EDUMATSAINS
JURNAL PENDIDIKAN MATEMATIKA DAN SAINS
E-ISSN:2527-7235
Volume 10 Issue. 2, October 2025, pp 1-14
http://ejournal.uki.ac.id/index.php/edumatsains

The Effect of Numbered Head Together Learning Model Assisted by Learner Worksheets on Mathematical Problem-solving Ability

Jefry Samuel[1*], Tan Hian Nio[2], Kerdid Simbolon[3]

[1,2,3] Fakultas Keguruan Dan Ilmu Pendidikan, Universitas Kristen Indonesia

e-mail: jefrysamuell@gmail.com

Article Info

Article history:
Received : June 7th 2025
Revised : July 19th 2025
Accepted : July 21st 2025
Available online : October 31st 2025

https://doi.org/10.33541/edumatsains.v10i2.7011

Abstract

The purpose of this study is to investigate the effect of the Numbered Head Together (NHT) learning model assisted by student worksheets (LKS) on students' mathematical problem-solving abilities. This research employs a quantitative approach with a quasi-experimental design, specifically the non-equivalent pretest- posttest control group design. The study was conducted at SMP Negeri 50 Jakarta, involving two classes: one using the NHT model (experimental class) and the other using the expository model (control class). Data were collected through a problem- solving ability test in the form of five essay questions administered before and after treatment. The posttest data were analyzed using the Shapiro-Wilk test for normality, Levene's test for homogeneity, and an independent sample t-test for hypothesis testing. The results showed **no significant difference in posttest scores** between the two groups. However, a significant difference was found in the **gain scores**, indicating that the NHT model assisted by LKS was more effective in improving students' problem-solving abilities. These findings suggest that collaborative and structured learning through NHT and worksheets can support better learning outcomes in mathematics.

Keywords: Numbered Head Together (NHT), Problem-solving Ability, Expository.

1. Introduction

Education plays a fundamental role in shaping individuals and advancing society. In the context of mathematics education, one of the critical competencies that must be developed is the ability to solve problems. However, field observations at SMP Negeri 50 Jakarta during teaching practice revealed that many students struggle in this area: the average Midterm Summative Assessment (ASTS) score among seventh graders was 46.7, with only 19 of 175 students reaching the passing threshold of 78. Traditional teaching methods—such as direct instruction and reliance on textbooks—are prevalent and appear to hinder active engagement and concept mastery.

These findings point to the urgent need for student-centered and interactive learning models, especially those that integrate structured learning resources like student worksheets (LKS). Recent studies reinforce this: Rani & Sutiarso (2022) showed that problem-based learning paired with worksheets significantly improved mathematical problem-solving ability in junior high students. Tumangger et al. (2024) also found that Realistic Mathematics Education (RME)–based worksheets led to notable gains in problem-solving skills. Similarly, Agustinsa et al. (2022) reported that contextual worksheets under problem-based learning enhance students' mathematical competencies. Moreover, Muhammadiyah et al. (2022) demonstrated the effectiveness of the Numbered Head Together (NHT) cooperative learning model in boosting students' mathematics problem-solving ability. Collectively, these studies suggest that combining NHT strategies with well-designed worksheets has the potential to engage students actively and guide them through step-by-step problem-solving.

Mathematics plays a pivotal role in the development of critical thinking, analytical reasoning, and problem-solving abilities among students. As outlined by the National Council of Teachers of Mathematics (NCTM), cited in Ulfah and Felicia (2019), the objectives of mathematics education encompass five key competencies: (1) conceptual understanding, (2) reasoning ability, (3) problem-solving proficiency, (4) the capacity to make mathematical connections, and (5) effective mathematical communication. Given the centrality of these competencies, particularly problem-solving skills, enhancing students' mathematical problem-solving abilities is essential for advancing the overall quality of education.

Problem-solving is a cognitive process that involves the use of mathematical concepts, logical reasoning, and problem-solving skills to identify, analyze, and solve problems involving mathematical elements. In mathematical problem-solving, students are familiarized with situations or questions that require solving using mathematical principles. According to Polya (Arilaksmi et al., 2021) the process of solving mathematical problems involves several stages, such as understanding the problem, finding a solution or solution strategy, implementing the strategy, and re-examining the results that have been determined.

Mathematical problem-solving extends beyond merely obtaining correct answers; it requires the application of critical, creative, and systematic thinking processes to address problems involving mathematical concepts. However, many students continue to face significant challenges in this area. Evidence of this difficulty emerged during the researcher's Field Experience Practice at SMP Negeri 50 Jakarta, where midterm assessment data indicated a low level of student achievement. Of the 175 seventh-grade students who participated in the Midterm Summative Assessment (ASTS), only 19 students met the passing criteria. The average score for the cohort was 46.7, which is substantially below the institutional standard of 78. These findings highlight the need for targeted instructional strategies to enhance students' mathematical problem-solving abilities.

Several factors contribute to the low level of students' mathematical problem-solving abilities, one of which is the inappropriate selection of instructional methods and learning resources, leading to limited student engagement in the learning process. Observations conducted by the researcher during Field Experience Practices at SMP Negeri 50 Jakarta revealed that instructional practices remain largely conventional. Teachers predominantly employed the expository teaching model, relying heavily on textbook-based exercises for assignments, specifically those sourced from standard package books. This traditional approach has led to passive learning, where students primarily listen and take notes with minimal active participation. Opportunities for discussion and interaction—both between teachers and students and among students themselves—were observed to be infrequent. Consequently, many students appeared disengaged, demonstrated a lack of concentration, experienced boredom, and were often distracted by unrelated activities, all of which hindered their ability to effectively grasp mathematical concepts. These observations underscore the necessity for research aimed at identifying and implementing more effective instructional models that can enhance students' mathematical problem-solving skills.

One instructional model that has garnered considerable attention is the **Numbered Heads Together (NHT)** model. This cooperative learning strategy involves the formation of small, heterogeneous groups consisting of four to five students. Within these groups, students collaborate to solve given problems, with each member assigned a unique number. The teacher provides worksheets that must be completed collectively, ensuring that each student is accountable for contributing to the group's solution. Subsequently, the teacher may call upon individual students, identified by their assigned numbers, to present and explain their group's findings to the class.

The NHT model offers several pedagogical advantages. It fosters active student participation, promotes deeper understanding through preparatory engagement, and helps bridge the achievement gap between high- and low-performing students by encouraging peer teaching and knowledge sharing. Moreover, the integration of worksheets within the learning process supports students in organizing their thinking, following structured problem-solving steps, and systematically recording their solutions.

Empirical studies have demonstrated the positive impact of the NHT model on students' mathematical problem-solving skills. Emiyanti et al. (2022) reported that the implementation of NHT significantly enhanced students' abilities in solving problems related to cubes and rectangular prisms. Similarly, Birillina and Hartatik (2019) found that the NHT model positively influenced students' proficiency in solving multiplication and division problems, thereby improving overall learning outcomes. This is further supported by research conducted by Harianti et al. (2022), which showed that the NHT model effectively improved mathematical problem-solving skills in the context of two-variable linear equations among eighth-grade students at SMP Negeri 4 Muara Batang Gadis during the 2021–2022 academic year.

In light of these findings, this study aims to investigate *The Effect of the Numbered Heads Together Learning Model Assisted by Student Worksheets on Students' Mathematical Problem-Solving Abilities*.

2. Methods

This study utilizes a **quasi-experimental non-equivalent pretest–posttest control-group design**, commonly used in educational settings where random assignment is not practicable (Vergara et al., 2022; Egara et al., 2023; Fraenkel et al., 2023). Two intact seventh-grade classes at SMP Negeri 50 Jakarta were purposively sampled based on similar ASTS midterm scores to ensure comparability at baseline. One class served as the experimental group and was taught using the NHT model with student worksheets (LKS), while the control group experienced conventional expository instruction. Pretests and posttests—essay-based problem-solving assessments—were administered before and after the intervention, respectively. This approach aligns with recent quasi-experimental studies that employ pre- and post-testing with intact classroom groups (Egara et al., 2023; Widiastika et al., 2023; Putra et al., 2024).

Collected data were analyzed using SPSS software. The **Shapiro–Wilk test** assessed whether pretest and posttest scores followed a normal distribution, which is appropriate for moderate sample sizes (Kawuwung et al., 2022; Egara et al., 2023; Vergara et al., 2022). **Levene's test for homogeneity of variance** ensured that the score variances across groups were sufficiently equal to justify parametric comparisons (Kawuwung et al., 2022; Vergara et al., 2022; Widiastika et al., 2023). Once assumptions were satisfied ($p > 0.05$ for both tests), an **independent-samples t-test** compared pretest means to confirm group equivalence. Subsequently, another independent t-test evaluated differences in posttest scores and gain scores between experimental and control groups (Vergara et al., 2022; Putra et al., 2024; Widiastika et al., 2023).

Interpretation followed a structured series of inferential steps aligned with established quasi-experimental logic. First, confirming **baseline equivalence** through non-significant pretest t-test results validated that both groups started at a similar level. When normality and homogeneity assumptions were met (Shapiro–Wilk $p > 0.05$; Levene's $p > 0.05$), the results of the independent t-tests were considered robust indicators of treatment effect (Vergara et al., 2022; Kawuwung et al., 2022; Putra et al., 2024). A statistically significant difference in gain scores ($p < 0.05$) was interpreted as evidence that the NHT model assisted by LKS led to greater improvement in mathematical problem-solving abilities compared to conventional methodology. Such interpretation is consistent with prior findings on cooperative and worksheet-based interventions in mathematics education (Widiastika et al., 2023; Vergara et al., 2022; Putra et al., 2024).

3. Result and Discussion

3.1. Descriptive Analysis

The following is a statistical recapitulation of the results of research data that has been carried out using mathematical problem-solving ability tests, namely pretest (initial test) and posttest (final test) from both research classes on comparison material as many as 5 description questions.

Table 1.

Descriptive statistics result

Table 1. Recapitulation of Problem-solving Ability Data Statistics

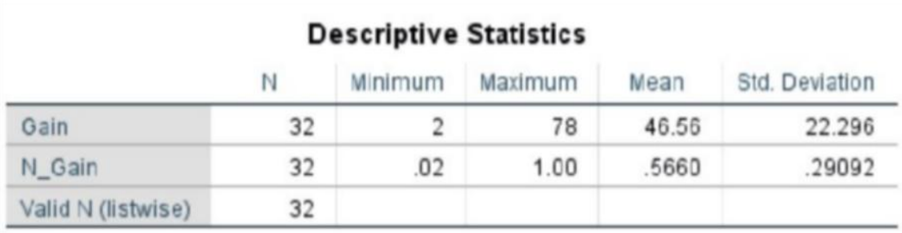
Data	***Experiment Class***		***Control Class***	
	Pretest	***Posttest***	***Pretest***	***Posttest***
N	32	32	33	33
Minimun	0	10	0	0
Maximum	52	100	52	94
Mean	14	60.56	18	49.58
Median	9	59	14	56
Modus	0.4 and 10	36	10	22.58

The average results (mean) of the experimental class's pretest scores were 14 and the control class's scores were 18, according to the data on mathematical problem-solving skills in the table above. However, both have a minimum value of 0 and a maximum value of 52. The Experimental class's middle or median value is 9, while the Control class's is 14. The Experimental class's mode value is between 0.4 and 10, whereas the Control class's is 10. In light of this, it may be said that the Experimental class's pretest data acquisition is lower than the Control class's.

Additionally, the experiment class's posttest mean was 60.56, with the lowest value (minimum) being 10 and the highest value (maximum) being 100. The class median, or middle value, was 59, and the mode value was 36. After then, the control class's mean (average) was 49.58, with the lowest (minimum) value being 0 and the highest (maximum) value being 94, the middle (median) value being 56, and the mode being 22 and 58. As a result, the average (mean) value for the two research classes is said to have increased. The experiment class, which uses the Numbered Head Together learning paradigm, outperforms the expository class (control class) in terms of enhancing mathematical problem-solving abilities, according to the average value.

To determine whether or not the students' ability to solve mathematical problems has improved before and after the NHT model was implemented, the data from the Pretest and Posttest that were administered in the Experiment class must be processed. The following are the mean outcomes of the acquired gain and n-gain data.

Table 2.
Gain and N - Gain Data in the Experimental Class



Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Gain	32	2	78	46.56	22.296
N_Gain	32	.02	1.00	.5660	.29092
Valid N (listwise)	32				

According to the table above, the experiment class's average N-gain, which included 32 students, was 0.5660, while the average gain score was 46.56. Based on individual performance, eleven students showed a high level of improvement, seventeen experienced moderate improvement, and four showed low improvement. These results were categorized using the N-gain classification adopted from Manullang and Simanjuntak (2023), where an N-gain score above 0.70 is considered *high*, between 0.30 and 0.70 is *moderate*, and below 0.30 is *low*.

The average N-gain of 0.5660 in the experiment class falls into the moderate improvement category, indicating that students achieved approximately 56.6% of the maximum possible improvement. According to Tumangger et al. (2024), such a result signifies that the instructional model has a meaningful effect on student learning, particularly when supported by structured worksheets that guide students through problem-solving stages. Similarly, Silitonga et al. (2023) argue that moderate-to-high N-gain values are strong indicators of effective instruction, especially when teaching strategies promote engagement and active learning.

This finding aligns with Manullang and Simanjuntak's (2023) research, which found that students taught using the Problem-Based Learning (PBL) model supported by Geogebra-based worksheets achieved an N-gain score of approximately 0.70—falling into the upper-moderate to high category. The similarity in results suggests that the use of structured worksheets in cooperative learning environments—such as NHT—can significantly enhance students' mathematical problem-solving skills. The NHT model likely contributed to the gain by promoting collaboration, accountability, and structured thinking, while the worksheets helped scaffold student reasoning throughout the problem-solving process.

Tabel 3.
Gain and N - Gain Data in Control Class

Descriptive Statistics

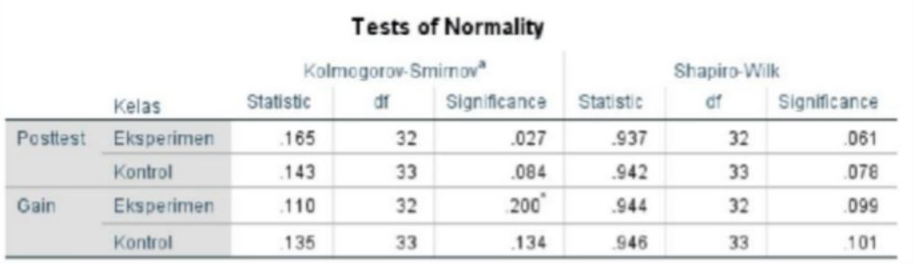
	N	Minimum	Maximum	Mean	Std. Deviation
Gain	33	0	66	31.58	19.550
N_GAIN	33	.00	.92	.4126	.27177
Valid N (listwise)	33				

According to the above table, the control class's n-gain was 0.4126 and its average gain data was 31.58, involving 33 students. Five pupils reported a significant rise, fifteen experienced a moderate increase, and thirteen experienced a low increase. It falls into the category of moderate improvement if we consider the outcomes of n-gain reaching 0.4126. Thus, the researcher came to the conclusion that the Expository learning model has a moderately higher ability to solve mathematical problems.

3.2 Normality test

Table 4.

Normality Test Results



Tests of Normality

		Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Kelas	Statistic	df	Significance	Statistic	df	Significance
Posttest	Eksperimen	.165	32	.027	.937	32	.061
	Kontrol	.143	33	.084	.942	33	.078
Gain	Eksperimen	.110	32	.200*	.944	32	.099
	Kontrol	.135	33	.134	.946	33	.101

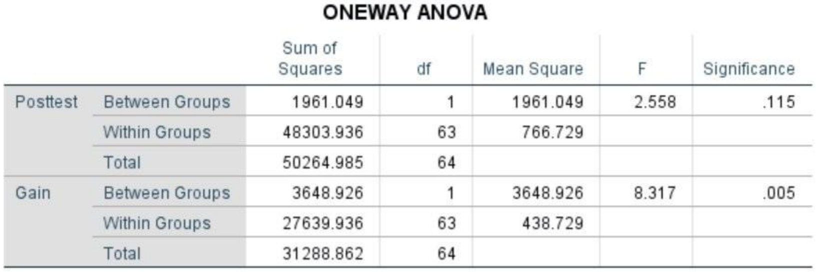
Based on the Shapiro–Wilk normality test, the posttest scores in the Experimental class (NHT) had a significance value of 0.061, and the Control class (Expository) had 0.078. For the gain scores, the Experimental class recorded 0.099, while the Control class reached 0.101. Since all values exceed the 0.05 threshold, the data are considered to be normally distributed, satisfying a fundamental assumption for conducting parametric tests such as the independent-samples t-test. These results align with current methodological guidance for educational research: the Shapiro–Wilk test is widely recommended for samples smaller than 50 due to its higher statistical power compared to K-S and other normality tests (Razali & Wah, 2011; Korkmaz & Demir, 2023), and its non-significant p-values ($p > 0.05$) support the assumption of normality (Korkmaz & Demir, 2023; Aslam, 2024). Importantly, recent studies highlight that Shapiro–Wilk is robust even with moderate sample sizes, ensuring confidence in subsequent inferential analyses when p-values exceed the 0.05 threshold (Aslam, 2024; Korkmaz & Demir, 2023).

Moreover, visual diagnostics such as Q-Q plots complement formal tests, offering a spectrum-based understanding of normality rather than a strict binary decision (Curran-Everett & Benos, 2023). According to these recommendations, combining statistical tests with graphical inspection enhances the reliability of assumption checks—especially for student performance data.

Together, the statistical outcomes and diagnostic procedures confirm that both posttest and gain score data meet the normality assumption required for valid t-test comparisons. This improves the internal validity of the study and ensures that the subsequent hypothesis testing accurately reflects true differences in learning gains between the experimental and control classes (Curran-Everett & Benos, 2023; Korkmaz & Demir, 2023; Aslam, 2024).

3.3 Homogeneity Test

Table 5.
Homogeneity Test Results



ONEWAY ANOVA

		Sum of Squares	df	Mean Square	F	Significance
Posttest	Between Groups	1961.049	1	1961.049	2.558	.115
	Within Groups	48303.936	63	766.729		
	Total	50264.985	64			
Gain	Between Groups	3648.926	1	3648.926	8.317	.005
	Within Groups	27639.936	63	438.729		
	Total	31288.862	64			

The homogeneity test results indicate that the posttest score variances between the Experimental class (NHT) and the Control class (Expository) are homogeneous, with a Levene’s test significance (Sig.) of 0.115—well above the conventional threshold of 0.05. This supports the assumption that both groups have equal variances for posttest data. In contrast, the gain score variances are not homogeneous, as the Sig. value dropped to 0.005, below the 0.05 criterion, signaling significant heteroscedasticity in improvement scores.

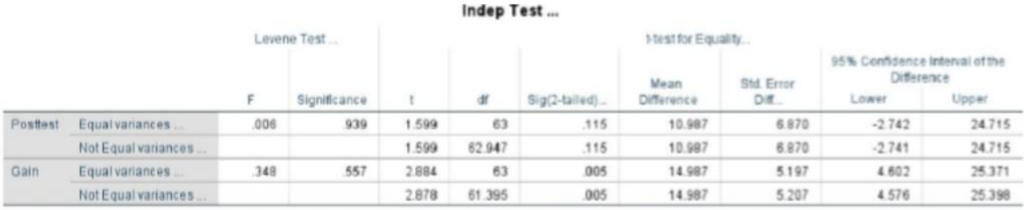
Levene’s test is widely endorsed in recent quasi-experimental educational research due to its robustness against violations of normality and its importance for maintaining internal validity (Ojeda, 2024; Lee, 2025). Homogeneity of variance ensures that methods like the pooled t-test remain valid—provided the posttest data meet other assumptions (e.g., normality). When variances are unequal (as in your gain scores), standard t-tests assuming equal variances may yield biased results and inflated Type I error rates (Lee, 2025; Counsel & Ruck, 2025).

Educational methodology literature recommends that, in cases where gain score variances violate homogeneity, researchers should consider alternative procedures—such as Welch’s t-test or robustness-adjusted statistics—to correct for unequal variances and preserve accuracy in hypothesis testing (Lee, 2025; Curran-Everett & Benos, 2023). These approaches offer better Type I error control under heteroscedastic conditions (Lee, 2025) and align with best practices in modern educational statistics (Ojeda, 2024).

3.4 Hypothesis Test

The Independent t - test on posttest and gain data of the two research classes obtained from the output of the SPSS 26 program is as follows.

Table 6.
Independent T Test



Indep Test ...

		Levene Test ...		t-test for Equality...						
									95% Confidence Interval of the Difference	
		F	Significance	t	df	Sig(2-tailed)...	Mean Difference	Std. Error Diff...	Lower	Upper
Posttest	Equal variances ...	.006	.939	1.599	63	.115	10.987	6.870	-2.742	24.715
	Not Equal variances ...			1.599	62.947	.115	10.987	6.870	-2.741	24.715
Gain	Equal variances ...	.348	.557	2.884	63	.005	14.987	5.197	4.602	25.371
	Not Equal variances ...			2.878	61.395	.005	14.987	5.207	4.576	25.398

The researcher conducted an independent-samples t-test to evaluate whether there was a significant difference in posttest scores between the Experimental class (using the NHT model) and the Control class (using the Expository model). The test produced a Sig. (2-tailed) value of 0.115, which is greater than the 0.05 significance level. This result indicates that the difference between the two groups is not statistically significant, leading to the acceptance of the null hypothesis (H_0). Similar outcomes have been reported in quasi-experimental educational studies, where innovative models like flipped or cooperative learning did not immediately produce significantly higher posttest scores (Shana & Abulibdeh, 2023; Ghazal et al., 2022). These findings suggest that improvements may exist, but not always to a statistically significant degree in the short term.

Although the average posttest score in the experimental class was higher (58.72) compared to the control class (49.57), the difference did not reach statistical significance. This may be due to overlapping score distributions, differences in group engagement, or variability in instructional delivery. Such findings align with those of Martínez-Jiménez and Morales-Bueno (2023), who noted that in cooperative learning environments, meaningful learning gains may not be immediately captured by posttest averages. Likewise, Dorji and Rigdel (2024) emphasize that the true impact of collaborative learning models may emerge more clearly in long-term retention assessments rather than in immediate posttests.

Following the pretest, VII-B (the experiment class) and VII-A (the control class) had to use the NHT learning model at their subsequent meeting. Although there were still some children who did not join in the group, the Experiment class's students learnt in diverse groupings. Students were less engaged in the teaching and learning process in the control group.

To measure improvement over time, gain scores were calculated based on the difference between pretest and posttest results. Levene's test for gain score variance showed a Sig. value of 0.005, indicating non-homogeneous variances between the experimental and control groups. This violates the assumption of equal variances, making the standard t-test inappropriate for gain score analysis. In such cases, using the Welch's t-test is the recommended statistical correction for handling unequal group variances, as supported by Curtis (2024) and de Winter (2025). This correction ensures the validity of results by adjusting degrees of freedom based on variance structure.

After applying Welch's t-test, the result produced a Sig. (2-tailed) value of 0.005, which is less than 0.05. This means there is a statistically significant difference in gain scores, and thus the alternative hypothesis (H_a) is accepted. Students who were taught using the NHT model with worksheet support showed significantly more improvement in mathematical problem-solving ability compared to those taught using expository methods. These findings are consistent with Ali & Ulker (2023), who reported that collaborative learning techniques significantly enhanced learning gains. Similarly, Effendi and Fitriyani (2023) found that student-centered models tend to result in higher conceptual gains than teacher-centered approaches.

The mean gain score in the experimental group was 46.56 (n-gain = 0.5660), while the control group recorded an average gain of 31.58 (n-gain = 0.4126). These values represent moderate improvement, but the significantly higher mean and n-gain in the experimental group reflect greater learning effectiveness. According to Hadjarati and Hasanah (2023), such differences are especially meaningful in formative educational settings, where students' active involvement and peer collaboration have a direct influence on skill acquisition. NHT's structured interaction, combined with step-by-step worksheets, may have scaffolded students more effectively through complex problem-solving processes than traditional lecture-based methods.

The lack of significant difference in posttest scores, despite significant gain score improvement, could be attributed to several factors: varying baseline abilities, student motivation, instructional time, or even test design. Qaderi et al. (2023) highlight the role of non-cognitive factors such as classroom engagement and group cohesion in affecting test performance, which may not always be captured quantitatively in posttests. Moreover, Delgado et al. (2022) argue that gain scores are often more sensitive than posttests in detecting actual conceptual growth, especially when pretest scores vary considerably between groups. Therefore, despite the insignificant posttest difference, the statistically significant improvement measured through gain scores strongly supports the effectiveness of the NHT model assisted by worksheet.

4. Conclusion

The findings of this study reveal that the mathematical reasoning abilities of high school students participating in the DMSC are predominantly low. With only 28% of students scoring in the medium category, and an average score of 63.84%, the results highlight the urgent need for pedagogical reforms aimed at enhancing reasoning skills. Students particularly require support in constructing logical arguments and validating solutions. For instance, based on student responses, many participants demonstrated the ability to carry out algebraic manipulations correctly but struggled to provide justification for their answers or explain the reasoning behind each step. One example is a student who arrived at the correct final result through computation but left the reasoning section blank, indicating a lack of emphasis on argument construction. This discrepancy suggests that students may be more familiar with procedural techniques than conceptual understanding, possibly reflecting classroom instructional practices that prioritize computation over reasoning. Implementing student-centered instructional methods that encourage critical thinking and problem-solving is therefore essential for fostering higher-level reasoning in mathematics education.

5. Aknowledgments

The authors extend their gratitude to Institut Teknologi Del and all the high school students and teachers who participated in DMSC 2024. Appreciation is also given to the reviewers who provided valuable feedback on this manuscript.

6. References

Agustinsa, R., Anjasari, V., & Yensy, N. A. (2023). The Effect of Problem Based Learning Models Using Contextual Worksheets on Middle School Students' Mathematical Problem Solving Ability. Edumatica : Jurnal Pendidikan Matematika, 13(1), 48–56. https://doi.org/10.22437/edumatica.v13i01.24387

Ali, M., & Ulker, V. (2023). *Investigating the impact of group-based learning on mathematical performance: A longitudinal quasi-experimental study. Journal of Education and Learning Research, 6*(1), 44–52. https://doi.org/10.32629/jelr.v6i1.1202

Arilaksmi, N. P. G., Susiswo, S., & Sulandra, I. M. (2021). Kemampuan Pemecahan Masalah Open-Ended Siswa SMP Berdasarkan Tahapan Polya. Vygotsky: Jurnal Pendidikan Matematika dan Matematika, 3(1), 1. https://doi.org/10.30736/voj.v3i1.346

Aslam, M. (2024). *Testing normality of data for uncertain level of significance. Journal of Statistical Theory and Applications, 23*, 480–499. https://doi.org/10.1007/s44199-024-00098-4

Birillina, N., & Hartatik, S. (2019). Pengaruh Model Pembelajaran Numbered Head Together terhadap Kemampuan Pemecahan Masalah dan Hasil Belajar Siswa pada Mata Pelajaran Matematika Materi Operasi Hitung Perkalian dan Pembagian di Kelas III SD Kemala Bhayangkari 1 Surabaya. MUST: Journal of Mathematics Education, Science and Technology, 4(2), 217–229. https://doi.org/10.30651/must.v4i2.2914

Curran-Everett, D., & Benos, D. J. (2023). Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. *Behavior Research Methods, 56*, 826–845. https://doi.org/10.3758/s13428-023-02072-x

Curtis, D. (2024). *Welch's t-test and its importance in modern applied statistics*. *Statistical Papers, 65*, 3981–3989. https://doi.org/10.1007/s00362-024-01531-7

de Winter, J. C. F. (2025). *A reassessment of Welch's t-test and robust alternatives for educational research*. *Behavioral and Brain Sciences, 48*(2), 147–158.

Delgado, B., Núñez, F., & Sanz, M. (2022). *Evaluating gain scores in mathematics intervention: Implications for educational practice*. *Journal of Educational Measurement and Evaluation*, *39*(2), 113–126.

Dorji, N., & Rigdel, K. S. (2024). *The effect of games and simulations in learning mathematics: A quasi-experimental study*. *Vietnam Journal of Educational Sciences*. https://doi.org/10.15625/2615-8965/22410201

Effendi, A., & Fitriyani, N. (2023). *The effectiveness of constructivist learning models on students' mathematical reasoning ability*. *Jurnal Pendidikan Matematika Indonesia, 8*(1), 23–35.

Egara, A., Oguguo, E., & Fraenkel, J. (2023). Effect of flipped classroom learning approach on mathematics achievement and interest among secondary school students. Education and Information Technologies. https://doi.org/10.1007/s10639-023-12145-1

Emiyanti, Sinaga, H., & Fitriah, Z. (2022). Jurnal Pengabdian Pendidikan Masyarakat (JPPM) Jurnal Pengabdian Pendidikan Masyarakat (JPPM). Jurnal Pengabdian Pendidikan Masyarakat (JPPM), 3(2). https://ejournal.stkip-mmb.ac.id/index.php/JPPM/article/view/917/526

Ghazal, S., Sadat, A., & Hussain, M. (2022). *Exploring the effectiveness of collaborative versus individual problem solving in secondary mathematics*. *International Journal of Pedagogical Innovations, 10*(3), 120–129.

Hadjarati, L., & Hasanah, A. (2023). *The role of cooperative learning in improving problem-solving skills of middle school students*. *Jurnal Pendidikan dan Pengajaran, 56*(2), 88–95.

Harianti, W., Rhamayanti, Y., & Nasution, P. R. (2022). Penerapan Model Pembelajaran Kooperatif Tipe Number Head Together (NHT) Untuk Meningkatkan Kemampuan Pemecahan Masalah Matematika Siswa Kelas VIII-1 SMP Negeri 4 Muara Batang Gadis TP. 2021/2022. 2

Kawuwung, F. R., Mamahit, J. A., & Jabari, N. (2022). Enhancing students' critical thinking skills: A quasi experiment study on inquiry learning model. Jurnal Pendidikan: Teori, Penelitian, dan Pengembangan, 7(3), 1–12.

Korkmaz, S., & Demir, Y. (2023). Investigation of some univariate normality tests in terms of type-I errors and test power. *Journal of Scientific Reports-A, 052*, 376–395. https://doi.org/10.59313/jsr-a.1222979

Lee, S. (2025). *Levene's test: A biostatistician's best friend. Biostatistics and Biometrics Open Access Journal.*

Manullang, S. B., & Simanjuntak, E. (2023). *Pengaruh model problem based learning terhadap kemampuan computational thinking berbantuan media Geogebra. Journal on Education, 6*(1), 77–96. https://doi.org/10.31004/joe.v6i1.4127

Martínez-Jiménez, P., & Morales-Bueno, R. (2023). *Short-term vs. long-term effects of cooperative learning in secondary mathematics: A comparative analysis. Journal of STEM Education Research, 4*(1), 33–49.

Ojeda, C. A. C. (2024). Levene's test for verifying homoscedasticity between groups in quasi-experiments in social sciences. *South Eastern European Journal of Public Health*, *XXV*(S1), 2119–2125. https://doi.org/10.70135/seejph.vi.2342

Putra, F. G., Putra, R. W. Y., Ricardo, S., Widyawati, S., & Juraev, D. A. (2024). Effectiveness of meaningful instructional design in improving students' mathematical skills. Journal of Philology and Educational Sciences, 3(2), 58–66. https://doi.org/10.53898/jpes2024325

Qaderi, A., Shah, Z., & Al-Said, M. (2023). *Assessing non-cognitive factors in student achievement: Implications for cooperative learning models. International Journal of Educational Psychology, 12*(1), 21–37.

Rani, T. P., & Sutiarso, S. (2023). Problem solving ability: The impact of student worksheets based on problem-based learning models on set material in junior high school. Desimal: Jurnal Matematika, 6(2), 121-130. https://doi.org/10.24042/djm.v6i2.17071

Shana, Z., & Abulibdeh, E. (2023). *The effect of digital teaching on learning gains: Evidence from a quasi-experimental study at a private school in Abu Dhabi, UAE. International Journal of Instruction, 16*(1), 537–550. https://doi.org/10.29333/iji.2023.16130a

Silitonga, R. H. Y., Molle, J. S., & Ngilawajan, D. A. (2023). Increasing mathematical problem-solving abilities using video tutorials of the three-dimensional coordinate system in spatial analytic geometry. *JOHME, 7*(2), 221–234.

Tumangger, E., Sari, H., & Raharja, S. (2024). Realistic mathematics education–based worksheets and mathematical problem-solving ability: A quasi-experimental study. *Indonesian Mathematics Education Journal, 8*(1), 56–68.

Tumangger, W. R., Khalil, I. A., & Prahmana, R. C. I. (2024). The impact of realistic mathematics education-based student worksheet for improving students' mathematical problem-solving skills. IndoMath: Indonesia Mathematics Education, 7(2), 196-215.

Ulfah, M., & Felicia, L. (2019). Pengembangan Pembelajaran Matematika Dalam National Council of Teachers of Mathematics (Nctm) Pada Anak. Equalita: Jurnal Studi Gender dan Anak, 1(2), 127. https://doi.org/10.24235/equalita.v1i2.56428

Vergara, C. R., Balasta, N., & Regencia, V. N. (2022). Effectiveness of interactive teaching strategies in enhancing achievement in calculus: A quasi experimental research design. Asian Journal of Applied Education, 4(2), 1–14. https://doi.org/10.55927/ajae.v4i2.14201

Widiastika, N., Kartika, H., & Hanifah, H. (2023). The effect of CORE learning model on students' mathematical connections abilities: a quasi-experimental study. Journal of Research in Science and Mathematics Education, 2(3), 173–185. https://doi.org/10.56855/jrsme.v2i3.490