PENERBIT UMT
Menjana Khazanah Ilmuwan

Journal of Mathematical Sciences and Informatics
Journal Homepage: https://journal.umt.edu.my/index.php/jmsi
eISSN: 2948-3697
DOI: http://doi.org/10.46754/jmsi.2023.06.002

JMSI
Journal of Mathematical Sciences and Informatics

ANALYSING THE EFFECTIVENESS OF SHORT-TERM MATHEMATICS PROGRAMME: IN THE CASE OF "REKREASI MATEMATIK MELALUI ALAM"

SYERRINA ZAKARIA[1*], NOR HAZIRAH MOHD ZAKI[1] AND NUR BAINI ISMAIL[1]

[1]Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia; syerrina@umt.edu.my, zirahhazirah00@gmail.com, baini@umt.edu.my

**Corresponding author: syerrina@umt.edu.my*

ARTICLE INFO

Article History:
Received 9 March 2023
Accepted 8 June 2023
Available online 30 June 2023

Section Editor:
Ummu 'Atiqah Mohd Roslan

Keywords:
Performance assessment; short term program; non-parametric methods; mathematics programme.

ABSTRACT

Analysing the effectiveness of a short-term programme is crucial to ensure its success and impact. The analysis allows for the determination of whether a programme has achieved its intended outcomes, the identification of areas for improvement, and enables necessary adjustments to be made. One method that can be used to analyse such effectiveness is statistical analysis, which can provide a more objective and accurate assessment of a programme's outcome. This study aims to assess the effectiveness of a short-term mathematics programme known as "Rekreasi Matematik Melalui Alam" by analysing the scores of 48 participants before (pre-test) and after (post-test) the programme, using various non-parametric statistical methods. The results showed that the programme had a positive impact on participants' mathematics performance. The Mann-Whitney U test revealed significant gender differences, with females obtaining higher scores in both tests. The Kruskal-Wallis test indicated no significant difference between the pre-test scores among schools, but vice versa for post-test scores. This suggests the programme benefits students from diverse backgrounds. The Wilcoxon Signed Rank test showed improved performance following students' participation in the programme, despite its short duration. Utilising different teaching approaches helps students better comprehend the subject and stay engaged, while real-life examples highlight the importance of mathematics.

2020 Mathematics Subject Classification:

INTRODUCTION

Analysing the effectiveness of a short-term programme is crucial to ensure its success and impact. Analysis allows for the determination of whether a programme has achieved its intended outcomes, the identification of areas for improvement, and enables necessary adjustments to be made. As reported by [33], schools, corporations, hospitals, businesses, and various other institutions regularly organise short-term training programmes. In post-secondary education institutions, these programmes are commonly utilised as part of faculty development initiatives aimed at enhancing instruction and teaching quality. According to [20], short-term training refers to training sessions that are less than a full day, but greater than two hours.

Analysing a programme's effectiveness provides valuable information on its impact, which can be used to make informed decisions on future programmes and funding allocations. For instance, as reported by [26], the Big Math for Little Kids programme was determined to be effective in assisting

children's mathematical development. A robust analysing process also establishes accountability and ensures efficient resource utilisation. In short, evaluating the effectiveness of a short-term programme is crucial to ensure that it is making a positive impact and meeting its objectives. One method that can be used to analyse such effectiveness is the statistical method.

Using statistical analysis to evaluate a programme's effectiveness can provide a more objective and accurate assessment than qualitative analysis. Statistical analysis enables the collection and examination of quantifiable data, providing a clearer picture of the programme's performance. The data collected can be analysed using statistical methods to determine if there are any significant relationships between variables, making it easier to determine if the programme has achieved its intended outcomes. Moreover, statistical analysis allows for the comparison of results to other programmes, providing a benchmark for evaluating the programme's performance. In contrast, qualitative analysis often relies on subjective assessments and personal opinions, which can be susceptible to biases and emotions. While qualitative analysis can provide valuable insights, combining it with statistical analysis can provide a more comprehensive and objective evaluation of a programme's effectiveness.

The effectiveness of the short-term programme evaluated in this study is assessed based on participants' performance, which involves scores for pre and post-tests. Performance assessments refer to substantive activities that yield solid and valid scores [16]. In general, performance assessments are measures of students' performance, wherein they are asked to complete specific tasks that have been constructed. Researchers typically use checklists or rubrics to assess students' scores. There are two major types of performance assessments: Performance events, where students have limited opportunities to enhance their early performance, and performance tasks, which provide more time for students to prepare their responses. To achieve the objective of measuring and predicting students' achievement of an intended knowledge or skill, teachers should have the ability to create valid curriculum-embedded performance assessments [8]. Curriculum-Embedded Performance Assessments (CEPAs) are essentially informational content-area units designed to ensure consistency between the content taught and assessed, and the preparation of teachers [3].

Mathematics performance assessments are closely connected to hands-on instructional activities that can be developed to deliver reliable measures of mathematics achievement that distinguish students' instructional histories and differ from traditional multiple-choice test scores[2]. However, achieving objective learning designs can be challenging due to the complexity of learning and other factors, including intelligence, interests, talents, preparedness, activeness, and awareness of the importance of the education process. Improvements in education quality can also influence learning outcomes. In a quasi-experimental study, researchers compared the outcomes of students treated with performance assessment and self-assessment. The results showed that both approaches yielded learning outcomes covering cognitive, affective, and psychomotor aspects. However, the hypothesis test indicated that the execution of performance assessment led to greater improvements in students' learning outcomes. Additionally, there are varying levels of maths competency among students. Math competency represents students' ability to formulate, apply, and interpret mathematics in various contexts [9]. The study concluded that identifying each student's potential for mathematics is essential to provide an individualised educational response, which is a priority of inclusive, high-quality education.

Properly planned short-term programmes can yield positive results. For instance, a study aimed at preventing excessive weight gain among adolescents achieved its objective through a well-planned moderate physical activity and nutritional alterations in the school curricula [35]. Short-term results can act as mediators for longer-term effects [6]. A study by [25] showed that a one-

day field had a significant effect on the attitudes towards and knowledge of biology among 6th grade students from Slovak schools, regardless of gender and school. The researchers reached this conclusion by examining the significant differences in pre-test and post-test scores using paired t-tests. They concluded that the short-term effect of informal learning in the field improved their attitudes from negative to neutral.

In their study, [15] compared several approaches, including a few parametric tests, to evaluate data from factorial experiments with additional treatments. However, before reaching a statistical conclusion when using a parametric test, the normality assumption needs to be verified by applying The Shapiro-Wilk test or the Kolmogorov-Smirnov test [19]. Similarly, [33] explained that parametric methods require the assumption of a specific data distribution in the population from which the data were sampled, while non-parametric methods usually do not assume a specific probability distribution. They also stated that non-parametric methods are commonly chosen when the data distribution assumptions of parametric tests are not satisfied.

The Wilcoxon Signed Rank test is a useful method for establishing preference between two competing food sources in certain experimental design settings [5]. The paper emphasises the importance of preference indicators in experiments where the items selected and the order of their selection are critical. The Wilcoxon Signed Rank test is suggested as the method of choice due to its advantage in describing the selections made in experimental trials. In another study, [29] used the Wilcoxon Signed Rank test to determine the effect of health education on anaemia in adolescent girls. Based on the results of their analysis, the p-value signified that the level of knowledge of adolescents about anaemia increased in line with education. Similarly, [1] applied a p-value < 0.05, as well as an array of statistically significant differences for all statistical tests. They found contrasting mean scores of qualities and quantity of food among subcategories of the different demographic variables, such as age group, gender, and nationality. Moreover, in a study by [12], the Wilcoxon Signed Rank test was used to analyse the average of pre-test and post-test scores, along with two independent Mann–Whitney U Sample tests, to compare the significant difference in self-esteem levels of vocational high school students experiencing psychological distress.

The collected data of certain samples can be tested using the Shapiro-Wilk method to identify the normal distribution of the data [32]. If the data is found not to be normally distributed, researchers can then opt to use the Mann-Whitney test as an alternative method. Unlike some other tests, Wilcoxon Mann-Whitney does not require specific data distribution, whether symmetric or asymmetric [30]. In the Mann-Whitney U test, the z-score obtained is compared at a given level of significance with an appropriate critical value obtained from a normal distribution table. This comparison is used to determine whether to reject or accept the null hypothesis, typically at a 5% significance level [25].

Analysis of Variance (ANOVA) is a statistical technique used to analyse a variable influenced by different types and combinations of factors. It allows researchers to determine whether there is a significant difference between groups by examining the p-value [18]. any research studies utilise ANOVA to analyse their data effectively. The application of the two-way ANOVA involves fitting the interaction model and testing the significance of the interaction term [4]. In a study, researchers found that the interaction between light and resin are statistically significant at an alpha level of 0.05 and $p < 0.001$. Subsequently, multiple comparisons can be performed using the post hoc method. This allows researchers to identify specific significant differences between means by comparing treatments [4].

The Kruskal-Wallis test is a non-parametric test that can serve as an alternative to one-way ANOVA, as it does not require the normality assumptions[14]. Research findings indicate that the Kruskal-Wallis test exhibits higher power compared with the traditional one-way ANOVA.

Additionally, when dealing with symmetrical distributions, the Kruskal-Wallis test has shown comparable performance to parametric ANOVA tests. This test is also valuable in assessing shifts in location [10]. The Kruskal-Wallis test shares the same objective as other non-parametric tests, which is to determine whether samples originate from the same population. It can be considered a generalization of the Wilcoxon Signed Rank test from two samples to multiple samples [11]. However, it is important to note that this test still assumes certain conditions, such as observations in each group coming from populations with the same shape of distribution, and the samples being random and independent [24]. Significant results from the Kruskal-Wallis test indicate that at least one of the samples is different from the other samples, but it does not specify the exact location and magnitude of the differences that occur.

Researchers often encounter major challenges during the qualitative data analysis process, including over-reliance on qualitative software packages, dealing with an overwhelming amount of data due to line-by-line approaches, and not fully exploiting the potential of the data [7]. One study proposed and discussed the Qualitative Analysis Guide of Leuven, which offers a comprehensive approach to guide qualitative data analysis. In another study by [22], content analysis and the MAXQDA software version 2020 were employed to analyse data collected through WhatsApp. Content analysis involves analysing written, spoken, or visual messages related to a specific concept ([21]). To facilitate data collection, the researchers utilised the WhatsApp mobile application, utilising video calls and voice calls to interact with and gather data from selected participants.

During the implementation of the "Rekreasi Matematik Melalui Alam" mathematics short-term programme, it becomes essential to gauge its effectiveness on the participants' performance. To achieve this, the present study aims to assess the programme's effectiveness by employing various statistical techniques to analyse the collected data.

DATA AND METHODS

The two-day "Rekreasi Matematik Melalui Alam" short-term programme was attended by 48 pupils aged 11 years old from four different primary schools, held at Universiti Malaysia Terengganu. Data for the study were collected based on the participants' scores in mathematical questions before (pre-test) and after (post-test) the programme. The collected data were analysed using Statistical Package for the Social Sciences (SPSS) and Microsoft Excel software.

Before the method to analyse the data is selected, it is essential to check for the assumption of normality. The normality assumption implies that the sample follows a normal distribution, which is a prerequisite for applying parametric statistical methods to analyse the data. To verify the normality assumption, the Shapiro-Wilk test or Kolmogorov–Smirnov test can be employed. In this study, the Shapiro-Wilk test was chosen to assess the normality assumption for both the pre-test and post-test score data using SPSS software version 28.

H_0 : The data are normally distributed.

H_1 : The data are not normally distributed.

The data is not normally distributed as the p-value is less than 0.05. Therefore, the null hypothesis is rejected. Hence, non-parametric methods are used for further analysis.

The Mann-Whitney U test, a non-parametric test, is commonly used to assess differences between two independent groups [13]. In this study, this test was employed to analyse whether there are statistically significant differences in the median scores between genders.

Below are the hypothesis statements of the Mann-Whitney U test:

H_0 : There is no significant difference of the median scores between genders.

H_1 : There is a significant difference of the median scores between genders.

The Mann-Whitney U test is as follows:

$$U_a = (n_a * n_b) + \frac{n_a * (n_a + 1)}{2} - T_a S$$

$$U_b = (n_a * n_b) - U_a$$

where

n_a : sample size of first sample

n_b : sample size of second sample

T_a : sum of the ranks for one of the samples

Next, the Kruskal-Wallis test (H) is used to determine whether there is a significant difference between the medians of the pre-test groups and post-test groups. This test serves as an alternative to one-way ANOVA in non-parametric conditions. The Kruskal-Wallis test is applied to the school variable because there are more than two independent groups in the variable. The hypotheses for the pre-test groups are:

H_0 : There is no difference in the median pre-test scores among the schools.

H_1 : There is at least one median difference in the pre-test scores among the schools.

The hypotheses for the post-test groups are:

H_0 : There is no difference in the median post-test scores among all schools.

H_1 : There is at least one median difference in the post-test scores among the schools.

The formula used to compute the test statistic H [28]:

$$H = [\frac{12}{n\,(n+1)} \sum_{j=1}^{c} \frac{T_j^2}{n_j}] - 3(n+1)$$

where

n = sum of sample sizes for all samples

c = number of samples

T_j = sum of ranks in the j^{th} sample

n_j = size of the j^{th} sample

The Wilcoxon Signed Rank test is used to investigate whether there is any difference in the median scores of the pre-tests and post-tests for all students.

The hypothesis statements are as follows:

H_0 : There is no significant differences between the pre-test and post-test median scores.

H_1 : There is a significant difference between the pre-test and post-test median scores.

The p-value serves as a benchmark to interpret the results. When the p-value is less than a specified threshold, typically set at 0.05, the null hypothesis is rejected in favour of the alternative hypothesis.

RESULTS AND DISCUSSION

Table 1 presents the descriptive statistics for the dependent variables in this study, which are the pre and post-test scores analysed by gender. The mean scores for males are higher than for females, as are the median value for both tests. To assess significant differences in scores between genders, the Mann-Whitney U test.

Table 1: Descriptive statistics for gender

Test	Gender	Mean	Median	Std. Deviation
Pre	F	1.328	1.000	1.4160
	M	2.368	2.000	1.8245
Post	F	8.155	9.000	2.8382
	M	9.289	10.500	2.6051

Table 2 presents the descriptive statistics for the pre and post-test scores categorised by school. The increase in median values is consistent with the mean values observed in each school. To determine the significant differences in medians among more than two independent groups, the Kruskal-Wallis test (H test) was utilised.

Table 2: Descriptive statistics for schools

Test	School	n	Mean	Median	Std. Deviation
Pre	1	14	2.607	2.250	2.0956
	2	13	2.192	2.000	1.3623
	3	11	0.409	0.000	0.5839
	4	10	1.400	1.500	1.1499
Post	1	14	9.036	9.000	2.0798
	2	13	8.654	10.500	3.3874
	3	11	8.455	9.000	3.2362
	4	10	8.100	8.750	2.5473

Table 3 provides the overall descriptive statistics for the pre- and post-test scores. The mean and median of students' scores for both tests demonstrate a significant improvement after participating in the programme. This positive impact aligns with the initial assumption of the programme's effectiveness.

Table 3: Descriptive statistics for pre and post-tests

Test	Mean	Median	Std. Deviation
Pre	1.740	1.500	1.6535
Post	8.604	9.000	2.7771

Before proceeding with the analysis, normality tests were conducted to assess the normality of the pre and post-test data.

Table 4 presents the results, indicating that both data sets do not follow a normal distribution, as evidenced by the p-values (sig.) being less than 0.05 for all categorical data. The S-W statistic values also suggest a poor fit to a normal distribution. Despite one value being close to 1, the significant p-value less than 0.05 confirms the non-normal distribution. Therefore, non-parametric methods were selected for further analysis.

Table 4: Normality test for gender

Mathematic Test	Gender	S-W Statistic	Sig.
Pre	F	0.851	<0.001
	M	0.817	0.002
Post	F	0.915	0.023
	M	0.884	0.025

Table 5 shows that certain values in pre and post-test are lower and more than 0.05 which means certain data are normally distributed and not normally distributed. Thus, no data transformation is required to analyse those data since a non-parametric method is used.

Table 5: Normality test for schools

Mathematic Test	School	S-W Statistic	Sig.
Pre	1	0.872	0.045
	2	0.916	0.218
	3	0.696	<0.001
	4	0.883	0.143
Post	1	0.954	0.619
	2	0.814	0.010
	3	0.875	0.089
	4	0.882	0.137

Table 6 presents the results of the Mann-Whitney U test on gender-based data. There is a significant difference between genders in the pre-test based at a 5% significance level and in the post-test at a 10% significance level. This indicates that female participants consistently scored higher in both tests. The findings suggest that females may have better performance in various situations, possibly due to their ability to pay more attention in class or during outdoor activities.

Table 6: Mann-Whitney (independent samples) test

Type of Test	W-value	Sig. (2-tailed)
Pre (F-M)	173.00	0.028
Post (F-M)	197.000	0.097

Table 7 presents the results of the Kruskal-Wallis test. The p-value (sig.) for the pre-test is 0.001, which is less than 0.05, indicating a rejection of the null hypothesis. Conversely, the p-value (sig.) for the post-test is 0.795, which is greater than 0.05, leading to a failure to reject the null hypothesis. In the pre-test, there was no significant difference in median scores between schools, whereas in the post-test, there was a significant difference in median scores between schools. The participants' performance showed a significant difference among different schools due to various factors such as diverse family backgrounds, different teaching and learning methods, and other relevant aspects. However, when they came together in a program with a unified delivery method, their performance increased and became approximately at the same level.

Table 7: Kruskal-Wallis Test for schools

Test	H Statistic	Sig.
Pre	15.480	0.001
Post	1.026	0.795

The result of the Wilcoxon Signed Rank test showed that there was enough evidence to reject the null hypothesis with the p-value (sig.) being less than 0.05. Hence, there is a significant difference between the median scores of the pre-and post-tests. Generally, it can be concluded that this short-term programme had a significant effect on participants' performance.

CONCLUSION

Analysing the effectiveness of a short-term programme is crucial to ensure its success and impact. By conducting an analysis, one can determine if said programme has achieved its intended outcomes, identify areas for improvement, and make the necessary adjustments.

During the implementation of this short-term program, evaluating its effectiveness on the participants' performance becomes essential. Hence, this study aims to assess the program's effectiveness in mathematics by analysing the collected data using various statistical techniques. As the data distribution did not meet the normality assumption, non-parametric tests were used to analyse the participants' pre and post-test scores. The results indicated a positive outcome in the participants' performance in mathematics. Specifically, the Mann-Whitney test revealed a significant difference in scores in terms of gender, with females obtaining higher scores in both tests.

Meanwhile, the Kruskal-Wallis test found no significant difference between pre-test scores among schools, but vice versa for post-test scores. This signifies that the programme made a fair contribution to all students from various backgrounds. The Wilcoxon Signed Rank test results further proved that students' performance had increased after participating in the program, although it was only for a short term. This study aligns with several previous studies on short-term programs ([17], [31], [34]). By employing diverse approaches to studying, students can gain a better understanding and avoid boredom. Moreover, relating mathematics to real-life examples can raise awareness about the importance of mathematics, as introduced in the "Rekreasi Matematik Melalui Alam" short-term programme. The significance of this study lies in showcasing the substantial

performance improvement among participants through a short-term program. This highlights the potential advantages of periodically providing comprehensive programmes to enhance the early mathematical skills of children. Nonetheless, it is important to acknowledge that the study's sample size was limited to only 48 participants. To offer more robust evidence and inform policymakers about required adjustments, future studies with larger sample groups should be undertaken to explore the long-term effects of such programmes on students' performance. These findings could potentially lead to modifications to the primary school curriculum.

REFERENCES

[1] Alhusseini, N., & Alqahtani, A. (2020). COVID-19 pandemic's impact on eating habits in Saudi Arabia. *Journal of Public Health Research*, *9*.

[2] Baxter, G. P., Shavelson, R. J., Herman, S. J., Brown, K. A., & Valadez, J. R. (1993). Mathematics performance assessment: Technical quality and diverse student impact. *Journal for Research in Mathematics Education*, *24*(3), 190. https://doi.org/10.2307/749344

[3] Best, J., & Winslow, E. (2015). Curriculum-Embedded Performance Assessments (CEPAs): Policy considerations for meaningful accountability. *McREL INTERNATIONAL*. https://www.mcrel.org

[4] Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 9: One-way analysis of variance. *Critical Care*, *8*(2), 130-136). https://doi.org/10.1186/cc2836

[5] Capanu, M., Jones, G. A., & Randles, R. H. (2006). Testing for preference using a sum of Wilcoxon signed rank statistics. *Computational Statistics and Data Analysis*, *51*(2), 793-796. https://doi.org/10.1016/j.csda.2005.12.017

[6] Constantine, N. A., Jerman, P., Berglas, N. F., Angulo-Olaiz, F., Chou, C. P., & Rohrbach, L. A. (2015). Short-term effects of a rights-based sexuality education curriculum for high-school students: A cluster-randomized trial. *BMC Public Health*, *15*(1). https://doi.org/10.1186/s12889-015-1625-5

[7] Dierckx de Casterle, B., Gastmans, C., Bryon, E., & Denier, Y. (2012). QUAGOL: A guide for qualitative data analysis. *International Journal of Nursing Studies*, *49*(3), 360-371. https://doi.org/10.1016/j.ijnurstu.2011.09.012

[8] French, D. (2017). The Future Is Performance Assessment. *Voices in Urban Education*, *2017*, 6-13.

[9] García Perales, R., & Palomares Ruiz, A. (2021). Comparison between performance levels for mathematical competence: Results for the sex variable. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.663202

[10] Gleason, J. (2013). *Comparative power of the anova, randomization anova, and Kruskal-Wallis Test* [Doctoral's Thesis, Wayne State University]. http://digitalcommons.wayne.edu/oa_dissertations

[11] Guo, S., Zhong, S., & Zhang, A. (2013). Privacy-preserving Kruskal-Wallis test. *Computer Methods and Programs in Biomedicine*, *112*(1), 135-145. https://doi.org/10.1016/j.cmpb.2013.05.023

[12] Habsy, B. A., Hidayah, N., Lasan, B. B., Muslihati, & Fudholi, A. (2019). The development model of semar counselling to improve the self-esteem of vocational students with psychological distress. *International Journal of Emerging Technologies in Learning*, *14*(10), 132-149. https://doi.org/10.3991/ijet.v14i10.10221

[13] Hoog, R. V., & Tanis, E. A. (2010). *Probability and statistical inference* (8th ed.). New Jersey, US: Pearson and Prentice Hall.

[14] Hecke, T. van. (2012). Power study of anova versus Kruskal-Wallis test. *Journal of Statistics and Management Systems*, *15*(2-3), 241-247. https://doi.org/10.1080/09720510.2012.10701623

[15] Hodges, J. L., & Lehmann, E. L. (2012). Rank methods for combination of independent experiments in analysis of variance. In *Selected Works of E. L. Lehmann* (pp. 403–418). Springer US. https://doi.org/10.1007/978-1-4614-1412-4_35

[16] Hoffman, P., Goodwin, B., & Kahl S. (2015). Re-balancing assessment: Placing formative and performance assessment at the heart of learning and accountability Denver, CO: McREL International. https://files.eric.ed.gov/fulltext/ED568906.pdf.

[17] Huivaniuk O., Stupnytska H., Fediv O., & Bocharov A. (2022). The effectiveness of short-term pulmonary rehabilitation program in patients with comorbid asthma, chronic obstructive pulmonary disease, and obesity. *Journal of Medicine and Life*, *15*(2), 196-201. DOI: 10.25122/jml-2021-0050. PMID: 35419107; PMCID: PMC8999100.

[18] Kim, H.-Y. (2014). Statistical notes for clinical researchers: Two-way analysis of variance (ANOVA)-exploring possible interaction between factors. *Restorative Dentistry & Endodontics*, *39*(2), 143. https://doi.org/10.5395/rde.2014.39.2.143

[19] Kyun Kim, T. (2015). T-test as a parametric statistic. *Korean Journal of Anesthesiology*, *68*(6), 540-546. https://doi.org/https://doi.org/10.4097/kjae.2015.68.6.540

[20] Law Insider, https://www.lawinsider.com/dictionary/short-term-education-program.

[21] Marjan, M.-H., Vanak, S., St, Z., Rezaee, N., Mardani-Hamooleh, M., & Seraji, M. (2020). Nurses' perception of ethical challenges in caring for patients with COVID-19: A qualitative analysis. *Journal of Medical Ethics and History of Medicine*, *13*, 23. https://doi.org/10.18502/jmehm.v13i23.4954

[22] Mortazavi, F., Salehabadi, R., Sharifzadeh, M., & Ghardashi, F. (2021). Students' perspectives on the virtual teaching challenges in the COVID-19 pandemic: A qualitative study. *Journal of Education and Health Promotion*, *10*(1). https://doi.org/10.4103/jehp.jehp_861_20

[23] Octaviani Putri, H., & Muchyidin, A. (2022). The comparison of students mathematics learning outcomes between using performance assessment and self-assessment. *Journal of Mathematics Instruction, Social Research and Opinion*, *1*(2), 11-19.

[24] Ostertagová, E., Ostertag, O., & Kováč, J. (2014). Methodology and application of the Kruskal-Wallis test. *Applied Mechanics and Materials*, *611*, 115-120. https://doi.org/10.4028/www.scientific.net/AMM.611.115

[25] Oti, E. U., Olusola, M. O., & Esemokumo, P. A. (2021). Statistical analysis of the median test and the Mann-Whitney U Test. *International Journal of Advanced Academic Research*, 7(9).

[26] Özlem Altındağ, K., & Cevriye. E. (2021). Effectiveness of the big math for little kids program on the early mathematics skills of children with risk group. *Athens Journal of Education*, *8*(4), 385-400. DOI:10.30958/aje.8-4-3.

[27] Prokop, P., Tuncer, G., & Kvasničák, R. (2007). Short-term effects of field programme on students' knowledge and attitude toward biology: A Slovak experience. *Journal of Science Education and Technology*, *16*(3), 247-255. https://doi.org/10.1007/s10956-007-9044-8

[28] Rice, J. A. (2007). *Mathematical statistics and data analysis* (3rd ed.). Thomson Higher Education, USA, 2007

[29] Riesmiyatiningdyah, R., Putra, K. W. R., & Sulistyowati, A. (2021). The effect of health education on adolescent knowledge about anemia. *Nurse and Health: Jurnal Keperawatan*, *10*(1), 59-64. https://doi.org/10.36720/nhjk.v10i1.273

[30] Rubarth, K., Sattler, P., Zimmermann, H. G., & Konietschke, F. (2022). Estimation and testing of Wilcoxon–Mann–Whitney effects in factorial clustered data designs. *Symmetry*, *14*(2). https://doi.org/10.3390/sym14020244

[31] Rashidi, Z., & Danesh Fard, K. (2011). Evaluating the internal efficiency of general and specthe short term training centers of universities. *International Journal of Trade, Economics and Finance*, 2(5), 430-436.

[32] Safie, N. B., Zakaria, S., Madhihah, S., Malik, A., Baini Ismail, N., Alias, A., & Idris, R. (2021). Statistical analysis on the effectiveness of short-term programs during COVID-19 pandemic: In the case of "Program Bijak Sifir 2020". *e-Proceedings of the 5th International Conference on Computing, Mathematics and Statistics* (iCMS 2021), 4-5 August 2021.

[33] Schober, P., & Vetter, T. R. (2020). Statistical minute: Nonparametric statistical methods in medical research. *Anesthesia & Analgesia*. https://bpb-us-w2.wpmucdn.com/sites.umassd.edu/dist/f/1316/files/2023/01/Nonparametric-Statistical-Methods-in-Medical-Research.pdf

[34] Sheets, K. J. (1985). Evaluating short-term training programs: A practical approach. *Journal of Instructional Development*, *8*(1), 26-32. http://www.jstor.org/stable/30220769.

[35] Sonja Singh, A., Jeannette Maidy Chin Paw, M. A., Brug, J., & van Mechelen, W. (2007). Short-term effects of school-based weight gain prevention among adolescents. *Archives of Pediatrics & Adolescent Medicine*, *161*(6), 565-571. https://doi.org/10.1001/archpedi.161.6.565