Indonesian Journal of Educational Science
ISSN 2662-6197 (online)

Volume 7 No 2, March 2025
ISSN 2655-4402 (print)

Exploring the Wilcoxon Test in Science Education: A Literature Review of Empirical Research

Fandi Ahmad[1], Nur Indah Sari[2*]
[1]Program Studi Pendidikan Kimia Universitas Negeri Makassar, Makassar, Indonesia
[2]Dosen Program Studi Pendidikan IPA Universitas Negeri Makassar, Makassar, Indonesia
[*]*e-mail: nur.indah.sari@unm.ac.id*

(Received:12 January 2025; Accepted: 08 March 2025)

Abstract

The Wilcoxon test is a non-parametric statistical method frequently used in educational research to analyze significant differences between two related data groups. This study explores the utilization of the Wilcoxon test in the context of science education research through a systematic literature review of empirical studies. The review encompasses empirical research articles published in accredited scientific journals that employ the Wilcoxon test as a data analysis technique. A total of 2700 articles were found in the first stage and 1790 in the second with specific searches. Screening continued, focusing on the relevance of titles and abstracts, resulting in 367 articles. Ten articles were selected based on SINTA 3, 4, and 5 accreditation criteria. Findings indicate a prevalence of using the Wilcoxon test in pre-test and post-test data analysis in experimental research designs, comparative research, exploratory research, and development research. Furthermore, this study identifies diverse variations in the application of the Wilcoxon test, including the Wilcoxon test for paired samples and the Wilcoxon test with continuity correction. This literature review provides a comprehensive description of the application of the Wilcoxon test in science education research and recommends careful consideration for researchers in selecting and applying the statistical test based on data characteristics and research objectives.
Keywords: Wilcoxon test, literature review, science education, non-parametric statistics, empirical research.

Abstrak

Uji Wilcoxon merupakan metode statistika nonparametrik yang kerap dimanfaatkan dalam penelitian pendidikan untuk menganalisis perbedaan signifikan antara dua kelompok data yang berhubungan. Studi ini mengeksplorasi pemanfaatan Uji Wilcoxon dalam konteks penelitian pendidikan IPA melalui pendekatan tinjauan literatur sistematis terhadap studi empiris. Tinjauan tersebut mencakup artikel-artikel penelitian empiris yang terpublikasi di jurnal ilmiah terakreditasi dan menerapkan Uji Wilcoxon sebagai teknik analisis data. Sebanyak 2700 artikel ditemukan pada tahap pertama dan 1790 pada tahap kedua dengan pencarian spesifik. Skrining dilanjutkan dengan fokus pada relevansi judul dan abstrak dan diperoleh 367 artikel. Sebanyak 10 artikel terpilih berdasarkan kriteria akreditasi SINTA 3,4 dan 5. Temuan menunjukkan prevalensi penggunaan Uji Wilcoxon dalam analisis data pre-test dan post-test pada desain penelitian eksperimen, penelitian komparatif, penelitian eksprolasi, dan penelitian pengembangan. Selain itu, studi ini mengidentifikasi beragam variasi penerapan Uji Wilcoxon, meliputi Uji Wilcoxon untuk sampel berpasangan dan Uji Wilcoxon dengan koreksi kontinuitas. Tinjauan literatur ini memberikan deskripsi komprehensif mengenai aplikasi Uji Wilcoxon dalam penelitian pendidikan IPA serta merekomendasikan pertimbangan yang cermat bagi peneliti dalam memilih dan mengaplikasikan uji statistik tersebut berdasarkan karakteristik data dan tujuan penelitian.
Kata Kunci: Wilcoxon test, literature review, pendidikan ipa, non-parametric statistics; penelitian empirik

INTRODUCTION

The purpose of science education is not only to teach scientific knowledge but also to develop a science-literate society capable of engaging in scientific reasoning and decision-making (Almasri et al., 2021; Ika Sari et al., 2024). One of the competencies that students must achieve in learning science, according to Permendiknas Number 21 of 2016, is formulating problems based on phenomena, formulating hypotheses, carrying out experiments, carrying out measurements, making tabulations and graphs, concluding based on data, and making reports both orally and in writing. Hypothesis testing plays an important role in this endeavor. Hypothesis testing provides a structured framework for investigating scientific phenomena, evaluating claims, and drawing conclusions based on evidence (Amaliah et al., 2024). By formulating testable hypotheses, designing experiments,

collecting data, and analyzing results, learners can develop critical thinking skills and improve their understanding and learning outcomes.

Quantitative research is a systematic approach to investigating phenomena using numerical data and statistical techniques (Ghanad, 2023). In science education, it plays a vital role in measuring variables, establishing relationships, and testing hypotheses related to teaching and learning (Hatlevik, 2017; Ika Sari et al., 2024). It allows researchers to collect data on a large scale (Subedi, 2021), providing a broad understanding of educational trends and the effectiveness of interventions (Chen & Cheng, 2022; Pregoner et al., 2024; Sari & Ahmad, 2022). By employing statistical analysis, quantitative research can offer objective and reliable insights into student performance, teaching methodologies, and curriculum design (Johnson & Larry, 2003).

The importance of quantitative research in science education lies in its ability to provide evidence-based insights that inform instructional practices and policy decisions (Schildkamp, 2019). Through quantitative studies, educators can identify effective teaching strategies, evaluate the impact of educational programs, and assess student learning outcomes (Rana et al., 2023). This data-driven approach enables educators and policymakers to make informed decisions that enhance the quality of science education and promote student success (McClure, 2020).

Quantitative research provides numerical data and statistical analysis that are frequently used to draw conclusions from experimental research designs. Experimental research is a specific type of research design where you manipulate variables to see their effect that focuses on collecting and analyzing numerical data (Johnson & Larry, 2003). Experimental research also plays a crucial role in science education, informing our understanding of effective teaching methods and student learning (Kooli, 2023). Hypothesis testing empowers individuals to distinguish credible information. In science education, hypothesis testing serves as a bridge between theoretical knowledge and practical applications, allowing learners to connect scientific concepts to real-world scenarios and make informed decisions about scientific issues that impact their lives and society (Narut & Wahyu, 2023). This ability to critically evaluate information and engage in evidence-based reasoning is important to increase responsibility and effective participation in Teaching and Learning Activities.

The Wilcoxon Signed-Rank test is a statistical test often used in quantitative research (Taheri & Hesamian, 2013). The Wilcoxon test is a non-parametric statistical method widely used in educational research, especially in science education (Atqiya et al., 2021). This non-parametric test allows researchers to compare two related samples or repeated measurements on a single sample without making assumptions about the underlying data distribution (Kim, 2014; Lu et al., 2024). The Wilcoxon test has been applied in a variety of educational studies, ranging from evaluating the effectiveness of instructional interventions to analyzing learners' learning outcomes and scientific reasoning patterns (Atqiya et al., 2021).

The Wilcoxon test is particularly useful when the assumptions of parametric tests, such as the t-test, are not met (Ahumada & Sanchez, 2019; Sarumaha et al., 2022). The Wilcoxon test has been widely applied in various educational settings, including Science Education, to analyze learner learning outcomes, evaluate the effectiveness of instructional interventions, and investigate patterns of scientific reasoning (Ahumada & Sanchez, 2019; Kim, 2014). Unlike parametric tests, the Wilcoxon test does not rely on normally distributed data, making it a powerful option for analyzing data that may not meet the assumptions of other statistical methods. (Ahumada & Sanchez, 2019; Kim, 2014). This flexibility has contributed to the widespread adoption of the Wilcoxon test in educational research, especially in science education, where data do not always fit the requirements of parametric analysis.

There are strong scientific justifications for a literature review on the Wilcoxon test in science education. First, experimental research is essential for understanding the impact of teaching methods, and the Wilcoxon Signed-Rank test is a valuable tool for analyzing data in these studies (Widodo,

2022). Because educational data often do not fit the requirements of parametric tests, the Wilcoxon test's non-parametric nature is especially helpful (du Prel et al., 2010). A literature review can show how the Wilcoxon test has been used in science education, what data types it's suited for, and what conclusions have been drawn from it. A literature review can reveal gaps in the current research and suggest directions for future studies.

A literature review on the Wilcoxon test in science education is scientifically justified because it helps to synthesize existing knowledge, identify best practices, and guide future research in the field. Reviewing existing literature can reveal how the Wilcoxon test has been applied in science education research, the types of data analyzed, and the conclusions drawn (Wilson, 2024). Furthermore, a literature review can identify gaps in current research and suggest directions for future studies, ultimately helping to synthesize knowledge, identify best practices, and guide further research in the field (Wilson, 2024). A systematic literature review can provide a comprehensive and balanced summary of primary research results (Setyawati et al., 2022).

Experimental research is frequently used in science education (Savec & Mlinarec, 2021; Treagust & Won, 2023). This systematic review was conducted to analyze the use of the Wilcoxon test in science education research. The study provides a comprehensive synthesis of how this non-parametric statistical test has been applied to generate new research findings in science teaching and learning. The focus of the analysis includes the research context, methodological approach, and key findings revealed in empirical studies that have used the Wilcoxon test. Research question:

1. How effective is the Wilcoxon Signed-Rank test in evaluating the impact of educational interventions on student performance in science education?
2. What are the methods and design used in studies on empirical research on science education?

METHOD

This research is a literature study on science education research. The articles were selected from journals using the Preferred Reporting Items for Systematic Review and Meta-Analysis systematic analysis (O'Dea et al., 2021). The research approach consists of various stages, namely, 1) searching for research journals, 2) systematic selection, and 3) article analysis. The steps in the research are described in detail in Figure 1.

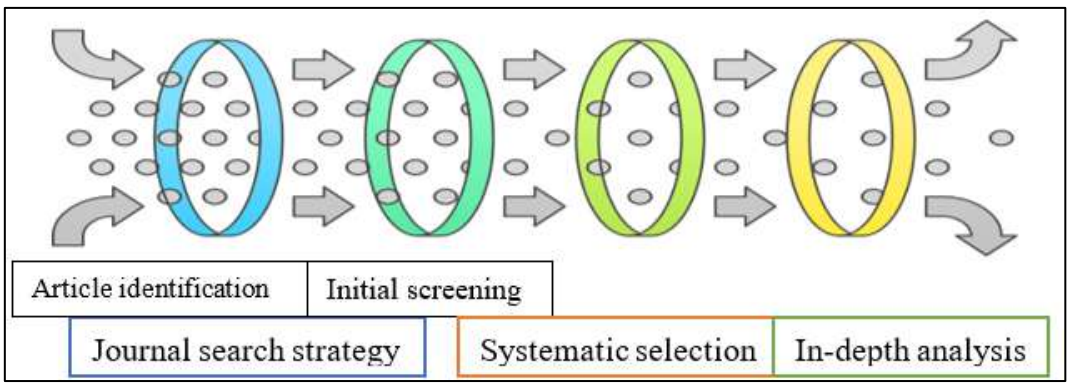

Figure 1. Research Steps Literature Review

1. Journal search strategy

The journal search began with the research topic of Science Education. The academic database used is Google Scholar, which verifies the source's credibility. The keywords used were "Wilcoxon signed-rank test in Science Education". The Wilcoxon signed-rank test is a non-parametric statistical test used to compare two related samples, or to compare one sample to a hypothetical population

median. 2700 articles from various sources, including journals, books, and conferences, were identified. then used the “since 2021” filter and 1790 articles were identified. The articles were then subjected to initial screening based on the title's relevance and abstract, and 367 articles were selected.

2. Selection process

After identifying potential journals through Google Scholar, a selection process was carried out to ensure the research objectives' relevance to the selected journals' quality, especially the journals' reputation and accreditation. The criteria researchers use in the selection process are journals published in SINTA 3, 4, and 5 from 2021 to 2024. SINTA is an Indonesian index for journals and research publications. Researchers focus on SINTA 3,4, and 5 journals to ensure a baseline level of quality and relevance, might be more accessible and relevant to research focusing specifically on Indonesian contexts or issues, and also provides a manageable scope for a literature review (Niam et al., 2023). The selection process was done by checking publisher affiliation, database indexation, and journal citation. A total of 10 journals were selected based on the criteria that have been stated.

3. Research Analysis

The literature review was conducted using systematic, critical, and structured analysis of data analysis techniques using the Wilcoxon test (Ünlü et al., 2024). The literature review in this study was conducted by adopting the *best-evidence synthesis* approach. *This method* emphasizes selecting and critically analyzing the most relevant and high-quality journal articles to answer the research questions. Each selected article was analyzed in depth and verified for data accuracy, namely the Wilcoxon statistical test, in analyzing the research data. The choice of this analysis method was supported by the data characteristics that did not meet the assumption of normality, as described in the articles. Thus, the final number of articles to be analyzed in the systematic review was obtaine in the PRISMA Flow Diagram (Hebebci, 2023) in Figure 2

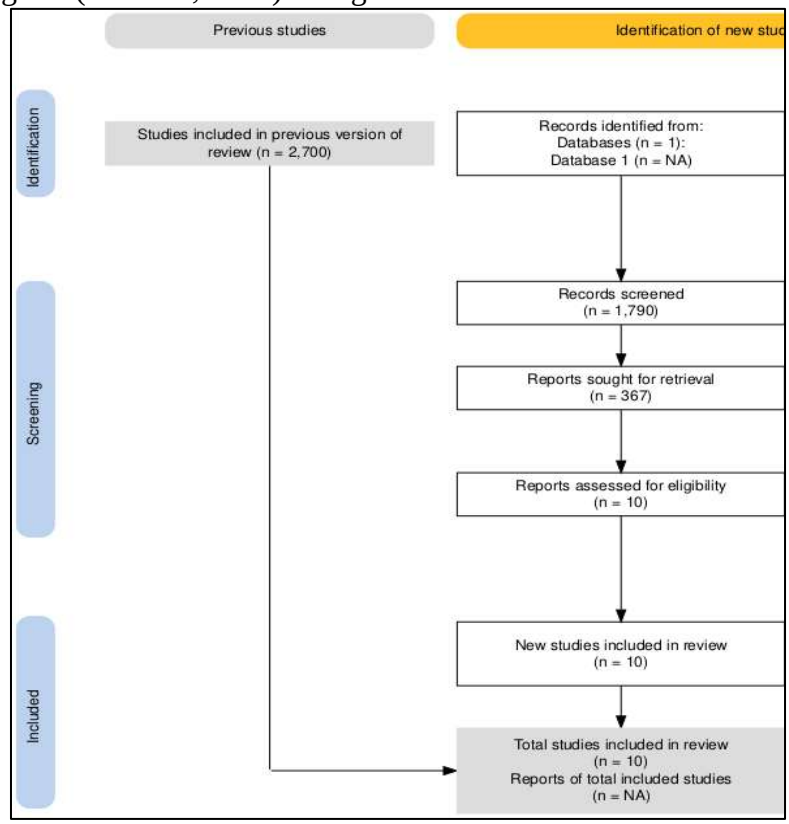

Figure 2. PRISMA flow diagram (Haddaway et al., 2022)

Based on Figure 2, ten articles that were reached that were reached as a result of the literature review and that met the research criteria were examined.

RESULT

In science education, the Wilcoxon test often used to analyze differences in student learning outcomes before and after specific treatments, such as applying innovative learning methods or using interactive learning media. The advantage of the Wilcoxon test lies in its ability to handle data that is not normally distributed or data with an ordinal scale, so it is an appropriate alternative when the assumptions of the parametric t-test are not met (Almasri et al., 2021; Kim, 2014; Zulkipli et al., 2024). According to Ahumada & Sanchez (2019), non-parametric analysis can be used in several specific conditions related to data and sample characteristics: 1) when the sample size used in the study is relatively small; 2) if the data being analyzed is ordinal, which allows the arrangement of data based on order or rank; 3) non-parametric statistics are also relevant when dealing with nominal data, which is data that can be grouped into categories and the frequency can be calculated: and 4) the use of non-parametric approaches is indicated when the data distribution and characteristics of the population from which the sample is taken are unknown or do not follow a normal distribution.

A literature review of empirical research using the Wilcoxon test in science education provides a comprehensive overview of current trends and findings in applying this test in science teaching and learning. Analysis of these empirical studies can reveal patterns of differences in learner outcomes influenced by learning interventions. In addition, the literature review can also identify factors that influence the successful application of a learning method or media. The information obtained from this literature review is helpful for educators and researchers in developing more effective and efficient learning strategies. Furthermore, according to Vierra et al. (2023), if the test data is not normally distributed, non-parametric statistics are less likely to make errors in interpreting the data compared to parametric statistics.

Studies on empirical research emphasize collecting and analyzing data obtained through observation, experimentation, surveys, and other empirical methods to test hypotheses, build theories, or answer research questions systematically and based on concrete evidence (Almasri et al., 2021; Zulkipli et al., 2024). Table 1 shows some empirical research findings in Science Education that use Wilcoxon non-parametric statistical analysis in making decisions based on research data.

Table 1. Wilcoxon analysis findings in science education research

No.	Researcher (Year/Sinta)	Research Methodology	Key findings
1	Masitoh et al., (2023/4)	a) Quasi-experimental research with *Pretest-Posttest* Design for 1 class b) Sample size 23 students	a) The Wilcoxon Signed Rank test showed statistically significant results, with an asymp.sig value of 0.002, which is less than the 0.05 threshold, leading to the rejection of the null hypothesis. b) The I-SETS approach positively influences learners' understanding and engagement in science.
2	Fazrina et al., (2023/5)	a) Quantitative research with a Quasi Experiment design with the *One Group*	a) Statistical tests confirmed that the *post-test* scores of both variables were significantly higher than the pre-test scores, indicating effective results.

No.	Researcher (Year/Sinta)	Research Methodology	Key findings
		Pretest-Posttest Design. b) Sample size = 18 students	b) *The Wilcoxon* Paired Test significantly affected concept understanding, and the T-Test confirmed significant effects on generic science skills.
3	Husni et al., (2022/4)	a) Pre-Experimental design with *One Group Pretest-Posttest* approach. b) Sample size = 15 students	a) There was a significant difference between the pretest and posttest scores, with the average pretest score being 45 and the average posttest score being 75. b) The Wilcoxon sign rank test showed a Z value of -3.521 with Asymp Sig. 0.000, confirming that the model had a significant impact on learning outcomes.
4	Reksi et al. (2024/4)	a) Quantitative research with experimental design. b) Sample size = 33 people	A significant reduction in the number of learners' errors after receiving remediation through *Google Classroom* feedback, with statistical analysis showing a significant difference before and after the intervention ($p < 0.05$).
5	Rahmadani (2024/4)	a) Pre-Experimental design with *One Group Pretest-Posttest* approach. b) Sample size = 32 students	a) The Wilcoxon rank test showed a statistically significant improvement, with an Asymp. Significance value (2-tailed) of 0.000, leading to the rejection of the null hypothesis (H0) and acceptance of the alternative hypothesis (H1) b) This teaching strategy can transform the learning experience into a more engaging and effective process.
6	Hariandi et al. (2023/3)	a) Pre-Experimental design with *One Group Pretest-Posttest* approach. b) Sample size = 33 students	a) The results of the Wilcoxon Sign Rank Test test, the z value obtained is -5.015 with a value (Asymp. Sig 2 tailed) of 0.000. b) There is a difference in KSP before and after applying the STEAM approach.
7	Damayanti et al., (2021/4)	a) Descriptive research with *one-group* design b) *Purposive* sampling = 36 learners	a) The Wilcoxon test results showed a significant difference between the pre-test and post-test scores, with the Z-count at 5.094 exceeding the Z-table at 2.892. b) PBL learning model effectively improves students' motivation and learning outcomes

No.	Researcher (Year/Sinta)	Research Methodology	Key findings
8	Amelia et al., (2021/3)	a) R&D research, ADDIE model and e-module effectiveness test with experimental method of *one-group pretest-postest design.* b) Research subjects consisted of 9 learners	The results of testing the effectiveness with the Wilcoxon test show the results of z = 2.673 and an *effect size* of r = 0.63 which means that there is a significant difference between the *pre-test* and post-test scores and the use of e-modules for misconceptions has a moderate effect in remediating students' misconceptions on the material Pressure of liquid substances.
9	Hartanti & Hartanti (2023/4)	a) Exploratory research with a quantitative approach b) The sample consisted of 192 students	The Wilcoxon test showed no significant difference between estimated and measured values for length and time, while a significant difference was found for mass indicating lower estimation ability.
10	Amalia et al., (2024/5)	a) A comparative study of TSTS and TGT cooperative models. b) Involving 2 classes with a sample size of 30 and 27 students	There is a difference in the value of learning outcomes by using the TGT type cooperative learning model better than the TSTS model in improving the science learning outcomes of VIII grade students of SMPN 1 Pattallassang.

Based on the literature review in Table 1, the use of non-parametric statistics Wilcoxon test can be identified with consideration of whether the data of empirical research results in science education research studies are normal or not. Wilcoxon test results are used to analyze differences between two groups of related data (paired data), such as pre-test and post-test. The Wilcoxon *Signed-Rank* test is used when the assumption of data normality is not met or when the data is ordinal. The Wilcoxon test has been applied in various educational studies, ranging from evaluating the effectiveness of instructional interventions to analyzing learning outcomes, exploration with quantitative, descriptive approaches, development of learning tools, and students' scientific reasoning patterns (Atqiya et al., 2021; Hartanti & Hartanti, 2024).

In line with the research's second question, the empirical research articles methods and design were disscussed can be seen in Figure 3

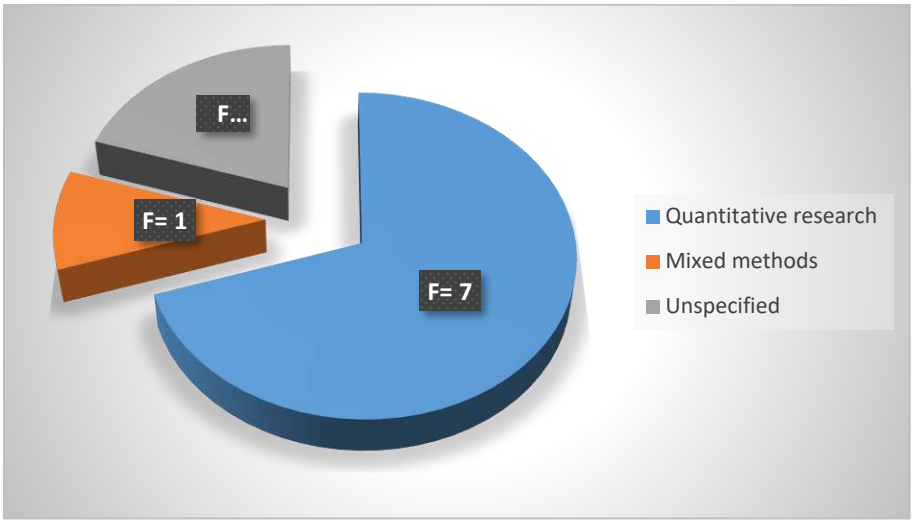

Figure 3. Distribution of articles by the methods used

As a result of the analysis in Figure 3, the distribution of research methods employed in the reviewed articles reveals a strong inclination towards quantitative approaches, with 7 articles utilizing such methods, while a smaller proportion adopted mixed methods in one article. This suggests that quantitative research remains a dominant paradigm in this field. A notable fraction of the reviewed studies (2 articles) did not specify the methodology used, which poses challenges in understanding and interpreting their findings (Tariq & Woodman, 2013)(. Moving forward, clearly specifying research methodology in publications is essential.

DISCUSSION

The Wilcoxon test, both the Wilcoxon signed rank test and the Wilcoxon rank sum test has been widely used in science education research with a variety of methodologies and sample sizes. This non-parametric test is chosen when the assumption of data normality is not met or when the data scale is ordinal. Based on Table 1, there are several examples of the application of the Wilcoxon test in empirical research in the study of Science Education. The Wilcoxon test can be used to compare pre-test and post-test results in experimental groups that are not randomly formed. The research compares the problem-solving ability of students before and after the application of a particular learning model or *treatment*. The application of this test can be seen in research conducted by Masitoh (2023) and Fazrina et al. (2023), which shows the results of statistical testing confirm that the *post-test* scores of both variables are significantly higher than the pre-test scores, showing effective results.

The compilation of quantitative research on science education utilizing the Wilcoxon Signed Rank test reveals a consistent trend, interventions, often implemented within pre-experimental or quasi-experimental designs, yield statistically significant improvements in student learning outcomes. Studies by (Masitoh, 2023), Fazrina et al. (2023), Husni et al. (2022), Rahmadani et al. (2024), and Hariandi et al. (2023). All report significant p-values ($p < 0.05$) from Wilcoxon tests, indicating that the interventions positively impacted concept understanding, generic science skills, and overall learning outcomes. This suggests that, within the scope of these studies, the applied educational strategies are effective in enhancing students' grasp of scientific concepts.

However, several limitations should be considered when interpreting these findings. The predominance of pre-experimental designs, particularly the One Group Pretest-Posttest approach, limits the ability to establish causality due to the absence of a control group (du Prel et al., 2010). Additionally, the relatively small sample sizes, ranging from 15 to 33 students, may restrict the generalizability of the results. While Reksi et al. (2024) also found significant improvements through

Google Classroom feedback, it is important to acknowledge these limitations and interpret the findings cautiously, recognizing the need for further research with more rigorous designs and larger, more diverse samples.

The application of the Wilcoxon test in the presented research demonstrates its utility in evaluating the impact of educational interventions, as seen in Damayanti et al. and Amelia et al. Both studies, employing a one-group pretest-posttest design, revealed statistically significant differences between pre- and post-test scores following the implementation of a PBL learning model and e-modules, respectively. Damayanti et al. (2021) found a $Z\text{-}_{count}$ of 5.094 exceeding the $Z\text{-}_{table}$ value, while Amelia et al. (2021) reported a significant effect size ($r = 0.63$), suggesting that these interventions effectively improved student motivation, learning outcomes, and addressed misconceptions. These findings align with previous research that highlights the effectiveness of various educational strategies in enhancing student learning (Coenen et al., 2018).

However, it's important to consider the context in which the Wilcoxon test is applied. Hartanti & Hartanti (2024) exploratory research, while also using the Wilcoxon test, yielded mixed results, indicating no significant difference in estimation abilities for length and time, but a significant difference for mass. This underscores the importance of considering the specific variables being measured and the potential limitations of relying solely on statistical significance. Furthermore, Amaliah et al. (2024) comparative study, while not directly using the Wilcoxon test, highlights the importance of comparing different instructional models to identify superior approaches for improving science learning outcomes. It's also important to acknowledge that the Wilcoxon Signed Rank test is a frequently used nonparametric test for paired data. It is important to choose the right type of Wilcoxon test according to the research design. If the data comes from two independent groups, then the analysis used is the Mann-Whitney U test. If the data comes from two related groups or paired data, use the *Wilcoxon Signed-Rank Test.*

CONCLUSION

Based on the analysis of several science education studies using the Wilcoxon test, it can be concluded that the Wilcoxon Signed Rank Test generally shows a significant difference before and after intervention or treatment when the sample taken is relatively small. In addition, the Wilcoxon test proved effective in testing comparative hypotheses and providing empirical evidence of the effect of the intervention or treatment given. The implication of this study is that the Wilcoxon Signed Rank test can be a useful tool in science education research, particularly when dealing with small sample sizes, to assess the impact of interventions or treatments.

REFERENCES

Ahumada, L. K., & Sanchez, I. D. (2019). Application of the Wilcoxon Test to correlate the results of the Saber 11 and Saber T&T Test. *IOP Conference Series: Materials Science and Engineering*, *519*(1). https://doi.org/10.1088/1757-899X/519/1/012034

Almasri, F., Hewapathirana, G. I., Ghaddar, F., Lee, N., & Ibrahim, B. (2021). Measuring attitudes towards biology major and non-major: Effect of students' gender, group composition, and learning environment. *PLoS ONE*, *16*(5 May), 251453. https://doi.org/10.1371/journal.pone.0251453

Amaliah, R., Amalia, A., Hasyim, A., Sari, N. I., & Ahmad, F. (2024). Perbandingan Hasil Belajar IPA Peserta Didik Melalui Model Pembelajaran Koperatif Tipe TSTS dan TGT. *Journal of Health, Education, Economics*, *7*(1).

Amelia, E. A., Maison, M., & Kamid, K. (2021). Pengembangan E-Modul Menggunakan 3D Pageflip Professional Untuk Meremediasi Miskonsepsi Siswa Smp Pada Materi Tekanan Zat Cair.

EduFisika: Jurnal Pendidikan Fisika, *6*(2), 72–83. https://doi.org/10.59052/edufisika.v6i2.14522

Arsyad, A. A., & Sartika, D. (2020). Analisis keterampilan proses sains mahasiswa calon guru fisika pada praktikum fisika dasar. *Indonesian Journal of Educational Science (IJES)*, *3*(1), 69-74.

Atqiya, N., Yuliati, L., & Diantoro, M. (2021). Argument-driven inquiry for STEM education in physics: Changes in students' scientific reasoning patterns. *AIP Conference Proceedings*, *2330*(1)). https://doi.org/10.1063/5.0043636

Chen, J., & Cheng, T. (2022). Review of research on teacher emotion during 1985–2019: a descriptive quantitative analysis of knowledge production trends. *European Journal of Psychology of Education*, *37*(2), 417–438. https://doi.org/10.1007/s10212-021-00537-1

Coenen, J., Cornelisz, I., Groot, W., Maassen van den Brink, H., & Van Klaveren, C. (2018). Teacher Characteristics and Their Effects on Student Test Scores: a Systematic Review. *Journal of Economic Surveys*, *32*(3), 848–877. https://doi.org/10.1111/joes.12210

Damayanti, L., Subagiyo, L., & Syam, M. (2021). Penerapan Model Problem Based Learning dengan Media Moodle dan Pengaruhnya terhadap Motivasi dan Hasil Belajar Kognitif Siswa. *Jurnal Literasi Pendidikan Fisika (JLPF)*, *2*(1), 53–62. https://doi.org/10.30872/jlpf.v2i1.354

du Prel, J.-B., Röhrig, B., Hommel, G., & Blettner, M. (2010). Choosing statistical tests: part 12 of a series on evaluation of scientific publications. *Deutsches Arzteblatt International*, *107*(19), 343–348. https://doi.org/10.3238/arztebl.2010.0343

Fazrina, H. N., Hidayat, O. S., & Hasanah, U. (2023). Pengaruh Pendekatan I-SETS (Islamic-Science, Environment, Technology, Society) terhadap Hasil Belajar IPA Siswa Sekolah Dasar. *Jurnal Basicedu*, *7*(2), 1348–1355. https://doi.org/10.31004/basicedu.v7i2.5130

Ghanad, A. (2023). An Overview of Quantitative Research Methods. *International Journal of Multidisciplinary Research and Analysis*, *06*(08), 3794–3803. https://doi.org/10.47191/ijmra/v6-i8-52

Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis. *Campbell Systematic Reviews*, *18*(2), 1230. https://doi.org/10.1002/cl2.1230

Hariandi, J., Sitompul, S. S., & Habellia, R. C. (2023). Peningkatan Keterampilan Proses Sains Dengan Menerapkan Pendekatan Steam. *Jurnal Pendidikan Fisika*, *11*(2), 157. https://doi.org/10.24127/jpf.v11i2.7945

Hartanti, S., & Hartanti, W. (2024). Analisis Kemampuan Siswa dalam Menaksir Ukuran sebagai Salah Satu Keterampilan Proses Sains pada Materi Pengukuran. *Unnes Physics Education Journal*, *13*(2), 133–146. https://doi.org/https://doi.org/10.15294/upej.v13i2.15186

Hatlevik, O. E. (2017). Examining the Relationship between Teachers' Self-Efficacy, their Digital Competence, Strategies to Evaluate Information, and use of ICT at School. *Scandinavian Journal of Educational Research*, *61*(5), 555–567. https://doi.org/10.1080/00313831.2016.1172501

Hebebci, M. T. (2023). A Systematic Review of Experimental Studies on STEM Education. *Journal of Education in Science, Environment and Health*, *9*(1), 56–73. https://doi.org/10.55549/jeseh.1239074

Husni, R., Suryawan, A., & Rahmawati, P. (2022). Pengaruh model pembelajaran discovery learning berbantuan media benda konkret terhadap hasil belajar IPA. *Borobudur Educational Review*, *2*(1), 10–22. https://doi.org/10.31603/bedr.6822

Ika Sari, G., Winasis, S., Pratiwi, I., Wildan Nuryanto, U., & Basrowi. (2024). Strengthening digital literacy in Indonesia: Collaboration, innovation, and sustainability education. *Social Sciences and Humanities Open*, *10*, 101100. https://doi.org/10.1016/j.ssaho.2024.101100

Johnson, B., & Larry, C. (2003). Quantitative, Qualitative, and Mixed Approaches. In *Educational Research*. Sage publications.

Kim, H.-Y. (2014). Statistical notes for clinical researchers: Nonparametric statistical methods: 2. Nonparametric methods for comparing three or more groups and repeated measures. *Restorative Dentistry & Endodontics*, *39*(4), 329. https://doi.org/10.5395/rde.2014.39.4.329

Kooli, C. (2023). Chatbots in Education and Research: A Critical Examination of Ethical Implications and Solutions. *Sustainability (Switzerland)*, *15*(7), 5614. https://doi.org/10.3390/su15075614

Lu, D., Li, H., Qiu, B., Liu, S., & Qi, S. (2024). Chemical simulation teaching system based on virtual reality and gesture interaction. *Virtual Reality and Intelligent Hardware*, *6*(2), 148–168. https://doi.org/10.1016/j.vrih.2023.09.001

Masitoh, S. (2023). Pengaruh Problem Based Learning (Pbl) Berbantuan Canva Terhadap Student Well Being Dan Hasil Belajar Ipa Siswa Sd Pada Materi Udara Bersih Bagi Kesehatan. *Pendas : Jurnal Ilmiah Pendidikan Dasar*, *8*(1), 509–523. https://doi.org/10.23969/jp.v8i1.7606

McClure, C. P. (2020). Development and assessment of a continuing education unit in quantitative literacy for high school stem teachers. *Numeracy*, *13*(2), 1–15. https://doi.org/10.5038/1936-4660.13.2.1300

Narut, Y. F., & Wahyu, Y. (2023). Pembelajaran Ipa Berorientasi Higher Order Thinking Skills (Hots) Di Sekolah Dasar. *Jurnal Literasi Pendidikan Dasar*, *4*(1), 85–91.

Niam, K., Nuril Huda, M., Sulanam, & Ridwan, A. (2023). Knowledge management practices in enhancing academic output: A Study of Indonesian Islamic Universities. *Ulumuna*, *27*(2), 735–761. https://doi.org/10.20414/ujis.v27i2.582

O'Dea, R. E., Lagisz, M., Jennions, M. D., Koricheva, J., Noble, D. W. A., Parker, T. H., Gurevitch, J., Page, M. J., Stewart, G., Moher, D., & Nakagawa, S. (2021). Preferred reporting items for systematic reviews and meta-analyses in ecology and evolutionary biology: a PRISMA extension. *Biological Reviews*, *96*(5), 1695–1722. https://doi.org/10.1111/brv.12721

Pregoner, J. D., Dave, J., Comparison, E. A., & Pregoner, J. D. (2024). Research Approaches in Education : A Comparison of Quantitative , Qualitative and Mixed Methods To cite this version : HAL Id : hal-04655066 Research Approaches in Education : A Comparison of Quantitative , Qualitative and Mixed Methods. *IMCC Journal of Science*, *4*(2), 31–36.

Qudratuddarsi, H., Meivawati, E., & Saputra, R. (2024). Pelatihan Penelitian Metode Kuantitatif dan Systematic Literature Review bagi Dosen dan Mahasiswa. *Beru'-beru': Jurnal Pengabdian kepada Masyarakat*, *3*(1), 22-32.

Rahmadani, R., Sitompu, S. S., & Oktavianty, E. (2024). Meningkatan Pemahaman Konsep Dengan Model Pbl Berbantuan Mind Mapping. *Jurnal Education And Development*, *12*(3), 58–63.

Rana, P., Dubey, M. K., Gupta, L. R., & Thakur, A. K. (2023). A model to combine qualitative and quantitative measures in education for better assessment of learning outcomes. *Interactive Learning Environments*, *32*(10), 5947–5977. https://doi.org/10.1080/10494820.2023.2243291

Reksi, M., Sitompul, S. S., & Hamdani, H. (2024). Remediasi Kesalahan Menyelesaikan Soal Menggunakan Umpan Balik Berbantuan Google Classroom di SMP. *Jurnal Ilmiah Profesi Pendidikan*, *9*(1), 69–75. https://doi.org/10.29303/jipp.v9i1.1649

Sari, N. I., & Ahmad, F. (2022). Effectiveness of Jigsaw Type II Cooperative Learning Model in Improving Self-Efficacy, Activities, and Learning Outcomes of STKIP Pembangunan Indonesia Makassar Students. *ARRUS Journal of Mathematics and Applied Science*, *2*(1), 14–22. https://doi.org/10.35877/mathscience714

Sarumaha, M., Harefa, D., Ziraluo, Y. P. B., Fau, A., Venty Fau, Y. T., Bago, A. S., Telambanua, T., Hulu, F., Telaumbanua, K., Lase, I. P. S., Laia, B., Ndraha, L. D. M., & Novialdi, A. (2022). Penggunaan Model Pembelajaran Artikulasi Terhadap Hasil Belajar Siswa Pada Mata Pelajaran

IPA Terpadu. *Aksara: Jurnal Ilmu Pendidikan Nonformal*, *8*(3), 2045. https://doi.org/10.37905/aksara.8.3.2045-2052.2022

Savec, V. F., & Mlinarec, K. (2021). Experimental work in science education from green chemistry perspectives: A systematic literature review using prisma. *Sustainability (Switzerland)*, *13*(23), 12977. https://doi.org/10.3390/su132312977

Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research*, *61*(3), 257–273. https://doi.org/10.1080/00131881.2019.1625716

Setyawati, R. D., Pramasdyahsari, A. S., Astutik, I. D., Aini, S. N., Arum, J. P., Widodo, W., Nusuki, U., Salmah, U., & Zuliah, N. (2022). Improving Mathematical Critical Thinking Skill through STEM-PjBL: A Systematic Literature Review. *International Journal on Research in STEM Education*, *4*(2), 1–17. https://doi.org/10.31098/ijrse.v4i2.1141

Subedi, K. R. (2021). Determining the Sample in Qualitative Research. *Scholars' Journal*, *4*, 1–13. https://doi.org/10.3126/scholars.v4i1.42457

Taheri, S. M., & Hesamian, G. (2013). A generalization of the Wilcoxon signed-rank test and its applications. *Statistical Papers*, *54*(2), 457–470. https://doi.org/10.1007/s00362-012-0443-4

Tariq, S., & Woodman, J. (2013). Using mixed methods in health research. *JRSM Short Reports*, *4*(6), 204253331347919. https://doi.org/10.1177/2042533313479197

Treagust, D. F., & Won, M. (2023). Paradigms in Science Education Research. In *Handbook of Research on Science Education: Volume III* (Vol. 3, pp. 3–27). Routledge. https://doi.org/10.4324/9780367855758-2

Ünlü, D., Şekercioğlu, A. G. Ç., & Yıldırır, H. E. (2024). *Systematic literature review of graduate thesis on instructional technologies in the fields of physics, chemistry, biology and science education*.

Vierra, A., Razzaq, A., & Andreadis, A. (2023). Chapter 27 - Continuous variable analyses: t-test, Mann–Whitney U, Wilcoxon sign rank. In *Handbook for Designing and Conducting Clinical and Translational Research* (pp. 165–170). Academic Press. https://www.sciencedirect.com/science/article/pii/B9780323903004000458

Widodo, W. (2022). Online Flipped Classroom: Developing Postgraduate Science Education Students' Critical Thinking Skills. *Journal of Science Learning*, *5*(3), 469–477. https://doi.org/10.17509/jsl.v5i3.43107

Wilson, P. (2024). An Updated Wilcoxon–Mann–Whitney Test. In *International Workshop on Statistical Modelling* (pp. 159–165). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-65723-8_25

Zulkipli, Z., Zulfachmi, Z., & Rahmad, A. (2024). Alasan Peneliti Menggunakan Analisis Statistik Wilcoxon (Non Parametrik. *Prosiding Seminar Nasional Ilmu Sosial Dan Teknologi (SNISTEK*, *6*, 119–125.