

THE USE OF FINAL TEST AND PORTFOLIO ASSESSMENT IN MEASURING THE LEARNING ACHIEVEMENT OF EIGHT GRADE STUDENTS AT SMPN 1 SUSOH

Zulfarkhannia¹, Sukri Adani²

^{1,2}STKIP Muhammadiyah Aceh Barat Daya

Email: zulfarkhannia402@gmail.com

Abstract

This study investigates the distinctions and interrelationships between Final Test and Portfolio Assessment in evaluating the academic achievement of eighth-grade students at SMP Negeri 1 Susoh. A quantitative, comparative-correlational design was employed, with a sample of 61 students selected through total sampling. Secondary data, consisting of Final Test and Portfolio Assessment scores from official school records, were analyzed. Data analysis included descriptive statistics, the Kolmogorov-Smirnov normality test, the Wilcoxon Signed-Rank test, and Spearman's correlation, utilizing IBM SPSS Statistics version 26. The results indicate a significant difference between Final Test and Portfolio Assessment scores ($p < 0.05$), with Final Test scores generally higher than Portfolio Assessment scores. Furthermore, a significant positive correlation was found between the two assessment types, with a Spearman correlation coefficient of 0.565 ($p < 0.05$), suggesting that students with higher Final Test scores also tended to achieve higher Portfolio Assessment scores. These findings demonstrate that, although the two assessment methods yield different results, they are complementary in providing a comprehensive evaluation of students' learning achievements. Therefore, integrating Final Test and Portfolio Assessment offers a more holistic approach to assessing student learning outcomes. The study's limitations include a relatively small sample size, focus on a single school, and reliance on secondary data, which may limit the generalizability of the findings.

Keywords: Final Test, Portfolio Assessment, Spearman's correlation

INTRODUCTION

The theoretical landscape of educational assessment has undergone a massive transformation over the past two decades. Historically, assessment was predominantly viewed as an activity that occurs after learning has taken place, commonly referred to as the 'assessment of learning' or summative assessment. This approach, heavily influenced by behaviorist learning theories, treats knowledge as a measurable product. However, the rise of constructivism, championed by theorists such as Vygotsky and Piaget, prompted educators to reevaluate this stance. Constructivism posits that learning is an active, deeply contextualized process of meaning-making (Boud & Falchikov, 2006). In light of this, the focus has broadened to include 'assessment for learning' (formative assessment) and 'assessment as learning' (where the assessment itself is a learning activity). This modern triadic framework demands that tools like portfolios, which inherently support formative feedback and self-reflection, be integrated alongside traditional summative tests to capture the full spectrum of intellectual growth.

Assessment is a critical component of the educational process, as it

collects information on students' competency attainment and serves as a basis for evaluating instructional effectiveness. Contemporary educational paradigms emphasize that assessment should address both learning outcomes and the learning process. Therefore, assessment systems must measure cognitive, affective, and psychomotor domains in a balanced manner to provide a comprehensive understanding of student development (Zhan et al., 2025). Selecting appropriate assessment instruments is essential for generating accurate data on student achievement. Nevertheless, many schools continue to rely primarily on final examinations as the main method for evaluating student learning (Hanafiah et al., 2024).

The final test, usually given at the end of a semester, remains a fundamental part of academic evaluation in many education systems around the world. Its lasting popularity is mainly because of its logistical simplicity and strong psychometric reliability. A well-designed final exam can quickly assess a large group's understanding of core concepts, providing consistent data that is easy to measure and compare across different groups (French et al., 2024). However, the educational limitations of relying solely on such tests are well-known. High-stakes final exams often cause significant test anxiety, which can hide a student's actual abilities. Additionally, they tend to promote 'cramming,' a superficial learning method where information is memorized just long enough to pass the test, rather than truly understood. Because final tests restrict access to learning resources and require quick recall, they generally do not measure skills that need time, practice, and teamwork skills that are arguably more important in the 21st century.

The dominance of the final test in the Indonesian educational landscape has deep historical roots. For decades, the system was heavily oriented toward high-stakes testing, conditioning both teachers and students to prioritize rote memorization and test-taking strategies over deep, meaningful learning. Final exams are undoubtedly effective for measuring mastery of material in a standardized, efficient manner at the conclusion of an instructional period. Still, they inevitably suffer from a fundamental limitation: they only provide a snapshot of students' abilities at a specific, often high-pressure point in time. As a result, they do not fully capture the nuanced, non-linear learning progress that occurs throughout the semester (French et al., 2024). Many recent studies and educational critiques indicate that an assessment framework heavily skewed toward final exams struggles to accommodate higher-order thinking skills, self-reflection, and the authentic, real-world competencies increasingly required in modern education (Ali, 2024). Therefore, assessment scholars universally recommend implementing authentic, process-oriented assessments to complement traditional summative methods, thereby ensuring that evaluation results are comprehensive and ecologically valid.

In the context of English as a Foreign Language (EFL) instruction, the arguments for authentic assessment become particularly compelling. Language acquisition is inherently a skill-based, developmental process rather than merely a knowledge-based one. Portfolio assessment in an EFL setting typically involves a curated collection of student work ranging from early drafts of essays and audio recordings of speaking tasks to peer-review notes and personal reflections on linguistic progress (Al-Hawamdeh et al., 2023). This method allows educators to

track improvements in writing complexity, grammatical accuracy, and communicative fluency over time. Unlike a multiple-choice grammar test, an EFL portfolio provides a rich narrative of a student's linguistic journey, revealing not just what they know, but how they apply the language to construct meaning. Moreover, portfolios shift the locus of control to the students, encouraging them to take ownership of their learning by actively selecting the pieces that best represent their capabilities (Naka, 2025). Despite its evident advantages, the successful implementation of portfolios requires a sophisticated level of assessment literacy from teachers, who must design clear rubrics, manage the increased grading workload, and provide continuous, constructive feedback.

One widely recommended form of authentic assessment is the portfolio assessment. Distinct from single-event tests, portfolio assessment allows teachers to systematically collect, curate, and review a variety of student work over time. This continuous gathering of evidence means that learning progress, rather than just final retention, can be observed and celebrated. In addition to providing rich, qualitative information on learning outcomes, portfolios encourage active student engagement, foster reflective thinking, instill a sense of ownership over the learning process, and stimulate the development of critical thinking skills (Zhan et al., 2025). Various literature reviews corroborate that utilizing portfolios paints a much more detailed and colorful picture of competency achievement compared to scenarios where teachers rely solely on final exams as the ultimate gatekeeper of academic success (Naka, 2025).

Despite these widely acknowledged benefits, the implementation of portfolio assessment at the junior high school level remains surprisingly limited. Practical constraints such as large class sizes, limited teacher time for providing detailed feedback, and students' unfamiliarity with autonomous learning often hinder its widespread adoption. More importantly, there is a notable scarcity of empirical studies that directly and quantitatively compare the results of final tests and portfolio assessments in measuring student learning achievement, particularly within the specific socio-cultural context of regional schools in Indonesia. Most existing literature focuses heavily on urban, well-resourced environments, leaving a gap in our understanding of how these assessments function in semi-urban or rural settings where educational resources and assessment cultures may differ.

Based on this theoretical and practical gap, rigorous research is needed to analyze the use of final tests and portfolio assessments in measuring the academic achievement of eighth-grade students at SMPN 1 Susoh. This research seeks to deliver concrete, empirical evidence regarding the degree to which these two types of evaluations offer distinct or interconnected insights into students' educational results. Ultimately, this study addresses two primary questions: (1) Is there a notable contrast in students' academic performance as evaluated by final examinations versus portfolio assessments? and (2) Is there a meaningful correlation between final exam results and portfolio scores when evaluating the educational success of eighth graders at SMPN 1 Susoh?

METHOD

This research utilized a numerical method with a comparative correlational framework. The reason for choosing this numerical method was to statistically analyze both the variations and the connection between Final Test results and

Portfolio Assessment results in evaluating the academic success of eighth-grade students at SMP Negeri 1 Susoh (Creswell, 2023). The research took place at SMP Negeri 1 Susoh, located in Aceh Barat Daya Regency, throughout the academic year of 2025/2026.

The population included all eighth graders at SMP Negeri 1 Susoh. Due to the population's small size, total sampling was utilized, where every member of the population participated in the research. Consequently, the sample consisted of 61 eighth-grade students (Creswell, 2023).

According to Mueller (2005), portfolio assessment is an authentic assessment that systematically documents students' learning through a collection of their work over time. It emphasizes evidence of learning, clear assessment criteria, student reflection, continuous feedback, and demonstration of learning progress. Based on these principles, the portfolio assessment indicators used in this study are presented in Table 1.

Table 1. Portfolio Assessment Indicators

Variable	Indicators
Portfolio Assessment	Collection of students' work
	Evidence of learning
	Student reflection
	Assessment using clear criteria (rubrics)
	Feedback and learning improvement

The study used secondary data obtained from the school's official assessment records. Two variables were analyzed: Final Test Score, representing students' final examination scores, and Portfolio Assessment Score, representing students' portfolio scores compiled from assignments, projects, classroom exercises, and other learning activities completed throughout one semester. The portfolio scores were determined by the subject teachers using a standardized rubric provided by the school, which assessed criteria including completeness of work, quality of content, creativity, and consistency of effort throughout the semester. Each student's portfolio was evaluated by a single rater, and the total score was calculated as the sum of rubric-based ratings for each component of the portfolio. Since the study relied on official school records, no additional validity or reliability testing of the research instrument was conducted. Data analysis was performed using IBM SPSS Statistics version 26. The examination was performed in four phases. Initially, descriptive statistics were computed to encapsulate the features of the data, encompassing the mean, median, standard deviation, and both minimum and maximum values for each variable (Field, 2018).

Second, to ascertain whether the data followed a normal distribution, a normality assessment was performed using the Kolmogorov–Smirnov test. If the significance value (Sig.) was greater than 0.05 ($p > 0.05$), the data were deemed normally distributed; if the significance value was less than 0.05 ($p < 0.05$), the data were deemed non-normally distributed (Pallant, 2020).

Third, because the data did not match the normality assumption, the Wilcoxon Signed-Rank Test was used to answer the original study question about the existence of a significant difference between Final Test scores and Portfolio Assessment scores. This non-parametric test can be used to compare two related

samples (Field, 2018).

Fourth, to address the second research question regarding the connection between Final Test scores and Portfolio Assessment scores, Spearman's Rank Correlation was applied since one of the variables did not follow a normal distribution. Spearman's correlation is a statistical technique that does not require a normal distribution, used to assess the strength and direction of the relationship between two variables (Pallant, 2020).

The significance level for all statistical analyses was set at $\alpha = 0.05$. A statistically significant difference or association was indicated by a significance value less than 0.05, while no statistically significant difference or relationship was indicated by a significance value greater than 0.05. Cohen's (1988) criteria were used to assess the correlation coefficient's strength: coefficients close to 0.10 suggest a weak association, 0.30 indicate a moderate association, and 0.50 or higher indicate a significant association

RESULT

Descriptive Statistics

A comprehensive statistical analysis was performed to present a clear overview of the central tendencies and dispersion traits of the research data, which encompassed the final exam results and portfolio evaluations of the eighth-grade cohort at SMP Negeri 1 Susoh. Table 2 displays the detailed findings derived from the descriptive statistical analysis.

Table 2. Descriptive Statistics for Final Test Scores and Portfolio

Statistics		Assessments	
		FINAL TEST	PORTFOLIO ASSESSMEN T
N	Valid	61	61
	Missing	0	0
Mean		83.9508	50.6393
Median		82.0000	55.0000
Std. Deviation		5.13299	31.81511
Minimum		79.00	.00
Maximum		98.00	100.00

As indicated in Table 2, the research sample consisted of exactly 61 students. The Final Test scores exhibited a relatively tight cluster, ranging from a minimum of 79 to a maximum of 98, yielding a high mean of 83.95 and a remarkably low standard deviation of 5.13. This suggests a strong, consistent baseline of theoretical knowledge retention among the students. Conversely, the Portfolio Assessment scores presented a drastically different landscape. With scores spanning the absolute minimum of 0 to a perfect maximum of 100, the mean settled much lower at 50.64, accompanied by a strikingly high standard deviation of 31.82.

The sheer magnitude of the standard deviation in the portfolio scores demands closer scrutiny. A portfolio score of 0 may indicate that a student did not

submit the required portfolio tasks or received no credit according to the school's assessment records. The inclusion of these zero scores contributed to the relatively large variability observed in the portfolio assessment results. This statistical reality points to a bimodal-like behavioral trend in the classroom: while some students excelled and fully engaged with the continuous demands of portfolio curation, a substantial subset struggled with the autonomy, organization, or motivation required, leading to non-participation. Ultimately, these descriptive statistics reveal that students' performance on continuous, process-oriented tasks is vastly more unpredictable and varied than their performance on structured, one-time final exams.

Normality Test

A normality test was conducted to determine whether the data met the assumptions of a normal distribution before performing inferential analysis. Given that the study sample consisted of 61 students, the Kolmogorov–Smirnov test was used to assess normality. The results of the normality test are presented in Table 3.

Table 3. Results of the Normality Test

	Tests of Normality					
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	Df	Sig.	Statistic	df	Sig.
FINAL TEST	.246	61	< .001	.802	61	< .001
PORTFOLIO ASSESSMENT	.112	61	.053	.918	61	.001

a. Lilliefors Significance Correction

The results of the Kolmogorov–Smirnov test indicated that the Final Test scores were not normally distributed ($p < .001$), whereas the Portfolio Assessment scores were normally distributed ($p = .053 > .05$), as presented in Table 3. Because one of the variables violated the assumption of normality, non-parametric statistical tests were employed. Accordingly, the Wilcoxon Signed-Rank Test was used to examine the differences between Final Test and Portfolio Assessment scores, while Spearman's Rank Correlation was applied to determine the relationship between the two variables.

Testing for Differences in Final Test and Portfolio Assessment Scores

The Wilcoxon Signed-Rank Test was applied to assess if a difference exists between the scores of the Final Test and Portfolio Assessment. Table 4 displays the results of the test.

Table 4. Results of the Wilcoxon Signed-Rank Test

Test Statistics	
	PORTFOLIO ASSESSMENT - FINAL TEST
Z	-6.318 ^b
Asymp. Sig. (2-tailed)	< .001
a. Wilcoxon Signed Ranks Test	
b. Based on positive ranks.	

The Wilcoxon test yielded a notable Z-value of -6.318 alongside an asymptotic significance level of $p < .001$. Because this p-value is decidedly less than the stringent 0.05 threshold, it confirms a highly significant statistical difference between the Final Test and Portfolio Assessment scores. Delving deeper into the ranked data uncovers a compelling pattern: out of the 61 participants, an overwhelming majority of 52 students received lower scores on their Portfolio Assessment compared to their Final Test. Only 8 students managed to outscore their Final Test with their Portfolio, and merely 1 student achieved perfectly equivalent scores across both formats. This incontrovertible evidence indicates that, within this specific educational cohort, Final Test scores are generally and substantially higher than their portfolio-based counterparts.

Testing the Relationship Between Final Test Scores and Portfolio Assessment Scores

To determine the relationship between final test scores and portfolio assessment scores, Spearman's correlation test was used. The test results are presented in Table 5.

Table 5. Results of the Spearman Correlation Test

Variables	Spearman's rho (ρ)	Sig. (2-tailed)	Interpretation
Final Test – Portfolio Assessment	0.565	< .001	Significant strong positive correlation

The results of the Spearman's rank correlation analysis revealed a correlation coefficient of 0.565 with a significance level of $p < .001$. These findings indicate a statistically significant strong positive correlation between the Final Test and Portfolio Assessment scores. This suggests that students who achieved higher scores on the Final Test also tended to obtain higher scores on the Portfolio Assessment.

DISCUSSION

The empirical Findings extracted from this investigation present a fascinating, multi-layered perspective on educational measurement at SMP Negeri 1 Susoh. First and foremost, the data unequivocally demonstrate a notable and statistically significant disparity between Final Test scores and Portfolio Assessment scores, with the former generally eclipsing the latter. This discrepancy is far from an anomaly; rather, it vividly illustrates how the structural characteristics and cognitive demands of different assessment formats directly shape student performance outcomes. The Final Test, by its very design, requires students to recall and synthesize material within a narrow, highly structured time frame. It leans heavily on short-term retention, test-taking strategies, and the ability to perform under pressure. Conversely, the Portfolio Assessment is an expansive, slow-burning process. It evaluates a student's stamina, intrinsic motivation, organizational skills, and capacity for continuous self-reflection over an entire semester.

The phenomenon of substantially lower portfolio scores, highlighted by a staggering standard deviation and the presence of absolute zero scores, warrants deep contextual analysis. In many Indonesian educational settings, particularly outside major metropolitan hubs, the learning culture has historically been heavily

skewed toward summative evaluations. One possible explanation is that students may have been more familiar with traditional summative assessments than with continuous portfolio assessment. Portfolio assessment requires sustained engagement, independent learning, and consistent completion of academic tasks throughout the semester. These findings are consistent with previous studies suggesting that students who are accustomed to summative assessment may experience challenges when adapting to portfolio-based assessment (Zaabalawi, 2024). These findings perfectly echo the assertions made by Zaabalawi (2024), who argues that while portfolios offer a profoundly comprehensive picture of intellectual development, their success relies heavily on a student's prior training in self-regulation and time management.

Furthermore, the complexities of implementing authentic assessments cannot be divorced from the teacher's reality. Evaluating 61 highly individualized portfolios using a multifaceted rubric is an astronomically more time-consuming task than grading multiple-choice or short-answer final exams. The cognitive load placed on the educator to provide fair, consistent, and meaningful feedback across so many disparate pieces of evidence is immense. As Hanafiah et al. (2024) pointed out, logistical bottlenecks in curriculum management often dilute the intended efficacy of portfolio systems. It is entirely plausible that the lower portfolio scores partially reflect the logistical friction of executing complex formative assessments within a traditional school infrastructure.

Despite these challenges and structural differences, the correlation test illuminates a critical alignment: a significant, positive relationship ($p = 0.565$) exists between the two assessments. This correlation is the conceptual bridge of the study. It proves that cognitive capability (as captured by the final test) and academic diligence (as captured by the portfolio) are not entirely isolated traits. A student who genuinely understands the material is likely to demonstrate that understanding regardless of the medium. Nevertheless, the correlation was not perfect, suggesting that the two assessment methods measure related but not identical aspects of student achievement. Therefore, the findings support the complementary use of Final Test and Portfolio Assessment in providing a more comprehensive evaluation of students' learning outcomes.

From a pedagogical perspective, this study strongly supports the integration policy in the Merdeka Curriculum currently in effect in Indonesia. This curriculum explicitly mandates a shift from high-stakes final assessments toward a balanced and holistic evaluation system that values the "process" as much as the "outcome." By combining data from rigorous final exams and rich narrative portfolio assessments, educators at SMP Negeri 1 Susoh and, of course, across the country, can build a far more objective, empathetic, and comprehensive profile of a student's true abilities.

CONCLUSION

In conclusion, this study demonstrates that Final Test scores and Portfolio Assessment scores provide distinct yet complementary perspectives on students' academic achievement. The statistically significant difference, characterized by higher scores on the Final Test, underscores the differing cognitive and behavioral demands of the two assessment methods. Concurrently, the strong positive correlation between the two assessment methods indicates that students who

performed well on the Final Test also tended to achieve higher Portfolio Assessment scores, suggesting that both methods measure related but not identical aspects of students' academic achievement. Rather than viewing them as competing approaches, educational stakeholders should recognize Final Tests and Portfolio Assessments as complementary components of a holistic assessment system. While the Final Test provides a reliable measure of students' mastery of learning content at the end of instruction, Portfolio Assessment documents students' learning progress, engagement, and reflective development throughout the learning process. Therefore, integrating both assessment methods provides teachers with a more comprehensive and multidimensional framework for evaluating students' competencies.

The practical implications of these findings are substantial. Teachers and school administrators are encouraged to integrate Portfolio Assessment with traditional Final Tests to ensure that student evaluation reflects both learning outcomes and the learning process. Furthermore, schools should provide adequate guidance to help students develop self-regulated learning skills and effectively complete portfolio tasks. Future studies are recommended to involve larger samples from multiple schools and to investigate additional factors, such as students' motivation, self-regulated learning, and critical thinking skills, to provide a broader understanding of the effectiveness of portfolio-based assessment.

REFERENCES

- Al-Hawamdeh, B.O., Hussen, N., & Abdelrasheed, N.S. (2023). Portfolio vs. summative assessment: impacts on EFL learners' writing complexity, accuracy, and fluency (CAF); self-efficacy; learning anxiety; and autonomy. *Language Testing in Asia*, 13, 1-29.
- Ali, Q. I. (2024). Towards more effective summative assessment in OBE: a new framework integrating direct measurements and technology. *Discover Education*, 3(1).
- Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. *Assessment & Evaluation in Higher Education*, 31, 399 - 413.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum Associates.
- Creswell, J. W., & Creswell, J. D. (2023). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (6th ed.). SAGE Publications.
- Field, A. (2018). Discovering Statistics Using IBM SPSS Statistics, 5th edition. SAGE Publications Ltd.
- French, S., Dickerson, A., & Mulder, R. (2024). A review of the benefits and drawbacks of high-stakes final examinations in higher education. *Higher Education*, 88(3).
- Hanafiah, H., Haryono, W., Juwanto, J., Pirngadi, H., & Ramadhan, F. R. (2024). Evaluation Of Curriculum Management And Assessment Systems At The Junior High School Level: Challenges And Solutions. *IJOBBA : International Journal of Bunga Bangsa Cirebon*. 3(2).
- Mueller, J. (2005). The Authentic Assessment Toolbox: Enhancing Student Learning through Online Faculty Development.
- Naka, L. (2025). Portfolio as an Assessment Tool: Impact on Student Participation



and Improvement in EFL Learning. *International Journal of Language Testing*, 15(2), 134-152.

Pallant, J. 2020. SPSS survival manual: A step-by-step guide to data analysis using IBM SPSS (7th ed.). London, UK: McGraw-Hill Education

Zaabalawi, R.S., Zaabalawi, J. (2024). Portfolios versus exams: a study to gauge the better student assessment tool. *Lang Test Asia* 14(1), 28.

Zhan, Y., Boud, D. & Du, Z. (2025). Designing for authentic assessment: a scoping review. *Higher Education*

