CONOCIMIENTO: Journal of Education and Social Sciences
Vol. 1, No. 1 (2026)
https://journal.ziacenter.org/index.php/conocimiento

Gender Differences in Numeracy Performance among Grade X Students: A Comparative Study in an Indonesian Islamic Senior High School

[1*]Muh. Safar, [1]Dahliah, [1]Andi Nabiah Armini
[1]Master of Mathematics Education, Graduate Program, Muhammadiyah University of Makassar, Makassar, Indonesia

Abstract

Numeracy is a fundamental component of mathematical competence, yet empirical evidence regarding gender-related differences in numeracy performance remains inconsistent across educational contexts. This study examined whether numeracy performance differed between male and female Grade X students at SMAIT Nurul Asafa Takalar, an Islamic Integrated Senior High School in Indonesia. A quantitative comparative design was employed involving the entire Grade X cohort of 60 students, comprising 30 male and 30 female students. Data were collected through a numeracy test and analyzed using JASP. The analysis included descriptive statistics, Shapiro–Wilk test (Shapiro & Wilk, 1965)s, the Brown–Forsythe test, an independent-samples t-test, and Cohen's d effect size. Female students obtained a higher mean numeracy score than male students (78.03 vs. 74.53), representing a mean difference of 3.50 points. However, the independent-samples t-test indicated that the difference was not statistically significant, $t(58) = -1.531$, $p = .131$. The standardized mean difference was $d = -0.3954$ ($|d| \approx 0.40$), indicating an observed difference between conventional small and medium effect-size benchmarks. These findings suggest that the descriptive difference in numeracy performance was not sufficient to establish a statistically significant gender difference within the studied cohort. The study contributes context-specific evidence to the mixed literature on gender and numeracy and highlights the importance of interpreting group differences alongside effect size, sample characteristics, and broader learner and instructional factors.

Keywords: Numeracy; Gender Differences; Numeracy Performance; Mathematics Education; Secondary School Students.
Corresponding author: E-mail address: muhsafar14@admin.sma.belajar.id

1. Introduction

Numeracy is widely regarded as an essential part of mathematical competence because it extends beyond the ability to perform calculations. It involves using mathematical, quantitative, and statistical information to understand situations, reason about relationships, interpret representations, and make informed decisions in everyday and social contexts (Gal et al., 2020; Geiger et al., 2015). This broader conception places numeracy within the wider goal of mathematical literacy, in which learners are expected not only to know mathematical procedures but also to formulate, apply, and interpret mathematics in meaningful situations (Kilpatrick, 2001; Yore et al., 2007). Such competencies are particularly relevant in contemporary education because mathematical

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

reasoning is closely connected with participation in social life, further education, employment, and responsible citizenship. Large-scale assessments such as PISA have consequently provided an important basis for examining mathematical literacy across educational systems and for understanding how student characteristics and learning contexts are associated with achievement (Anderson et al., 2007; Stacey, 2011). Recent research likewise continues to position numeracy as a central concern in mathematics education, including work addressing mathematical literacy, assessment, instructional practices, and opportunities for students to develop numeracy in different contexts (Hodgen & Noyes, 2026; Tran, 2026).

As attention to numeracy has grown, research has moved beyond defining the construct toward examining how students develop and demonstrate it. Numeracy is embedded in contextual practices rather than being limited to decontextualized mathematical procedures, making the relationship between students, learning opportunities, instructional approaches, and assessment particularly important (Gal et al., 2020; Geiger et al., 2015). Research has shown that mathematical literacy can be supported through approaches that connect mathematics with authentic and culturally meaningful situations. For example, recent studies have investigated ethnomathematics, local wisdom, technology-supported learning, and contextual assessment as ways of strengthening students' opportunities to engage with mathematical ideas meaningfully (Deda et al., 2026; Maysarah et al., 2026; Payadnya et al., 2026; Sagita et al., 2026). Other studies have examined technological and pedagogical innovations, including augmented reality, adaptive e-learning, and mobile-based learning, demonstrating the continuing effort to create learning environments that support mathematical literacy development (Leton et al., 2026; Muhaimin et al., 2026; Talib et al., 2026). These developments indicate that numeracy performance should be understood as a multidimensional educational outcome rather than as a simple reflection of students' demographic characteristics.

Against this broader background, gender remains an important but unresolved issue in mathematics education research. Studies have not produced a uniform account of how gender is related to mathematical performance or to the factors associated with learning. For instance, Hemmings et al. (2011), using data from Australian secondary school students, found that females tended to report more positive attitudes toward mathematics, while gender itself was no longer a significant predictor of mathematics achievement after prior achievement and attitudes were taken into account. Similarly, research using large-scale TIMSS data has shown that gender differences can emerge in mathematics-related psychological constructs even when achievement and other background variables are considered. Mejía-Rodríguez et al. (2021), for example, identified significant gender differences in mathematics self-concept across many of the countries examined, generally

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

favoring boys, and found that these differences persisted after controlling for achievement and parental characteristics. These findings suggest that gender-related patterns in mathematics education cannot be interpreted solely through differences in test scores. Instead, gender may intersect with attitudes, self-concept, prior achievement, learning experiences, and other contextual factors.

This complexity also means that group averages alone may provide only a partial picture of gender-related differences in mathematics learning. Research has increasingly examined additional variables that may explain variation in mathematical and numeracy outcomes. The home numeracy environment, for example, has been shown to contribute uniquely to early numeracy after accounting for child characteristics and the home literacy environment (Segers et al., 2015). Research on mathematics achievement has similarly emphasized the role of prior achievement and attitudes rather than treating gender as an isolated explanatory factor (Hemmings et al., 2011). More recent studies have examined factors associated with numeracy through data-driven approaches, while others have focused on students' mathematical thinking processes and decision-making characteristics (Roza et al., 2026; Rosita et al., 2026). At the instructional level, studies have explored differentiated learning, contextual tasks, and culturally situated learning as approaches for responding to differences among learners (Deda et al., 2026; Siregar et al., 2026). Collectively, these approaches indicate that investigating demographic differences is most informative when it is situated within a broader understanding of the factors that shape numeracy development and performance.

Recent intervention studies further illustrate the importance of learning context. Maysarah et al. (2026) examined an Ethno-Project-Based Learning model supported by GeoGebra and reported improved mathematical literacy among senior high school students compared with conventional instruction. Talib et al. (2026) investigated Android-based modules integrating deep learning and intrapersonal intelligence to support secondary students' mathematical literacy. Similarly, Muhaimin et al. (2026) examined augmented reality as a technological development for mathematics learning, while Leton et al. (2026) combined augmented reality with local wisdom in teaching exponents. Other research has emphasized the importance of designing meaningful numeracy tasks and assessment opportunities, including contextual assessment frameworks (Lestari et al., 2026; Sagita et al., 2026). These studies do not directly resolve the question of whether male and female students differ in numeracy performance; rather, they demonstrate that students' numeracy outcomes are situated within instructional, cognitive, technological, and contextual conditions. Consequently, a gender comparison should be interpreted cautiously and should not be assumed to represent a causal effect of gender.

The literature therefore leaves a specific empirical question unresolved. Much of the

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

recent work represented in the Scopus literature has focused on improving mathematical literacy through instructional innovation, examining contextual opportunities for numeracy development, or investigating student characteristics associated with mathematics learning (Deda et al., 2026; Maysarah et al., 2026; Payadnya et al., 2026; Roza et al., 2026; Tran, 2026). Other studies have investigated gender differences in mathematics-related constructs or the relationship between gender and mathematics achievement while accounting for additional variables (Hemmings et al., 2011; Mejía-Rodríguez et al., 2021). However, these strands of research do not establish whether differences in numeracy performance are consistently observable across educational contexts, particularly in a Grade X Islamic Integrated Senior High School setting. The present study addresses this gap by examining numeracy performance directly through a comparative analysis of male and female students within the same school context. This focus is important because a within-school comparison provides a more contextually bounded examination of the observed distribution of numeracy performance while avoiding claims that gender itself causes differences in achievement. In addition to statistical significance, examining the magnitude of the observed difference through an estimated effect size provides a more complete interpretation than relying on the p-value alone.

Accordingly, this study examines whether numeracy performance differs between male and female Grade X students at SMAIT Nurul Asafa Takalar. The study compares the two naturally formed gender groups using numeracy test scores and evaluates both the statistical significance and standardized magnitude of the observed difference. The study is guided by the following research question: Is there a difference in numeracy performance between male and female students in Grade X at SMAIT Nurul Asafa Takalar? By addressing this question, the study contributes empirical evidence from a specific secondary-school context to the broader discussion of gender and numeracy, while recognizing that observed differences in performance should be considered alongside the wider instructional and learner-related factors identified in previous research.

2. Methodology

Research Design

A quantitative comparative design was used in this study. The design examined numeracy performance across male and female students who were already grouped by gender; no experimental or control conditions were assigned, and no intervention was administered. This approach was appropriate for the research question because it compares the numeracy performance of two naturally formed groups (Sugiyono, 2019).

The study involved Grade X students at SMAIT Nurul Asafa Takalar. Gender was treated as the grouping variable, while numeracy performance, measured through a

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

numeracy test, served as the outcome variable. The analysis compared the central tendency of numeracy scores across the two naturally formed groups.

Participants

All Grade X students enrolled at SMAIT Nurul Asafa Takalar participated in the study. A total of 60 students participated in the study, consisting of 30 male students and 30 female students. The two comparison groups were therefore equal in size. Because the entire Grade X cohort was included, the participants represented the full cohort available at the school rather than a selected subsample.

Including the full Grade X cohort allowed the study to compare male and female students within the same school context. The analysis was therefore focused specifically on the observed differences between these two groups of Grade X students.

Data Collection and Instrument

A numeracy test was administered to all 60 participants. The test provided individual scores representing students' numeracy performance. The same assessment procedure was used for both groups, providing a common basis for comparison.

The manuscript identifies the instrument only as a numeracy test and does not report its number of items, item format, scoring rubric, development or adaptation process, or evidence of validity and reliability. These details have therefore not been inferred here. Such information should be incorporated if documentation of the instrument and its psychometric properties is available.

Data Analysis

The data were analyzed in JASP. The analysis first described the numeracy scores and then examined the difference between the two gender groups.

Descriptive statistics were first calculated separately for male and female students. The descriptive analysis included the number of valid observations, mean, standard deviation, median, minimum, and maximum scores. These measures describe the performance of each group and indicate the direction and size of the raw difference between the group means.

The Shapiro–Wilk test was then used to examine distributional assumptions. Normality was assessed for the relevant data before the comparative analysis. The analysis reported both group-specific Shapiro–Wilk results and a Shapiro–Wilk test for the residuals. Because the reported results include both forms of assessment, they are presented separately rather than treated as interchangeable indicators of normality.

The Brown–Forsythe test was used to examine whether the variances differed between the two groups. This procedure was used to assess whether the variability in

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

numeracy scores differed statistically between male and female students. The resulting test statistic and probability value were considered when evaluating the suitability of the subsequent mean-comparison procedure.

An independent-samples t-test was then used to compare the mean numeracy scores of male and female students. This analysis directly addressed the research question: whether numeracy performance differed between male and female Grade X students at SMAIT Nurul Asafa Takalar. The analysis was conducted using a significance level of $\alpha = .05$. The null hypothesis stated that there was no difference in mean numeracy scores between male and female students, whereas the alternative hypothesis stated that a difference existed between the two groups.

Cohen's d was also calculated to describe the standardized magnitude of the observed difference. Effect size was included to provide a standardized estimate of the difference between the two groups rather than relying solely on the probability value from the t-test. The sign of Cohen's d was retained to indicate the direction of the group comparison, while its absolute magnitude was considered when describing the size of the standardized difference. The calculation followed the effect-size framework reported by Cohen (1988).

The statistical results were interpreted in relation to the research question and the observed data. Statistical significance and effect size were treated as distinct forms of evidence: the p-value was used to evaluate evidence against the null hypothesis at the prespecified $\alpha = .05$ level, whereas Cohen's d was used to quantify the standardized magnitude of the observed group difference.

3. Results and Discussion

Descriptive Statistics

Table 1 summarizes students' numeracy performance by gender. The analysis involved 60 Grade X students, comprising 30 male students and 30 female students. The descriptive statistics include the mean, standard deviation, median, minimum, and maximum scores for each group.

Table 1: Descriptive Statistics of Numeracy Performance by Gender

Descriptive Statistics	**Value**	
	Male	Female
Valid	30	30
Missing	0	0
Median	77.50	80.00
Mean (arithmetic)	74.53	78.03
Std. Deviation	9.719	7.889

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Descriptive Statistics		
	Value	
Shapiro-Wilk	0.9348	0.9253
P-value of Shapiro-Wilk	.066	.037
Minimum	54.00	60.00
Maximum	89.00	89.00

Female students recorded a higher mean numeracy score than male students. The mean score for female students was 78.03 (SD = 7.889), compared with 74.53 (SD = 9.719) for male students. The difference between the two group means was therefore 3.50 points in favor of female students. The median score was also higher for female students (80.00) than for male students (77.50). The observed scores ranged from 60 to 89 among female students and from 54 to 89 among male students. The female group therefore had the higher central tendency, although the score ranges overlapped substantially.

Normality Test

Normality was examined with the Shapiro–Wilk test, with the results reported in Table 2.

Table 2: Shapiro–Wilk Test of Normality

Test of Normality (Shapiro-Wilk)		
Residuals	W	p
Value	0.943	.007

Note. Significant results suggest a deviation from normality.

The Shapiro–Wilk results indicated that the male group did not show a statistically significant deviation from normality, W = .935, p = .066. In contrast, female students showed a statistically significant deviation from normality, W = .925, p = .037. The analysis of the residuals also produced a statistically significant result, W = .943, p = .007, indicating a deviation from normality in the residual distribution. This result is reported transparently because it represents an assumption relevant to the subsequent mean-comparison analysis.

Test of Equality of Variances

The Brown–Forsythe test was used to assess the equality of variances between the male and female groups; the result is reported in Table 3.

Table 3: Test of Equality of Variances (Brown–Forsythe)

Test of Equality of Variances (Brown-Forsythe)				
	F	df_1	df_2	p
Value	1.171	1	58	.284

The Brown–Forsythe test produced $F(1, 58) = 1.171$, $p = .284$. Because the p-value was greater than .05, there was no statistically significant evidence of a difference in score variances between the male and female groups. The variability of numeracy scores can

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

therefore be considered comparable across the two groups for the purpose of the subsequent comparison.

Comparison of Numeracy Scores

An independent-samples t-test was conducted to determine whether the observed difference in numeracy performance between male and female students was statistically significant (Table 4).

Table 4: Independent-Samples *t*-Test

Independent Samples T-Test					
	t	df	p	Cohen's d	SE Cohen's d
Value	-1.531	58	.131	-0.3954	0.2632

Note. Student's t-test.

The independent-samples t-test produced t(58) = −1.531, p = .131. Because p exceeded the prespecified .05 significance level, the null hypothesis was not rejected. Thus, the analysis did not detect a statistically significant difference in numeracy performance between male and female Grade X students at SMAIT Nurul Asafa Takalar.

Female students had the higher observed mean, but the 3.50-point difference was not statistically significant. Therefore, the descriptive difference should not be interpreted as evidence of a statistically established gender difference in numeracy performance within this sample.

Effect Size

Cohen's d was calculated alongside the significance test to estimate the standardized magnitude of the observed difference. As reported in Table 4, Cohen's d was −0.3954, with a standard error of 0.2632. Because the sign indicates the direction of the group comparison, the absolute magnitude of the effect size was approximately 0.40.

Using Cohen's (1988) conventional reference points, the observed standardized difference falls between the commonly used small (d = 0.20) and medium (d = 0.50) benchmarks. This estimated effect size indicates that the observed difference in mean numeracy scores was not negligible in standardized terms, but it should be interpreted together with the nonsignificant p-value and the characteristics of the present sample.

Overall, female students had the higher observed mean numeracy score, with a mean difference of 3.50 points. However, the independent-samples t-test did not provide sufficient statistical evidence to conclude that the two groups differed in numeracy performance, t(58) = −1.531, p = .131. The estimated effect size (d = −0.3954) indicates a standardized difference of approximately 0.40 in the observed direction. These results are considered below in relation to previous research on gender and numeracy performance.

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Discussion

This study compared numeracy performance between male and female Grade X students at SMAIT Nurul Asafa Takalar. The findings become clearer when the group means, inferential test, and effect size are considered together. Although female students obtained a higher mean numeracy score than male students, with a difference of 3.50 points, the independent-samples t-test did not provide sufficient statistical evidence to conclude that the two groups differed in their mean numeracy performance, $t(58) = -1.531$, $p = .131$. The standardized mean difference was Cohen's $d = -0.3954$, indicating an observed difference that was larger than a negligible difference but remained below Cohen's conventional benchmark for a medium effect (Cohen, 1988). Taken together, the results do not support interpreting the higher female mean as evidence of a systematic gender difference in numeracy within this cohort. Rather, the results indicate that male and female students in this particular school displayed broadly comparable numeracy performance, while the observed difference in means warrants further investigation.

This result illustrates why descriptive differences, statistical evidence, and effect magnitude need to be distinguished. The higher female mean suggests a difference in the observed scores, but the p-value does not provide sufficient evidence to reject the null hypothesis at the .05 level. At the same time, the estimated effect size shows that the difference is not necessarily zero in practical terms. Cohen's (1988) framework places an effect of approximately .40 between the conventional small and medium benchmarks. However, the effect size should be interpreted together with the sample size and uncertainty surrounding the estimate rather than as independent evidence that gender has an effect on numeracy. The present study included 60 students, equally divided between the two groups. Consequently, the findings are most appropriately understood as evidence concerning this particular Grade X cohort rather than as evidence of a general gender pattern. This distinction is especially important in numeracy research because performance may reflect multiple interacting characteristics of students and their learning environments rather than a single demographic grouping.

The findings are consistent with literature showing that gender patterns in mathematics-related outcomes vary across contexts. Hemmings et al. (2011), for example, reported that female students in an Australian secondary-school context tended to express more positive attitudes toward mathematics, but gender was not a significant predictor of mathematics achievement after prior achievement and attitudes were taken into account. This comparison is relevant because it shows that an observed gender-related pattern can change when other learner characteristics are taken into account. Similarly, Mejía-Rodríguez et al. (2021), using TIMSS 2015 data across countries, identified gender differences in mathematics self-concept, generally favoring boys, but their analysis also showed that such

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

differences should be considered alongside achievement and parental characteristics. Together, these findings suggest that gender-related patterns may depend on the outcome being measured. A difference in mathematics self-concept, for example, does not necessarily translate into a statistically detectable difference in numeracy performance.

The present result is also similar to Mellyzar et al. (2021), who reported somewhat higher numeracy performance among female students but no statistically significant difference between the groups. More importantly, their study identified a strong relationship between self-efficacy and numeracy performance. This result helps place the current finding in context: gender should not be considered apart from psychological characteristics that may be more closely related to students' engagement with mathematical tasks. If students with different gender identities possess varying levels of self-efficacy, learning experience, motivation, or persistence, these characteristics may contribute to performance differences within each group and potentially reduce the explanatory value of gender as a simple grouping variable. The present study did not measure these variables, so their contribution cannot be tested here. Nevertheless, the evidence from Mellyzar et al. (2021) indicates why a non-significant gender comparison should not be interpreted as evidence that individual differences in numeracy are unimportant.

The result can also be considered alongside Indonesian studies conducted at different educational levels. Mariamah et al. (2021) reported a higher proportion of high numeracy performance among female elementary-school students, while Maryam et al. (2025) found a significant gender difference in numeracy among junior high school students, with female students obtaining higher performance. These findings differ from the absence of a statistically significant difference in the present Grade X cohort. Rather than treating these results as contradictory, the differences across studies suggest that the relationship between gender and numeracy may vary with educational level, student characteristics, assessment conditions, and local learning contexts. A result observed among elementary or junior high school students cannot automatically be generalized to senior high school students. The transition to higher levels of schooling may involve changes in mathematical content, instructional expectations, assessment demands, and students' prior learning experiences, all of which may affect the extent to which group differences become visible in test performance.

The qualitative evidence reported by Khotimah and Siswono (2025) provides an additional explanation for why differences in final scores may be limited even when differences between male and female students exist in mathematical activity. Their study indicated that gender-related differences could be more apparent in the problem-solving process than in the final outcome. In particular, differences were observed in how students approached stages of problem solving, including planning and organizing their solutions.

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

This perspective is particularly relevant to the present study because the outcome variable was the score obtained from a numeracy test. A score captures the correctness or performance represented by the assessment but does not necessarily reveal how students interpreted the problem, selected a strategy, monitored their reasoning, or arrived at an answer. Therefore, the absence of a statistically significant difference in total scores does not establish that male and female students necessarily use identical cognitive or problem-solving approaches. It instead indicates that such possible process differences were not examined by the present outcome measure.

The multidimensional nature of numeracy provides another important consideration. Geiger et al. (2015) conceptualized numeracy broadly, emphasizing that it involves more than procedural calculation and encompasses the use of mathematical knowledge in different situations and practices. Gal et al. (2020) similarly emphasized the contextual and functional dimensions of numeracy. From this perspective, an overall test score captures only one observable expression of numeracy competence. Students may perform similarly on an aggregate score while differing in specific aspects of reasoning, interpretation, representation, or application. Research on contextual numeracy has likewise demonstrated the importance of the opportunities students receive to engage with meaningful mathematical situations (Deda et al., 2026; Sagita et al., 2026). Accordingly, the absence of a statistically significant gender difference should be interpreted as a finding about the measured numeracy outcome, not as evidence that gender is unrelated to every dimension of numeracy development.

Research on factors associated with numeracy also supports this interpretation. Roza et al. (2026), using educational assessment data from Indonesia and decision-tree analysis, found that numeracy proficiency was associated with multiple factors, with literacy skills emerging as the dominant predictor in their model. Their analysis also identified psychological well-being, learning-environment quality, teacher competence, and school inclusivity as associated factors. Although that study used a different methodology and should not be used to infer causal relationships in the present context, its findings illustrate the complexity of numeracy performance and provide a useful contrast to analyses that focus on a single demographic characteristic. Similarly, Segers et al. (2015) demonstrated the contribution of the home numeracy environment to early numeracy after accounting for other child and home characteristics. Taken together, these studies place students' numeracy performance within a wider set of educational, psychological, familial, and contextual conditions. The current finding therefore supports a more cautious interpretation of gender differences: gender may describe group membership, but it does not provide a sufficiently comprehensive account of why students perform differently.

This interpretation also accords with recent research emphasizing instructional

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

opportunities and learning-environment design. Studies involving ethnomathematics, contextual learning, technology-supported instruction, and differentiated learning have demonstrated that mathematical literacy and numeracy can be approached through diverse pedagogical conditions (Deda et al., 2026; Maysarah et al., 2026; Payadnya et al., 2026; Siregar et al., 2026). Other studies have explored augmented reality, mobile-based learning, and adaptive digital environments as means of expanding students' opportunities to engage with mathematical ideas (Leton et al., 2026; Muhaimin et al., 2026; Talib et al., 2026). These studies do not directly explain the gender comparison observed at SMAIT Nurul Asafa Takalar, but they strengthen an important interpretation of the present findings: comparable numeracy performance between male and female students may be understood within a learning environment in which instructional opportunities, rather than gender alone, provide important conditions for students' performance. At the same time, because the present study did not measure instructional practices or learning opportunities, this interpretation remains theoretically informed rather than an empirically tested explanation.

From a methodological standpoint, the findings also show the value of reporting effect size alongside significance testing. The p-value answers a specific inferential question concerning the evidence against the null hypothesis, whereas Cohen's d describes the standardized magnitude and direction of the observed group difference. Reporting both therefore provides a fuller account of the result than reporting the t-test alone. The current combination of a non-significant p-value and an effect size of approximately .40 illustrates why statistical significance and substantive magnitude should not be treated as interchangeable concepts. Nevertheless, the effect should not be described as evidence of a practically important gender difference without considering its uncertainty and the study's limited sample. Future studies with larger and more diverse samples could determine whether the observed direction and magnitude are replicated across schools and educational settings.

The findings have several implications for mathematics education. First, the findings provide no empirical basis within this cohort for treating male and female students as fundamentally different groups in terms of overall numeracy performance. Teachers may therefore need to avoid assuming that one gender requires systematically different numeracy instruction solely because of gender membership. Second, the findings reinforce the value of examining learner characteristics beyond demographic categories. Variables such as self-efficacy, motivation, prior achievement, learning experience, and problem-solving processes may provide more informative explanations of individual variation in numeracy performance, as suggested by the literature discussed above. Third, assessment research should consider whether aggregate scores adequately capture the dimensions of numeracy that are educationally relevant. Combining quantitative performance measures

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

with process-oriented or qualitative evidence could reveal differences that a final score alone cannot detect.

The main contribution of this study is its context-specific evidence that a descriptive difference in numeracy scores between male and female Grade X students does not necessarily correspond to a statistically significant group difference. The study also contributes by reporting the magnitude of the observed difference rather than relying exclusively on statistical significance. This is particularly relevant given the inconsistent findings across previous studies and the multidimensional character of numeracy. However, the contribution should be interpreted within the boundaries of the research design. The study involved the complete Grade X cohort of a single school, and the numeracy instrument is described in the manuscript only as a numeracy test; information about its item structure, validity, reliability, and scoring procedure is not provided. In addition, the study did not measure psychological, instructional, family, or contextual variables that may account for individual variation in performance. The normality analysis also indicated a deviation from normality for the reported residuals ($p = .007$), although the groups were balanced and their variances were not significantly different. These considerations limit the extent to which the findings can be generalized beyond the studied cohort.

Taken together, the findings suggest that gender-related differences in numeracy are better treated as context-dependent empirical questions than as assumed characteristics of male and female students. The female group obtained a higher observed mean, but the evidence from the inferential test was insufficient to establish a statistically significant difference. In combination with prior research showing mixed patterns across educational levels, psychological constructs, and learning contexts, the present result supports a broader view of numeracy in which performance emerges from multiple interacting conditions. Future research could extend this work through larger multi-school samples and by combining gender comparisons with measures of self-efficacy, prior achievement, motivation, learning opportunities, and problem-solving processes. Such approaches would make it possible to determine not only whether group differences occur, but also under what educational conditions they emerge and which factors are most closely associated with students' numeracy performance.

4. Conclusion

This study examined the relationship between students' perceptions of field-based learning and mathematical problem-solving ability among Grade XI students at SMAN 6 Gowa identified as having a predominantly kinesthetic learning style. Based on data from 50 students, perceptions of field-based learning were positively and significantly associated with mathematical problem-solving ability ($r = .5339$, $p < .001$), with the magnitude of the

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

association falling within the moderate range. The finding indicates that students who reported more favorable perceptions of the field-based learning experience tended to obtain higher scores on the mathematical problem-solving assessment. The learning activities, which integrated problem-based learning with direct observation, measurement, data collection, physical representation, calculation, and verification, provided the instructional context in which this association was observed. Importantly, the result should be interpreted as an association rather than a causal effect, because the correlational design does not establish that field-based learning or students' perceptions of it caused differences in problem-solving performance.

The findings should be interpreted in light of several limitations. The study involved a relatively small purposive sample from a single school and focused exclusively on students classified as predominantly kinesthetic using the VAK screening instrument available through the Aku Pintar application; therefore, the findings should not be generalized to other learning-style groups or educational contexts without further evidence. In addition, the instructional design combined problem-based learning and field-based learning, making it impossible to isolate the specific association attributable to the field-based component. Future research should therefore employ larger and more diverse samples, independently validated measures of learner characteristics, and comparative or experimental designs that can distinguish the effects of instructional approaches and examine whether the observed pattern is sustained across different learning-style groups and mathematical topics. Longitudinal and mixed-methods studies could also investigate how students' perceptions develop and how specific features of field-based activities relate to different dimensions of mathematical problem solving. Despite these limitations, the present study contributes an empirical perspective that shifts attention from simply comparing learning-style groups toward examining how students' perceptions of a specific instructional context are related to their mathematical problem-solving performance within a defined learner group.

5. References

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Khotimah, K., & Siswono, T. Y. E. (2025). Kemampuan siswa SMP dalam memecahkan masalah numerasi ditinjau berdasarkan gender. *MATHEdunesa, 14*(2), 600–610. https://doi.org/10.26740/mathedunesa.v14n2.p600-610

Mariamah, Suciyati, & Hendrawan. (2021). Kemampuan numerasi siswa sekolah dasar ditinjau dari jenis kelamin. *TUNAS: Jurnal Penelitian Pendidikan Dasar, 1*(2), 17–19.

Maryam, S., Yulianto, D., & Juniawan, E. A. (2025). Studi komparatif kemampuan numerasi siswa SMP berdasarkan jenis kelamin di Kabupaten Lebak Banten. *JKPM (Jurnal Kajian*

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Pendidikan Matematika), 10(2), 165–174. https://doi.org/10.30998/jkpm.v10i2.26034

Mellyzar, Unaida, R., Muliani, & Novita, N. (2021). Hubungan *self-efficacy* dan kemampuan literasi numerasi siswa: Ditinjau berdasarkan gender. *Lantanida Journal, 9*(2), 93–182.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika, 52*(3–4), 591–611. https://doi.org/10.2307/2333709

Sugiyono. (2019). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Anderson, J. O., Lin, H.-S., Treagust, D. F., Ross, S. P., & Yore, L. D. (2007). Using large-scale assessment datasets for research in science and mathematics education: Programme for International Student Assessment (PISA). *International Journal of Science and Mathematics Education, 5*(4), 591–614. https://doi.org/10.1007/s10763-007-9090-y

Deda, Y. N., Amsikan, S., & Tanujaya, B. (2026). Integrating dryland agriculture into student worksheets to enhance students' numeracy skills. *Mathematics Education Journal, 20*(2), 349–374. https://doi.org/10.22342/mej.v20i2.pp349-374

Gal, I., Grotlüschen, A., Tout, D., & Kaiser, G. (2020). Numeracy, adult education, and vulnerable adults: A critical view of a neglected field. *ZDM–Mathematics Education, 52*(3), 377–394. https://doi.org/10.1007/s11858-020-01155-9

Geiger, V., Goos, M., & Forgasz, H. (2015). A rich interpretation of numeracy for the 21st century: A survey of the state of the field. *ZDM–International Journal on Mathematics Education, 47*(4), 531–548. https://doi.org/10.1007/s11858-015-0708-1

Hemmings, B., Grootenboer, P., & Kay, R. (2011). Predicting mathematics achievement: The influence of prior achievement and attitudes. *International Journal of Science and Mathematics Education, 9*(3), 691–705. https://doi.org/10.1007/s10763-010-9224-5

Hodgen, J., & Noyes, A. (2026). System-level improvement in mathematics education: Re-considering the National Numeracy Strategy in England. *Research in Mathematics Education, 28*(2), 267–286. https://doi.org/10.1080/14794802.2026.2628162

Kilpatrick, J. (2001). Understanding mathematical literacy: The contribution of research. *Educational Studies in Mathematics, 47*(1), 101–116. https://doi.org/10.1023/A:1017973827514

Leton, S. I., Lakapu, M., Jagom, Y. O., Djong, K. D., Dosinaeng, W. B. N., & Kusuma, D. A. (2026). Integrating augmented reality and local wisdom in teaching exponents: A strategy to improve students' mathematical literacy. *Infinity Journal, 15*(3), 901–920. https://doi.org/10.22460/infinity.v15i3.p901-920

Lestari, N. D. S., Pambudi, D. S., Murtafiah, W., Suwarno, Kurniati, D., & Sunardi. (2026). Decision-making skills of prospective mathematics teachers in designing numeracy tasks: A study of medium level pedagogical content knowledge. *Mathematics Teacher Education and Development, 28*(1), Article 2.

Maysarah, S., Armanto, D., & Saragih, S. (2026). Integrating ethnomathematics and GeoGebra: A project-based learning approach to improve mathematical literacy. *Mathematics Teaching-Research Journal, 18*(2), 161–191.

Mejía-Rodríguez, A. M., Luyten, H., & Meelissen, M. R. M. (2021). Gender differences in mathematics self-concept across the world: An exploration of student and parent data

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

of TIMSS 2015. *International Journal of Science and Mathematics Education, 19*(6), 1229–1250. https://doi.org/10.1007/s10763-020-10100-x

Muhaimin, L. H., Dasari, D., Hendriyanto, A., Andriatna, R., & Sahara, S. (2026). Can augmented reality enhance students' mathematical literacy? A study on technological development for learning practice. *International Journal of Mathematical Education in Science and Technology, 57*(5), 909–943. https://doi.org/10.1080/0020739X.2025.2502398

Payadnya, P. A. A., Puspadewi, K. R., & Noviana, L. P. R. (2026). AI-assisted adaptive geometry e-learning: Integrating ethno-realistic mathematics education to boost students' numeracy abilities. *STEM Education, 6*(3), 392–417. https://doi.org/10.3934/steme.2026017

Roza, Y., Adnan, A., Indra, Z., & Alawiyah, T. (2026). Correlation analysis of factors associated with students' numeracy skills using decision tree algorithms. *Numeracy, 19*(2), Article 1. https://doi.org/10.5038/1936-4660.19.2.1494

Rosita, C. D., Setiyani, Asmara, A. S., Sumarwati, S., & Suprayo, T. (2026). Mathematical thinking processes based on decision-making types in Grade VIII numeracy: Foundations for differentiated instruction in geometry. *Journal on Mathematics Education, 17*(1), 1–26. https://doi.org/10.22342/jme.v17i1.pp1-26

Sagita, L., Utami, N. W., Dwipa, N. M. S., Zulkardi, Putri, R. I. I., & Fauzan, A. (2026). Designing a contextual numeracy assessmen

Segers, E., Kleemans, T., & Verhoeven, L. (2015). Role of parent literacy and numeracy expectations and activities in predicting early numeracy skills. *Mathematical Thinking and Learning, 17*(2–3), 219–236. https://doi.org/10.1080/10986065.2015.1016819

Siregar, P. S., Sofyan, H., Nazurty, & Rosmiati. (2026). Boosting students' abilities in understanding, application, and reasoning of numeracy: Didactical differentiated learning. *Mathematics Teaching-Research Journal, 18*(2), 215–241.

Talib, A., Sappaile, B. I., Rahman, M. S., & Nasir, N. (2026). Enhancing secondary students' mathematical literacy through Android-based modules integrating deep learning and intrapersonal intelligence. *Eurasia Journal of Mathematics, Science and Technology Education, 22*(3), Article em2794. https://doi.org/10.29333/ejmste/18069

Tran, D. (2026). Measuring opportunities to develop mathematical literacy and their prediction on mathematical literacy outcomes. Mathematical Thinking and Learning. https://doi.org/10.1080/10986065.2026.2652846

Yore, L. D., Pimm, D., & Tuan, H.-L. (2007). The literacy component of mathematical and scientific literacy. International Journal of Science and Mathematics Education, 5(4), 559–589. https://doi.org/10.1007/s10763-007-9089-4

This content is covered by a Creative Commons license, Attribution-ShareAlike 4.0 International (CC BY-SA 4.0). This license allows re-users to copy, distribute, remix, transform, and build upon the material in any medium or format, for any purpose, including commercially, provided that appropriate credit is given to the creator, a link to the license is provided, and any changes are indicated. If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.