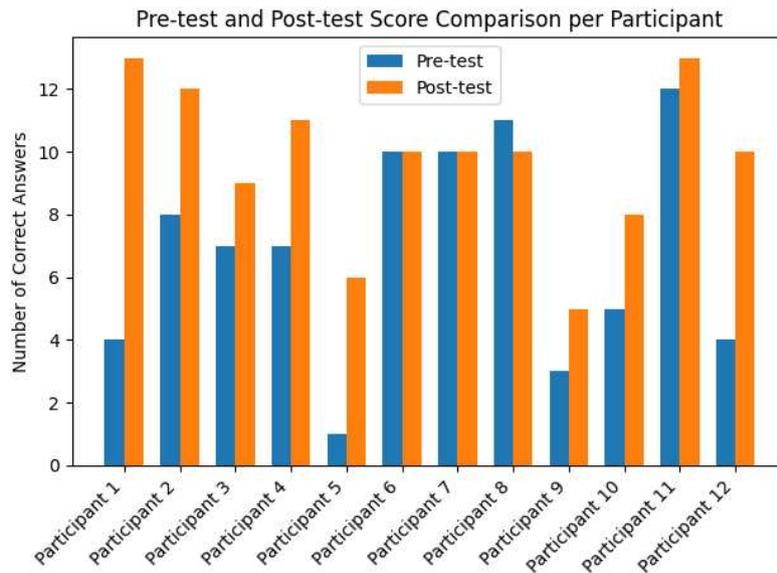
Figure 1: Comparison of Pre-test and Post-test Scores per Participant

This comparison chart clearly illustrates the effectiveness of the instructional intervention across 12 participants, with the majority showing an increase in the number of correct answers from the pre-test to the post-test. The most pronounced improvement is observed for Participant 1, whose score increased substantially from 4 to 13. Meanwhile, several participants, such as Participants 6 and 7, maintained their already high performance levels. Although a minor score decrease is observed for Participant 8, the overall trend strongly indicates that the instructional program was effective in enhancing participants' understanding.
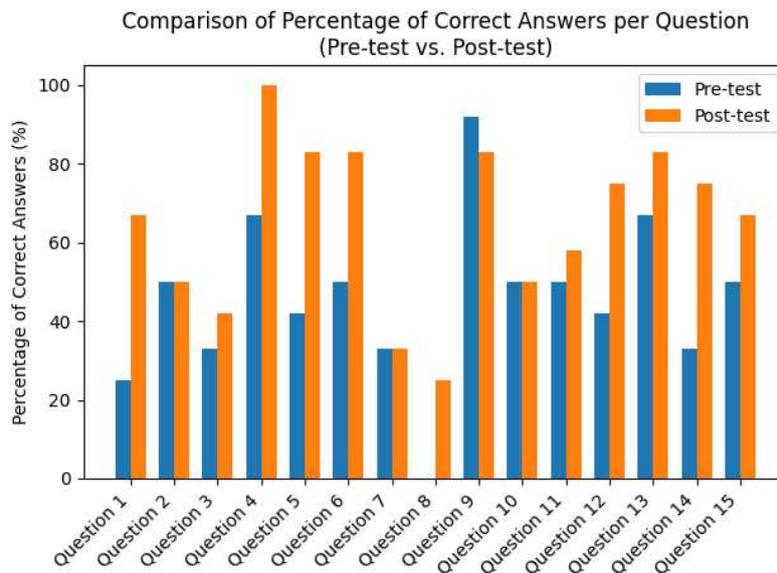
Figure 2: Comparison of the Percentage of Correct Answers per Question (Pre-test and Post-test)

This figure presents a detailed analysis of instructional effectiveness at the item level by comparing the percentage of correct responses between the pre-test and post-test. Overall, the trend indicates improved understanding across most test items, with several particularly notable gains. For instance, Question 4 achieved a perfect post-test accuracy of 100%, while Question 14 showed a substantial increase from 33% in the pre-test to 83% in the post-test.

However, the results also reveal several important anomalies. Question 9 experienced a slight decrease in performance, whereas Questions 2, 7, and 10 showed no change between the pre-test and post-test results, suggesting that the instructional material associated with these items may have been less effective or that the questions themselves may have been ambiguous. Additionally, Question 8, despite improving from 0%, remained the most challenging item, with only 25% correct responses in the post-test, indicating an area requiring the greatest instructional refinement.

Table 2: Summary of Significance Testing (Paired t-test)

Data Category	t-Statistic	p-Value
Total Score	-3.514407	0.004846
Introduction to Excel & Data Literacy	-2.111743	0.058400
Data Analysis & Logical Functions	-1.317893	0.214326
Data Visualization	-2.128241	0.056753

The increase in total scores from the pre-test to the post-test was statistically significant ($p = 0.0048$), which is below the commonly accepted significance threshold of $\alpha = 0.05$. This provides strong evidence to reject the null hypothesis of no difference and to conclude that the instructional intervention had a meaningful and positive overall impact on participants' understanding.

At the category level, however, none of the observed improvements reached conventional statistical significance. The Introduction to Excel & Data Literacy ($p = 0.058$) and Data Visualization ($p = 0.057$) categories yielded marginally significant results. This outcome is likely attributable to reduced statistical power when analyzing smaller subscale scores. In other words, despite the presence of clear improvement trends, the sample size ($n = 12$) may have been insufficient to detect statistically significant effects at the category level. The Data Analysis & Logical Functions category ($p = 0.214$) did not exhibit a significant change.
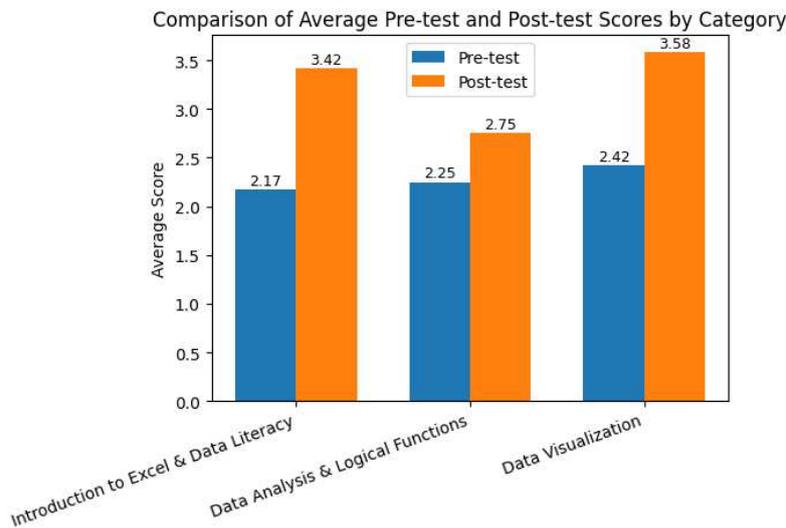
Figure 3: Comparison of Mean Pre-test and Post-test Scores by Category

Based on the comparison of mean scores across categories, the results indicate that the instructional intervention successfully improved participants' understanding in all content areas. The most substantial increase in mean score occurred in Introduction to Excel & Data Literacy, which rose from 2.17 to 3.42. The Data Visualization category achieved the highest post-test mean score (3.58) and also demonstrated a strong improvement from an already relatively high pre-test baseline. In contrast, Data Analysis & Logical Functions exhibited a smaller increase compared to the other two categories, suggesting that this topic may have been the most challenging for participants.

Table 3: Summary of Effect Size (Cohen's *d*)

Category	Cohen's *d*
Total Score	-0.962747
Introduction to Excel & Data Literacy	-0.920338
Data Analysis & Logical Functions	-0.399275
Data Visualization	-0.732338

The effect size analysis provides insight into the practical magnitude of the observed changes. The Cohen's *d* value for the total score ($d = -0.96$) is conventionally classified as a very large effect, indicating that the difference between pre-test and post-test scores is not only statistically significant but also highly meaningful in practical terms. Large effect sizes were also observed for Introduction to Excel & Data Literacy ($d = -0.92$) and Data Visualization ($d = -0.73$). These findings reinforce the notion that, although some category-level improvements did not reach

conventional statistical significance—likely due to the limited sample size—the intervention produced strong practical impacts in these areas.

In contrast, Data Analysis & Logical Functions exhibited a small-to-moderate effect size ($d = -0.40$), which is consistent with the paired t-test results and suggests that the impact of the intervention on this category was more limited compared to the others.

Overall, these findings strongly support the hypothesis that the instructional program effectively enhanced participants' knowledge of Microsoft Excel. The observed improvements were statistically significant at the overall level and demonstrated large effect sizes, indicating a substantial practical impact.

4. Conclussion

This study demonstrates that the spreadsheet-based instructional intervention was effective in enhancing high school students' data literacy and basic data analysis skills. The primary objective of improving students' competence in data processing and analysis was successfully achieved, as evidenced by a statistically significant increase in total scores from the pre-test to the post-test ($p < 0.05$) accompanied by a large effect size. These results indicate that the observed learning gains were not only statistically meaningful but also practically substantial.

Despite the overall positive outcomes, the analysis identified Data Analysis & Logical Functions as the most challenging content area for participants. Although improvement was observed, the relatively smaller effect size suggests that students require deeper conceptual support and extended practice to fully master conditional logic and analytical reasoning within spreadsheet environments. This finding highlights an important area for pedagogical refinement in future implementations.

The second objective, which focused on developing students' ability to visually represent data, was achieved with particularly strong results. The Data Visualization category exhibited the highest post-test mean score (3.58) and a large effect size, indicating a substantial improvement in students' ability to interpret and communicate data graphically. These quantitative findings are further reinforced by qualitative feedback, in which many participants identified chart selection and visualization techniques as the most impactful and accessible components of the learning experience.

Overall, the findings provide robust evidence that spreadsheet-assisted instruction can serve as an effective educational approach for strengthening students' numeracy and data literacy skills. The combination of statistically significant learning gains and large practical effect sizes underscores the value of integrating spreadsheet-based tools into mathematics and data-oriented instruction at the secondary school level.

Based on these conclusions, several recommendations can be proposed for future implementations. Given that Data Analysis & Logical Functions posed the greatest challenge, it is advisable to allocate additional instructional time to this module, introduce simpler and more contextually relevant case studies, and provide increased opportunities for guided, hands-on practice—particularly for functions such as *IF* and *COUNTIF*. These refinements are expected to improve conceptual understanding and reduce cognitive load for learners.

Future studies are also encouraged to employ larger sample sizes and comparative research designs to further validate the effectiveness of spreadsheet-based learning interventions and to strengthen the generalizability of the findings.

Acknowledgments

References

Bhargava, R. (2019). Data Literacy. In *The International Encyclopedia of Media Literacy* (pp. 1–5). John Wiley & Sons, Ltd. https://doi.org/https://doi.org/10.1002/9781118978238.ieml0049

Formby, S. K., Medlin, B. D., & Ellington, V. (2017). Microsoft Excel®: Is It an Important Job Skill for College Graduates?. *Information Systems Education Journal*, *15*(3), 55–63.

Griffin, G. W., & Holcomb, D. (2023). Data Literacy and Skills Development. In *Building a Data Culture: The Usage and Flow Data Culture Model* (pp. 109–122). Apress. https://doi.org/10.1007/978-1-4842-9966-1_6

Jeong, J., & Lee, Y. (2024). Development of Data Literacy Competency System for K-12. *International Journal on Advanced Science, Engineering and Information Technology*, *14*(6 SE-Articles), 2130–2140.

https://doi.org/10.18517/ijaseit.14.6.12373

Leon-Urrutia, M., Taibi, D., Pospelova, V., Splendore, S., Urbsiene, L., & Marjanovic, U. (2022). Data Literacy: An Essential Skill for the Industry. In B. Lalic, D. Gracanin, N. Tasic, & N. Simeunović (Eds.), *Proceedings on 18th International Conference on Industrial Systems -- IS'20* (pp. 326–331). Springer International Publishing.

Liengme, B. V. (2000). A guide to Microsoft Excel for business and management. *(No Title)*.

Miloradov, M., Rakic, S., & Marjanovic, U. (2022). Towards digital economy through data literate workforce. *Companion Proceedings of the Web Conference 2022*, 417–422. https://doi.org/10.1145/3487553.3524696

Mustafy, T., & Rahman, M. T. U. (2024). *Statistics and data analysis for engineers and scientists*. Springer.

Simaremare, S., & Siregar, R. (2024). Effectivity of the Microsoft Excel Application on Student Understanding in Statistics Courses. *Indonesian Journal of Education and Social Humanities*, *1*(2), 9–19.

Talib, Y. Y. A., & Mohamad, R. (2020). *Microsoft Excel: Preparing Data, Analysing Data and Designing a Business Model–A Practical Guide (UUM Press)*. UUM Press.