MALAYSIAN JOURNAL OF LEARNING AND INSTRUCTION
https://e-journal.uum.edu.my/index.php/mjli

How to cite this article:
Veloo, A., Shanmugam, K. S., & Revindran, S. (2025). A comparative study of cognitive processing in oral mathematics tests among Malaysian *Orang Asli* pupils. *Malaysian Journal of Learning and Instruction, 22*(2), 119-139. https://doi.org/10.32890/mjli2025.22.2.6

A COMPARATIVE STUDY OF COGNITIVE PROCESSING IN ORAL MATHEMATICS TESTS AMONG MALAYSIAN *ORANG ASLI* PUPILS

[1]Arsaythamby Veloo, [2]S. Kanageswari Suppiah Shanmugam & [3]Suheysen Revindran
[1,2]School of Education, Universiti Utara Malaysia
[3]Awang Had Salleh Graduate School of Arts and Sciences, Universiti Utara Malaysia

[1]Corresponding Author: arsay@uum.edu.my

Received: 1/1/2024 Revised: 18/4/2024 Accepted: 17/6/2025 Published: 31/7/2025

ABSTRACT

Purpose – This study investigates the mathematics performance of *Orang Asli* pupils across two cognitive domains—Knowing and Applying—using three oral mathematics tests developed in different languages: the academic language (*Bahasa Melayu*) (OBM), the academic native language (*Bahasa Semai*) (OSL), and the tribal native language (*Bahasa Temiar*) (OTL).

Methodology – Adopting a quantitative approach, the study involved 225 Year 4 *Orang Asli* pupils who were assessed using a 20-item mathematics computation testlet. The analysis focused on pupils' performance across the Knowing and Applying domains in the three oral test formats (OBM, OSL, and OTL). One-way Analysis of Variance (ANOVA) was used to compare performance differences across the cognitive domains and test formats.

Findings – In the Knowing domain, pupils performed best in the OSL test, followed by OBM and OTL. Conversely, in the Applying domain, pupils excelled in the OTL test, followed by OSL and OBM. ANOVA results indicated significant differences between OSL and OTL, as well as OBM and OTL in the Knowing domain. However, no significant difference was found between OSL and OBM. For the Applying domain, a significant difference was observed only between OTL and OBM, with no significant differences between OTL and OSL, or between OSL and OBM.

Significance – The findings highlight the potential of oral mathematics tests in native languages—particularly *Bahasa Semai*—as a culturally responsive and equitable assessment method. This approach enables *Orang Asli* pupils to better demonstrate their mathematical proficiency helping to address linguistic barriers in mathematics assessment.

Keywords: Oral mathematics test, computation, academic language, native language, knowing, applying.

INTRODUCTION

Early mastery of cognitive skills plays a crucial role in an individual's personal development and future economic success (Spaull & Kotze, 2015). Bloom’s taxonomy, a widely referenced framework in education, categorises learning into three domains: cognitive, affective, and sensory. It serves as a foundation for curriculum design (Zapalska et al., 2018) and the development of assessments that track pupils’ academic progress (Radmehr & Drake, 2018). Similarly, the Trends in Mathematics and Science Study (TIMSS) framework underscores the importance of cognitive skills by evaluating pupils’ mathematical achievement every four years. TIMSS categorizes cognitive processes into three levels: Knowing, Applying, and Reasoning—each representing a different degree of cognitive complexity (Mullis & Martin, 2017). This hierarchical model aligns with the revised version of Bloom’s taxonomy (Krathwohl, 2002) and reflects a continuum from lower-order to higher-order thinking skills (Jansen & Möller, 2022; Kosasih et al., 2022). However, while these structured frameworks provide a valuable tool for measuring learning outcomes, they can be challenging for pupils who struggle with abstract reasoning and complex problem-solving (Mullis et al., 2017).

These challenges are even more pronounced for Indigenous pupils, such as the Malaysian *Orang Asli*, due to linguistic barriers. Studies indicate that Indigenous pupils often face difficulties in abstract reasoning and higher-order problem-solving, which are further complicated by the linguistic demands of standard mathematics assessments (Kūkea Shultz & Englert, 2021; Matang & Owens, 2014). Limited proficiency in the academic language can create a mismatch with the cognitive demands outlined in the TIMSS framework, potentially resulting in inaccurate assessments of their true mathematical abilities (Yushau & Omar, 2015). This highlights the importance of equitable testing practices that minimize language-related bias and more accurately reflect learners’ actual performance.

In Malaysia, *Orang Asli* pupils continue to face significant challenges in both literacy and numeracy. For instance, Sani and Idris (2018) reported that only one out of 28 *Orang Asli* schools met the minimum proficiency standards of the national Literacy and Numeracy Screening (LINUS) programme. Furthermore, only 1.5% (48 out of 3,011) of *Orang Asli* pupils passed the Primary School Achievement Test (*Ujian Pencapaian Sekolah Rendah* [UPSR]). A major contributor to their underperformance in mathematics is limited language proficiency (Abd Jalil et al., 2023; Shanmugam et al., 2023).

The Malaysian primary school curriculum reinforces a hierarchical classification of mathematical tasks following the structure of Knowing, Applying, and Reasoning skills (Tan et al., 2018). Although the revised Bloom’s taxonomy underpins curriculum development to foster higher-order thinking skills, national assessments outcomes suggest that many pupils continue to struggle with complex tasks. These difficulties are often rooted in the challenge of integrating computation, reading comprehension, and analytical reasoning (Misrom et al., 2020; Shanmugam et al., 2021; Susanto et al., 2019). Mathematics tests, which include computational tasks requiring proficiency in basic arithmetic (Boonen et al., 2016; Pongsakdi et al., 2020; Tambunan, 2019), can be overwhelming when combined with higher-level problem-solving, particularly for pupils with limited foundational skills.

A promising approach to these challenges is the use of orally administered mathematics tests delivered in both the academic language and Indigenous languages. Oral assessments have been shown to reduce the linguistic demands associated with written tests (Cawthon et al., 2012; Cohen et al., 2017) and help alleviate cognitive overload by allowing students to focus more fully on mathematical concepts (Chow & Ekholm, 2019; Peng et al., 2020). Building on this premise, the present study seeks to isolate and address the cognitive challenges that stem from language mismatches in mathematics assessments. Specifically, it examines *Orang Asli* pupils' performance in the Knowing and Applying cognitive domains. As prior research indicates that language barriers disproportionately affect performance in the Reasoning domain (Ismail et al., 2020), this study excludes items from that category to ensure that language-related issues do not distort assessments of mathematical ability. By distinguishing between lower-order thinking (Knowing) and the application of knowledge (Applying), this study seeks to offer deeper insights into the cognitive challenges experienced by *Orang Asli* pupils.

In summary, this study investigates the mathematical computation performance of *Orang Asli* pupils under three distinct oral mathematics test conditions: (1) an oral test in the academic language, *Bahasa Melayu* (OBM); (2) an oral test in the academic native language, *Bahasa Semai* (OSL); and (3) an oral test in the tribal native language, *Bahasa Temiar* (OTL). While the OBM and OSL tests reflect the formal academic languages currently taught in their primary schools, the OTL test represents the pupils' first, everyday language spoken within their communities. The study focuses on performance in the Knowing and Applying cognitive domains, with two primary objectives: (1) to examine overall mathematics computation performance within these domains, and (2) to explore differences in cognitive processing as revealed by the three oral test modalities. Ultimately, the findings aim to provide valuable insights into how language medium influences mathematics assessment outcomes across different levels of cognitive complexity, and to inform the development of more culturally responsive and inclusive educational practices for *Orang Asli* pupils.

LITERATURE REVIEW

Oral Mathematics Test

Orally administered mathematics tests have garnered much attention in educational research, particularly in relation to pupils with linguistic limitations—both within special education and general student populations (Penner, 2016; Rogers et al., 2019; Spiel et al., 2019). For instance, Cho et al. (2020) emphasized that oral tests can serve as a valid measure of mathematical ability by reducing the influence of language and reading proficiency. Similarly, Bolt and Thurlow (2007) found that pupils with reading difficulties performed better in oral mathematics tests. This suggests that oral mathematics test may enhance comprehension for learners with limited decoding skills and reduce the cognitive load associated with written text.

In the Malaysian context, studies specific to *Orang Asli* pupils have also demonstrated the effectiveness of oral tests in overcoming language-related challenges. Ismail et al. (2020) reported significantly higher performance among young *Orang Asli* pupils in oral mathematics tests compared to written assessments. Similarly, Veloo et al. (2021) noted a marked difference in performance between written and oral tests among *Orang Asli* pupils, underscoring the importance of alternative assessment methods to accurately evaluate mathematical abilities among Indigenous learners. While existing literature supports the use of oral tests to reduce the impact of language in mathematics assessments (Ercikan & Por, 2020; Shanmugam et al., 2024), there remains a significant gap in research concerning how these assessments cater to different levels of cognitive domains. Therefore, there is a growing need to design oral mathematics tests that not only address language barriers but also promote critical thinking and problem-solving. Such assessments would provide a more comprehensive and equitable evaluation of pupils’ mathematics performance.

Mathematics Computation Items

Computation items in mathematics assessments are generally designed to allow students to demonstrate their mathematical knowledge and problem-solving skills through tasks based on real-world contexts (Tambunan, 2019). Fuchs et al. (2018) describe a computation item as a mathematical question that requires arithmetic calculations and the application of skills in both "number combination" and "procedural computation." These two skills are key to solving computation tasks, as learners must consider numerical relationships and combinations in order to perform the correct calculation and arrive at the appropriate solution (Brezovszky et al., 2019). Braeuning et al. (2021) introduced the concepts of “symbols” and “magnitude” in relation to numerical understanding, arguing that computation items assess a learner’s ability to estimate quantities or perform comparisons between numbers. This indicates the importance of students’ familiarity with magnitude representation—often presented as Arabic numerals—which supports the use of appropriate solution strategies. Barcelos et al. (2018) and Wijns et al. (2021) contend that the numerical abilities assessed in computation items are important predictors of mathematical achievement in the later years of primary education.

Furthermore, Abedi (2011) discussed the linguistic aspects of computation items, asserting that such tasks typically involve minimal language load. This characteristic makes computation items valuable tools for assessing learners’ mathematical ability without complications due to language or linguistic complexity. A number of studies have explored how linguistic factors impact the difficulty level of mathematics test items (Boonen et al., 2016; Cho et al., 2020; Daroczy et al., 2015; Pongsakdi et al., 2020), indirectly addressing the disparity in difficulty between computation tasks and word problems—largely due to variations in language load.

Cognitive Domains in Mathematics

According to the TIMSS 2015 Mathematics Framework, the cognitive domain refers to the level of thinking required for students to solve mathematical tasks (Grønmo et al., 2015). At the primary level, TIMSS categorizes these cognitive domains into Knowing and Applying. The Knowing domain involves low cognitive demand and encompasses students’ recall of facts, procedures, and basic concepts (Mullis & Martin, 2017). Tasks within this domain typically focus on memorisation and basic calculations. On the other hand, the Applying domain entails a higher level of cognitive demand (Tan et al., 2018). It requires students to use acquired knowledge—such as mathematical concepts, procedures, and facts—to solve tasks or routine problems.

Some frameworks also differentiate this from the Reasoning domain, which involves more advanced thinking processes, including analyzing, generalizing, synthesizing, justifying, and solving non-routine problems (Philpot et al., 2021) Grønmo et al. (2015) elaborated on the Knowing and Applying domains by identifying specific aspects associated with each domain. The aspects within the Knowing domain include Recall, Recognize, Classify/Order, Compute, Retrieve and Measure. Meanwhile, the Applying domain includes the aspects Determine, Represent/Model and Implement. Table 1 presents the definitions of each aspect under the Knowing and Applying domains as defined by Grønmo et al. (2015).

Table 1

Definitions of Each Aspect under the Knowing and Applying Domains

Domain	Aspect	Meaning
Knowing	Recall	Recall definitions, terminology, number properties, units of measurement, geometric properties, and notation (e.g., a × b = ab, a + a + a = 3a).
	Recognize	Recognize numbers, expressions, quantities, and shapes. Recognize entities that are mathematically equivalent (e.g., equivalent familiar fractions, decimals, and percents; different orientations of simple geometric figures).
	Classify/Order	Classify numbers, expressions, quantities, and shapes by common properties.
	Compute	Carry out algorithmic procedures for +, –, ×, ÷, or a combination of these with whole numbers, fractions, decimals, and integers. Carry out straightforward algebraic procedures.
	Retrieve	Retrieve information from graphs, tables, texts, or other sources.
	Measure	Use measuring instruments; and choose appropriate units of measurement.
Applying	Determine	Determine efficient/appropriate operations, strategies, and tools for solving problems for which there are commonly used methods of solution.
	Represent/Model	Display data in tables or graphs; create equations, inequalities, geometric figures, or diagrams that model problem situations; and generate equivalent representations for a given mathematical entity or relationship.
	Implement	Implement strategies and operations to solve problems involving familiar mathematical concepts and procedures.

Note. Adapted from https://timssandpirls.bc.edu/timss2019/frameworks/framework-chapters/mathematics-framework/mathematics-cognitive-domains-fourth-and-eighth-grades/index.html#side. Copyright 2019 by TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College, and International Association for the Evaluation of Educational Achievement.

Tan et al. (2018) highlighted that the proportion of mathematical tasks within the Knowing domain remains relatively high, reflecting the emphasis on building pupils' foundational competence during the early stages of learning. However, there has been a gradual increase in tasks within the Applying domain, indicating a growing focus on developing pupils' problem-solving skills (Tan et al., 2018). Generally, the primary-level pupils encounter a greater number of tasks in the Knowing domain, as they are learning new mathematical concepts. These are often complemented with tasks in the Applying domain, particularly in topics with which pupils are more familiar, in order to help consolidate their learning and understanding (Francis, 2024; Nordby et al., 2022).

Cognitive Load Theory

Cognitive Load Theory is a psychological framework that examines the amount of mental effort required by learners during the learning process (Sweller et al., 2011). The theory stipulates that when learners are presented with too much information, they may become cognitively overloaded, which hinders their ability to process and retain information effectively. This issue is particularly relevant for Indigenous learners, who often encounter additional challenges such as linguistic and cultural barriers, as well as limited access to educational resources (Abd Jalil et al., 2023). These factors may further exacerbate cognitive overload, particularly when engaging with mathematics test items (Shanmugam et al., 2024).

According to Sweller et al. (2011), there are three types of cognitive load that influence learning. The first is intrinsic cognitive load, which refers to the inherent difficulty or complexity of the material being learned. For example, mathematics topics such as algebra or calculus are naturally more cognitively demanding than simpler arithmetic concepts. The second type, extraneous cognitive load, is imposed by the way information is presented. For example, poor instructional design—such as presenting too much information at once, using confusing or ambiguous language, or displaying content in a disorganized manner—can increase this form of load, diverting cognitive resources away from actual learning. Finally, germane cognitive load refers to the mental effort that contributes positively to learning. It supports the construction of mental schema and the integration of new information with existing knowledge. In this sense, germane load is beneficial, as it enhances the learner's ability to make sense of and retain new concepts.

Cognitive Load Theory has increased awareness of how learners process information. Although the theory is more commonly applied to guide teaching and learning in the classroom, its principles are equally relevant to assessment practices. Effective comprehension occurs only when intrinsic cognitive load is appropriately managed, and extraneous cognitive load is minimized, allowing learners to allocate more of their cognitive resources toward germane cognitive load. In other words, instructional designers and teachers should strive to present information in ways that minimizes unnecessary cognitive load, while enhancing the relevance and meaningfulness of the material being assessed (Blackley et al., 2021; Ellerton, 2022; Sweller, 2020). Figure 1 presents an overview of the three types of cognitive load and their influence on pupils' performance in mathematics.

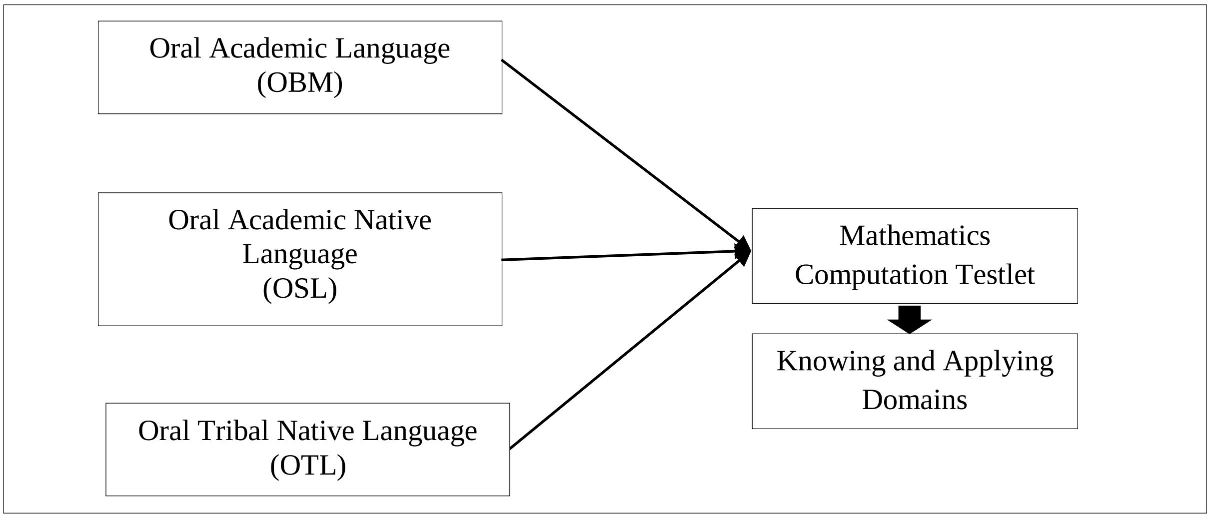
For Indigenous learners, reducing cognitive load is particularly important due to the additional challenges they often face. Compared to their mainstream peers, these learners may require more scaffolding and support. In this context, orally administered mathematics test represent an innovative form of test accommodation that may potentially reduce the cognitive demands associated with conventional mathematics tests (Roschmann et al., 2021; Zhang & Rivera, 2021). Such measures have the potential to help pupils more accurately demonstrate their mathematical knowledge and skills when being assessed.

Conceptual Framework for Oral Mathematics Test

This study employed three independent variables and a dependent variable to examine the mathematics computation performance of *Orang Asli* pupils across the Knowing and Applying cognitive domains, as assessed through a Mathematics Computation testlet. The three independent variables are the modes of oral test administration: (1) Oral test in the academic language (OBM); (2) Oral test in the academic native language (OSL) and (3) Oral test in the tribal native language (OTL). The dependent variable is the pupils' performance on the Mathematics Computation Oral Test, which focuses on the Knowing and Applying domains. Based on these variables of the study, a conceptual framework was constructed to guide the assessment of *Orang Asli* pupils' mathematics performance. Figure 1 presents the structure of this framework, illustrating the relationships between the oral test conditions and the cognitive domains being assessed.

Figure 1

Conceptual Framework for Oral Test Accommodation



This study responds to the urgent need to support *Orang Asli* pupils—especially those in Level One—who often struggle to understand languages other than their own native tongue. It underscores the importance of providing scaffolding to facilitate their learning of mathematics. Based on previous studies, the present study investigates the use oral computation testlets in both academic and native languages to better accommodate these learners. The study aims to: (1) Identify the level of mathematics computation performance across the Knowing and Applying domains. (2) Examine differences in cognitive domain performance across the three oral mathematics tests conditions: oral test in the academic language (OBM), oral test in the academic native language (OSL) and oral test in the tribal native language (OTL).

METHODOLOGY

Research Design

This study employed a quantitative approach to examine the validity of the orally administered mathematics tests in both the academic and native languages of the *Orang Asli* pupils. A random equivalent group design was employed to control for potential stress-inducing effects and to minimize order effects that stem from the administration of the three different test booklets (Livingston & Kim, 2010). By employing this research design, the groups of test-takers were made comparable, thus, ensuring that their experiences across the different oral mathematics test conditions were balanced and consistent.

Sample

A total of 225 *Orang Asli* pupils participated in this study. The participants were drawn from two primary schools in the district of Sungai Siput, Perak, a primary school in Gua Musang, Kelantan and a primary school in Kuala Lipis, Pahang. The sample consisted of 106 (47%) male pupils, and 119 (53%) female pupils. The sample included pupils from two native *Orang Asli* groups: The *Temiar* pupils were from schools in Sungai Siput, Perak and Gua Musang, Kelantan while the *Semai* pupils were from a school in Kuala Lipis, Pahang. Table 2 shows the distribution of *Orang Asli* pupils in the study by gender and ethnic group.

Table 2

Sample Distribution by Gender and Ethnic Group

Gender	Ethnic Group		Total
	Temiar	Semai	
Male	72	34	106
Female	80	39	119
Total	152	73	225

Mathematics Tests

The test used three specially constructed mathematics test booklets, each representing a different oral test condition: the oral test in the academic national language (OBM), the oral test in the academic native language (OSL) and the oral test in the tribal native language (OTL). Each test booklet included a computation testlet designed for Year 4 pupils. A total of 20 computation items were selected based on their relevance and appropriateness for assessing pupils' mathematics skills and knowledge. The test blueprint for the computation testlet was structured according to the TIMSS cognitive domains—Knowing and Applying. Of the 20 items, 16 items (80%) targeted the Knowing domain, while four items (20%) addressed the Applying domain. The mathematics computation testlet consisted of seven out of eight topics found in the Year 4 mathematics syllabus. These topics include: Whole Numbers up to 100, Basic Operations, Fractions, Money, Time, Space and Data Management, as presented in Table 3 which shows the breakdown of the computation items by topic and cognitive domain.

Table 3

Cognitive Domains for the Mathematics Computation Test

Topic	Cognitive Domain		Total
	Knowing	Applying	
Whole Numbers up to 100	C1, C2, C3, C4, C5, C6	C17, C18	8
Basic Operations	C7, C8, C9, C10, C11	C19	6
Fractions	C12		1
Money	C13, C14		2
Time	C15		1
Space	C16		1
Data Management		C20	1
Total	16	4	20

Oral Test in Academic Language (OBM)

The OBM test booklet was developed based on a standard written mathematics test in the academic language, Bahasa Melayu. Each of the 20 computation items in the test was individually audio recorded. An expert teacher—who is a native speaker and fluent in *Bahasa Melayu—*read each item aloud for recording. The expert involved in the development of the OBM test is from the *Temiar* native tribe in Gua Musang, Kelantan. The OBM test was compiled as an audio booklet containing audio recordings of all 20 computation items. To ensure validity of the oral test, all items in the OBM test were reviewed and validated by the expert to confirm their equivalence to the original standard written mathematics test in the academic language (Bahasa Melayu). The reliability of the OBM test was assessed using Cronbach's alpha, yielding a value of $\alpha = .71$. According to Kline (2015), a minimum value of 0.70 is considered acceptable for test reliability. Therefore, the OBM test was deemed reliable and consistent for use in this study.

Oral Test in Academic Native Language (OSL) and Tribal Native Language (OTL)

For the development of the OSL (*Bahasa Semai*) and OTL (*Bahasa Temiar*) tests, the original mathematics test items—initially developed in the standard written academic language (Bahasa Melayu)—were first translated into the respective native languages of the *Orang Asli* pupils. These translated items were produced as written scripts and underwent a process of transadaption to ensure both linguistic and cultural equivalence appropriate to the context of the *Temiar* and *Semai* communities. Each item was then read aloud and audio recorded in the respective native language by native speakers of *Semai* and *Temiar*. These speakers, who are also primary school teachers in *Orang Asli* schools, are fluent in their native languages and familiar with both the educational context and the specific computation test items used in the study.

The audio recordings of the 20 computation items were saved as two distinct sets: one in Bahasa Semai (OSL) and the other in Bahasa Temiar (OTL), representing the two primary native languages spoken by the *Orang Asli* pupils. The reliability of the tests was evaluated using Cronbach's alpha. The OSL test yielded a value of $\alpha = .69$, and the OTL test $\alpha = .68$. While slightly below the conventional threshold, both values are close to the minimum acceptable standard of $\alpha = .70$ as suggested by Kline (2015), and thus the OSL and OTL tests were considered to demonstrate acceptable reliability. Figure 2 presents examples of Item 2 and Item 4 from the OBM, OSL, and OTL tests, illustrating computation items for the Knowing and Applying domains in each of the three languages.

Figure 2

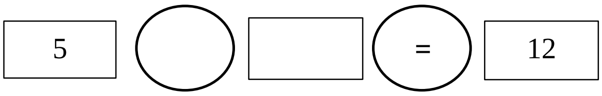
Example of Computation Test Items for the Knowing and Applying Domains

Item 2 (Knowing – C7)
English Version (For publication purposes)
Write the numbers and symbols in the box provided. Five plus seven equals twelve.

Oral Test in Academic Language (OBM)
Tuliskan nombor dan simbol dalam kotak yang disediakan. Lima tambah tujuh sama dengan dua belas.

Oral Test in Academic Native Language (OSL)
Tules numbor ru simbol ku kateh kutak kerom. Limak tambah tujuh samak dengan duak belas.

Oral Test in Tribal Native Language (OTL)
Te'elkan numbor nyot simbol keloj kutak yang disediakan. Limak tabé tujuh samak dengan dua belas.

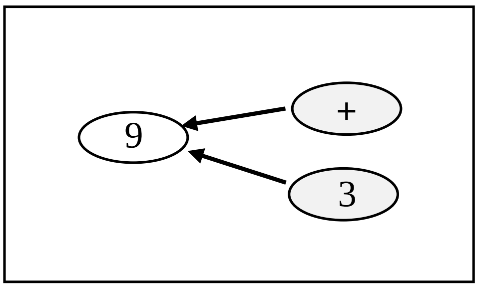


Item 4 (Applying – C9)
English Version (For publication purposes)
Fill in the appropriate number in the empty circle.

Oral Test in Academic Language (OBM)
Isikan nombor yang sesuai dalam bulatan. (OBM)

Oral Test in Academic Native Language (OSL)
Isik numbor de mentul kateh gambar bulat. (OSL)

Oral Test in Tribal Native Language (OTL)
*Te'elkan numbo yang sesuai keloj bulatan. (*OTL*)*



Data Collection

The Year 4 *Orang Asli* pupils were administered three mathematics computation tests—OBM, OTL, and OSL using a spiral administration method. Within each primary school, pupils were divided randomly into two groups. One group was administered the OBM test, while the second group received either the OSL or OTL test, depending on the dominant ethnic group represented in the school. Specifically, pupils in primary schools located in Sungai Siput, Perak, and Gua Musang, Kelantan took the OTL test (Bahasa Temiar), while those in Kuala Lipis, Pahang took the OSL test (Bahasa Semai). During the administration of the tests, audio recordings of each item (for OBM, OSL, and OTL) were played twice to the respective groups. An additional two minutes was allocated for each item to allow for sufficient time for pupils to respond in the test booklet. A printed version of the standard mathematics test (in *Bahasa Melayu*) was also provided to the two groups in each school. This booklet served as a reference for the 20 mathematics computation items and as the answer booklet, where pupils recorded their answers. The total duration for completing the mathematics computation testlet was approximately one hour.

Data Analysis

In this study, data were analysed using item analysis and a one-way ANOVA on the test items in the mathematics computation testlet. Item analysis was conducted based on Classical Test Theory (CTT) framework, which is a statistical technique commonly used to evaluate the quality and performance of individual test items (Bichi, 2016). Specifically, this study employed the item difficulty index to assess the ease or difficulty of each item.

The item difficulty index (p) is calculated by dividing the number of participants who answered an item correctly by the total number of participants (Boopathiraj & Chellamani, 2013). A higher index indicates that the item was relatively easy, while a lower index suggests that the item was more difficult for the pupils.

To examine performance differences across the three oral test conditions—OBM, OSL, and OTL —a one-way ANOVA was conducted using the two cognitive domains (Knowing and Applying) as dependent variables. This analysis was used to determine whether there were statistically significant differences in pupils' performance across the three tests conditions for each cognitive domain. Following the ANOVA, a Tukey HSD post hoc test was performed to determine pairwise comparisons between the groups for both the Knowing and Applying domains (Abdi & Williams, 2010). All statistical analyses were conducted using the Statistical Package for the Social Sciences (SPSS), Version 28.

RESULTS

Mathematics Computation Performance Across Knowing and Applying Domains

The first objective of this study was to examine the level of mathematics computation performance among *Orang Asli* pupils based on the two cognitive domains: Knowing and Applying. Results indicated that pupils performed considerably better in the Knowing domain, with 74% of pupils answering the test items correctly ($p = .74$). On the other hand, only 30% of the pupils were able to correctly answer test items classified under the Applying domain ($p = .30$), suggesting that these in the Applying domain were more difficult for the *Orang Asli* pupils compared to those in the Knowing domain. Table 4 shows the overall item difficulty levels across the two cognitive domains as assessed in the oral computation testlet.

Table 4

Mathematics Oral Computation Performance by Cognitive Domain

Cognitive Domain	N	p
Knowing	225	0.74
Applying	225	0.30

Based on the three oral mathematics tests administered—OBM, OSL, OTL—the findings revealed that pupils found the OSL test to be the easiest overall, with a difficulty index of p = .68. This was followed by the OBM test (p = .65) and the OTL test (p = .62). For the Knowing domain specifically, pupils performed best on the OSL test (p = .78), followed by the OBM test (p = .74) and the OTL test (p =.68). In contrast, for the Applying domain, pupils found the OTL test to be the least difficult (p = .35), followed by the OSL test (p = .29) and the OBM test (p = .26). Table 5 summarises the item difficulty index for each cognitive domain across the three oral test conditions (OBM, OTL and OSL).

Table 5

Item Difficulty Index by Cognitive Domain and Oral Test Condition

Cognitive Domain	OBM	OTL	OSL
Knowing	0.74	0.68	0.78
Applying	0.26	0.35	0.29
Total	0.65	0.62	0.68

Mathematics Performance between Cognitive Domains and Oral Mathematics Tests

The second research question aimed to examine whether there were significant differences in mathematics computation performance across the two cognitive domains (Knowing and Applying) and the three oral test conditions: OBM, OTL, and OSL. The findings are presented separately for each cognitive domain as follows.

Knowing Domain

Levene's Test for Equality of Variances indicated that the assumption of homogeneity of variances was met for the Knowing domain ($p > 0.05$) suggesting comparable variance across the three oral mathematics tests. The results of one-way ANOVA showed a statistically significant difference in pupil performance across the three test conditions for the Knowing domain (F (2, 222) = 6.56, $p < 0.05$. As presented in Table 6, pupils achieved the highest mean score on the OSL test (M = 12.50, SD = 2.05), followed by the OBM test (M = 11.88, SD = 2.50), while the OTL test recorded the lowest performance (M = 10.94, SD = 2.27).

Table 6

One-way ANOVA Results for Mathematics Computation in the Knowing Domain

Oral Mathematics Test	N	Mean	SD	df1	df2	F	p
OSL	36	12.50	2.05	2	222	6.56	0.02*
OBM	108	11.88	2.50				
OTL	81	10.94	2.27				

*p < 0.05

In addition, the Tukey HSD post hoc test revealed significant differences in pupil performance between specific test conditions for the Knowing domain. There was a statistically significant difference between OBM and OTL ($p < 0.05$), with pupils performing better on the OBM test, indicating that OBM was easier compared to OTL. Similarly, a significant difference was found between OSL and OTL ($p < 0.05$), with the mean scores indicating that OSL was also easier than OTL for pupils. However, no statistically significant difference was found between OSL and OBM ($p > 0.05$) suggesting that *Orang Asli* pupils performed comparably on these two tests in the Knowing domain. Table 7 shows the detailed results of the Tukey HSD post hoc test for the Knowing domain.

Table 7

Tukey HSD Post Hoc Test Results for Mathematics Computation - Knowing Domain

Mathematics Test		Mean Difference	P
OSL	OTL	1.56	0.00*
OBM	OTL	.94	0.02*
OSL	OBM	.62	0.36

*p < 0.05, and p > 0.05

Applying Domain

As shown in Table 8, Levene's Test for the oral mathematics computation tests in the Applying domain was not significant ($p > 0.05$), indicating that the assumption of homogeneity of variance was met across the three test conditions: OBM, OTL and OSL. The results of one-way ANOVA revealed a statistically significant difference in pupils' performance across the three oral mathematics tests, $F (2, 222) = 4.16$, $p < 0.05$. Based on the mean scores presented in Table 8, pupils performed best on OTL test ($M = 1.38$, $SD = 0.68$), followed by OSL test ($M = 1.14$, $SD = 0.68$), while the OBM test recorded the lowest mean score ($M = 1.06$, $SD = 0.83$).

Table 8

One-way ANOVA Results for Mathematics Computation - Applying Domain

Oral Mathematics Test	N	Mean	SD	df1	df2	F	p
OTL	81	1.38	.68	2	222	4.16	.02*
OSL	36	1.14	.68				
OBM	108	1.06	.83				

*p < 0.05

The Tukey HSD post hoc test was conducted to identify pairwise differences between the three oral mathematics computation tests for the Applying domain. The analysis revealed a statistically significant difference between OTL and OBM ($p < 0.05$), with pupils performing significantly better on the OTL test. This finding suggests that the test administered in the tribal language (OTL) was easier for pupils compared to the test delivered in the academic national language (OBM). However, no significant differences were observed between OTL and OSL, or between OSL and OBM ($p > 0.05$). These results indicate that *Orang Asli* pupils performed best in the Applying domain when assessed in their tribal native language (OTL). Table 9 presents the detailed results of the Tukey HSD post hoc test for Applying domain.

Table 9

Tukey HSD Post Hoc Test Results for Mathematics Computation - Applying Domain

Mathematics Test		Mean Difference	p
OTL	OBM	.32	0.01*
OTL	OSL	.24	0.25
OSL	OBM	.74	0.87

*p < 0.05, and p > 0.05

DISCUSSION

One of the key findings of this study is that *Orang Asli* pupils performed significantly better in the Knowing domain of the mathematics computation testlet, while the Applying domain posed the most difficult challenge. The computation test items were generally direct in nature, which likely allowed pupils to better demonstrate their mathematical knowledge and problem-solving skills when responding to lower-order thinking tasks compared to those requiring higher-order thinking skills. The pupils' weaker performance in the Applying domain is consistent with previous studies by Jansen and Möller (2022), and Kosasih et al. (2022), which highlight the ongoing difficulties faced by *Orang Asli* pupils in tackling test items that require more complex reasoning and problem-solving abilities.

Furthermore, as most of the items in the mathematics computation testlet focused on basic arithmetic skills—such as simple calculations and number classification—the findings suggest that *Orang Asli* pupils did not find computation items in the Knowledge domain too difficult. Tasks in this domain typically require the reproduction of previously learned knowledge, which may be sufficiently manageable even with a rudimentary grasp of the relevant mathematical concepts and procedures, as highlighted by Francis (2024) and Mullis and Martin (2017). This suggests that test items in the Knowing domain place relatively low cognitive demands on learners. In contrast, test items in the Applying domain likely required more advanced integration of knowledge and cognitive effort, aligning with the perspectives of Blackley et al. (2011), Ellerton (2022), and Sweller (2020).

Alternatively, the difficulties faced by *Orang Asli* pupils with tasks in the Applying domain underscore their challenges in fully grasping mathematical concepts to the extent that they can apply them in practical contexts. This domain typically involves higher-order tasks such as problem-solving using a variety of mathematical tools and implementing appropriate strategies, all of which place a higher cognitive load on learners. For *Orang Asli* pupils, applying learned concepts to real-life scenarios can be particularly daunting due to the mismatch between the academic language used in instruction and their cultural and linguistic context. This disconnect often widens the gap between knowledge acquisition and meaningful understanding. Language further complicates this issue. The cognitive differences observed across various language test conditions suggests that language proficiency significantly impacts performance. When mathematical tasks are presented in a language that pupils understand well, they are more likely to perform effectively due to improved comprehension and cognitive alignment, as supported by the findings of Kūkea Shultz and Englert (2021), and Matang and Owens (2014).

The use of oral mathematics tests presents a promising approach to addressing the linguistic challenges commonly associated with traditional written assessments. By incorporating language-inclusive strategies, such tests can potentially reduce the language load on students, allowing for a more accurate evaluation of their mathematical abilities. Recent studies involving *Orang Asli* pupils, including those by Shanmugam et al. (2021) and Veloo et al. (2021) suggest that oral mathematics tests conducted in native languages may help mitigate the linguistic challenges often faced by these pupils. Interestingly, findings from this study reveal varying levels of difficulty across different native language test conditions, with participants performing better in *Bahasa Semai* compared to *Bahasa Temiar*. This suggests that performance among *Orang Asli* pupils may differ depending on the specific native language used in the assessment. While these findings offer valuable insights into the influence of native languages on mathematics performance among *Orang Asli* pupils, caution must be exercised when extrapolating these findings to other *Orang Asli* subgroups or broader educational contexts without considering the distinct linguistic, cultural and educational characteristics of each indigenous community when designing equitable and inclusive assessment strategies.

On the other hand, oral mathematics tests also demonstrate potential in enhancing pupils' computational skills and enabling them to showcase their mathematical abilities more effectively. This was evident in the present study, where pupils found the oral test in their tribal native language (OTL) to be the easiest for computational tasks. In contrast, the greatest difficulty was observed in the OBM test, particularly in the Applying domain, where higher cognitive demands are required. The findings align with those of Abdullah (2022) and Shanmugam et al. (2024), who emphasized the significance of linguistic familiarity in facilitating problem comprehension among *Orang Asli* pupils during mathematics assessments.

While the OSL and OBM tests may be more suitable for assessing basic computational tasks due to their alignment with academic instruction, the OTL test proves especially beneficial for addressing more complex problems. This advantage likely stems from pupils' greater fluency and comfort in their tribal native language, which facilitates comprehension and supports cognitive processing during problem-solving. The use of oral mathematics tests can therefore reduce the cognitive burden associated with deciphering the language of the test item, enabling pupils to focus on identifying the embedded mathematical problem (Cho et al., 2020; Rogers et al., 2019; Spiel et al., 2019). This approach recognizes the critical role of linguistic familiarity in facilitating both comprehension and application, thereby addressing the diverse cognitive needs of learners. By employing oral mathematics tests, educators can help reduce linguistic barriers and improve access to mathematical content, ultimately fostering deeper engagement with mathematical concepts across different cognitive domains (Roschmann et al., 2021; Zhang & Rivera, 2021).

Research Implications and Limitations

The findings of this study on the mathematics performance of *Orang Asli* pupils across different cognitive domains present significant implications for educational practice and policy. The observation that *Orang Asli* pupils performed better in the Knowing domain—characterized by tasks involving basic arithmetic skills and factual recall—suggests that educational interventions should focus on strengthening foundational mathematical concepts. However, the persistent difficulties observed in the Applying domain indicate the need for more culturally relevant and contextualized teaching strategies. This finding supports the call by Van den Heuvel-Panhuizen and Drijvers (2020), and Nugraheni and Marsigit (2021) for instructional practices that incorporate real-life problem-solving contexts, helping bridge the gap between abstract mathematical concepts and its practical applications. Moreover, the study underscores the importance of considering the pupils' sociocultural backgrounds and daily experiences in curriculum and instructional design to facilitate deeper understanding and improve engagement.

Another significant implication of the study concerns the role of transadaptation—the cultural and linguistic adaption—of test materials. The findings indicate that pupils performed better on computation tests delivered in academic languages such as *Bahasa Melayu* and *Bahasa Semai*, compared to tribal languages like *Bahasa Temiar*. Accordingly, there is need to develop linguistically accessible and culturally sensitive assessment tools that align with pupils' proficiency levels and linguistic contexts (Robertson & Graven, 2020; Tai, 2022).

However, a notable limitation of this study arises from the disparity in performance observed in the Applying domain between academic and tribal language tests. Pupils struggled more with academic language tests, suggesting a potential mismatch between language proficiency and the cognitive demands of the tasks. This highlights that direct translation of test materials alone may not be sufficient to support effective learning and assessment. Future studies should consider exploring innovative approaches to language integration in mathematics education—such as bilingual instruction or translanguaging practices—to more effectively address the diverse learning needs of *Orang Asli* pupils across varying domains of mathematical proficiency.

In addition to these implications, the study has several limitations that warrant consideration in future research. Firstly, the focus on *Orang Asli* pupils from specific geographical areas may limit the generalizability of the findings to other indigenous or marginalized populations. Future research should adopt a more diverse and inclusive sampling approach to capture a wider range of sociocultural contexts and educational experiences. Additionally, the exclusive use of oral tests in this study may overlook variations in literacy levels among participants, which could have influenced their performance outcomes. A comprehensive assessment framework—including both oral and written components—would provide a more holistic understanding of students' mathematical abilities as well as insights into how language and literacy skills interact in mathematics learning. Overall, addressing these limitations in future research can enrich our understanding of the complex interplay between language, culture, and cognition in mathematics education. Such efforts are crucial for the development of more equitable, culturally responsive, and inclusive mathematics assessment practices for indigenous learners.

CONCLUSION

This study highlights the significance of linguistic accommodation—particularly through the implementation of oral mathematics tests in both academic and tribal languages—as a promising and practical test accommodation for *Orang Asli* pupils. While fully adapting the national mathematics curriculum to the specific cultural context of Indigenous communities may be impractical, prioritizing linguistic accommodation offers a feasible and practical solution to support learning and assessment. In an increasingly interconnected and inclusive educational landscape, integrating Indigenous learners into mainstream education systems is imperative for promoting social cohesion and equity.

In line with UNESCO's Education for All initiative, inclusive education must prioritize the needs of marginalized populations, especially those who face economic, geographic, or linguistic disadvantages. Although considerable emphasis has often been directed towards curriculum design and pedagogical strategies, this study highlights the equal necessity of re-evaluating assessment practices to ensure that all pupils—regardless of background—are given fair opportunities to demonstrate their academic potential. The oral mathematics test emerges as a viable, equitable and culturally responsive assessment approach that reduces linguistic and literacy barriers. It provides *Orang Asli* pupils with a more accessible platform to demonstrate their mathematical knowledge, particularly in computation. Ultimately, addressing these barriers is essential not only for assessing pupils' true mathematical performance, but also for advancing the broader goal of equitable and inclusive education for all.

ACKNOWLEDGMENT

This research was supported by the Ministry of Higher Education (MoHE) of Malaysia through the Fundamental Research Grant Scheme (FRGS/1/2021/SSI01/UUM/02/1).

CONFLICT OF INTEREST

The authors declare no conflicts of interest related to this study.

REFERENCES

Abd Jalil, A. H., Abdullah, A. H., & Hamzah, M. H. (2023). Does language matter in the mathematics classroom for Orang Asli students in Malaysia? *Malaysian Journal of Social Sciences and Humanities, 8*(2), e002061. https://doi.org/10.47405/mjssh.v8i2.2061

Abdi, H., & Williams, L. J. (2010). Tukey's honestly significant difference (HSD) test. In N. J. Salkind (Ed.), *Encyclopedia of research design, 1,* 1-5. Sage Publication.

Abdullah, A. H. (2022). A systematic review of what Malaysia can learn to improve Orang Asli students' mathematics learning from other countries. *Sustainability, 14*(20), 13201. https://doi.org/10.3390/su142013201

Abedi, J. (2011). Language issues in item development. In T. M. Haladyna & S. M. Downing (Eds.), *Handbook of test development* (pp. 391-412). Routledge.

Barcelos, T. S., Muñoz-Soto, R., Villarroel, R., Merino, E., & Silveira, I. F. (2018). Mathematics learning through computational thinking activities: A systematic literature review. *Journal of Universal Computer Science*, *24*(7), 815-845.

Bichi, A. A. (2016). Classical test theory: An introduction to linear modeling approach to test and item analysis. *International Journal for Social Studies, 2*(9), 27-33.

Blackley, C., Redmond, P., & Peel, K. (2021). Teacher decision-making in the classroom: The influence of cognitive load and teacher affect. *Journal of Education for Teaching*, *47*(4), 548-561. https://doi.org/10.1080/02607476.2021.1902748

Bolt, S. E., & Thurlow, M. L. (2007). Item-level effects of the read-aloud accommodation for students with reading disabilities. *Assessment for Effective Intervention*, *33*(1), 15-28. https://doi.org/10.1177/15345084070330010301

Boonen, A. J., de Koning, B. B., Jolles, J., & Van der Schoot, M. (2016). Word problem solving in contemporary math education: A plea for reading comprehension skills training. *Frontiers in Psychology*, *7*, 155518. https://doi.org/10.3389/fpsyg.2016.00191

Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*, *2*(2), 189-193.

Braeuning, D., Hornung, C., Hoffmann, D., Lambert, K., Ugen, S., Fischbach, A., Schiltz, C., Hübner, N., Nagengast, B., & Moeller, K. (2021). Long-term relevance and interrelation of symbolic and non-symbolic abilities in mathematical-numerical development: Evidence from large-scale assessment data. *Cognitive Development*, *58*, 101008. https://doi.org/10.1016/j.cogdev.2021.101008

Brezovszky, B., McMullen, J., Veermans, K., Hannula-Sormunen, M. M., Rodríguez-Aflecht, G., Pongsakdi, N., Laakkonen, E., & Lehtinen, E. (2019). Effects of a mathematics game-based learning environment on primary school students' adaptive number knowledge. *Computers & Education*, *128*, 63-74. https://doi.org/10.1016/j.compedu.2018.09.011

Cawthon, S. W., Kaye, A. D., Lockhart, L. L., & Beretvas, S. N. (2012). Effects of linguistic complexity and accommodations on estimates of ability for students with learning disabilities. *Journal of School Psychology*, *50*(3), 293-316. https://doi.org/10.1016/j.jsp.2012.01.002

Cho, E., Fuchs, L. S., Seethaler, P. M., Fuchs, D., & Compton, D. L. (2020). Dynamic assessment for identifying Spanish-speaking English learners’ risk for mathematics disabilities: Does language of administration matter? *Journal of Learning Disabilities*, *53*(5), 380-398. https://doi.org/10.1177%2F0022219419898887

Chow, J. C., & Ekholm, E. (2019). Language domains differentially predict mathematics performance in young children. *Early Childhood Research Quarterly*, *46*, 179-186. https://doi.org/10.1016/j.ecresq.2018.02.011

Cohen, D. J., Blanc-Goldhammer, D., & Quinlan, P. T. (2018). A mathematical model of how people solve most variants of the number-line task. *Cognitive Science*, *42*(8), 2621-2647. https://doi.org/10.1111/cogs.12698

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology, 6*, 348. https://doi.org/10.3389/fpsyg.2015.00348

Ellerton, P. (2022). On critical thinking and content knowledge: A critique of the assumptions of cognitive load theory. *Thinking Skills and Creativity*, *43*, Article 100975. https://doi.org/10.1016/j.tsc.2021.100975

Ercikan, K., & Por, H. H. (2020). Comparability in multilingual and multicultural assessment contexts. In A. I. Berman, E. H. Haertel, & J. W. Pellegrino (Eds.), *Comparability of large-scale educational assessments: Issues and recommendations* (pp. 205-225). National Academy of Education.

Francis, D. I. C. (2024). Mathematics teaching and learning. In P. A. Schutz & K. R. Muis (Eds.), *Handbook of educational psychology* (pp. 480-508). Routledge.

Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Martin, B. N. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading*, *22*(2), 152-166. https://doi.org/10.1080/10888438.2017.1398259

Grønmo, L. S., Lindquist, M., Arora, A., & Mullis, I. V. (2015). *TIMSS 2015 mathematics framework*. https://timssandpirls.bc.edu/timss2015/downloads/t15_fw_chap1.pdf

Ismail, Z., Ching, T. Y., & Muda, N. A. (2020). Numeracy competency of year 5 aboriginal students using written and oral tests. *The Mathematics Enthusiast*, *17*(1), 32-62. https://doi.org/10.54870/1551-3440.1479

Jansen, T., & Möller, J. (2022). Teacher judgments in school exams: Influences of students' lower-order-thinking skills on the assessment of students' higher-order-thinking skills. *Teaching and Teacher Education*, *111*, 103616. https://doi.org/10.1016/j.tate.2021.103616

Kline, P. (2015). *A handbook of test construction (psychology revivals): Introduction to psychometric design*. Routledge.

Kosasih, A., Supriyadi, T., Firmansyah, M. I., & Rahminawati, N. (2022). Higher-order thinking skills in primary school. *Journal of Ethnic and Cultural Studies*, *9*(1), 56-76.

Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, *41*(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2

Kūkea Shultz, P., & Englert, K. (2021, July). Cultural validity as foundational to assessment development: An indigenous example. *Frontiers in Education*, *6*, 1-11. https://doi.org/10.3389/feduc.2021.701973

Livingston, S. A., & Kim, S. (2010). Random-groups equating with samples of 50 to 400 test takers. *Journal of Educational Measurement*, *47*(2), 175-185. https://doi.org/10.1111/j.1745-3984.2010.00107.x

Matang, R. A., & Owens, K. (2014). The role of indigenous traditional counting systems in children's development of numerical cognition: Results from a study in Papua New Guinea. *Mathematics Education Research Journal*, *26*, 531-553. https://doi.org/10.1007/s13394-013-0115-2

Misrom, N. B., Muhammad, A., Abdullah, A., Osman, S., Hamzah, M., & Fauzan, A. (2020). Enhancing students' higher-order thinking skills (HOTS) through an inductive reasoning strategy using Geogebra. *International Journal of Emerging Technologies in Learning*, *15*(3), 156-179.

Mullis, I. V., & Martin, M. O. (2017). *TIMSS 2019 assessment frameworks*. TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Mullis, I.V.S., Martin, M.O., Hooper, M. (2017). Measuring changing educational contexts in a changing world: Evolution of the TIMSS and PIRLS questionnaires. In M. Rosén, K. Yang Hansen, & U. Wolff (Eds.), *Cognitive abilities and educational outcomes*. Springer. https://doi.org/10.1007/978-3-319-43473-5_11

Nordby, S. K., Bjerke, A. H., & Mifsud, L. (2022). Computational thinking in the primary mathematics classroom: A systematic review. *Digital Experiences in Mathematics Education*, *8*(1), 27-49. https://doi.org/10.1007/s40751-022-00102-5

Nugraheni, L. P., & Marsigit, M. (2021). Realistic mathematics education: An approach to improve problem solving ability in primary school. *Journal of Education and Learning, 15*(4), 511-518. https://doi.org/10.11591/edulearn.v15i4.19354

Peng, P., Lin, X., Ünal, Z. E., Lee, K., Namkung, J., Chow, J., & Sales, A. (2020). Examining the mutual relations between language and mathematics: A meta-analysis. *Psychological Bulletin*, *146*(7), 595-634. https://psycnet.apa.org/doi/10.1037/bul0000231

Penner, K. (2016). Oral administration as an effective accommodation for students with ADHD. *Literacy Information and Computer Education Journal, 8*(2), 2570-2577.

Philpot, R., Lindquist, M., Mullis, I. V., & Aldrich, C. E. (2021). TIMSS 2023 mathematics framework. In I. V. S. Mullis, M. O. Martin, & M. von Davier (Eds.), *TIMSS 2023 assessment frameworks* (pp. 5-18). TIMSS & PIRLS International Study Centre, Lynch School of Education and Human Development, Boston College and International Association for the Evaluation of Educational Achievement (IEA).

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM*, *52*, 33-44. https://doi.org/10.1007/s11858-019-01118-9

Radmehr, F., & Drake, M. (2018). An assessment-based model for exploring the solving of mathematical problems: Utilizing revised Bloom's taxonomy and facets of metacognition. *Studies in Educational Evaluation*, *59*, 41-51. https://doi.org/10.1016/j.stueduc.2018.02.004

Robertson, S. A., & Graven, M. (2020). Language as an including or excluding factor in mathematics teaching and learning. *Mathematics Education Research Journal*, *32*(1), 77-101. https://doi.org/10.1007/s13394-019-00302-0

Rogers, C. M., Thurlow, M. L., Lazarus, S. S., & Liu, K. K. (2019). *A summary of the research on effects of test accommodations: 2015-2016. (NCEO Report 412).* National Center on Educational Outcomes.

Roschmann, S., Witmer, S. E., & Volker, M. A. (2021). Examining provision and sufficiency of testing accommodations for English learners. *International Journal of Testing*, *21*(1), 32-55. https://doi.org/10.1080/15305058.2021.1884872

Sani, N., & Idris, A. R. (2018). Implementation of Linus programme based on the model of Van Meter and Van Horn. *The Malaysian Online Journal of Educational Sciences*, *1*(2), 25-36.

Shanmugam, S. K. S., Veloo, A., & Md-Ali, R. (2021). Culturally responsive assessment: Assessing mathematics performance of indigenous pupils in Malaysia using trilingual test. *Diaspora, Indigenous, and Minority Education*, *15*(2), 113-136. https://doi.org/10.1080/15595692.2020.1846515

Shanmugam, S. K. S., Veloo, A., & Bin Jusoh, Y. A. (2023). The challenges of virtual learning: An exploratory study among Orang Asli pupils from the mathematics teachers' perspectives. *Diaspora, Indigenous, and Minority Education*, 1-19. https://doi.org/10.1080/15595692.2023.2212823

Shanmugam, S., Veloo, A., & Yusoff, Y. A. B. J. (2024). Examining utility of oral-administered test accommodation in assessing aboriginal pupils' mathematics performance using score comparability. *International Journal of Science and Mathematics Education*, 1-24. https://doi.org/10.1007/s10763-024-10451-9

Spaull, N., & Kotze, J. (2015). Starting behind and staying behind in South Africa: The case of insurmountable learning deficits in mathematics. *International Journal of Educational Development*, *41*, 13-24. https://doi.org/10.1016/j.ijedudev.2015.01.002

Spiel, C., Evans, S. W., & Harrison, J. R. (2019). Does reading standardized tests aloud meet the scientific definition of an accommodation? *Journal of Applied School Psychology*, *35*(4), 380-399. https://doi.org/10.1080/15377903.2019.1601145

Susanto, A., Nengsih, R., Akhirina, T. Y., Nulhakim, A. L., & Awaludin, A. A. R. (2019, March). How to improve students mathematics problem solving by implementing Indonesian realistics mathematics education (IRME) approach. In *Journal of Physics: Conference Series* (*Vol. 1175*, No. 1, p. 012046). IOP Publishing.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1-16. https://doi.org/10.1007/s11423-019-09701-3

Tai, K. W. (2022). Translanguaging as inclusive pedagogical practices in English-medium instruction science and mathematics classrooms for linguistically and culturally diverse students. *Research in Science Education*, *52*(3), 975-1012. https://doi.org/10.1007/s11165-021-10018-6

Tambunan, H. (2019). The effectiveness of the problem-solving strategy and the scientific approach to students' mathematical capabilities in high order thinking skills. *International Electronic Journal of Mathematics Education*, *14*(2), 293-302.

Tan, K. J., Ismail, Z., & Abidin, M. (2018). A comparative analysis on cognitive domain for the Malaysian primary four textbook series. *EURASIA Journal of Mathematics, Science and Technology Education*, *14*(4), 1273-1286. https://doi.org/10.29333/ejmste/82625

Van den Heuvel-Panhuizen, M., & Drijvers, P. (2020). Realistic mathematics education. In S. Lerman (Ed.), *Encyclopedia of mathematics education* (pp. 713-717). Springer.

Veloo, A., Shanmugam, S. K. S., Md-Ali, R., Yusoff, Y. A. J., & Awang-Hashim, R. (2021). Grade five indigenous (Orang Asli) pupil's achievement in bilingual versions of mathematics test. *Journal of Language and Linguistic Studies*, *17*(4), 1863-1872.

Wijns, N., Verschaffel, L., De Smedt, B., & Torbeyns, J. (2021). Associations between repeating patterning, growing patterning, and numerical ability: A longitudinal panel study in 4-to 6-year-olds. *Child Development*, *92*(4), 1354-1368. https://doi.org/10.1111/cdev.13490

Yushau, B., & Omar, M. H. (2015). Mathematics performance and its relation to English language proficiency level of bilingual Arab university students. *Indian Journal of Science and Technology*, *8*(13), 1-15.

Zapalska, A. M., McCarty, M. D., Young-McLear, K., & White, J. (2018). Design of assignments using the 21st century Bloom's revised taxonomy model for development of critical thinking skills. *Problems and Perspectives in Management*, *16*(2), 291-305.

Zhang, D., & Rivera, F. D. (2021). Predetermined accommodations with a standardized testing protocol: Examining two accommodation supports for developing fraction thinking in students with mathematical difficulties. *The Journal of Mathematical Behavior*, *62*, 100861. https://doi.org/10.1016/j.jmathb.2021.100861