Scilight

Journal of Modern Applied Statistical Methods
https://jmasm.com/index.php/jmasm/index

Article

Independent Samples *t* Test and the Mann-Whitney-Wilcoxon Test to Know the Effect of the Drill Method on Mathematics Learning Outcomes

Nurwiani * and Lia Budi Tristanti

Mathematics Department, PGRI Jombang University, Jombang 61418, Indonesia
* Correspondence: nurw_13iem64@yahoo.com

How To Cite: Nurwiani; Tristanti, L.B. Independent Samples *t* Test and the Mann-Whitney-Wilcoxon Test to Know the Effect of the Drill Method on Mathematics Learning Outcomes. *Journal of Modern Applied Statistical Methods* **2025**, *24*(1), 1. https://doi.org/10.56801/Jmasm.V24.i1.1.

Abstract: This study aims to analyze the effect of the drill method on the mathematics learning outcomes of seventh-grade students at SMP Negeri 5 Jombang. The pre-test results were tested for normality using the Shapiro-Wilk test and for homogeneity using the Bartlett test. The pre-test results showed the data were normally distributed and homogeneous. The *t*-test indicated that the pre-test mean scores between the control and experimental classes were the same. However, the post-test data for the experimental class were not normally distributed, so homogeneity was tested using the Levene test. The Mann-Whitney test showed a significant difference, proving that the drill method affects students' learning outcomes.

Keywords: independent samples *t* test; Mann-Whitney-Wilcoxon test; the drill method; mathematics learning outcomes; State Junior High School 5 Jombang

1. Introduction

Independent samples *t*-test is a statistical tool often used in educational research to compare means from two different groups [1]. Independent samples *t*-test helps educational researchers or practitioners to determine whether or not there are significant differences between two groups in terms of certain variables, such as academic achievement, response to intervention, or other characteristics [2]. Some researchers use the independent samples *t*-test by comparing the calculated *t*-value with the *t*-table or comparing the significance value with the significance level value (α) determined by the researcher [3]. For example, an independent samples *t*-test to compare the average test results between students who studied with method A and students who studied with method B. The main purpose of the independent samples *t*-test is to determine whether there is a significant difference between the means of the two groups [4]. To do this, the *t*-test tests whether or not the differences between the means of the groups are statistically significant. The process involves calculating a test statistic (*t*-statistic) which compares the difference between the means of the two groups with the variation or variability within each group [5]. The results of the *t*-test are then compared with the critical value to determine whether the difference between the means is statistically significant. The results of the independent samples *t*-test are interpreted by referring to the p-value. A low p-value indicates that the difference between the means of the two groups is most likely not the result of pure chance, while a high p-value indicates that there is not enough evidence to reject the null hypothesis (no significant difference between the groups the).

The main conditions that must be met to carry out an independent samples *t*-test are that the data in the two groups must have a normal distribution and the variances of the two groups must be the same or almost the same. If one or more of these conditions are not met, you can consider using other alternatives, such as non-parametric tests (for example the Mann-Whitney-Wilcoxon test) which do not require the assumption of normal distribution or homogeneity of variance [6]. The Mann-Whitney-Wilcoxon test, often referred to as the Mann- Whitney U test or Wilcoxon rank-sum test, is a non-parametric statistical test used to compare two independent groups [7,8].

A learning process needs to use learning methods, so that it can provide opportunities for students to learn well [9]. With appropriate learning methods, it is hoped that various student learning activities will grow, in other words good learning interactions will be created between teachers and students. Therefore, a good learning method is needed where the method can foster student learning activities and is appropriate to the learning material. There are several methods in learning. One method that can be used in learning is the drill method [10].

One way in the teaching and learning process is to use the practice method. The training method is a teaching approach where students are invited to a training location to observe and understand how a skill is created, how it is used, the purpose of the creation, its benefits, and other related matters [11]. Using the practice method in learning has several advantages, namely (a) students can develop motor skills such as writing, pronouncing letters, making, and using tools. (b) encourage the development of mental skills such as multiplication, addition, subtraction, division, recognition of signs/symbols, and so on. (c) helps in forming good habits and increasing accuracy and speed in implementing these skills [12]. The drill method learning steps include): a. provide an explanation to students regarding the meaning, benefits, and objectives of the exercise, b. do the exercises in stages, starting from simple ones and then increasing to more complex or difficult levels, c. pay attention to parts of the exercise that students find difficult when they do it, d. Prioritize accuracy in doing exercises so that students can do them correctly, then pay attention to speed, so that students can hone their skills according to the time set, e. adjust the training period so that it is not boring but can change the situation to be fun, f. pay attention to individual differences so that the student's abilities and needs can be channeled well [13].

The drill method is a method that teaches students to carry out practice activities repeatedly so that students have higher dexterity or skills from the things they have learned [14,15]. In mathematics learning, the drill method emphasizes practice activities which can take the form of practicing calculating and operating numbers and understanding and solving problems with your abilities. To make it easier for students to understand the material, practice solving questions is needed because, if students do practice questions gradually and continuously, they can increase their understanding and memory of formulas and technical skills in solving questions [16,17]. Apart from that, practice questions can be in the form of multiple-choice questions or description questions which are questions of understanding, application, or analysis. Through these training activities, skills and understanding of the material being studied can be produced so that good and satisfying learning results are obtained.

Based on the results of previous research conducted by Novyanti (2022) [18], the use of the drill method can work well, it can be seen that students become more active and understand the material being studied more easily so that the learning process is more effective. Based on the results of this research, the average pre-test score was 45.37, and for the post-test was 79.68. Rusyani et al. (2022) [19] used the abacus-assisted dribbling method to improve arithmetic subtraction operation skills in deaf children. The research results of Karnes et al., (2021) [20] also show an increase in students' addition and subtraction abilities through practice methods. Derrick et al.'s (2017) [21] research compared two samples for paired data and independent data. This research used paired sample *t*-test and independent sample *t*-test. These two statistics tests belong to Type I error robust, and they are more powerful than standard tests.

In this study, researchers determined how much influence the drill method had on student learning outcomes. Research data was obtained by knowing the student's initial abilities (pre-test scores) and the students' final abilities after being given treatment (post-test scores). To determine the effect of the drill method on student learning outcomes, researchers used the independent sample *t*-test, there is an assumption that the two standard deviations are equal [22] and the coefficient of determination.

2. Methodology

This research is a type of quantitative research used to analyze data measuring the influence of the drill method on the learning outcomes of mathematics. The population consists of seventh-grade students from SMP Negeri 5 Jombang. The data pattern for the research on treatment *k* is presented in Table 1 below:

Table 1. Research Data Pattern.

	j			
i	**1**	**2**	$\cdots$	k
1	x_{11}	x_{11}	$\cdots$	x_{11}
2	x_{21}	x_{21}	$\cdots$	x_{21}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
n_j	$x_{n_1 1}$	$x_{n_2 1}$	$\cdots$	$x_{n_k 1}$

Table 1. *Cont.*

	j			
i	**1**	**2**	**⋯**	***k***
Data *ij*-th	x_{*1}	x_{*2}	⋯	x_{*k}
The number of samples	n_1	n_2	⋯	n_k
Mean of the *j*-th sample	$\bar{x}_1$	$\bar{x}_2$	⋯	$\bar{x}_k$
Variance s_j^2	s_1^2	s_2^2	⋯	s_k^2

where:

n_j is the number of samples for treatment j for $j = 1, 2, \cdots, k$

k is the number of treatments

$i = 1, 2, \cdots, n_j$

The researcher determines the significance level at ∝ = 0.05 and calculates the Sig or P-value to conduct statistical test analysis. A P-value is the lowest level (of significance) at which the observed value of the test statistic is significant [23]. In making a decision, the researcher compares the P-value and the significance level ∝. Before conducting the independent samples *t*-test on the sample data, the researcher performs a normality test and a homogeneity test.

Lin and Mudholkar (1980) stated that the Shapiro-Wilk test is more powerful than the Kolmogorov-Smirnov test for exponential distributions [24]. The researcher performs the normality test using the Shapiro-Wilk test. The Shapiro-Wilk test uses the Shapiro-Wilk test coefficient for $20 < n \leq 50$ [25].

The first step the researcher took before giving treatment to the selected random sampling cluster was to pay attention to the students' initial abilities by providing pre-test. The pre-test was carried out in two classes to find out that the initial abilities of the two classes (control class and experimental class) were the same. After it was stated that the initial abilities of the two classes were the same, different treatment was given to the two classes. To determine the effect of the drill method on students' mathematics learning outcomes, researchers used coefficient determination with independent sample *t* test.

The research instrument used test sheets before (pre-test) and after (post-test) the intervention. The test sheet is a sheet used to determine the learning outcomes of students who were treated using the drill method (experimental class) and without using the drill method (control class).

Researchers determine the magnitude of ∝ and calculate the Sig value (P value) to carry out statistical analysis of the test. A P-value is the lowest level (of significance) at which the observed value of the test statistic is significant [23]. When making a decision, researchers compare the P-value and significance level α. Before carrying out an independent sample *t* test on the sample data, researchers carried out a normality test and homogeneity test. Nurwiani et al. (2014) conducted normality testing using the Kolmogorov- Smirnov test [26]. If the data were not normally distributed, they applied Johnson Transformation. However, in this study, the normality of the data was assessed using the Shapiro-Wilk test. The research results of Razali and Wah (2011) show that the Shapiro-Wilk test is the strongest normality test [27]. Researchers carried out a normality test with the Shapiro-Wilk test with the following steps.

2.1. Normality Test with the Shapiro-Wilk Test

The test statistic is given by the following test procedure:

a. determine the proposed hypothesis as follows:

*H*0: the data is normally distributed

*H*1: the data is not normally distributed

b. calculate:

$$t_{3j} = \begin{cases} \dfrac{1}{d_j}\left[\sum_{i=1}^{m} a_i\left(x_{(n_j-i+1)} - x_{ij}\right)\right]^2 & , n_j \text{ odd } i = 1,2,\cdots, m = \dfrac{n_j+1}{2} \\ \dfrac{1}{d_j}\left[\sum_{i=1}^{m} a_i\left(x_{(n_j-i+1)} - x_{ij}\right)\right]^2 & , n_j \text{ even } i = 1,2,\cdots, m = \dfrac{n_j+1}{2} \end{cases} \quad (1)$$

where:

a_i: coefficient Shapiro−Wilk test for sample data of n_j in the j-th treatment

x_{n_j-i+1}: data to $(n_j - i + 1)$ or sample data of n_j in the j-th treatment

x_{ij}: data i-th where $i = 1, 2, \cdots, m$ for sample data of n_j in the j-th treatment

$d_j = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2$,

$\bar{x}_j$: average of j-th treatment data

Determine the value of Sig_j using Equation (1) with the rejection criteria H_0 for the following:

If $Sig_j < \propto$, then reject H_0

If $Sig_j \geq \propto$, then accept H_0

By using the Shapiro-Wilk test probability table, the value of Sig_j can be determined.

c. Apart from using Equation (1), determining the Shapiro-Wilk test can also be determined using the following Equation (2):

$$g_j = b_{n_j} + c_{n_j} + ln\left(\frac{t_{3_j} - d_{n_j}}{1 - t_{3_j}}\right) \quad (2)$$

where:

g_j: identical to the Z value of a normal distribution;

$b_{n_j}, c_{n_j}, d_{n_j}$: statistical conversion of the Shapiro-Wilk test with a normal distribution approach.

Determine the value of Sig_j using the two tails test obtained from Equation (2) to determine the rejection criteria for H_0 as follows:

If $Sig_j < \propto$, then reject H_0

If $Sig_j \geq \propto$, then accept H_0

with the normal distribution with one tail the values can be determined [28]:

$$\begin{aligned} Sig_j &= P(Z \leq g_j) \\ &= N(g_j) \\ &= \int_{-\infty}^{g_j} \frac{1}{\sqrt{2\pi}} e^{-\omega/2} d\omega \\ &[N(-g_j) = 1 - N(g_j)] \end{aligned}$$

2.2. Homogeneity Test

The researcher determines homogeneity testing using Bartlett's Test of Homogeneity and Levene's Test of Homogeneity. The researcher formulates the hypothesis for the homogeneity test as follows:

$$H_0: \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 \neq \cdots \neq \sigma_k^2$$

The researcher conducted a test of equality of *k* variances to determine the homogeneity of *k* treatments in the samples.

2.2.1. Homogeneity Bartlett Test

The Bartlett test can be used if the data has been tested for normality and the results are normal. The next step, the researcher carried out an equality of variance test. $\sigma_1^2, \sigma_2^2 \cdots \sigma_k^2$ variance equality test is used to find out whether the k samples are homogeneous. If there are several normally distributed populations with variance $\sigma_1^2, \sigma_2^2 \cdots \sigma_k^2$. then a two tails (two sided) homogeneity test is carried out.

The Bartlett test can be used if the data has been tested for normality and the results are normal. The next step, the researcher carried out an equality of variance test. The *k* variance equality test is used to find out whether the *k* samples are homogeneous. If there are several populations with normal distribution with variance, a two tails (two sided) homogeneity test is carried out. The test statistic is given by the following test procedure:

a. determine the variance $s_1^2, s_2^2, \cdots, s_k^2$ on Table 1 with the following Equation (3):

$$s_j^2 = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{ij} - \bar{x}_j\right)^2}{n_j - 1} \quad (3)$$

b. determine the combined variance with using Equation (3), thus obtaining the following Equation (4):

$$s^2 = \frac{\sum_{j=1}^{k} (n_j - 1)s_j^2}{\sum_{j=1}^{k} n_j - k} \tag{4}$$

c. The Bartlett test can be determined using Equations (3) and (4), so that the following Equation (5) is obtained:

$$B = (\log s^2)\left(\sum_{i=1}^{n_j} (n_j - 1)^2\right) \tag{5}$$

d. Using Equation (5), the Barlett test value can be determined using the following Chi-Square test:

$$\chi^2 = (\ln 10)\left[B - \sum_{j=1}^{k} (n_j - 1)\log s_j^2\right] \tag{6}$$

e. Determining the value of Sig for two-tailed testing obtained using Equation (6) and establishing the criteria for rejectin H_0 as follows:
If $Sig < \propto$, then reject H_0
If $Sig \geq \propto$, then accept H_0
With the Chi-Square distribution, the values can be determined as follows [28]:

$$\text{Sig} = P(X \leq \chi^2)$$
$$= \int_0^{\chi^2} \frac{1}{\Gamma(r/2)2^{r/2}} \omega^{r/2} e^{-\omega/2} d\omega$$

2.2.2. Homogeneity Levene Test

If the data is not normally distributed, Levene's homogeneity test can be used. Levene's test (O'Neill et al. 2002, Lemeshko et al., 2010) is used to test the equality of variances of several populations [29,30]. The Levene test is an alternative test to the Bartlett test, if the data is not normally distributed. If there is strong evidence that the data is normally or near-normally distributed, then the Bartlett test is better to use. Levene's test uses one-way analysis of variance. The data is transformed by looking for the difference between each score and the group average. The Levene test steps are as follows:

a. determine the statistical value of the Levene test with the following test:
b. determine the statistical value of the Levene test using the following Equation (7)

$$\ell = \frac{(\sum_{j=1}^{k} n_j - k)\sum_{j=1}^{k} n_j(\bar{y}_{.j} - \bar{y}_{..})^2}{(k-1)\sum_{j=1}^{k}\sum_{i=1}^{n_j} (y_{ij} - \bar{y}_{.j})^2} \tag{7}$$

where:

$$y_{ij} = |x_{ij} - \bar{x}_{\cdot j}|$$

$\bar{x}_{\cdot j}$ is the average of the j-th group
$\bar{y}_{.j}$ is the group mean $y_{.j}$.
$\bar{y}_{..}$ overall average of y_{ij}

c. Determine the Sig value for the two tails test using Equation (7) with the following H_0 rejection criteria:
If $Sig < \propto$, then reject H_0
If $Sig \geq \propto$, then accept H_0
With the F distribution, values can be determined [27]:

$$Sig = P(F \leq \ell)$$
$$= \int_0^{\ell} \frac{\Gamma[(r_1 + r_2)/2](r_1/r_2)^{r_1/2}\omega^{r_1/2-1}}{\Gamma(r_1/2)\Gamma(r_2/2)} d\omega$$

2.3. Independent Samples t Test

The more prevalent situations involving tests on two means are those in which variances are unknown. If the scientist involved is willing to assume that both distributions are normal and that $\sigma_1 = \sigma_2 = \sigma$, the pooled *t*-test (often called the two-sample *t*-test) may be used. The test statistic is given by the following test procedure.

a. determine the hypothesis:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

b. Independent sample *t* test for $\sigma_1 = \sigma_2 = \sigma$, but σ is unknown, it is determined with the computed *t*-statistic using the following Equation (8):

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \tag{8}$$

with:

s can be obtained by Equations (3) and (4) for $j = 1, 2$
t: the computed *t*-statistic independent samples *t* test
$\bar{x_j}$: average value of sample j for $j = 1, 2$
n_j: number of samples j for $j = 1, 2$

c. Determine the *Sig* value for the two tails test using Equation (8) with the H_0 rejection criteria as follows
If $Sig < \propto$, then reject H_0
If $Sig \geq \propto$, then accept H_0
with the *t* distribution can be determined [28]:

$$\text{Sig} = P(T \leq t)$$
$$= \int_{-\infty}^{t} \frac{\Gamma[(r+1)/2]}{\sqrt{\pi r}\Gamma(r/2)(1+\omega^2/r)^{(r+1)/2}} d\omega$$
$$P(T \leq -\, t) = [1 - P(T \leq t)]$$

2.4. Statistical Nonparametric Tests

If the post-test scores are not normally distributed, then the researcher compares the average post-test scores for the control class and experimental class using the Mann-Whitney-Wilcoxon test. Data tested using the Mann-Whitney-Wilcoxon test must be on a ratio, interval or ordinal scale (if the data is on an interval or ratio scale if the normality assumption is not met), there are 2 groups of data being tested, it is not affected by the normality of the data, the data is not paired or different. groups, and the groups tested have the same variance, aka homogeneous.

The Mann-Whitney-Wilcoxon test is a non-parametric test used to determine the difference in medians from two independent samples [31]. This test is used when the data does not meet the normality assumption. However, some experts still state that the Mann-Whitney-Wilcoxon test does not only test median differences, but also tests the mean. This is because in various cases, the median of the two groups may be the same, but the resulting P value is small, namely <0.05, which means there is a difference. The reason is because the means of the two groups are significantly different. So it can be concluded that this test not only tests the Median difference, but also the Mean difference.

The Mann-Whitney-Wilcoxon test steps are as follows:

a. determine the hypothesis:

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

b. combine the two independent groups (samples);
c. sort each member of the observation value starting from the smallest value to the largest value;
d. if there are two or more observation values, then the ranking given to each sample member is the average ranking;
e. calculate the number of ratings for each sample group (R_1 and R_2);
f. determine the values of u_1 and u_2 with the following Equation (9).

$$u_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$
$$u_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2 \quad (9)$$

with
u_1: number of ratings 1
u_2: number of ratings 2
n_1: number of samples 1
n_2: number of samples 2
R_1: number of rankings in sample 1
R_2: number of rankings in sample 2
Equation (8) is used for values $n \le 20$ with u: the smallest value of u_1 or u_2

g. for the value $n > 20$ the test value is determined using the following Equation (10):

$$z = \frac{u - E(u)}{\sqrt{var(u)}} \quad (10)$$

with
$E(u) = \frac{n_1 n_2}{2}$
$var(u) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$

h. determine the *Sig* value using Equation (10) with the following H_0 rejection criteria:
If $Sig < \propto$, then reject H_0
If $Sig \ge \propto$, then accept H_0

with the normal distribution value can be determined [28]:

$$Sig = P(Z \le z)$$
$$= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\omega^2/2} d\omega$$

3. Results

This research design uses Matching pre-test—post-test Control Group Design. Matching pre-test—post-test Control Group Design is a study with two classes where both are given a pre-test to determine the initial abilities between the experimental class and the control class. The experimental class was treated using the drill method while the control class did not use the drill method. After completion, both classes were given a post-test.

This research design used Matching pre-test—post-test Control Group Design. Matching pre-test—post-test Control Group Design is a study with two classes where both are given a pre-test to determine the initial abilities between the experimental class and the control class. The experimental class was treated using the drill method while the control class did not use the drill method. After completion, both classes were given a post-test. Table 2 below presents the two types of validation and the description.

Table 2. Validity type

Type of Validity	Description
Content validity	Suitability of questions to the material
	It is possible that the questions can be solved
	Instructions for writing questions are formulated clearly
	The questions are formulated briefly and clearly
Language validity	Conformity of the language used in the questions follows Indonesian language rules
	The question sentence does not show ambiguity
	The question sentences use simple language for students
	The question sentences are easy for students to understand

After the pre-test and post-test questions were declared valid by the validator, a pre-test was then carried out on two samples for the control class and the experimental class. Before being given different treatment in the control class and experimental class, several statistical tests were carried out to determine the students' initial abilities. The statistical tests carried out were normality, homogeneity test, independent *t* test on pre-test scores. Researchers used research data [32] for data analysis.

Using Equations (1) and (2) and SPSS software, the results of the Shapiro-Wilk normality test of students' initial abilities (pre-test scores) in the control class and experimental class are presented in Table 3 below.

Table 3. Normality test Shapiro-Wilk pre-test scores.

Tests of Normality			
	Shapiro-Wilk		
	Statistic	df	Sig.
Pre-test Score of Control Class	0.966	32	0.406
Pre-test Score of Experiment Class	0.984	32	0.908

Based on the results of data processing, the pre-test scores for the control class and experimental class in Table 3 obtained $Sig_1 = P(Z \leq 0.966) = 0.406$ and $Sig_2 = P(Z \leq 0.984) = 0.908$. In other words, the control class pre-test score $Sig_1 = 0.406$ and the experimental class pre-test score $Sig_2 = 0.908$. The Sig_1 and Sig_2 Shapiro-Wilk normality test values are greater than 0.05, meaning that H_0: the data is normally distributed is accepted. It can be concluded that the pre-test scores for the control class and the experimental class are both normally distributed.

Based on the calculation results of the average pre-test score for the control class and experimental class with $n_1 = 32$, $n_2 = 32$ is $\bar{x_1} = 55.47$, $\bar{x_2} = 56.09$, and the standard deviation value is $s_1 = 12.513$, $s_2 = 14.079$. Before giving treatment to students in carrying out learning in the control class and in the experimental class, a test of homogeneity of variance of pre-test scores was carried out. The results of calculating the s^2 value using Equations (3) and (4) are:

$$s^2 = \frac{(32-1)12.513^2 + (32-1)14.079^2}{32+32-2} = 177.396705 \quad (11)$$

Using Equations (5) and (6), the results in Equation (11) and SPSS software, a test of homogeneity of variance of pre-test scores for the control class and experimental class with the Bartlett test is obtained in Table 4 below.

Table 4. Results of homogeneity of variance Bartlett test for pre-test scores.

Box's M		0.430
F	Approx.	0.423
	df1	1
	df2	11,532.000
	Sig.	0.515

Based on the output results of the test of homogeneity of variance of pre-test scroes with the Bartlett test in Table 4, it appears that the value of $Sig = P(X \leq 0.430) = 0.515$. Based on the hypothesis test of homogeneity of variance with a Sig value > 0.05, it can be concluded that H0 which states $\sigma_1^2 = \sigma_2^2$ is accepted. This means that the control class and experimental class data observed in the pre-test scores come from populations that have the same or homogeneous variance.

Next, an independent samples *t*-test was carried out for the pre-test scores. This was done to test whether the average pre-test scores for the control class and the experimental class were the same. By using Equations (10) and (11) and data processing with SPSS software, the following results are obtained in Table 5.

Table 5. Test results, average pre-test scores for control class and experimental class.

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Pre-test Scores	Equal variances assumed	−0.188	62	0.852
	Equal variances not assumed	−0.188	61.158	0.852

Based on the output of Table 5, the value of $t = -0.188$ and $Sig = 0.852$ (2—tailed. Value of $Sig > 0{,}05$ and H_0 which states $\mu_1 = \mu_2$ is accepted. It can be concluded that the average pre-test scores for the control class and experimental class is the same. Because the average pre-test scores for the control class and experimental class are the same, the two classes can be given different treatments. The control class was given no drill method and the experimental class was given the drill method. After students were given integer material using the drill method in learning in the experimental class, students were given a post-test in the control class and experimental class.

Based on data processing of post-test scores for the control class and experimental class with $n_1 = 32$, $n_2 = 32$ the average value $\bar{x_1} = 69.2188$, $\bar{x_2} = 82.8750$, and a standard deviation value of $s_1 = 8.66485$, $s_2 = 10.10828$. To find out how much influence the drill method has on students' mathematics learning outcomes (post-test scores), several statistical tests must be carried out. First, a normality test was carried out for the post-test scores for the control class and experimental class. By using Equations (1) and (2) and SPSS software, the Shapiro-Wilk Normality test output post-test scores are obtained in Table 6 below.

Table 6. Normality test Shapiro-Wilk post-test scores.

	Shapiro-Wilk		
	Statistic	df	Sig.
Post-test Score of Control class	0.965	32	0.380
Post-test Score of Experimental class	0.883	32	0.002

Based on the results in Table 6 the control class post-test score $Sig_1 = 0.380$ and the experimental class post-test score $Sig_2 = 0.002$ were obtained. Sig_1 is the post=test score for the control class and Sig_2 is the post-test score for the experimental class. It can be concluded that the post-test scores of the control class has a normal distribution, while the post-test score of the experimental class does not have a normal distribution. Because one of the post-test classes is not normally distributed, the next step is to carry out a test of homogeneity of variance for the post-test scores with the Levene test. Table 7 below is the result of homogeneity of variance Levene test for post-test scores.

Table 7. Results of homogeneity of variance Levene test for post-test scores.

Test of Homogeneity of Variances					
		Levene Statistic	df1	df2	Sig.
Post-test score	Based on Mean	0.106	1	62	0.746
	Based on Median	0.031	1	62	0.862
	Based on Median and with adjusted df	0.031	1	56.357	0.862
	Based on trimmed mean	0.059	1	62	0.808

Based on data processing using Equation (6) and SPSS software, the results of the test of homogeneity of variance of post-test scores with the Levene test are obtained in Table 7. It can be seen that based on mean the value of $\ell = 0.106$ and $Sig = P(F \leq 0.106) = 0.746$, so the value of $Sig > 0.05$. Based on the test of homogeneity of variance hypothesis, it can be concluded that H_0 which states $\sigma_1^2 = \sigma_2^2$ is accepted, which means that the control class and experimental class data observed in the post-test scores come from populations that have the same or homogeneous variance.

Because the post-test scores in the experimental class were not normally distributed, the Mann-Whitney test was used to compare the two averages for the control class and the experimental class. Based on the results of data processing using Equation (8) and SPSS software, the results obtained for the R_1 and R_2 values are in Table 8 below.

Table 8. R_1 dan R_2 values of the Mann-Whitney-Wilcoxon test.

Ranks				
	Class	N	Mean Rank	Sum of Ranks
Post-test Scores	control class	32	21.11	675.50
	exprerimental class	32	43.89	1404.50
	Total	64		

Based on the results of data processing in Table 8, the value $R_1 = 675.50$ and the value $R_2 = 1404.50$. The next step uses Equation (9) and SPSS software to obtain the Mann-Whitney-Wilcoxon test in Table 9 below:

Table 9. Mann-Whitney-Wilcoxon test.

Test Statistics	
	Post-test Score
Mann-Whitney-Wilcoxon	147.500

Wilcoxon W	675.500
Z	−4.902
Asymp. Sig. (2-tailed)	0.000

Based on the output in Table 9, the values obtained are z = 147.500 and Sig = 0.000. The Sig < 0,05 and H_0 which states $\mu_1 = \mu_2$ are rejected. It can be concluded that the average post-test score for the control class and experimental class is not the same. So it can be concluded that there is an influence of the Drill Method on learning outcomes in mathematics learning.

4. Conclusions

The influence of the drill method on students' mathematics learning outcomes can be determined by comparing the average scores of the control class and the experimental class pre-test scores using an independent t-test if the data is normally and homogeneously distributed. Test normality using the Shapiro-Wilk test and homogeneity using the Barlett test. After being given different treatment in mathematics learning to compare post-test scores, a normality test was carried out using the Shapiro-Wilk test and a homogeneity test using the Levene test, this was done because the experimental class was not normally distributed. To compare the average post-test scores for the control class and experimental class, the Mann-Whitney-Wilcoxon test was used. So that the learning method used has a greater influence on students' mathematics learning outcomes, further research is recommended using learning methods other than the drill method and larger samples.

Author Contributions

All authors contributed to the design of the study. The first author (N.) was responsible for the statistics-related content, while the second author (L.B.T.) was responsible for the education-related content. Both authors wrote the first draft, and the first author checked the final versions. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

The study was conducted according to the guidelines of the University of PGRI Jombang, and approved by the Ethics Committee (number 014/TKIA/KL/2024 and it was approved on 4 October 2024).

Informed Consent Statement

Informed consent was obtained from the parents of all subjects in this study

Data Availability Statement

The data used in this study was secondary data obtained by Rahayu [32] during her study for Bachelor's degree under the supervision of Nurwiani. However, the authors employed different analysis methods to Rahayu.

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

1. Hair, J.F.; Black, W.C.; Babin, B.J.; et al. *Multivariate Data Analysis*, 8th ed.; Cengage Learnin: Boston, MA, USA, 2019; pp. 355–360.
2. Rumsey, D.J. *Statistics for Dummies*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2016.
3. Gravetter, F.J.; Wallnau, L.B. *Statistics for the Behavioral Sciences*, 10th ed.; Cengage Learning: Boston, MA, USA, 2016.
4. Field, A. *Discovering Statistics Using IBM SPSS Statistics*, 5th ed.; SAGE Publications: London, UK, 2018.
5. Moore, D.S.; McCabe, G.P.; Craig, B.A. *Introduction to the Practice of Statistics*, 9th ed.; W.H. Freeman and Company: New York, NY, USA, 2017.
6. Daniel, W.W. *Applied Nonparametric Statistics*, 2nd ed.; Georgia State University: Atlanta, GA, USA, 1990.
7. Conover, W.J. *Practical Nonparametric Statistics*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1999.

8. McKnight, P.E.; Najab, J. Mann-Whitney U Test. In *The Corsini Encyclopedia of Psychology*, 4th ed.; John Wiley & Sons: Hoboken, NJ, USA, 2010.
9. Arends, R.I. *Learning to Teach*, 9th ed.; McGraw-Hill: New York, NY, USA, 2012.
10. Joyce, B.; Weil, M.; Calhoun, E. *Models of Teaching*, 9th ed.; Pearson: Boston, MA, USA, 2015.
11. Lufri; Fitri, R.; Yogica, R. Effectiveness of Concept-Based Learning Model, Drawing and Drill Methods to Improve Student's Ability to Understand Concepts and High-level Thinking in Animal Development Course. *J. Phys. Conf. Ser.* **2018**, *1116*, 052040.
12. Gan, C.; Cao, W.; Wu, M.; et al. Two-Level Intelligent Modeling Method for the Rate of Penetration in Complex Geological Drilling Process. *Appl. Soft Comput.* **2019**, *80*, 592–602.
13. Kumar, R.; Hynes, N.R.J.; Pruncu, C.I.; et al. Multi-objective Optimization of Green Technology Thermal Drilling Process using Grey-Fuzzy Logic Method. *J. Clean. Prod.* **2019**, *236*, 117711.
14. Killen, R. *Effective Teaching Strategies: Lessons from Research and Practice*, 4th ed.; Thomson Social Science Press: Melbourne, Australia, 2007.
15. Brown, H.D. *Teaching by Principles: An Interactive Approach to Language Pedagogy*, 2nd ed.; Longman: White Plains, NY, USA, 2001.
16. Gagne, R.M. *The Conditions of Learning and Theory of Instruction*, 4th ed.; Holt, Rinehart and Winston: New York, NY, USA, 1985.
17. Silberman, M. *Active Learning: 101 Strategies to Teach Any Subject*; Allyn and Bacon: Boston, MA, USA, 1996.
18. Novyanti, Y. The Effect of Drill Method on Numerical Ability of Grade V Students of Batulaccu State Elementary School Makassar. Doctoral Dissertation. Bosowa University Makassar, Indonesia 2022 (In Indonesian)
19. Rusyani, E.; Ratnengsih, E.; Putra, A.S.; et al. The Drilling Method Application using Abacus to Arithmetic Operations Skills in Student With Hearing Impairment at Special School. *Indones. J. Community Spec. Needs Educ.* **2022**, *2*, 1–10.
20. Karnes, J.; Barwasser, A.; Grünke, M. The Effects of a Math Racetracks Intervention on the Single-Digit Multiplication Facts Fluency of Four Struggling Elementary School Students. *Insights Into Learn. Disabil.* **2021**, *18*, 53–77.
21. Derrick, B.; Russ B; Toher, D.; et al. Test Statistics for the Comparison of Means for Two Samples that Include both Paired and Independent Observations. *J. Mod. Appl. Stat. Methods* **2017**, *16*, 137–157.
22. Ross, A.; Willson, V.L. *Basic and Advanced Statistical Tests: Writing Results Sections and Creating Tables and Figure*; Sense Publishers: Rotterdam, The Netherlands, 2017.
23. Walpole, R.E.; Myers, R.H.; Myers, S.L.; et al. *Probability and Statistics for Engineers and Scientists*, 9th ed.; Prentice-Hall: Hoboken, NJ, USA, 2016.
24. Lin, C.C.; Mudholkar, G.S. A Simple Test for Normality Against Asymmetric Alternatives. *Biometrika* **1980**, *67*, 455–461.
25. Yazici, B.; Yolacan, S. A comparison of Various Tests of Normality. *J. Stat. Comput. Simul.* **2007**, *77*, 175–183.
26. Nurwiani; Sunaryo, S.; Setiawan; et al. Ridge Regression in Calibration Models with Symmetric Padding Extension-Daubechies Wavelet Transform Preprocessing. *J. Mod. Appl. Stat. Methods* **2014**, *13*, 255–266.
27. Razali, N.M.; Wah, Y.B. Power Comparison of the Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *J. Stat. Model. Anal.* **2011**, *2(1)*, 21–32
28. Hogg, R.V.; McKean, J.W.; Craig, A.T. *Introduction to Mathematical Statistics*, 8th ed.; Pearson Education: London, UK, 2018.
29. O'Neill, M.E.; Mathews, K. Theory & Methods: A Weighted Least Squares Approachch to Levene's Test of Homogeneity of Variance. *Aust. N. Z. J. Stat.* **2002**, *42(1)*, 81–100.
30. Lemeshko, B.Y.; Lemeshko, S.B.; Gorbunov, A.A. General Problems of Metrology and Measurement Technique: Application and Power of Criteria for Testing the Homogeneity of Variance. Part I. Parametric Criteria. *Meas. Tech.* **2010**, *53(3)*, 237–246.
31. Qolby, B.S. Mann Whitney Test in Non-Parametric Statistics between The Different Level of using the Public Transport with the Private Vehicle. 2014. (In Indonesian)
32. Rahayu, D.D. The Impact of Drill Method to the Students' Mathematics Learning Outcome in the Seventh Grade at SMP Negeri 5 Jombang. Thesis, PGRI Jombang University, Jombang, Indonesia, 2023. (In Indonesian)