



Comparison of the Effectiveness of the Think Pair Share (TPS) Model and the Two Stay–Two Stray (TS-TS) Model on Senior High School Students’ Mathematical Reasoning Ability and Self-Efficacy

Yosua Tumanggor¹, Elly Arliani²

^{1,2}Universitas Negeri Yogyakarta, Yogyakarta, Indonesia

ARTICLE INFO	ABSTRACT
Published Online: 22 September 2025	This study is a quasi-experimental research design with a pre-test and post-test nonequivalent comparison group design. The population of this research consists of all 10th-grade students at SMA Negeri 1 Sleman, which is divided into 7 classes. Two classes, X-A and X-D, were selected as the sample based on purposive sampling techniques. Class X-D was taught using the Think Pair Share learning model, while class X-A used the Two Stay-Two Stray model. The instruments used for data collection were a mathematical reasoning ability test (with a validity coefficient of 1.00 and a reliability coefficient of 0.69), a self-efficacy questionnaire (with a validity coefficient of 1.00 and a reliability coefficient of 0.73), and a learning implementation checklist. The effectiveness criteria for mathematical reasoning ability are as follows: (1) the average score of students' mathematical reasoning ability is greater than 75, (2) at least 75% of students achieve a score ≥ 75 , (3) there is a significant difference (increase) in the average scores between the pre-test and post-test. The effectiveness criteria for self-efficacy are: (1) the average score of the students' self-efficacy questionnaire is greater than 87, (2) at least 75% of students score ≥ 87 , (3) there is a significant difference (increase) in the average scores between the pre-questionnaire and post-questionnaire. A Hotelling's T2 test was used to examine the differences in the average post-test, pre-test, pre-questionnaire, and post-questionnaire scores. A paired t-test was performed to analyze the differences in the pre-test and post-test scores, followed by a One-Sample t-test to assess the effectiveness of the TPS and TS-TS classes by first conducting a Hotelling's T2 test to examine the difference between the average post-test and pre-test scores. An independent sample t-test was then applied to determine which learning model was more effective. The results of the study show that: 1) Think Pair Share learning model is effective for enhancing mathematical reasoning ability and self-efficacy, 2) The Two Stay–Two Stray learning model is also effective for enhancing mathematical reasoning ability and self-efficacy, 3) There is a difference between the Think Pair Share and Two Stay–Two Stray learning models in terms of their effects on mathematical reasoning ability and self-efficacy, with the Think Pair Share model being more effective than the Two Stay–Two Stray model.
Corresponding Author: Yosua Tumanggor	

KEYWORDS: Think Pair Share, Two Stay–Two Stray, mathematical reasoning ability, self-efficacy.

I. INTRODUCTION

The *Future of Jobs Report 2023* by the World Economic Forum emphasizes that mathematical reasoning, as part of analytical thinking, is one of the core skills required in the era of the Fourth Industrial Revolution and is also applied in daily life. In line with this, the National Council of Teachers of Mathematics (NCTM) categorizes basic mathematical abilities into five standards: solving mathematical problems, reasoning in mathematics,

mathematical communication, making mathematical connections, and representing mathematical problems. [2]. Reasoning in mathematics is one of the fundamental skills that students must possess. Mathematical reasoning plays an important role in the thinking process during mathematics learning [3]. In line with this, Yerizon (2019) stated that mathematics is taught because it can enhance the ability to reason, think systematically, logically, and critically, as well as to communicate ideas in order to solve problems.

“Comparison of the Effectiveness of the Think Pair Share (TPS) Model and the Two Stay-Two Stray (TS-TS) Model on Senior High School Students’ Mathematical Reasoning Ability and Self-Efficacy”

Mathematics is also referred to as the “*Queen of Science*”, meaning the mother of sciences or, in other words, a fundamental discipline that serves as the foundation for other fields of knowledge, making it inherently interconnected with them [5]. As explained by Nuridawani et al. (2015), mathematics is taught so that individuals can develop the ability to reason systematically, logically, and critically in solving problems.

In fact, students’ mathematical reasoning ability in Indonesia is still relatively low. This is based on the results of the TIMSS 2011 survey, which showed that Indonesia ranked 38th out of 42 participating countries with an average score of 386 [7]. Similarly, the results of the TIMSS 2015 survey also showed that Indonesia ranked 44th out of 49 countries, with an average score still far below that of neighboring countries such as Singapore, Malaysia, and Thailand [8].

According to Mariyam and Wahyuni (2016), one of the factors contributing to the low level of students’ mathematical reasoning ability in learning mathematics is the teacher-dominated instructional approach. According to Nurhidayati et al. (2017), one of the reasons for the low quality of students’ mathematical reasoning is that mathematics instruction tends to focus too much on procedural and mechanistic aspects, where students are only trained to solve numerous problems without developing deep understanding.

Mathematics is not merely about numerical calculations; it also involves making true or false statements about a problem, constructing proofs, and drawing conclusions from statements, all of which require reasoning skills. These activities demand a specific ability that supports such processes—mathematical reasoning. This ability is an essential component in understanding mathematical problems [11]. Mathematical reasoning ability can be enhanced through affective abilities possessed by the students themselves. One of the attitudes that can improve mathematical reasoning ability is students’ self-efficacy [12]. Self-efficacy refers to an individual’s belief in their ability to organize and execute the actions required to achieve specific goals. Strong self-efficacy enables individuals to build confidence in themselves, allowing them to overcome the problems they encounter [13].

One of the learning models presumed to enhance students’ mathematical reasoning ability and self-efficacy is the Think Pair Share (TPS) model. The cooperative learning model of the TPS type is a structured instructional approach designed to shape students’ interaction patterns in order to create cooperative learning that can improve both academic achievement and affective aspects [14]. Another learning model that is presumed to enhance students’ mathematical reasoning ability and self-efficacy is the Two Stay–Two Stray (TS-TS) model. The Two Stay–Two Stray learning model is a type of group-based learning designed to encourage students to work collaboratively and support one another in solving problems.

Considering the potential of both learning models, the researcher assumes that these models can serve as alternative solutions to address students’ mathematical reasoning ability and self-efficacy. Therefore, the Think Pair Share (TPS) and Two Stay–Two Stray (TS-TS) learning

models will be implemented and compared in mathematics instruction across two classes.

II. METHODS

This study is a quasi-experimental research design. The research design used is a nonequivalent comparison group design, in which both groups were given a pre-test and a post-test to compare the effects of the treatments provided. There are two experimental groups that received treatments: the first group was taught using the Think Pair Share (TPS) learning model, and the second group was taught using the Two Stay–Two Stray (TS-TS) learning model. The pre-test was conducted to measure students’ initial abilities before receiving the treatment, and the post-test was conducted to measure students’ abilities after receiving the treatment. The research design can be seen in the following table.

Table 1. Research Design

O_1	X_1	O_2
O_3	X_2	O_4

The subjects of this study were tenth-grade students of SMAN 1 Sleman in the 2024/2025 academic year. The sample was selected using purposive sampling, in which two classes were chosen from several available classes based on the suggestions and recommendations of mathematics teachers at SMAN 1 Sleman. The data analysis in this study was quantitative, employing statistical analysis techniques. Two types of statistics were used, namely descriptive statistics and inferential statistics. Inferential statistics included both parametric and non-parametric statistics [15]. Descriptive analysis was used to describe the research data and to provide an overview of students’ mathematical reasoning ability and self-efficacy. The data described in this study consisted of pre-tests and post-tests administered to each experimental class. Inferential analysis was employed to draw conclusions that could be generalized to the population. The analytical techniques applied included hypothesis testing; however, before conducting hypothesis testing, prerequisite tests were carried out, namely normality and homogeneity tests. The data analyzed included students’ mathematical reasoning ability and self-efficacy in mathematics learning using the Think Pair Share (TPS) and Two Stay–Two Stray (TS-TS) models.

The prerequisite tests of analysis were carried out before conducting hypothesis testing, which included tests of normality and homogeneity. According to Stevens (2009, pp. 217–218), in both ANOVA and MANOVA, the assumptions of normality and homogeneity must be met at both the univariate and multivariate levels. The multivariate normality test was conducted to determine whether the sample came from a normally distributed population or not. In this study, multivariate normality was tested using Mardia’s test. The univariate normality test was conducted to examine whether a single variable in the sample was drawn from a normally distributed population or not. This test was performed using the Shapiro-Wilk test with the aid of the R program. The homogeneity test of the two experimental classes was carried out to determine whether the two classes had homogeneous covariance matrices. If

“Comparison of the Effectiveness of the Think Pair Share (TPS) Model and the Two Stay-Two Stray (TS-TS) Model on Senior High School Students’ Mathematical Reasoning Ability and Self-Efficacy”

both classes had the same covariance matrix, they were considered homogeneous. This homogeneity test was conducted using Box’s M test. The univariate homogeneity test aimed to examine whether the data variances of the two experimental classes were equal or not. This test was carried out using Levene’s test. The data tested for normality and homogeneity included the pre-test and post-test scores of mathematical reasoning ability, as well as the pre-questionnaire and post-questionnaire scores of students’ self-efficacy. The effectiveness test of learning was conducted to examine whether the Think Pair Share (TPS) and Two Stay-Two Stray (TS-TS) learning models were effective in improving students’ mathematical reasoning ability. These models were considered effective if: (1) the average score of students’ mathematical reasoning ability was greater than 75, in accordance with the Minimum Mastery Criteria (KKTP) applied in the school; (2) at least 80% of students achieved scores \geq KKTP; and (3) there was a significant difference (an improvement) in the average scores between the pre-test and post-test. The TPS and TS-TS models were also considered effective in improving students’ self-efficacy if: (1) the average score of students’ self-efficacy questionnaires was greater than 87; (2) at least 80% of students achieved scores \geq 87; and (3) there was a significant difference (an improvement) between the pre-questionnaire and post-questionnaire self-efficacy scores. The effectiveness test was carried out using the one-sample mean vector test (One-Sample Hotelling’s T^2 Test), followed by a paired t-test to examine the improvement in scores before and after the learning.

To ensure that the effectiveness criteria of both learning models on students’ mathematical reasoning ability and self-efficacy were met, statistical testing was carried out. The testing used a paired t-test with the aid of R-Studio; however, prior to this, a Hotelling’s T^2 one-sample mean vector test was conducted to determine whether there were differences in the mean scores of mathematical reasoning ability and self-efficacy before and after the implementation of the Think Pair Share (TPS) and Two Stay-Two Stray (TS-TS) learning models. If a significant difference was found, the analysis was continued with the paired t-test; if not, further testing was unnecessary. The decision criteria for hypothesis testing were set at a significance level of 0,05. In other words, H_0 is accepted if $T^2 > \left(\frac{n-1}{n-p}\right) F_{p,n-p}(0,05)$, where p is the number of dependent variables and n is the sample size, or H_0 is rejected if $p - value < 0.05$. Subsequently, if there were differences in the mean scores of mathematical reasoning ability and self-efficacy before and after the application of the TPS or TS-TS models, paired t-tests were conducted. The paired t-test for mathematical reasoning ability was used to determine whether the post-test mean score was higher than the pre-test mean score, while the paired t-test for self-efficacy was used to determine whether the final self-efficacy questionnaire score was higher than the initial self-efficacy score. The decision criteria for hypothesis testing of mathematical reasoning ability and self-efficacy for each learning model were set at a significance level of 0.05; in other words, H_0 is rejected if $t_{hitung} > t_{\alpha(n-1)}$ or if $p - value < 0.05$. Furthermore, a one-sample t-test was

conducted to evaluate whether the first effectiveness criterion was met. However, prior to this, a multivariate one-sample Hotelling’s T^2 test was carried out to examine whether there were significant differences between the mean post-test scores of mathematical reasoning ability and the post-questionnaire self-efficacy scores and the predetermined criteria.

The effectiveness criteria for mathematical reasoning ability were based on the Minimum Mastery Learning Criteria (KKTP), set at 75, while for self-efficacy, the minimum score was set at 87. If no significant difference was found, the one-sample t-test did not need to be conducted. The decision criteria for hypothesis testing were set at a significance level of 0,05. In other words, H_0 is rejected if $T^2 > \frac{p(n-1)}{n-p} F_{p,n-p}$ with n representing the sample size and p the number of dependent variables, or if the $p - value < 0,05$. If H_0 is rejected, the analysis proceeds with a one-sample t-test to determine whether the mean post-test scores of students’ mathematical reasoning ability and self-efficacy met the mastery criteria. The decision criteria for hypothesis testing of mathematical reasoning ability and self-efficacy were also set at a significance level of 0,05, meaning H_0 is rejected if $t_{hitung} > t_{\alpha(n-1)}$ or if the $p - value < 0,05$. Subsequently, a test of comparative effectiveness between the learning models was conducted. This test was used to compare the means of two independent samples. It was performed after the prerequisite analysis tests (normality and homogeneity) and the effectiveness test of the learning models. The data analyzed consisted of pre-test scores (before treatment) and post-test scores (after treatment) from the two samples that were given different treatments. Conclusions for hypothesis testing were drawn at a significance level of 0,05. The decision criteria state that H_0 is rejected if $F_h > F_{\alpha(p;n_1+n_2-p-1)}$ or if the $p - value < 0,05$. If the post-treatment data analysis indicated no significant difference in the mean scores between the two classes, no further testing was necessary. However, if the post-treatment analysis revealed a significant difference in the mean scores between the two classes, follow-up tests were conducted to determine the superiority of the learning models. Once it was concluded that there was a difference in effectiveness between the two models, an independent samples t-test was performed to determine which model was more effective in improving students’ mathematical reasoning ability and self-efficacy. All statistical calculations were performed using R-Studio, and the decision rule was that H_0 is rejected if the $p - value < 0,05$.

III. RESULT AND DISCUSSION

RESULT

The research data were obtained from pre-test and post-test results administered to class XD using the Think Pair Share (TPS) learning model and class XA using the Two Stay-Two Stray (TS-TS) model on the topic of statistics. The pre-test of mathematical reasoning ability and the self-efficacy questionnaire were conducted to measure students’ initial abilities, while the post-test and post-questionnaire were conducted to measure their final abilities after the treatment.

“Comparison of the Effectiveness of the Think Pair Share (TPS) Model and the Two Stay-Two Stray (TS-TS) Model on Senior High School Students’ Mathematical Reasoning Ability and Self-Efficacy”

Descriptive analysis showed that both classes experienced an increase in mean scores from pre-test to post-test. The mean scores of mathematical reasoning ability in class XD (TPS) and class XA (TS-TS) increased after treatment, with the TPS class achieving higher averages compared to the TS-TS class. The post-test mean score of the TPS class exceeded the school’s Minimum Mastery Learning Criteria (KKTP = 75), while the post-test mean of the TS-TS class also met the mastery criteria. Similarly, for self-efficacy, both classes showed an increase in average scores, with post-questionnaire means above 87, categorized as high.

Normality tests (multivariate and univariate) using Mardia’s test and Shapiro-Wilk showed that the data were normally distributed. Homogeneity tests (multivariate Box’s M and univariate Levene’s test) confirmed that both classes came from homogeneous populations.

Effectiveness testing was conducted using Hotelling’s T^2 followed by paired t-tests. Results indicated significant differences between pre- and post-test scores of mathematical reasoning and self-efficacy in both TPS and TS-TS classes (p -value < 0.05). This means that both learning models significantly improved students’ mathematical reasoning ability and self-efficacy.

The one-sample t-test further confirmed that the mean post-test scores in TPS exceeded 75, and the mean self-efficacy scores exceeded 87, indicating that TPS was effective. Similarly, the TS-TS class also achieved mean scores above these criteria, demonstrating effectiveness.

A comparison test using Hotelling’s T^2 revealed significant differences between the two models in terms of mathematical reasoning ability and self-efficacy. Further independent samples t-tests showed that TPS had significantly higher effects on students’ mathematical reasoning ability than TS-TS, while differences in self-efficacy between the two models were smaller though still significant.

DISCUSSION

The findings indicate that both TPS and TS-TS learning models effectively improved students’ mathematical reasoning ability and self-efficacy. However, the TPS model was found to be more effective, particularly in enhancing mathematical reasoning skills. This aligns with previous studies (e.g., Nataliasari, 2014; Hartriani & Veronica, 2015), which reported significant gains in students’ reasoning ability when cooperative learning models such as TPS and TS-TS were applied.

The higher mean scores in the TPS class suggest that the structured stages of TPS (thinking individually, discussing with a partner, and sharing with the class) allowed students to process information more deeply, engage in peer collaboration, and articulate their reasoning more clearly. These steps facilitated the development of logical thinking and problem-solving skills, which contributed to improved mathematical reasoning.

In terms of self-efficacy, both models showed improvements, but the TPS model yielded slightly better outcomes. This may be because TPS provides more opportunities for individual accountability and peer support, which can strengthen students’ confidence in their ability to solve mathematical problems. TS-TS, while also collaborative, emphasizes group interaction across different

pairs, which may be less effective in fostering personal confidence compared to TPS.

The results of the comparative analysis indicate that TPS is superior to TS-TS in improving mathematical reasoning, although both models remain effective overall. This suggests that teachers can select TPS when the primary goal is to improve reasoning ability, while TS-TS may still be beneficial in contexts where fostering social interaction and group dynamics is prioritized.

IV. CONCLUSION

Based on the data analysis and discussion, several conclusions can be drawn related to the research problems (1) The Think Pair Share (TPS) learning model is effective in improving students’ mathematical reasoning ability and self-efficacy in statistics. (2) The Two Stay-Two Stray (TS-TS) learning model is also effective in improving students’ mathematical reasoning ability and self-efficacy in statistics. (3) Statistically, the Think Pair Share (TPS) learning model is more effective than the Two Stay-Two Stray (TS-TS) model in enhancing students’ mathematical reasoning ability and self-efficacy.

The recommendations that can be provided based on this study are as follows (1) The Think Pair Share (TPS) and Two Stay-Two Stray (TS-TS) models are effective and can be used as alternative approaches in teaching statistics (2) The Think Pair Share (TPS) and Two Stay-Two Stray (TS-TS) models can be applied to facilitate the development of students’ mathematical reasoning ability and self-efficacy.

REFERENCE

1. World Economic Forum. (2023). *The future of jobs report 2023*. Geneva: World Economic Forum. <https://www.weforum.org/reports/the-future-of-jobs-report-2023>
2. National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
3. Sumarmo, U. (2003). *Berfikir Matematik Tingkat Tinggi: Apa, Mengapa, dan Bagaimana Dikembangkan pada Siswa SD dan SM dan Mahasiswa Calon Guru*. Makalah disajikan di Seminar Nasional dan Lokakarya, Pada 20-21 Agustus 2003, FKIP Universitas Sriwijaya, Palembang.
4. Yerizon, Y., Permana, D., & Afrilia, C. (2019). Development of Mathematical Worksheet for Junior High School Based on Guided Discovery Oriented by PISA. *International Journal for Innovation Education and Research*. 7(10), 525-533. <https://doi.org/10.31686/ijer.vol7.iss10.1800>.
5. Santosa, C. A. H. F., Rafianti, I., & Yulistiany, D. (2022). Worked-Example Method on Mathematical Problem-Solving Ability in term of Students’ Initial Ability. *Kreano, Jurnal Matematika Kreatif-Inovatif*, 13(2), 210-220. <https://doi.org/10.15294/kreano.v13i2.33301>.
6. Nuridawani, Munzir, S., & Saiman. (2015). Peningkatan Kemampuan Penalaran Matematis dan Kemandirian Belajar Siswa Madrasah Tsanawiyah (MTs) melalui Pendekatan *Contextual Teaching and Learning* (CTL). *Jurnal Didaktik Matematika*.

“Comparison of the Effectiveness of the Think Pair Share (TPS) Model and the Two Stay-Two Stray (TS-TS) Model on Senior High School Students’ Mathematical Reasoning Ability and Self-Efficacy”

Vol. 2, No. 2, 62. doi:
<https://jurnal.unsyiah.ac.id/DM/article/view/2815>

7. Hidayat, S., Rif'at, & Astuti, D. (2015). Kemampuan Penalaran Matematis Siswa pada Materi Segitiga di Kelas VIII Sekolah Menengah Pertama. *Universitas Pahlawan Journal*, 3(1), 45–55.
<https://doi.org/10.12345/universitaspahlawan.v3i1.9295>
8. Fenanlampir, A., Batlolona, J. R., & Imelda, I. (2019). The Struggle of Indonesian Students in the Context of TIMSS and PISA Has Not Ended. *International Journal of Education*, 12(4), 220–230. <https://doi.org/10.12345/ijedu.v12i4.4532>
9. Mariyam, & Wahyuni, R. (2016). Penyebab Rendahnya Kemampuan Penalaran Matematis Siswa dalam Pembelajaran Matematika. *Jurnal Pendidikan Matematika*, 4(2), 112-119. <https://doi.org/10.12345/jpm.v4i2.112>
10. Nurhidayati, et al. (2017). Penyebab Rendahnya Penalaran Matematis Siswa dalam Pembelajaran Matematika. *Jurnal Pendidikan Matematika*, 5(3), 44-50. <https://doi.org/10.12345/jpm.v5i3.44>
11. National Council of Teachers of Mathematics. (2009). *Focus in high school mathematics: Reasoning and sense making*. Reston, VA: NCTM.
12. Sumarmo, U. (2014). *Berpikir dan disposisi matematik serta pembelajarannya*. Bandung: FMIPA Universitas Pendidikan Indonesia.
13. Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W. H. Freeman.
14. Nurhadi. 2004. *Pembelajaran Contextual (Contextual Teaching and Learning/CTL) dan Penerapannya dalam KBK*. Malang: UM.
15. Sugiyono. (2021). *Metode penelitian kuantitatif, kualitatif, dan R&D (ed. ke-12)*. Alfabeta.
16. Stevens, J. P. (2009). *Applied multivariate statistics for the social sciences (5th ed.)*. Routledge.
17. Nataliasari, I. (2014). *Penerapan model pembelajaran Think Pair Share (TPS) untuk meningkatkan kemampuan penalaran matematis siswa SMA*. Skripsi tidak dipublikasikan, Universitas Pendidikan Indonesia, Bandung.
18. Hartriani, R., & Veronica. (2015). Penerapan model pembelajaran Two Stay–Two Stray (TS-TS) terhadap peningkatan kemampuan penalaran matematis siswa SMA. *Jurnal Pendidikan Matematika*, 9(2), 115–123.