ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Linguistic Facilitation in Scientific Literacy: A Comparative Study of English and Tagalog-Translated HOTS Questions in Life Science

Alvin A. Pabores
Gov. Felicisimo T. San Luis National Agro-Industrial Integrated High School
alvin.pabores@deped.gov.ph
orcid.org/0009-0007-0832-7533

Date Submitted:
March 13, 2026

Date Accepted:
April 21, 2026

Date Published:
May 11, 2026

DOI:
10.5281/zenodo.20072411

ABSTRACT

The landscape of modern science learning in the Philippines is shaped by a bilingual medium of instruction, with English as the primary language of assessment. This study addressed the potential linguistic barriers Filipino students face when demonstrating Higher Order Thinking Skills (HOTS) in Life Science. The objective is to determine whether Tagalog-translated instructions enhance student performance and to identify significant differences in scores between Tagalog-facilitated and traditional English-based assessments. Utilizing a quantitative approach, the study employed a counterbalanced between-subjects experimental design with 80 Grade 11 students from a Tagalog-speaking community in Laguna, Philippines. Participants engaged with a researcher-developed test comprising 28 HOTS-based questions, in which instructional syntax was translated into Tagalog while technical terms remained in English. The instrument underwent content validation by subject-matter experts and statistical validation through test parallelism, confirming equivalence in difficulty and reliability across test versions. Analysis focuses on comparing performance across these two linguistic media to isolate language comprehension from scientific aptitude. Findings are contextualized within the Tagalog-speaking population and may not directly generalize to students whose primary language is Cebuano, Ilocano, or other regional languages. The results revealed a clear pattern where students achieve higher scores and demonstrate more consistent performance when assessed in Tagalog. Students answer more items correctly in their primary language, suggesting that beginning assessments with Tagalog facilitates better cognitive processing and conceptual understanding. Statistical analysis confirms that the language of instruction significantly impacts student outcomes, with a large effect size indicating that the performance gap is practically meaningful. In conclusion, aligning assessments with a learner’s primary language is essential for accurately measuring scientific literacy.

Keywords: *Bilingual Assessment; Higher Order Thinking Skills (HOTS); Life Science; Linguistic Facilitation; Scientific Literacy; Taglish*

INTRODUCTION

The landscape of modern science learning in the Philippines is uniquely shaped by a bilingual medium of instruction in schools, especially in public schools, where English has traditionally been the primary medium of instruction. As students transition into the Senior High School level, the complexity of science necessitates not only the memorization of facts but the application of Higher Order Thinking Skills

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

(HOTS) such as analysis, evaluation, and creation. Based on the study of Nouri et al. (2022), scientific literacy, within this context, is indistinguishably linked to linguistic competence, as the ability to decode complex scientific scenarios depends heavily on the student's mastery of the language in which the problem is presented. As stated by Rose et al. (2022), current educational discourse emphasizes that while English serves as the global lingua franca of science. However, Haiyao and Abd Halim (2026) pointed out that the cognitive load required to process a second language may inadvertently hinder a student's ability to demonstrate their true analytical potential. As a final point, Karunanithi and Sivanadhan (2025) have long argued that a learner's primary language provides a more direct cognitive pathway to conceptual understanding, yet the shift toward mother-tongue-based instruction remains a point of contention in specialized secondary sciences.

Despite the theoretical emphasis on cognitive development, a significant disparity remains in international performance benchmarks, most notably highlighted in the results of the 2022 Programme for International Student Assessment (PISA). A critical observation of the PISA rankings reveals that the top-performing nations in science—including Japan, Macao, Taiwan, South Korea, Estonia, Switzerland, Slovenia, Latvia, Denmark, and the Czech Republic—administer their assessments in the students' respective mother tongues (San Juan, 2022; Globalen LLC, 2022; Organisation for Economic Co-operation and Development, 2023). In stark contrast, the Philippines continues to struggle within the lower quintiles of scientific literacy, utilizing English as the primary medium for both instruction and standardized assessment (OECD, 2023). According to Pineda (2024), Tagalog is the most widely spoken native language in the Philippines, with approximately 26.3 million speakers. Filipino, which is based on Tagalog, serves as the national language. The prevalence of Tagalog shapes the country's linguistic identity and forms the basis of students' education. This creates a potential linguistic barrier where Filipino students may possess the underlying scientific logic but fail to express it due to the complexities of a second language. The studies of Igarashi et al. (2024), Dagalea et al. (2022), and Lumanlan et al. (2024) explored the general benefits of Mother Tongue-Based Multilingual Education (MTB-MLE) in primary years, albeit there is a notable scarcity of research focusing on how localized language affects the specialized, high-level analytical demands of Senior High School Life Science.

This research addresses this gap by investigating whether the use of Tagalog-translated instructions, a code-switching as a pedagogical tool, integrated with essential scientific concepts, significantly enhances the performance of Senior High School students in answering HOTS-based Life Science questions compared to traditional English-only assessments. By employing a counterbalanced assessment design, the study seeks to isolate the variable of language comprehension from scientific aptitude. The primary hypothesis posits that students will demonstrate a higher level of proficiency and clearer logical pathways when the linguistic structure of the question aligns with their primary language. This investigation specifically limits its scope to the translation of instructional syntax and common phrasing, maintaining technical terms in their original English form to reflect the contemporary "Taglish" linguistic reality of the Philippine classroom.

To ensure a rigorous investigation into the linguistic factors affecting science literacy, this study is guided by three pivotal research questions. First, the study seeks to determine the level of proficiency achieved by students when engaging with Life Science assessments delivered through Tagalog-translated instructions. Second, it aims to establish a baseline of student performance using the traditional English language assessment format currently mandated in the curriculum. Finally, the research culminates in a comparative analysis to identify whether a significant difference exists between the scores obtained in the Tagalog-facilitated assessment and those in the English-based version. By addressing these questions, the study provides a statistical basis for evaluating the extent to which the medium of instruction influences the manifestation of higher-order thinking skills among Senior High School students.

This study lies in the urgent need to refine pedagogical approaches and assessment strategies in the Philippine science classroom. If the study reveals a significant performance disparity in favor of Tagalog-based instructions, it provides a compelling argument for the integration of bilingual assessments to ensure that scientific talent is not stifled by language constraints. Understanding this linguistic influence is vital for curriculum developers and educators who aim to foster a more inclusive and effective learning environment where the goal is the mastery of science, not merely the mastery of a foreign language. Ultimately, this research connects the specific cognitive challenges faced by students to the broader national goal of improving scientific literacy, ensuring that the evaluation of a student's intellect is a true reflection of their capacity to think, rather than their capacity to translate.

METHODS

Research Design

The study employed a quantitative research approach. It used a counterbalanced between-subjects experimental design to examine the effect of language presentation on students' performance in answering Higher Order Thinking Skills (HOTS) questions in Life Science. Corriero (2024) explained that this design helps control potential order effects and minimizes bias from the sequence of language exposure. By varying the test's starting language across two independent groups, the design ensures that any performance differences can be more confidently attributed to the language condition rather than to practice effects, fatigue, or familiarity. Pollatsek and Well (1995) noted that the counterbalanced design enhances internal validity by distributing extraneous variables evenly across conditions. This allows a more reliable comparison of the impact of English and Tagalog on students' comprehension and analytical skills.

Research Population and Sampling

The participants in this study consisted of 80 Grade 11 students enrolled in a public senior high school in Laguna, Philippines, during the 2025–2026 academic year. According to Ramos (2022), Laguna is located in the Luzon region of Calabarzon, where Tagalog is the predominant dialect. A cluster sampling technique was utilized. Simkus (2023) notes that cluster sampling involves selecting naturally occurring groups, such as intact classes or sections, rather than individual participants. The students ranged in age from 16 to 19 years. All selected students were born and raised in the Philippines, and their primary language of daily communication is Tagalog. This sampling approach was appropriate because respondents were drawn from existing Grade 11 sections, ensuring both practicality and efficiency in data collection while maintaining population representativeness. All participants were eligible as they had completed Life Science and had prior exposure to both English and Tagalog as languages of instruction.

Data Collection and Analysis

The primary data collection instrument was a researcher-developed test with 28 HOTS-based Life Science questions, presented in two parallel forms. Both versions contained identical items. They differed only in sequence: one alternated English and Tagalog questions starting with English, while the other began with Tagalog. Common instructional language was used in translation, and key scientific terms remained in English for clarity. The instrument underwent a multi-step validation process. First, content validity was established through expert review: three subject-matter specialists in Life Science and one bilingual language expert evaluated each item for content accuracy, cognitive level alignment with Bloom's Taxonomy, and translational equivalence between the English and Tagalog versions. Items flagged for ambiguity or translational bias were revised before administration. Second, to detect potential item-level bias attributable to translation, the panel reviewed each Tagalog item for naturalness and lexical appropriateness for a Tagalog-speaking Grade 11 population in Laguna. Third, statistical equivalence of the two forms was validated through test parallelism. Hassan (2023) emphasized that validating test parallelism is essential in counterbalanced designs like this to ensure equivalence in difficulty, content, and reliability.

Data collection followed a systematic process. Respondents were divided into two groups of 40 students. Group 1 completed Set A, while Group 2 completed Set B. The assessment took place in a controlled classroom environment with standardized instructions. Students had 30 minutes to complete the 28-item test, ensuring consistency and minimizing external influences. The counterbalanced design further controlled for sequence effects and ensured fair representation of both language conditions.

For data analysis, the collected scores were encoded and processed using appropriate statistical software. Descriptive statistics were used to summarize the performance of each group. To determine whether a significant difference existed between the two groups, an independent samples t-test was applied if the data satisfied the assumption of normality; otherwise, the Mann–Whitney U test was employed as a nonparametric alternative. A significance level of 0.05 was set as the criterion for statistical significance, ensuring that the analysis provided a valid and objective basis for interpreting the effect of language on students' scientific literacy performance.

Ethical Considerations

Ethical considerations were strictly observed throughout the conduct of the study. Informed consent was obtained from all participants, and they were assured that their participation was voluntary. The confidentiality and anonymity of the respondents were maintained by assigning codes to their responses and ensuring that no personal identifiers were disclosed. All collected data were used solely for academic and research purposes.

RESULTS AND DISCUSSION

Comparative Analysis of Students' Performance in English and Tagalog Versions of HOTS-Based Life Science Questions

The results of the two test versions reveal a clear pattern in students' performance across language conditions. In the Tagalog–English version (Set A), the total score reached 660, which is notably higher than the total score of 557 obtained in the English–Tagalog version (Set B). This already suggests that students performed better when the test began with Tagalog, then alternated to English. When examining the breakdown, the total score for Tagalog items in Set A is 395 compared to 265 for English items, while in Set B, Tagalog items still obtained a higher score of 357 compared to 200 for English. This consistent trend indicates that, across test versions, students achieved higher scores on Tagalog items than on English items.

The mean scores further support this observation. In Set A, the mean score for Tagalog items is 9.875, substantially higher than the English mean of 6.625, yielding a combined mean of 16.5. In contrast, Set B shows lower overall performance, with a Tagalog mean of 8.925 and an English mean of 5, yielding a combined mean of 13.925. The difference between the total mean scores for Sets A and B (16.5 versus 13.925) suggests that students performed better overall when the assessment began with Tagalog. Numerically, this difference of approximately 2.575 points in the mean score is quite meaningful given that the test consists of only 28 items, indicating a noticeable advantage in comprehension or processing when Tagalog is presented first.

The variability of scores, as reflected by the standard deviation, also provides insight into the consistency of student performance. In Set A, the standard deviation for Tagalog items (2.9888) is lower than that for English items (3.9137), indicating that students' scores in Tagalog were not only higher but also more consistent. In Set B, the standard deviation for Tagalog (3.6961) is slightly lower than English (3.8763), but both are relatively close, suggesting more varied performance when the test begins with

English. The combined standard deviation is also slightly higher in Set B (7.2089) than in Set A (6.7406), reinforcing the finding that scores were more dispersed and less stable when English was presented first.

Looking at the overall results across both groups, the total score for Tagalog items is 752, significantly higher than the 465 obtained for English items. The overall mean also shows a clear difference: Tagalog at 9.4, compared to 5.8125 for English, resulting in a combined mean of 15.2. This indicates that, on average, students answered approximately 3.6 more items correctly in Tagalog than in English. The standard deviation values for English (3.9558) and Tagalog (3.3738) suggest that while performance varies in both languages, scores in Tagalog are slightly more consistent.

Overall, the numerical data strongly suggest that students perform better in Tagalog than in English, as evidenced by higher scores and more consistent results. Furthermore, the higher total and mean scores in Set A imply that beginning the test in Tagalog may facilitate better comprehension and cognitive processing, thereby positively influencing performance even when switching to English. These findings highlight the potential role of language familiarity in enhancing students' ability to engage with and correctly answer HOTS-based science questions.

Table 1. *Summary Statistics of Students' Scores in English and Tagalog MCQs Across Test Versions A and B*

Set A: Tagalog - English version					Set B: English -Tagalog version			
Total	265	395	660		Total	200	357	557
Mean	6.625	9.875	16.5		Mean	5	8.925	13.925
SD	3.9137	2.9888	6.7406		SD	3.8763	3.6961	7.2089
		TOTAL	465	752	1217			
		MEAN	5.8125	9.4	15.2125			
		SD	3.9557	3.3738	7.0544			

Normality of Data Distribution and Validation of Test Parallelism

To establish the reliability of the research instrument and ensure that results are not biased by the sequence of language presentation, it is mathematically necessary to conduct a validation check between the two test versions. This step confirms that the two groups of students are comparable and that the "Set A" (Tagalog-English) and "Set B" (English-Tagalog) question formats did not create artificial performance gaps based solely on the starting language. Without proving that Set A and Set B are statistically equivalent, any final comparison between English and Tagalog scores could be criticized as being a result of "test version difficulty" rather than "language ability." Therefore, the following analysis serves as a crucial baseline to demonstrate that both versions of the 28-item test were equally challenging and that the 80 respondents demonstrated similar baseline capabilities across both sets.

Before proceeding with the main comparative analysis, the data were subjected to the Shapiro-Wilk test to determine the appropriate statistical treatment. For the English components, Set A yielded a normality index of 0.06576, whereas Set B showed a non-normal distribution (index = 0.002417). Similarly, the Tagalog components showed indices of 0.01197 for Set A and 0.09324 for Set B, while the total scores for both sets also failed the normality assumption with values of 0.0407 and 0.01744, respectively. Because the data did not consistently follow a normal distribution across all variables, the Mann-Whitney U test, a nonparametric alternative to the independent t-test, was used to assess differences between the two groups.

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

The results of the Mann-Whitney U test across all three categories indicate no significant differences between the two test versions. For the English items, the p-value of 0.07627 is greater than the 0.05 alpha level, leading to the acceptance of the null hypothesis that the populations are equal. For the Tagalog items, the p-value of 0.2619 further supports the equality of the two sets, indicating that any difference in score is not statistically significant. Finally, the comparison of total scores between Set A and Set B resulted in a p-value of 0.1097, which also exceeds the significance threshold. These findings collectively demonstrate that the counterbalanced design was successful; since there is no significant difference between the outcomes of Sets A and B, the researcher can confidently proceed to the final analysis comparing English versus Tagalog performance without fear that the specific test version influenced the data.

Table 2. *Normality and Homogeneity Analysis of Test Versions A and B via Shapiro-Wilk and Mann-Whitney U Tests*

Assessment Category	Set	Shapiro-Wilk Normality (p-value)	Mann-Whitney U Statistic	Z-score	P-value (2-tailed)	Interpretation
English HOTS MCQ	A	0.06576	984	1.7728	0.07627	Not Significant
	B	0.002417				
Tagalog HOTS MCQ	A	0.01197	916.5	1.1219	0.2619	Not Significant
	B	0.09324				
Total Test Score	A	0.0407	966.5	1.5995	0.1097	Not Significant
	B	0.01744				

Linguistic Determinants of Cognitive Performance in Life Science Assessments

The initial data analysis began by assessing score distributions to determine the most appropriate comparison method. Scores from the HOTS MCQs in Life Science were tested for normality using the Shapiro-Wilk test. This yielded p-values of 0.001262 for the English assessment and 0.0092 for the Tagalog version. Both values are well below the standard 0.05 alpha level, so we rejected the null hypothesis of normality. This indicated a non-normal distribution for both language conditions. As a result, the researcher used the Mann-Whitney U test, a nonparametric test, rather than the independent t-test. This approach ensures valid findings despite the irregular distribution among the 80 student respondents.

The Mann-Whitney U test was then applied to evaluate the primary research question regarding the difference in performance between the two linguistic media. The null hypothesis for this calculation posited that there was no significant difference between the English and Tagalog-translated scores, suggesting that any observed variation in student performance was merely due to chance. However, the resulting p-value was extremely small, falling well below the 0.05 threshold, leading to the definitive rejection of the null hypothesis. This level of significance indicates that the probability of the performance gap occurring by sheer coincidence is nearly zero, thereby supporting the alternative hypothesis that the language of instruction significantly impacts student outcomes. For the test statistics, the calculated U-value was 1706, with a corresponding Z-score of 4.9525. The Z-score represents how many standard deviations the observed mean is from the population mean, and a value of 4.9525 indicates a highly significant deviation that falls well outside the typical region of acceptance. Furthermore, the analysis of effect size yielded a value of 0.55, which indicates a moderate and meaningful difference between groups.

These results clearly show that linguistic facilitation has a decisive role in scientific literacy. The large effect size demonstrates that the difference between English and Tagalog scores is both statistically significant and practically meaningful. In education, this means students meet the cognitive demands of analysis and evaluation more effectively when the language barrier is reduced. Sanchez and Miralles (2025) have noted significant local evidence supporting Mother Tongue-Based education, especially in specialized subjects like Life Science. By showing that students score higher and more consistently when assessed in Tagalog, the research suggests changes in teaching methods that emphasize understanding concepts over strict language use. This study identifies a localized way to address challenges raised by international benchmarks such as PISA. It suggests that aligning a student's primary language with academic assessment can help unlock their cognitive potential in science.

Table 3. *Differential Analysis of Student Scores in English versus Tagalog-Mediated Life Science Questions*

Assessment Variable	Sample Size (n)	Shapiro-Wilk (p-value)	Mann-Whitney U	Z-score	P-value (2-tailed)	Effect Size (r)	Interpretation
English HOTS MCQ	80	0.001262	1706	4.9525	<0.001	0.55	Significant
Tagalog HOTS MCQ	80	0.0092					(Medium Effect)

Legend for Effect Size: < 0.20 → Very small or negligible effect; 0.20 – 0.49 → Small effect; 0.50 – 0.79 → Medium effect; ≥ 0.80 → Large effect

The primary objective of this study was to determine whether Tagalog-translated instructions in Higher Order Thinking Skills (HOTS) assessments enhance Senior High School students' performance in Life Science. The Mann-Whitney U test results provided a clear answer: students achieved significantly higher scores when assessments were administered in their mother tongue than when administered in English-only versions. The null hypothesis was rejected, as indicated by a very small p-value and a substantial effect size of 0.55. These results indicate that the medium of instruction is a critical determinant of scientific literacy. The findings are consistent with those of Shah and Nizamani (2025), who reported that reducing second-language barriers enables students to engage more effectively in complex analytical and evaluative processes in Life Science. Similarly, Mufori et al. (2025) found that localized language scaffolding allows students to demonstrate their scientific abilities more accurately, as cognitive resources are otherwise devoted to basic language decoding.

These findings align closely with the global trends observed in the 2022 PISA results, which showed that top-performing nations such as Japan, South Korea, and Estonia, as well as other performing countries that administered scientific assessments in students' primary languages (San Juan, 2022; Globalen LLC, 2022; OECD, 2023). The disparity in performance between the Tagalog and English versions of the test, mirroring the challenges faced by the Philippines on the international stage, suggests that the "language of science" in the local context is often more a hurdle of vocabulary than a lack of logical reasoning. By integrating Taglish—keeping technical terms in English while using Tagalog for instructional syntax—this study aligns with Nouri et al. (2022), who argued that scientific literacy is indistinguishably linked to linguistic competence. The results confirm that students do not necessarily lack the Higher Order Thinking Skills required for Life Science; rather, their ability to manifest these skills is often bottlenecked by the traditional insistence on English-only assessments.

The practical implications of this research are significant for both pedagogical practice and educational policy in the Philippines. The data suggest that for Senior High School students, scientific comprehension is significantly deeper when instructional scaffolding is provided in a familiar language. This approach is aligned with the study by Juky and Nur (2025), which proposes a more flexible bilingual approach in science classrooms, encouraging teachers to use functional transliteration to ensure conceptual mastery. In terms of policy, these results provide an empirical basis for re-evaluating the linguistic format

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

of localized and national standardized tests. According to Bakogiannis et al. (2026), by bridging the gap between a student's primary language and academic evaluation, educational institutions can foster a more inclusive environment that prioritizes the development of critical thinkers over linguistic conformity, potentially improving the country's standing in future international scientific benchmarks. However, it is essential to recognize that the Philippines is a richly multilingual nation and Tagalog, while the basis of the national language Filipino, is not the primary mother tongue of all Filipino learners (Bustos-Orosa & Symaco, 2025). Students in the Visayas and Mindanao regions, for instance, may speak Cebuano or Hiligaynon as their first language, and for them, Tagalog-mediated assessments may introduce a different, though comparably reduced, cognitive load relative to English. Policymakers should therefore consider region-specific bilingual scaffolding rather than a uniform national Tagalog-facilitation model. Furthermore, while this study demonstrates the value of mother-tongue facilitation at the Senior High School level, educators and curriculum designers must remain attentive to the transitional demands of tertiary education, where English remains the dominant academic medium. Bilingual facilitation strategies should be designed not as a replacement for English academic literacy, but as a pedagogical scaffold that builds students' conceptual confidence while simultaneously supporting a deliberate and graduated transition toward proficiency in English academic discourse.

Despite the clear significance of the findings, several limitations must be acknowledged to maintain the study's credibility. First, with only 80 students drawn from a single Tagalog-speaking community in Laguna, the generalizability of these findings to the broader Filipino student population is constrained. The Philippines is a highly multilingual nation with over 180 languages; students whose primary language is Cebuano, Ilocano, Hiligaynon, or other regional mother tongues may respond differently to Tagalog-mediated assessments, as Tagalog itself functions as a second language for a substantial portion of the country. Future studies should replicate this design using region-specific translations (e.g., Cebuano or Ilocano) to determine whether the benefits of mother-tongue facilitation hold across the full spectrum of Philippine linguistic diversity. Second, although the researcher-developed instrument underwent content validation and statistical test-parallelism checks, the equivalence of cognitive demand between the English and Tagalog versions remains an inherent challenge. Subtle differences in translation naturalness, regional lexical familiarity, and item-level bias may have influenced responses in ways that were not fully captured by the Mann–Whitney U parallelism analysis. Future instruments should incorporate independent expert panels for both linguistic and content equivalence reviews, along with differential item functioning (DIF) analysis, to strengthen cross-form validity. Third, the findings illuminate the short-term advantage of primary-language assessments but do not address the longitudinal challenge of preparing students for tertiary education and specialized professional fields where English remains the dominant medium of instruction. Students who perform well in Tagalog-mediated HOTS assessments may still face significant transitions when confronted with purely English-based academic materials in college. This study thus identifies an important pedagogical gap: bilingual facilitation at the Senior High School level should be viewed not as a terminal solution but as a scaffolded bridge toward graduated English academic language proficiency. Future longitudinal research should examine whether Taglish-mediated instruction at the secondary level translates into stronger English-language academic performance at the tertiary level, or whether explicit bridging strategies—such as metacognitive language transfer exercises—are needed to close this gap. Finally, while the counterbalanced design successfully mitigated test-version bias, factors such as students' varying levels of bilingual proficiency and prior exposure to English-heavy curricula could still confound the results. The study also focused solely on Life Science MCQs; therefore, the impact of linguistic facilitation on other science domains or on subjective, essay-based assessments remains to be explored. These constraints suggest that while the findings are robust for this specific group, they should be interpreted as a localized indicator of a broader systemic phenomenon warranting large-scale, multi-regional investigation.

Based on these outcomes, future research should expand the scope of this study by involving a larger, more diverse student population across different academic strands, such as STEM versus non-STEM, to determine whether academic background influences the level of linguistic facilitation required. Critically, replication studies using other regional Philippine languages—particularly Cebuano, Ilocano, and Hiligaynon—are strongly recommended to determine whether the performance benefits observed here are specific to Tagalog or are generalizable to mother-tongue facilitation more broadly. Further inquiry into other subject areas, such as Physical Science or Earth Science, would help determine if these findings are consistent across all scientific disciplines. It is also recommended that longitudinal studies be conducted to examine the critical transition from secondary to tertiary education: specifically, whether students who receive Taglish-mediated HOTS instruction in Senior High School develop stronger or weaker English academic language skills upon entering college, where English-only instruction is predominantly used. Such research should also investigate targeted bridging strategies—such as bilingual scaffolding that explicitly models the shift from Tagalog to English academic discourse—to ensure that linguistic facilitation at the secondary level functions as a stepping stone toward full academic English proficiency rather than as a potential hindrance to it. This study provides critical evidence that the mother tongue is a powerful tool for unlocking scientific potential. By demonstrating that Tagalog-translated assessments lead to significantly higher cognitive performance, this research underscores the necessity of linguistic inclusivity in the pursuit of genuine scientific literacy for Filipino students.

CONCLUSION

This research investigated the impact of Tagalog-translated instructions on the Higher Order Thinking Skills (HOTS) performance of Senior High School students in Life Science. Statistical analysis using the Mann-Whitney U test showed that students achieved significantly higher scores and demonstrated greater analytical proficiency when assessments were delivered in their mother tongue rather than in English. The findings indicate that the complexity of a second language acts as a cognitive bottleneck, hindering scientific reasoning. The study rejected the null hypothesis with a highly significant p-value and a large effect size. Results confirm that aligning assessments with students' primary language is essential for scientific inquiry.

The findings of this study contribute to the body of knowledge on Mother Tongue-Based Multilingual Education (MTB-MLE), especially in specialized Senior High School subjects. These results support models in top-performing PISA (Program for International Student Assessment) nations that prioritize the use of the mother tongue in scientific assessments. This research offers empirical guidance for Filipino practices and policymakers. Integrating functional transliteration—Taglish (a blend of Tagalog and English)—with technology integration in English and localized instructional language structures can enhance scientific literacy. By demonstrating that linguistic barriers, rather than intellectual limitations, often impede student success, this study emphasizes the need for inclusive, linguistically responsive teaching strategies in Philippine science education.

Educational practitioners should gradually incorporate localized instructional scaffolding in high-stakes science assessments. This ensures that evaluations reflect conceptual mastery rather than translation skills. Future research should examine the long-term effects of this bilingual approach and its applicability to other scientific disciplines, such as Chemistry and Physics. Investigating the impact of linguistic facilitation on a broader demographic can help refine strategies for national implementation. This study highlights the importance of ensuring that assessment language bridges understanding, enabling Filipino thinkers to be evaluated for the depth of their reasoning and the scientific clarity of their arguments.

Acknowledgments

The authors would like to acknowledge the Grade 11 students of Gov. Felicisimo T. San Luis National Agro-Industrial Integrated High School, Academic Year 2025–2026, for their unwavering support and willingness to participate in this study.

Ethical statements
The researcher addressed the ethical considerations of this study by obtaining informed consent from all participants."

Declaration of conflicts of interest
This paper holds no conflict of interest.

Data availability statement
The data from this study were captured in an Excel sheet and analyzed using an online statistical calculator.

Funding statement
The author solely funds this paper.

References

Bakogiannis, A., Lorrimer, S., & Papavasiliou, E. (2026). Inclusive pedagogies and practices of English for Academic Purposes (EAP) in higher education (HE): An exploratory survey-based study. *Journal of English for Academic Purposes*, *80*, 101649. https://doi.org/10.1016/j.jeap.2026.101649

Bustos-Orosa, M. A., & Symaco, L. (2025). Language Policy in the Philippines: The Ongoing Narrative of Mother Tongue-Based Multilingual Education. *Moussons*, *45*, 149–160. https://doi.org/10.4000/147jr

Corriero, E. (2024). *Counterbalancing*. Sagepub.com. https://methods.sagepub.com/ency/edvol/the-sage-encyclopedia-of-communication-research-methods/chpt/counterbalancing

Dagalea, A. J. L., Peralta, S. B., & Abocejo, F. T. (2022). Evaluation of the Mother Tongue-Based Multilingual Education Program in the Philippines. *Budapest International Research and Critics in Linguistics and Education (BirLE) Journal*, *5*(4), 422–431. https://doi.org/10.33258/birle.v5i4.7269

Globalen LLC. (2022). *PISA science scores by country, around the world | TheGlobalEconomy.com*. TheGlobalEconomy.com. https://www.theglobaleconomy.com/rankings/pisa_science_scores/

Haiyao, Z., & Abd Halim, N. D. (2026). Impact of Different Types of Cognitive Load in Foreign Language Learning. *Environment and Social Psychology*, *11*(1). https://doi.org/10.59429/esp.v11i1.3538

Hassan, M. (2023, May 11). *Parallel Forms Reliability - Methods, Example and Guide*. Research Method. https://researchmethod.net/parallel-forms-reliability/

Igarashi, T., Maulana, S., & Suryadarma, D. (2024). Mother Tongue-based Education in a Diverse Society and the Acquisition of Foundational Skills: Evidence from the Philippines. *Labour Economics*, *91*, 102641. https://doi.org/10.1016/j.labeco.2024.102641

Juky, S., & Nur, S. (2025). The Impacts of Bilingualism Learning Strategies in Science Education: A Systematic Literature Review. *International Journal of Academic Research in Progressive Education and Development*, *14*(3). https://doi.org/10.6007/IJARPED/v14-i3/26196

Karunanithi, P., & Sivanadhan, I. (2025). The Role of Mother Tongue in Shaping Educational Outcomes. *International Journal of Academic Research in Business and Social Sciences*, *15*(8). https://doi.org/10.6007/ijarbss/v15-i8/26307

Lumanlan, J. S., De La Torre, R., Asistio, F. V., De Guia, L. G., Do, H., Ramos , E. T. Jr., & Acuña, Ma. J. B. (2024). Perspectives on Mother Tongue-Based Multilingual Education Implementation in the Philippines. *Journal of Interdisciplinary Perspectives*, *2*(12), 1–1. https://ejournals.ph/article.php?id=24854

Manipuspika, Y. S. (2021). The Effectiveness of Translation Quality Assessment (TQA) Models in the Translation of Indonesian Students. *International Journal of Linguistics, Literature and Translation*, *4*(5), 287–297. https://doi.org/10.32996/ijllt.2021.4.5.32

Mufori, S. N., Mpofu, N., & Simuja, C. (2025). Exploring pedagogical translanguaging strategies in ESL primary classrooms: enhancing reading comprehension in multilingual learners. *International Multilingual Research Journal*, 1–16. https://doi.org/10.1080/19313152.2025.2585817

Nouri, N., Rodríguez, A. D., & Saberi, M. (2022). Proposing a framework for science teachers' competencies regarding translanguaging in multicultural settings. *Springer International Handbooks of Education*, 617–651. https://doi.org/10.1007/978-3-030-83122-6_24

OECD. (2023, December 5). *PISA 2022 results (volume I and II) - country notes: Philippines*. OECD. https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/philippines_a0882a2d-en.html

Organisation for Economic Co-operation and Development, (2023). *PISA 2022 Results (Volume I)*. OECD. https://www.oecd.org/en/publications/pisa-2022-results-volume-i_53f23881-en.html

Pineda, A. (2024, September 21). *Tagalog: The Most Widely Spoken Language in the Philippines - Fluent Filipino*. Fluent Filipino. https://fluentfilipino.com/tagalog-the-most-widely-spoken-language-in-the-philippines/

Pollatsek, A., & Well, A. D. (1995). On the use of counterbalanced designs in cognitive research: A suggestion for a better and more powerful analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(3), 785–794. https://doi.org/10.1037/0278-7393.21.3.785

Ramos, E. R. (2022, August 19). *The Most Common Among the 187 Dialects in the Philippines*. Your Guide to the Big City. https://villagepipol.com/the-most-common-among-the-187-dialects-in-the-philippines/

Rose, H., Sahan, K., & Zhou, S. (2022). Global English Medium Instruction: Perspectives at the crossroads of Global Englishes and EMI. *Asian Englishes*, 1–13. https://doi.org/10.1080/13488678.2022.2056794

San Juan, D. M. M. (2022, July 4). *What is The Best Medium of Instruction for Philippine Schools?* ResearchGate. https://doi.org/10.13140/RG.2.2.27080.80641

Sanchez, M. J. C., & Miralles, A. C. (2025). The role of mother tongue based-multilingual education in enhancing early language development: Impact study. *International Journal of Science and Research Archive*, *15*(2), 789–812. https://doi.org/10.30574/ijsra.2025.15.2.1430

Shah, S. H. R., & Nizamani, D. L. (2025). *Analysing Linguistic Barriers faced by Secondary Level Students while Speaking English as a Second Language*. *4*(3), 2083–2097. https://doi.org/10.63056/acad.004.03.0511

Simkus, J. (2023). *Cluster Sampling: Definition, Method and Examples*. Www.simplypsychology.org. https://www.simplypsychology.org/cluster-sampling.html

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

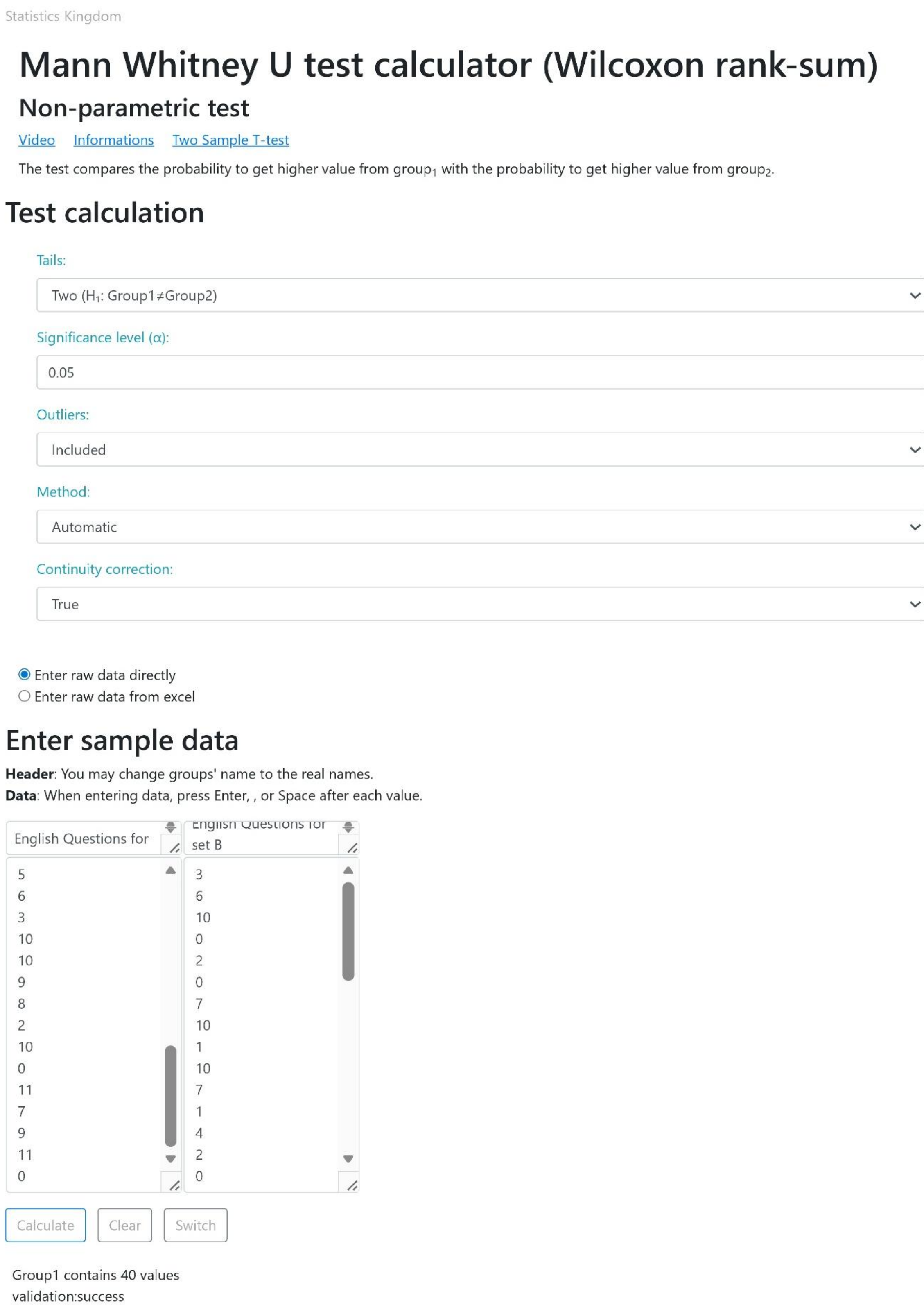
APPENDIX

Set	Student	English Questions	Tagalog Questions	TOTAL
A	1	4	7	11
A	3	1	5	6
A	5	7	11	18
A	7	1	5	6
A	9	10	14	24
A	11	13	14	27
A	13	3	7	10
A	15	4	9	13
A	17	6	9	15
A	19	9	13	22
A	21	7	9	16
A	23	2	7	9
A	25	7	10	17
A	27	9	14	23
A	29	5	10	15
A	31	2	7	9
A	33	9	9	18
A	35	14	12	26
A	37	2	5	7
A	39	11	14	25
A	41	2	7	9
A	43	14	13	27
A	45	8	13	21
A	47	9	12	21
A	49	5	10	15
A	51	5	9	14
A	53	6	10	16
A	55	3	9	12
A	57	10	12	22
A	59	10	13	23
A	61	9	12	21
A	63	8	11	19
A	65	2	6	8
A	67	10	13	23
A	69	0	5	5
A	71	11	11	22
A	73	7	9	16
A	75	9	11	20
A	77	11	14	25
A	79	0	4	4
	Total	265	395	660
	Mean	6.625	9.875	16.5
	SD	3.913733217	2.988760999	6.740616

Set	Student	English Questions	Tagalog Questions	TOTAL
B	2	3	7	10
B	4	6	9	15
B	6	10	13	23
B	8	0	3	3
B	10	2	6	8
B	12	0	3	3
B	14	7	12	19
B	16	10	14	24
B	18	1	7	8
B	20	10	9	19
B	22	7	12	19
B	24	1	5	6
B	26	4	10	14
B	28	2	4	6
B	30	0	2	2
B	32	9	12	21
B	34	4	9	13
B	36	7	11	18
B	38	5	10	15
B	40	0	4	4
B	42	12	14	26
B	44	5	10	15
B	46	1	3	4
B	48	2	16	18
B	50	9	12	21
B	52	2	6	8
B	54	3	9	12
B	56	1	5	6
B	58	3	8	11
B	60	7	11	18
B	62	10	13	23
B	64	11	10	21
B	66	0	4	4
B	68	1	7	8
B	70	6	11	17
B	72	7	12	19
B	74	10	14	24
B	76	2	7	9
B	78	12	13	25
B	80	8	10	18
	Total	200	357	557
	Mean	5	8.925	13.925
	SD	3.876292175	3.696065129	7.208924

	English	Tagalog	Total
Overall total	465	752	1217
Mean	5.8125	9.4	15.2125
SD	3.955748	3.373763	7.05439

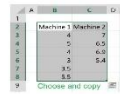
ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)



ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Enter sample data

You may copy data from Excel, Google sheets or any tool that separate data with **Tab** and **Line Feed**.
Copy the data, **one block of 3 consecutive columns includes the header**, and paste below.
Copy the data,



Calculate Clear Validate

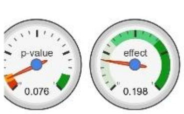
It is okay to leave empty cells, empty cells or non numeric cells won't be counted

Results

Sample average (x̄):	6.625	5
Sample size (n):	40	40
Sample SD (S):	3.913733	3.876292
Median:	7	4.5
Skewness:	-0.0111855	0.286385
Skewness Shape:	Potentially **Symmetrical**	Potentially **Symmetrical**
Normality:	0.06576	0.002417
Outliers:		
Outlier count:	0	0
Rank:	1804	1436
U:	616	984
Ties Correction:	0.007934833567744961	
Exact:	false	

The difference between the expected SD and the sample SD

How to do with R?



Two sample mann-whitney u, using Normal distribution (two-tailed) (validation)

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

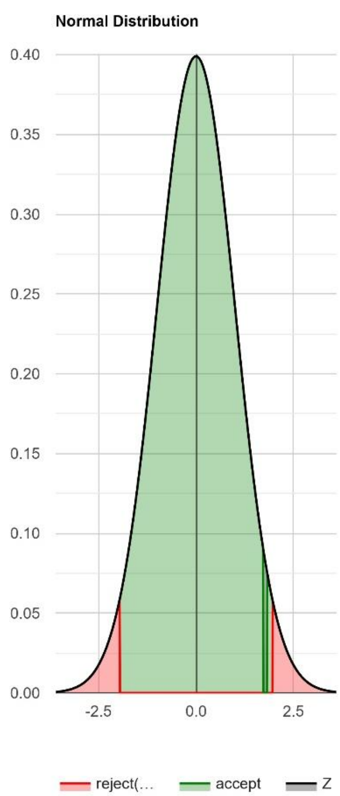
The **normal approximation** is used. The statistic's distribution is $N(800, 103.51^2)$.
The data contains ties, identical values, it is recommended to use the **normal approximation** that uses the ties correction. Since n1 > 20 or n2 > 20, the tool cannot use the **exact calculation**.

1. H_0 hypothesis
Since p-value > α, H_0 cannot be rejected.
The randomly selected value of **English Questions for set A's** population is assumed to be **equal to** the randomly selected value of **English Questions for set B's** population.
In other words, the difference between the randomly selected value of **English Questions for set A** and the **English Questions for set B** populations is not big enough to be statistically significant.

2. P-value
The p-value equals 0.07627, ($p(x \le Z) = 0.9619$). It means that the chance of type I error, rejecting a correct H_0, is too high: 0.07627 (7.63%).
The larger the p-value the more it supports H_0.

3. The statistics
The test statistic Z equals 1.7728, which is in the 95% region of acceptance: [-1.96 : 1.96].
U=984, is in the 95% region of acceptance: [597.1243 : 1002.8757].

4. Effect size
The observed **standardized effect size**, $Z/\sqrt{(n_1+n_2)}$, is **small** (0.2). That indicates that the magnitude of the difference between the value from **English Questions for set A** and the value from **English Questions for set B** is small.
The observed **common language effect size**, $U_1/(n_1 n_2)$, is **0.61**, this is the probability that a random value from **English Questions for set A** is greater than a random value from **English Questions for set B**.
If you like the page, please share or like. Questions, comments and suggestions are appreciated. (statskingdom@gmail.com)



ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Test validation

The requested test was calculated, it is likely you chose the right test.

- **Outliers** +

  Outliers' detection method: Tukey Fence, k=1.5
  The data doesn't have outliers.

Information

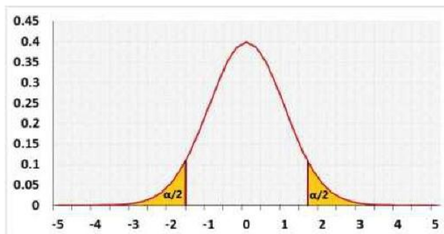
Hypotheses

H_0: $Group_1 = Group_2$
H_1: $Group_1 \neq Group_2$

Test statistic

$$Z = \frac{U - \mu + c}{б}$$

Normal distribution



C - continuity correction, when U > μ: C = -0.5, when U < μ: C = 0.5

Target

Unlike t-test that compares the means, the Mann-Whitney U test compares a randomly selected value from group1 to a randomly selected value from group2.
When the two distributions have a similar **shape** you can use the test to compare also the **medians**.
When the two distributions have a similar **symmetrical shape**, you can use the test to compare also the **means**. more

Method

The mann whitney u test calculator may use three methods. We reccomend to use the "Automatic" method.
Automatic - when $n_1 \leq 20$ and $n_2 \leq 20$ and the data doesn't have ties, the tool uses the exact value, otherwise the tool uses the z approximation.
Exact - when $n_1 \leq 20$ and $n_2 \leq 20$ the tool uses the exact value, calculated base on all the possible combinations, otherwise the tool uses the z approximation.
Z approximation - the tool uses the z approximation.

Continuity correction - relevant only for the Z approximation. It is used since a continuous distribution is used to calculate a discrete distribution.

R Code

The following R code should produce similar results:

```
x1<-c(4,1,7,1,10,13,3,4,6,9,7,2,7,9,5,2,9,14,2,11,2,14,8,9,5,5,6,3,10,10,9,8,2,10,0,11,7,9,11,0)
x2<-c(3,6,10,0,2,0,7,10,1,10,7,1,4,2,0,9,4,7,5,0,12,5,1,2,9,2,3,1,3,7,10,11,0,1,6,7,10,2,12,8)
wilcox.test(x1, x2, alternative = "two.sided", paired = FALSE, exact = FALSE, correct = TRUE)
```

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Statistics Kingdom

Mann Whitney U test calculator (Wilcoxon rank-sum)

Non-parametric test

Video Informations Two Sample T-test

The test compares the probability to get higher value from $group_1$ with the probability to get higher value from $group_2$.

Test calculation

Tails:

Two (H_1: Group1≠Group2)

Significance level (α):

0.05

Outliers:

Included

Method:

Automatic

Continuity correction:

True

(•) Enter raw data directly
( ) Enter raw data from excel

Enter sample data

Header: You may change groups' name to the real names.
Data: When entering data, press Enter, , or Space after each value.

Tagalog Questions	Tagalog Questions
10	9
9	5
12	8
13	11
12	13
11	10
6	4
13	7
5	11
11	12
9	14
11	7
14	13
4	10

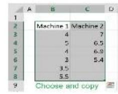
Calculate Clear Switch

Group1 contains 40 values
validation:success

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Enter sample data

You may copy data from Excel, Google sheets or any tool that separate data with **Tab** and **Line Feed**.
Copy the data, **one block of 3 consecutive columns includes the header**, and paste below.
Copy the data,



Calculate Clear Validate

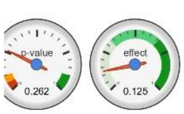
It is okay to leave empty cells, empty cells or non numeric cells won't be counted

Results

Sample average (x̄):	9.875	8.925
Sample size (n):	40	40
Sample SD (S):	2.988761	3.696065
Median:	10	9.5
Skewness:	-0.303575	-0.206931
Skewness Shape:	Potentially **Symmetrical**	Potentially **Symmetrical**
Normality:	0.01197	0.09324
Outliers:		
Outlier count:	0	0
Rank:	1736.5	1503.5
U:	683.5	916.5
Ties Correction:	0.01004453820909517	
Exact:	false	

The difference between the expected SD and the sample SD

How to do with R?



Two sample mann-whitney u, using Normal distribution (two-tailed) (validation)

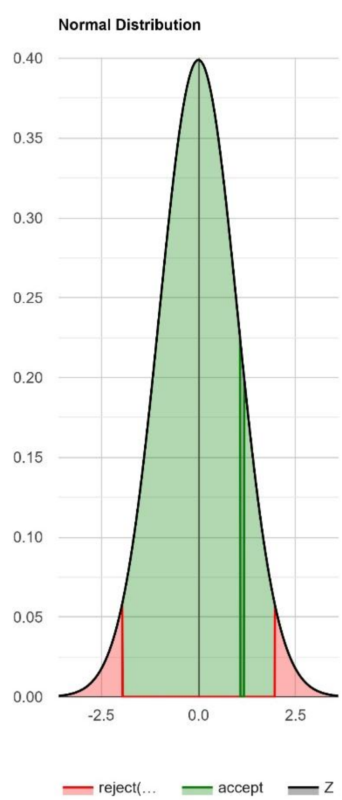
The **normal approximation** is used. The statistic's distribution is $N(800,103.4^2)$.
The data contains ties, identical values, it is recommended to use the **normal approximation** that uses the ties correction. Since n1 > 20 or n2 > 20, the tool cannot use the **exact calculation**.

1. H_0 hypothesis
Since p-value > α, H_0 cannot be rejected.
The randomly selected value of **Tagalog Questions for set A's** population is assumed to be **equal to** the randomly selected value of **Tagalog Questions for set B's** population.
In other words, the difference between the randomly selected value of **Tagalog Questions for set A** and the **Tagalog Questions for set B** populations is not big enough to be statistically significant.

2. P-value
The p-value equals 0.2619, ($p(x \le Z) = 0.869$). It means that the chance of type I error, rejecting a correct H_0, is too high: 0.2619 (26.19%).
The larger the p-value the more it supports H_0.

3. The statistics
The test statistic Z equals 1.1219, which is in the 95% region of acceptance: [-1.96 : 1.96].
U=916.5, is in the 95% region of acceptance: [597.3401 : 1002.6599].

4. Effect size
The observed **standardized effect size**, $Z/\sqrt{(n_1+n_2)}$, is **small** (0.13). That indicates that the magnitude of the difference between the value from **Tagalog Questions for set A** and the value from **Tagalog Questions for set B** is small.
The observed **common language effect size**, $U_1/(n_1n_2)$, is **0.57**, this is the probability that a random value from **Tagalog Questions for set A** is greater than a random value from **Tagalog Questions for set B**.
If you like the page, please share or like. Questions, comments and suggestions are appreciated. (statskingdom@gmail.com)



ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Test validation

The requested test was calculated, it is likely you chose the right test.

- **Outliers** +

  Outliers' detection method: Tukey Fence, k=1.5
  The data doesn't have outliers.

Information

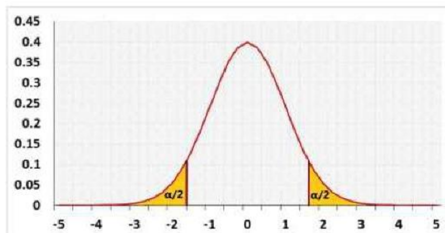
Hypotheses

H_0: $Group_1 = Group_2$
H_1: $Group_1 \neq Group_2$

Test statistic

$$Z = \frac{U - \mu + c}{\sigma}$$

Normal distribution



C - continuity correction, when U > μ: C = -0.5, when U < μ: C = 0.5

Target

Unlike t-test that compares the means, the Mann-Whitney U test compares a randomly selected value from group1 to a randomly selected value from group2.
When the two distributions have a similar **shape** you can use the test to compare also the **medians**.
When the two distributions have a similar **symmetrical shape**, you can use the test to compare also the **means**. more

Method

The mann whitney u test calculator may use three methods. We reccomend to use the "Automatic" method.
Automatic - when $n_1 \leq 20$ and $n_2 \leq 20$ and the data doesn't have ties, the tool uses the exact value, otherwise the tool uses the z approximation.
Exact - when $n_1 \leq 20$ and $n_2 \leq 20$ the tool uses the exact value, calculated base on all the possible combinations, otherwise the tool uses the z approximation.
Z approximation - the tool uses the z approximation.

Continuity correction - relevant only for the Z approximation. It is used since a continuous distribution is used to calculate a discrete distribution.

R Code

The following R code should produce similar results:

```
x1<-c(7,5,11,5,14,14,7,9,9,13,9,7,10,14,10,7,9,12,5,14,7,13,13,12,10,9,10,9,12,13,12,11,6,13,5,11,9,11,14,4)
x2<-c(7,9,13,3,6,3,12,14,7,9,12,5,10,4,2,12,9,11,10,4,14,10,3,16,12,6,9,5,8,11,13,10,4,7,11,12,14,7,13,10)
wilcox.test(x1, x2, alternative = "two.sided", paired = FALSE, exact = FALSE, correct = TRUE)
```

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Statistics Kingdom

Mann Whitney U test calculator (Wilcoxon rank-sum)

Non-parametric test

Video Informations Two Sample T-test

The test compares the probability to get higher value from $group_1$ with the probability to get higher value from $group_2$.

Test calculation

Tails:

Two (H_1: Group1≠Group2)

Significance level (α):

0.05

Outliers:

Included

Method:

Automatic

Continuity correction:

True

◉ Enter raw data directly
○ Enter raw data from excel

Enter sample data

Header: You may change groups' name to the real names.
Data: When entering data, press Enter, , or Space after each value.

Total Test Score for set A	Total Test Score for set B
14	8
16	12
12	6
22	11
23	18
21	23
19	21
8	4
23	8
5	17
22	19
16	24
20	9
25	25
4	18

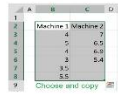
Calculate Clear Switch

Group1 contains 40 values
validation:success

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Enter sample data

You may copy data from Excel, Google sheets or any tool that separate data with **Tab** and **Line Feed**. Copy the data, **one block of 3 consecutive columns includes the header**, and paste below.



Copy the data,

Calculate Clear Validate

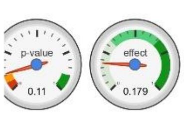
It is okay to leave empty cells, empty cells or non numeric cells won't be counted

Results

Sample average (x̄):	16.5	13.925
Sample size (n):	40	40
Sample SD (S):	6.740616	7.208924
Median:	16.5	15
Skewness:	-0.232922	-0.0883955
Skewness Shape:	Potentially **Symmetrical**	Potentially **Symmetrical**
Normality:	0.0407	0.01744
Outliers:		
Outlier count:	0	0
Rank:	1786.5	1453.5
U:	633.5	966.5
Ties Correction:	0.0026957337083919363	
Exact:	false	

The difference between the expected SD and the sample SD

How to do with R?



Two sample mann-whitney u, using Normal distribution (two-tailed) (validation)

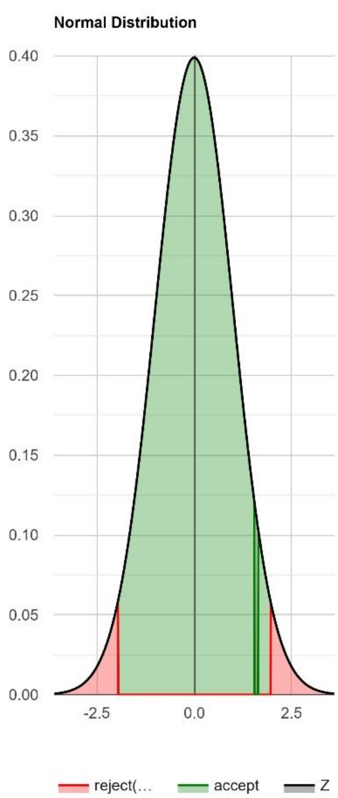
The **normal approximation** is used. The statistic's distribution is N(800,103.783^2).
The data contains ties, identical values, it is recommended to use the **normal approximation** that uses the ties correction. Since n1 > 20 or n2 > 20, the tool cannot use the **exact calculation**.

1. H_0 hypothesis
Since p-value > α, H_0 cannot be rejected.
The randomly selected value of **Total Test Score for set A's** population is assumed to be **equal to** the randomly selected value of **Total Test Score for set B's** population.
In other words, the difference between the randomly selected value of **Total Test Score for set A** and the **Total Test Score for set B** populations is not big enough to be statistically significant.

2. P-value
The p-value equals 0.1097, ($p(x \le Z) = 0.9451$). It means that the chance of type I error, rejecting a correct H_0, is too high: 0.1097 (10.97%).
The larger the p-value the more it supports H_0.

3. The statistics
The test statistic Z equals 1.5995, which is in the 95% region of acceptance: [-1.96 : 1.96].
U=966.5, is in the 95% region of acceptance: [596.5893 : 1003.4107].

4. Effect size
The observed **standardized effect size**, $Z/\sqrt{(n_1+n_2)}$, is **small** (0.18). That indicates that the magnitude of the difference between the value from **Total Test Score for set A** and the value from **Total Test Score for set B** is small.
The observed **common language effect size**, $U_1/(n_1 n_2)$, is **0.6**, this is the probability that a random value from **Total Test Score for set A** is greater than a random value from **Total Test Score for set B**.
If you like the page, please share or like. Questions, comments and suggestions are appreciated. (statskingdom@gmail.com)



ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Test validation

The requested test was calculated, it is likely you chose the right test.

- **Outliers** +

Outliers' detection method: Tukey Fence, k=1.5
The data doesn't have outliers.

Information

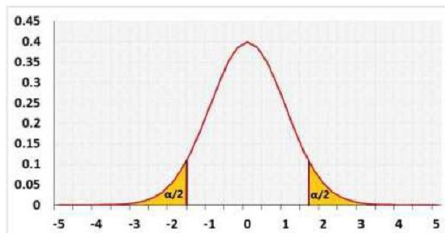
Hypotheses

H_0: $Group_1 = Group_2$
H_1: $Group_1 \neq Group_2$

Test statistic

$$Z = \frac{U - \mu + c}{\sigma}$$

Normal distribution



C - continuity correction, when U > μ: C = -0.5, when U < μ: C = 0.5

Target

Unlike t-test that compares the means, the Mann-Whitney U test compares a randomly selected value from group1 to a randomly selected value from group2.
When the two distributions have a similar **shape** you can use the test to compare also the **medians**.
When the two distributions have a similar **symmetrical shape**, you can use the test to compare also the **means**. more

Method

The mann whitney u test calculator may use three methods. We reccomend to use the "Automatic" method.
Automatic - when $n_1 \leq 20$ and $n_2 \leq 20$ and the data doesn't have ties, the tool uses the exact value, otherwise the tool uses the z approximation.
Exact - when $n_1 \leq 20$ and $n_2 \leq 20$ the tool uses the exact value, calculated base on all the possible combinations, otherwise the tool uses the z approximation.
Z approximation - the tool uses the z approximation.

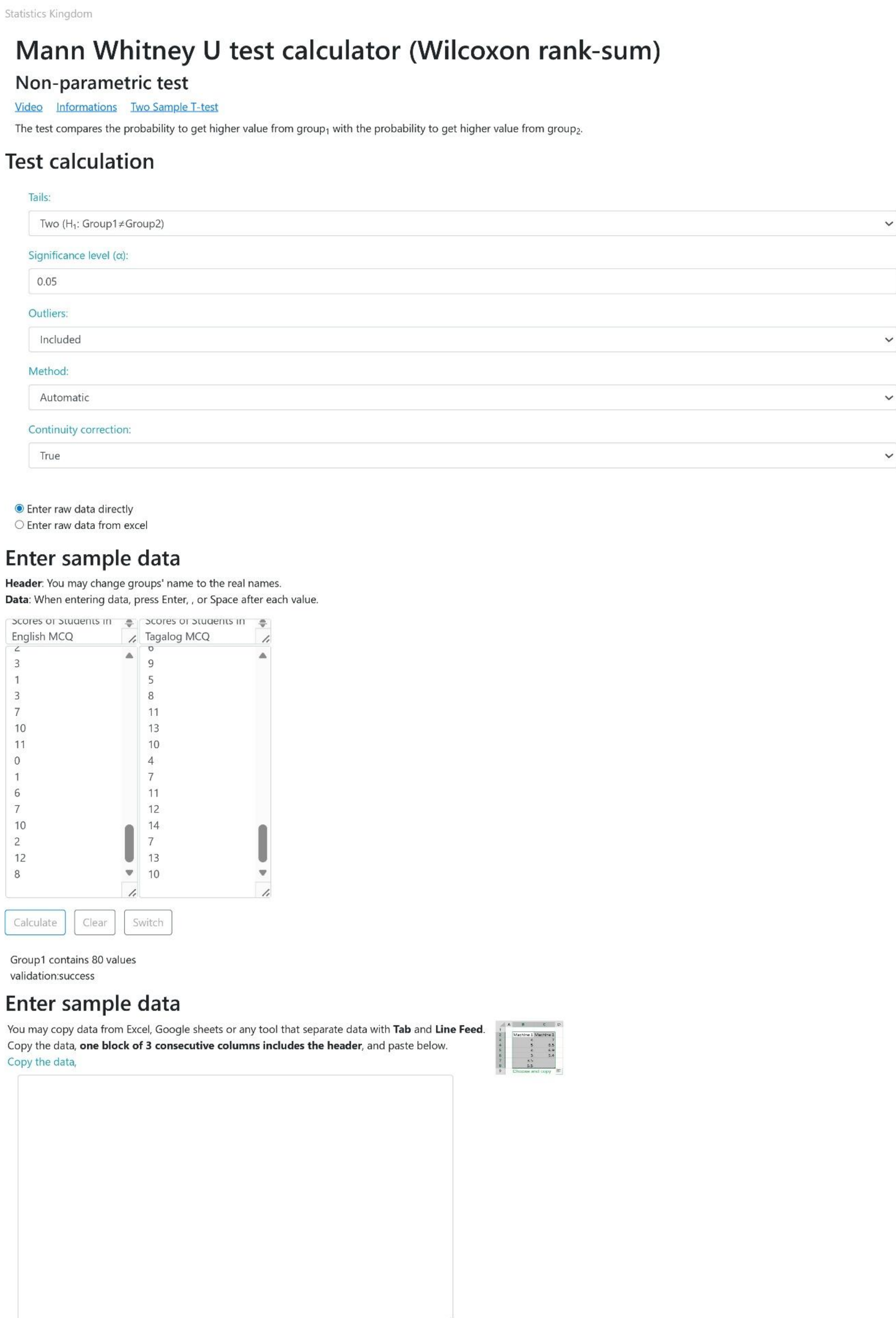
Continuity correction - relevant only for the Z approximation. It is used since a continuous distribution is used to calculate a discrete distribution.

R Code

The following R code should produce similar results:

```
x1<-c(11,6,18,6,24,27,10,13,15,22,16,9,17,23,15,9,18,26,7,25,9,27,21,21,15,14,16,12,22,23,21,19,8,23,5,22,16,20,25,4)
x2<-c(10,15,23,3,8,3,19,24,8,19,19,6,14,6,2,21,13,18,15,4,26,15,4,18,21,8,12,6,11,18,23,21,4,8,17,19,24,9,25,18)
wilcox.test(x1, x2, alternative = "two.sided", paired = FALSE, exact = FALSE, correct = TRUE)
```

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)



ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)

Calculate Clear Validate

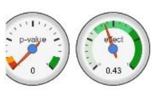
It is okay to leave empty cells, empty cells or non numeric cells won't be counted

Results

Sample average (x̄):	5.8125	9.4
Sample size (n):	80	80
Sample SD (S):	3.955748	3.373763
Median:	6	10
Skewness:	0.130224	-0.321234
Skewness Shape:	Potentially **Symmetrical** (pval=0.628)	Potentially **Symmetrical** (pval=0.232)
Normality:	0.001262	0.0092
Outliers:		
Outlier count:	0	0
Rank:	4850	8030
U:	4790	1610
Ties Correction:	0.006722430563693894	
Exact:	false	

The difference between the expected SD and the sample SD

How to do with R?



Two sample mann-whitney u, using Normal distribution (two-tailed) (validation)

The **normal approximation** is used. The statistic's distribution is $N(3200, 292.044^2)$.
The data contains ties, identical values, it is recommended to use the **normal approximation** that uses the ties correction. Since n1 > 20 or n2 > 20, the tool cannot use the **exact calculation**.

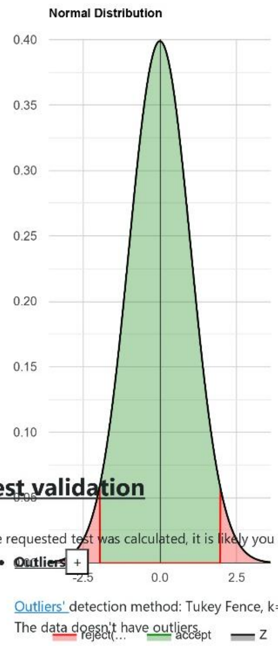
1. H_0 hypothesis
Since p-value < α, H_0 is rejected.
The randomly selected value of **Scores of Students in English MCQ's** population is considered to be **not equal to** the randomly selected value of **Scores of Students in Tagalog MCQ's** population.
In other words, the difference between the randomly selected value of **Scores of Students in English MCQ** and the **Scores of Students in Tagalog MCQ** populations is big enough to be statistically significant.

2. P-value
The p-value equals 5.248e-8, ($p(x \le Z) = 2.624e\text{-}8$). It means that the chance of type I error (rejecting a correct H_0) is small: 5.248e-8 (0.0000052%).
The smaller the p-value the more it supports H_1.

3. The statistics
The test statistic Z equals -5.4427, which is not in the 95% region of acceptance: [-1.96 : 1.96].
U=1610, is not in the 95% region of acceptance: [2627.6052 : 3772.3948].

4. Effect size
The observed **standardized effect size**, $Z/\sqrt{(n_1+n_2)}$, is **medium** (0.43). That indicates that the magnitude of the difference between the value from **Scores of Students in English MCQ** and the value from **Scores of Students in Tagalog MCQ** is medium.
The observed **common language effect size**, $U_1/(n_1 n_2)$, is **0.25**, this is the probability that a random value from **Scores of Students in English MCQ** is greater than a random value from **Scores of Students in Tagalog MCQ**.
If you like the page, please share or like. Questions, comments and suggestions are appreciated. (statskingdom@gmail.com)

ISSN: 3116-3475
International Journal
of Education, Research, and Innovation Perspectives
Volume 2 Issue 5 (May 2026)



Test validation

The requested test was calculated, it is likely you chose the right test.

- **Outliers**

Outliers' detection method: Tukey Fence, k=1.5
The data doesn't have outliers.

Information

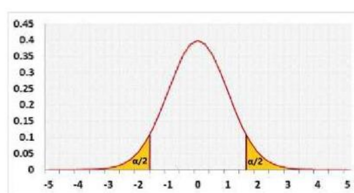
Hypotheses

H_0: $Group_1 = Group_2$
H_1: $Group_1 \neq Group_2$

Test statistic

$$Z = \frac{U - \mu + c}{\sigma}$$

Normal distribution



C - continuity correction, when $U > \mu$: $C = -0.5$, when $U < \mu$: $C = 0.5$

Target

Unlike t-test that compares the means, the Mann-Whitney U test compares a randomly selected value from group1 to a randomly selected value from group2.
When the two distributions have a similar **shape** you can use the test to compare also the **medians**.
When the two distributions have a similar **symmetrical shape**, you can use the test to compare also the **means**. more

Method

The mann whitney u test calculator may use three methods. We reccomend to use the "Automatic" method.
Automatic - when $n_1 \leq 20$ and $n_2 \leq 20$ and the data doesn't have ties, the tool uses the exact value, otherwise the tool uses the z approximation.
Exact - when $n_1 \leq 20$ and $n_2 \leq 20$ the tool uses the exact value, calculated base on all the possible combinations, otherwise the tool uses the z approximation.
Z approximation - the tool uses the z approximation.

Continuity correction - relevant only for the Z approximation. It is used since a continuous distribution is used to calculate a discrete distribution.

R Code

The following R code should produce similar results:

```
x1<-c(4,1,7,1,10,13,3,4,6,9,7,2,7,9,5,2,9,14,2,11,2,14,8,9,5,5,6,3,10,10,9,8,2,10,0,11,7,9,11,0,3,6,10,0,2,0,7,10,1,10,7,1,4,2,0,9,4,7,5,0,12,5,1,2,9,2,3,1,3,7,10,11,0,1,6,7,10,2,12,8)
x2<-
c(7,5,11,5,14,14,7,9,9,13,9,7,10,14,10,7,9,12,5,14,7,13,13,12,10,9,10,9,12,13,12,11,6,13,5,11,9,11,14,4,7,9,13,3,6,3,12,14,7,9,12,5,10,4,2,12,9,11,10,4,14,10,3,16,12,6,9,5,8,11,13,10,4,7,11,12,14,7,13,10)
wilcox.test(x1, x2, alternative = "two.sided", paired = FALSE, exact = FALSE, correct = TRUE)
```