A peer reviewed, open-access electronic journal: ISSN 1531-7714

Indigenizing Mathematics Assessment: Developing A Culturally Inclusive Mathematics Test Framework for Indigenous Pupils

S. Kanageswari Suppiah Shanmugam, *School of Education*, *Universiti Utara Malaysia*
Arsaythamby Veloo, *School of Education*, *Universiti Utara Malaysia*
Suheysen Revindran, *Awang Had Salleh Graduate School*, *Universiti Utara Malaysia*

Abstract: Conventional mathematics testing often fails to reflect the diverse cultural backgrounds and lived experiences of Indigenous pupils. While efforts to improve educational access for Indigenous communities have increased, less emphasis has been placed on adapting test development processes to align with Indigenous learners' linguistic backgrounds and cultural context. This study aims to develop a mathematics test framework that considers linguistic proficiency, knowledge systems, and everyday realities of Malaysia's Orang Asli pupils. By adopting a mixed-methods approach, this study was conducted in two phases. In the first phase, a panel consisting of two mathematics education experts and three experienced teachers collaboratively designed 40 mathematics test items grounded in the linguistic and cultural realities of Orang Asli pupils. In the second phase, the test was administered to 109 Grade Four pupils across four Orang Asli primary schools. Item analysis showed acceptable levels of item difficulty and discrimination, and internal consistency was acceptable using Cronbach's alpha. The phases undertaken in this study provided evidence-based support for the development of a test development framework that offers practical guidance for designing culturally inclusive mathematics assessments that are better aligned with the learning needs of Indigenous pupils as an equitable testing practice in diverse educational settings.

Keywords: Culturally Responsive Test, Test Development Framework, Inclusive Testing Practice, Mathematics Assessment, Indigenous Pupils

Introduction

Mathematics assessment plays a central role in the teaching and learning process as it functions not only as a tool to measure pupil understanding, but also to guide instructional decisions and evaluate the effectiveness of teaching strategies (Nortvedt & Buchholtz, 2018). For these assessments to be meaningful, their design must closely align with the intended learning outcomes. Such alignment ensures that test score inferences accurately reflect what pupils are expected to know and can do (Dündar et al., 2016). To this end,

the present study focuses on developing an inclusive mathematics test framework that considers curricular goals, and the diverse cultural and linguistic backgrounds of the pupils. In achieving this, mathematics tests must meet rigorous standards of validity and reliability while minimizing bias. For instance, Abedi (2011) and Kubiszyn and Borich (2024) have emphasized that any test should be sensitive to pupils' cultural and language abilities to ensure equitable measurement of learning. It is therefore important to ensure that these tests are free from the cultural and linguistic biases that may potentially hinder pupils' ability to effectively engage with the test.

Biases in test items can arise from several sources, including complex or unfamiliar language, culturally specific references, or assumptions about pupils' prior knowledge and experiences (Smith et al., 2018). When left unaddressed, such biases distort assessment results, leading to inaccurate interpretations of pupils' mathematical abilities and potentially perpetuating educational inequities among these pupils (Jenkins & Leung, 2019). While these concerns about bias have prompted improvements in test design, most existing frameworks guiding the development of mathematics assessments have seldom addressed the educational realities and learning contexts of diverse learners—especially Indigenous pupils (Klenowski, 2009; Preston & Claypool, 2021). As such, there is a growing demand for frameworks that ground assessment design in the sociocultural realities of the learners they serve, highlighting the need for mathematics assessments that are better aligned with the ways Indigenous pupils acquire and express mathematical understanding when being assessed (Trumbull & Nelson-Barber, 2019).

Despite growing awareness about fairness in assessment, most conventional mathematics assessments continue to be developed within dominant knowledge traditions that reflect specific cultural and linguistic norms. These assessments are typically based on Western models of mathematics education, which prioritize abstract reasoning, formal procedures, and individualistic problem-solving approaches (Dreher et al., 2021; Louie, 2017; Rigney et al., 2020). Such frameworks often assume that pupils share the same background knowledge and cultural reference points, which can marginalize those from different sociocultural contexts. Scholars have long critiqued the dominance of Western paradigms in mathematics education, highlighting how they fail to represent or value the knowledge systems of Indigenous communities (Han, 2019; Milne, 2017). Recent calls for more inclusive practices, such as advocated by Xu and Ball (2022) to bridge Western and Indigenous knowledge systems, or Ball's (2024) framework for *mathematics without borders*, highlight Australian efforts to incorporate Aboriginal and Torres Strait Islander mathematical traditions. Similar initiatives in New Zealand emphasize integrating Māori mathematical concepts and Indigenous languages into curricula, reflecting broader efforts to decolonize education (Allen & Trinick, 2022).

Nonetheless, the orientation of conventional mathematics assessments toward abstract and decontextualized problem-solving continues to marginalize Indigenous pupils whose learning is grounded in lived, experiential knowledge. Aikenhead (2017) and Meaney et al. (2021) argue that these assessments ignore the contextual, hands-on learning experiences that are central to Indigenous ways of knowing. As these tests are designed around Western knowledge frameworks, this practice may prioritize content, language choice and cultural references that are rooted in the mainstream culture, thereby potentially creating barriers for pupils of diverse cultural backgrounds (Dreher et al., 2021). For instance, word problems embedded in urban, Western lifestyles—involving shopping malls, bank transactions, or subway systems—are unlikely to resonate with Indigenous pupils who have never encountered such environments. In contrast, daily activities like fishing, crafting, or foraging in Indigenous communities involve complex mathematical reasoning in estimation, measurement, and spatial awareness. However, such practices are seldom adopted when developing mathematics assessments.

Stoffer (2017) observed the absence of a standardized tool capable of meaningfully assessing Inuit pupils' learning, pointing to a systemic disconnect between assessment content and Indigenous pupils' lived experiences. Similarly, Nelson-Barber and Trumbull (2007) highlighted how biases in most conventional

testing extend beyond item content to include language use, test format, administration, scoring, and interpretation. These systemic mismatches not only hinder the accurate evaluation of Indigenous pupils' abilities but also reinforce educational inequalities by privileging those whose experiences align more closely with mainstream norms (Wilkinson, 2019). These patterns suggest a persistent misalignment between mainstream assessment practices and the diverse ways in which Indigenous learners engage with and demonstrate mathematical understanding, an issue that continues to surface in various educational systems around the world (Aikenhead, 2017; Parra & Trinick, 2018; Revindran et al., 2025).

In the Malaysian context, the situation is further complicated by the linguistic and educational experiences of the Indigenous *Orang Asli* communities. These Indigenous groups often face a significant language transition in school, where the Malay language, which is the official language of instruction and assessment, is a second or even third language for many *Orang Asli* pupils (Abdullah, 2022; Simamora & Saragih, 2019). This language gap presents a major barrier in mathematics assessments, as pupils may struggle not with the mathematical concepts themselves but with the academic language used to present them (Shanmugam et al., 2025; Veloo et al., 2024). The issue is not limited to vocabulary; it includes unfamiliar sentence structures, idiomatic expressions, and implicit cultural assumptions that reduce test accessibility (Revindran et al., 2025). Consequently, mathematics assessments risk measuring language proficiency more than mathematical understanding, exacerbating disparities and misrepresenting what this group of pupils knows. Addressing this requires not only simplifying the language used in these tests but also maintaining fairness and test validity.

Another important element that needs to be considered is the cultural context used in assessments. In this study, cultural context refers to the values, knowledge systems, daily practices, language norms, and socio-educational environments that shape how pupils understand and engage with learning (Ottmann & Jeary, 2016). This consideration is important in establishing cultural validity as Indigenous pupils' understanding and response to test items are shaped by their everyday experiences, values, and ways of knowing (Solano-Flores & Nelson-Barber, 2001). Many Indigenous communities acquire mathematical understanding through real-world experiences, such as navigating environments, crafting tools, or managing communal resources. These activities require mathematical reasoning—estimation, measurement, and pattern recognition—yet are rarely reflected in most conventional mathematics assessments (Kadonsi et al., 2023). This notion is similar among *Orang Asli* pupils who possess rich cultural knowledge and unique approaches to learning that are often overlooked in mainstream educational settings (Renganathan, 2016). Growing up in rural, community-centered environments, *Orang Asli* pupils acquire mathematical knowledge informally through daily activities. By embedding culturally meaningful contexts into assessments, test developers can create opportunities for Indigenous pupils to demonstrate their mathematical abilities authentically (Meaney et al., 2021).

Several existing frameworks have attempted to make assessment more culturally and linguistically responsive. Klenowski (2009) advocates for the incorporation of cultural context into assessment design, while Wiese and Nortvedt (2023) emphasize the importance of language simplification for pupils from diverse linguistic backgrounds. Loewen et al. (2017) also propose equity-oriented approaches that aim to make assessment fairer for marginalized populations. While these frameworks offer valuable insights, they often focus on isolated elements, such as language or cultural context, without integrating them into a holistic model that accounts for the complex realities of Indigenous learners. For instance, simplified language may still carry culturally irrelevant content, while culturally relevant examples may use academic registers that remain inaccessible to Indigenous pupils who may be more proficient in their unique native languages (Abdulrahim & Orosco, 2020; Preston & Claypool, 2021). These limitations suggest the need for a more comprehensive framework that simultaneously addresses linguistic accessibility and cultural context, particularly for *Orang Asli* pupils, whose educational challenges are shaped by multiple intersecting factors.

The Malaysian education system currently lacks a framework that adequately reflects the language use and cultural experience of *Orang Asli* pupils in the design of mathematics assessments (Abdullah, 2022). This gap results in assessments that are not only linguistically inaccessible but also culturally disconnected, thereby failing to capture pupils' true capabilities when being assessed. Although recent curricular reforms and policy initiatives have aimed to improve equity in education within Malaysia (Nordin et al., 2020), there remains little guidance on how to develop mathematics tests that cater to the *Orang Asli* pupils. These limitations underscore the need for a bespoke framework tailored to the specific linguistic, cultural, and educational contexts of *Orang Asli* pupils. Such a framework should address linguistic barriers through clear, simplified language and avoid culturally alienating content by incorporating examples rooted in *Orang Asli* traditions, daily activities, and environmental knowledge (Revindran et al., 2025; Shanmugam et al., 2025). Moreover, it should go beyond surface-level modifications by embedding principles of cultural sustainability and respect for Indigenous epistemologies, which many existing frameworks fail to do.

This study seeks to address this overarching gap by developing a culturally responsive mathematics test framework that is tailored to the specific needs of *Orang Asli* pupils in Malaysia. By embedding accessible language and culturally meaningful problem scenarios for Indigenous pupils (Aikenhead, 2017; Trumbull & Nelson-Barber, 2019), the framework aims to provide teachers with a structured approach for creating mathematics assessments that are aligned with *Orang Asli* pupils' lived experiences and knowledge systems. The study is guided by one overarching research question: How can a culturally inclusive mathematics test framework be developed that addresses the language and cultural context of *Orang Asli* pupils to ensure the validity and fairness of mathematics assessments? This study not only proposes a new framework but also tests its practical viability among *Orang Asli* pupils. The intended outcome is the development of a validated mathematics test framework that caters to the educational needs of the *Orang Asli* pupils in mathematics testing, which can be further adapted for use in other Indigenous pupils' learning environments.

Malaysian Indigenous *Orang Asli*

The *Orang Asli*, also known as the Indigenous community, are among the earliest inhabitants of Peninsular Malaysia. However, they currently represent a minority within the Malaysian population. The *Orang Asli* community comprises three primary tribes—Negrito, Senoi, and Proto-Malay—encompassing a total of eighteen ethnic groups (*Jabatan Kemajuan Orang Asli* (Department of *Orang Asli* Development) [JAKOA], 2022). The Negrito tribe includes six ethnic groups: *Kensiu, Kintaq, Jahai, Mendriq, Bateq,* and *Lanoh*. The Senoi tribe consists of *Temiar, Semai, Jah Hut, Che Wong, Semoq Beri,* and *Mah Meri*. The Proto-Malay tribe includes *Temuan, Jakun, Semelai, Orang Kuala, Orang Seletar*, and *Orang Kanaq*.

The *Orang Asli* ethnic groups exhibit significant linguistic and cultural diversity, which can profoundly influence their educational experiences. In the mainstream Malaysian educational system, *Orang Asli* pupils often face challenges transitioning to the Malay language, which is the primary medium of instruction (Abdullah, 2022; Shanmugam et al., 2024). Culturally, their learning styles are shaped by traditional oral practices, communal knowledge-sharing, and experiential, environment-based learning, which contrast with the more formal, text-centric methods employed in mainstream schooling (Renganathan, 2016; Shah et al., 2020). This interplay between language and culture affects how *Orang Asli* pupils engage with educational content and frequently results in disparities in performance on mathematics tests designed for the mainstream pupil population.

Consequently, there is a need to incorporate support measures that reflect their linguistic proficiencies and cultural experiences in assessment design. For instance, using simple and clear language to improve comprehension and incorporating real-world examples aligned with the lived experiences of *Orang Asli* pupils can make mathematical problems more meaningful. By aligning test development with these linguistic

and cultural dimensions, the resulting assessments can be more inclusive and equitable, offering a more accurate representation of pupils' mathematical abilities.

Universal Design for Assessment

Assessments play a crucial role in evaluating pupils' academic performance and tracking their development. However, concerns remain about the fairness and equity of conventional assessments, which may disadvantage certain groups, such as Indigenous pupils (Forsyth & Evans, 2019). For *Orang Asli* pupils, these disadvantages often arise from culturally irrelevant test content, language barriers, and unfamiliar test formats that may impact how these pupils respond to the test items and subsequently fail to capture their true abilities. To address these concerns, the concept of Universal Design for Assessment (UDA) promotes the development of inclusive, accessible tests that accommodate all pupils regardless of their backgrounds or individual characteristics (Ketterlin-Geller, 2005).

By applying UDA principles, it is possible to design mathematics tests that fairly and accurately assess pupils' mathematical abilities (Nieminen, 2022; Russel et al., 2009). These principles emphasize the use of multiple means of representation, expression, and engagement to ensure that assessments are accessible to all learners (Dolan et al., 2005). In practice, this could involve adapting the language, test structure, and context of test items, integrating culturally relevant context, and offering pupils multiple ways to demonstrate their mathematical understanding when being assessed (Rose et al., 2018; Tai et al., 2022).

For example, mathematics problems can be contextualized using scenarios familiar to pupils' daily lives and cultural experiences, such as community or environmental activities. A word problem could ask pupils to calculate the number of fish caught over several days using data from a traditional fishing practice in their village, or to measure land areas based on a communal farming activity. These types of problems not only integrate relevant cultural and environmental knowledge but also make abstract mathematical concepts more tangible and meaningful (Abrams et al., 2013; Aikenhead, 2017). Additionally, pupils might be allowed to respond through oral explanations, drawings, or alternative formats that reflect their preferred modes of expression. For instance, a pupil could describe their problem-solving process verbally in their native language or illustrate a visual representation of quantities using objects or drawings, rather than relying solely on written responses. This flexibility acknowledges diverse ways of knowing and communicating, which are central to Indigenous learning traditions, and helps ensure that assessments capture pupils' true mathematical understanding rather than their proficiency in written language alone (Reid O'Connor & Norton, 2022).

By embracing UDA, mathematics assessments can eliminate irrelevant variations, such as unfamiliar cultural references or language-based obstacles, and avoid test characteristics that do not reflect actual ability, such as an overreliance on timed, written responses (Haladyna & Downing, 2004). These changes help ensure that assessments focus on the intended mathematical constructs, rather than extraneous factors that may distort results. For instance, replacing abstract or unfamiliar scenarios with culturally meaningful contexts can enhance engagement and improve accessibility for all pupils. More importantly, UDA also serves as a framework for culturally responsive test design, ensuring that assessments reflect the backgrounds and experiences of Indigenous pupils (James, 2018; Trumbull & Nelson-Barber, 2019). Integrating UDA principles into assessment development enables educators to accommodate diverse needs while maintaining accuracy in measuring mathematical knowledge. Ultimately, this approach supports the assessment of Indigenous pupils' mathematical abilities and reinforces broader goals of educational equity and inclusion in testing (Seipel et al., 2022).

Test Development Process

To ensure test validity, various studies have proposed frameworks and guidelines for test development over the years. These processes follow rigorous protocols to ensure that the test effectively measures the intended constructs. For example, Hambleton and Jones (1993) developed an eight-step test development process, which includes: developing test specifications, creating an item pool, conducting a pilot test, compiling norm references (for norm-referenced tests), establishing pass score criteria (for criterion-referenced tests), testing reliability, assessing item validity, and finalizing the test.

Additionally, Ramadhan et al. (2019) proposed a nine-step process, which involves developing test specifications, writing the test, conducting a test review, piloting the test, analyzing the items, revising the test, reconstructing it, administering it, and interpreting the results. However, a more comprehensive and recent version of the test development process was proposed by Lane et al. (2015), consisting of twelve steps. This process is applicable to all types of tests, regardless of mode (paper-and-pencil or computer-based) or test format (selected-response, constructed-response, or performance tasks). The twelve steps include: overall planning, defining content, specifying test requirements, developing items, designing and assembling the test, producing the test, administering the test, scoring responses, determining passing scores, reporting results, item banking, and preparing the technical report.

A widely accepted set of guidelines for test development is provided by the Standards for Educational and Psychological Testing, a joint publication by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (AERA et al., 2014). These standards outline general procedures applicable to most kinds of tests, including specifying the test's purpose, constructing the test specification, developing test items and scoring rubrics, reviewing and validating test items, and evaluating the overall quality and validity of the items. In addition, the International Test Commission (ITC) established guidelines for adapting tests to ensure linguistic, cultural, and conceptual equivalence with the original instrument (ITC, 2017). The ITC framework consists of twenty guidelines grouped into six categories: precondition (construct evaluation), test development (linguistic and cultural adaptation, pilot testing), confirmation (data validation and reliability), administration (test delivery to the target group), score scales and interpretation (results comparison and evaluation), and documentation (reviewing variations and applications).

While these frameworks offer valuable insights, the integration of principled assessment design, particularly evidence-centered design (ECD), has significantly enhanced the test development process by emphasizing the alignment between assessment tasks and the constructs they are intended to measure (Mislevy & Haertel, 2006; Mislevy, 2007). This framework guides test developers from identifying the knowledge and skills to be assessed (domain analysis) to designing tasks and scoring models (conceptual assessment framework) and implementing and interpreting the assessment (assessment delivery). By focusing on the explicit connections between the targeted knowledge, the evidence required, and the tasks designed to elicit this evidence, ECD ensures that assessments are coherent and provide a strong foundation for evaluating students' abilities in complex and diverse educational contexts (Foster & Piacentini, 2023).

The value of principled assessment design goes beyond theoretical coherence; it also addresses practical challenges related to validity and content representation. Hartell and Strimel (2019) critiqued the broad, unitary conception of validity and emphasized the need for targeted efforts that focus on content validity during test construction. By organizing the assessment process around explicit assessment arguments, principled design provides a conceptual foundation for both conventional and innovative assessments. This approach integrates validity throughout the design and delivery process, ensuring that assessments are transparent, reusable, and communicable (Mislevy, 2007). These arguments connect the substantive content of the domain to the operational processes of administration, scoring, and reporting—essential components for developing assessments that are both theoretically sound and practically effective.

While many test development frameworks exist, there remains a need for one that is specifically tailored to the design of mathematics tests for Indigenous pupils, such as the *Orang Asli* pupils in Malaysia. Such a framework must take into account their cultural context, lived experiences, and knowledge systems, with the goal of creating culturally responsive and inclusive assessments that are accessible and could more accurately measure their mathematical performance. Therefore, there is a need for a well-designed framework that could potentially bridge the gap between most conventional test development processes and the unique educational needs of *Orang Asli* pupils, ultimately fostering equitable opportunities for demonstrating their mathematical abilities.

Materials and Methods

This study seeks to develop a comprehensive framework for constructing mathematics tests tailored to *Orang Asli* pupils. To achieve this aim, this study employed both qualitative and quantitative approaches, following systematic procedures for developing a standardized mathematics test. The methodology detailed in this section underscores the validity and reliability of the proposed framework, ensuring its cultural responsiveness and relevance to the target population.

Research Design

This study adopted a mixed-methods research design, incorporating both qualitative and quantitative approaches to develop a culturally responsive mathematics test framework for *Orang Asli* pupils. The qualitative phase involved the systematic development of the test specification table, followed by validation of the mathematics test by two mathematics experts and three experienced mathematics teachers. This process involved identifying the key mathematical concepts, skills, and competencies to be assessed, ensuring alignment with the cultural and educational context of *Orang Asli* pupils.

In the quantitative phase, the constructed mathematics test was administered to *Orang Asli* pupils to evaluate its psychometric properties, focusing on validity and reliability. The pilot testing was aimed at assessing the test's performance in a real-world setting and identify any potential issues with the test items. This integrated approach ensured a rigorous and culturally responsive framework for the test's development.

Participants and Sample

For the qualitative approach, two panels of experts were employed as the main participants of this study for the development and validation of the mathematics test items. The first panel comprised mathematics education experts who are subject-matter experts in the cross-cultural test development process and mathematics education. On the other hand, the second panel consisted of mathematics teachers who are currently teaching in rural *Orang Asli* primary schools within the states of Perak and Pahang in Malaysia. These teachers possess firsthand experience and contextual knowledge of the educational needs, cultural background and test-taking experience of *Orang Asli* pupils (Shanmugam et al., 2024). Their unique perspective allowed them to evaluate the test items not only for mathematical accuracy but also for their cultural relevance, appropriateness, and accessibility for the pupils. Table 1 presents the profile of the participants who provided their insights, opinions, and expertise in developing the framework for constructing the mathematics test.

A pilot test was conducted with a sample of 109 Grade Four *Orang Asli* pupils from four primary schools located in the states of Perak, Kelantan, and Pahang to examine the psychometric properties of the developed mathematics test. To ensure the sample was representative and comparable, information about the *Orang Asli* pupils' language and mathematics proficiency was obtained through enculturation interviews with teachers from the selected schools. These interviews revealed that the students generally shared similar

levels of proficiency in both the Malay language and foundational mathematical concepts, which was critical for meaningful participation in the study.

Table 1. Profile of Participants for Panel of Experts

Participant	Profile
R1	• Nearly two decades of teaching experience as a mathematics educator. • Currently holds the position of senior lecturer at a public university. • Extensive knowledge and proficiency in test development, psychometrics, and evaluation. • Focuses on educational measurement, particularly in assessing the reliability and validity of cross-cultural tests. • Specialized in creating mathematics tests specifically designed for *Orang Asli* pupils, with six years of experience in this area of research.
R2	• Possesses an extensive 30-year background as a mathematics educator. • Currently holds the position of professor at a public university. • Acquired specialized expertise in the domains of psychometrics, testing and evaluation, as well as mathematics education. • Has a notable 20-year track record in the development of mathematics tests.
MT1	• Holds the position of Assistant Headmaster overseeing academic affairs. • Brings 29 years of teaching experience as a mathematics teacher, including four years spent in an *Orang Asli* primary school. • Achieved a Bachelor of Education (B Ed.) degree with a specialization in mathematics.
MT2	• Currently serving as the Head of the mathematics department at a primary school. • Has 17 years of valuable teaching experience as a mathematics teacher. • Presently teaching mathematics to Grade Two, Four, Five, and Six *Orang Asli* pupils. • Holds a Bachelor of Education (B Ed.) degree with a specialization in mathematics.
MT3	• Has close to 10 years of teaching experience as a mathematics teacher. • Holds a Bachelor of Education (B Ed.) degree with specialization in special education and Malay language (*Bahasa Melayu*).

Moreover, these students were selected due to their stronger familiarity with formal school culture and relatively consistent instructional experiences across the four selected primary schools. This was significant because the mathematics curriculum implemented in these schools aligns with the national curriculum used by all primary schools in Malaysia, ensuring uniformity in educational objectives for mathematics education. This consistency enabled the study to assess mathematics performance within a standardized framework, while minimizing variability due to differing instructional methods or curricular content.

Instrument

Through the content and cultural validation process carried out by the panel of experts, a single written mathematics test consisting of 40 items was developed from the Level 1 mathematics curriculum, which consisted of eight topics; whole numbers up to 100 (10 items), basic operations (12 items), fractions (3 items), money (5 items), time (3 items), measurement (2 items), space (2 items), and data management (3 items). The test items developed are also categorized according to the cognitive domains of knowing (14 items), applying (16 items) and reasoning (10 items).

Procedure

The test development process for this study was guided by the Standards for Educational and Psychological Testing (AERA et al., 2014) and adapted to develop a mathematics test suitable for *Orang Asli* pupils. The process was implemented in six interrelated phases: defining the test purpose and specifications, developing and adapting test items, conducting expert and teacher panel reviews, validating the items through a focus group discussion, conducting psychometric analysis, and finalizing the item pool. This multistage process was designed to ensure content alignment with the mathematics curriculum, cultural and linguistic relevance for the pupils, and to ensure the validity of the mathematics test items.

The first phase involved clarifying the purpose of the test and determining the constructs to be measured. The mathematics test was designed to assess *Orang Asli* pupils' proficiency in number operations and word problem-solving through items that were deliberately constructed to be linguistically accessible and culturally relevant, reflecting an indigenised and context-responsive approach to assessment design. The overarching aim was to evaluate these competencies while minimising construct-irrelevant variance attributable to language barriers or culturally unfamiliar content, thereby enhancing the validity and equity of the test outcomes. In order to embed these design principles into the assessment framework, a comprehensive test specification (test blueprint) was prepared. This blueprint defined the content domains, item formats, and cognitive levels to be assessed, and was anchored in the Malaysian national Level 1 mathematics curriculum for the relevant topics, as described in the earlier section. In doing so, it ensured that the test remained curriculum-aligned while incorporating the linguistic and cultural considerations essential for a fair and valid assessment of *Orang Asli* pupils.

A pool of 68 draft test items was constructed and adapted by a panel of two mathematics education experts who were experienced in test development and Indigenous education. This panel formed the first expert group. Each item was reviewed independently by the two mathematics education experts based on four guiding criteria: (1) alignment with the intended content domain, (2) correspondence with the cognitive demand (knowing, applying, or reasoning), (3) linguistic accessibility for *Orang Asli* pupils, and (4) cultural context grounded in pupils' lived experiences. The reviewers provided qualitative feedback for each item, focusing on the appropriateness of the mathematical content, clarity of sentence structure, familiarity of vocabulary, and cultural resonance of the context or scenario presented. Items that were found to contain unclear language, culturally unfamiliar examples, or mismatches between the task and the intended construct were flagged for revision.

In cases where the two reviewers offered differing interpretations of an item's suitability, such as whether a word problem scenario would be easily understood by pupils, the items were brought forward for collaborative discussion. These discussions served to clarify each expert's rationale, identify the root of disagreement, and generate modifications that aligned with both content goals and cultural considerations. The aim of these iterative discussions was to reach consensus through dialogue, whereby both reviewers agreed that the revised item adequately reflected the intended construct, used linguistically appropriate phrasing, and was culturally inclusive for *Orang Asli* pupils. Items for which agreement could not be reached were either revised further and revisited later or excluded from the item pool.

After incorporating initial expert feedback, the adapted item pool was reviewed by a second panel consisting of three mathematics teachers who had direct teaching experience with *Orang Asli* pupils in rural schools. These teacher-practitioners assessed each test item using the same four qualitative criteria: content relevance to the mathematics curriculum, cognitive demand, linguistic clarity, and cultural familiarity. The review process was conducted qualitatively, with teachers providing detailed written comments and suggestions for each item. Their feedback emphasized the practical challenges faced by *Orang Asli* pupils in classroom settings, such as unfamiliar vocabulary, confusing sentence constructions, or problem scenarios that did not align with the pupils' lived experiences.

Items that raised concerns were identified for further discussion and revision. In instances where teacher feedback diverged, such as when one teacher viewed an item as accessible while another flagged potential misunderstanding, the discrepancies were documented and analyzed. These disagreements were typically related to perceived differences in pupil exposure to certain terms or test-taking formats. When consensus could not be reached through written feedback, items were then brought forward for discussion in the subsequent focus group session involving both panels. The goal throughout was to ensure that each item was both instructionally appropriate and linguistically comprehensible for *Orang Asli* pupils, based on the collective insights of experienced classroom teachers.

To consolidate the evaluation process and integrate the distinct perspectives of both panels, a final focus group discussion was conducted with all five panel members, the two mathematics education experts and the three teacher-practitioners. This joint validation session was conducted online and followed a semi-structured format, guided by a predefined discussion protocol. The discussion was designed to validate the content, language, and cultural appropriateness of all test items by considering the earlier independent reviews. The focus group allowed the participants to revisit each item collaboratively, compare feedback, and deliberate on the strengths and weaknesses of specific wordings, representations, and contexts. The discussion protocol included guiding questions such as: Does the item accurately reflect the mathematical construct being measured? Are the instructions and language used accessible to Grade 4 *Orang Asli* pupils? Does the item context reflect lived experiences that are familiar and culturally relevant? Are there any unintended biases or ambiguities that could disadvantage the pupils?

This focus group was critical in refining the test for both cultural and content validity. During the session, consensus was reached through iterative deliberation, often involving suggestions for simpler phrasing, replacement of culturally unfamiliar references, or adjustment of item format. In instances where agreement could not be reached initially, the item was provisionally revised, and the revised version was re-reviewed asynchronously by the same participants until final consensus was documented. Items that continued to elicit concern or ambiguity were either substantially restructured or excluded from the final item bank. This triangulated process ensured the selection of 40 test items that were included in the final version of the mathematics test, which was subsequently administered to the *Orang Asli* pupils.

Data Analysis

In this study, the data analysis was carried out via item analysis on the validated test items. Item analysis is a statistical technique used to evaluate the performance of individual test items to ensure the validity and reliability of the test (Fatimah et al., 2020). The specific item analysis techniques employed in this study included the item difficulty index, item discrimination index, and Cronbach's alpha coefficient.

The item difficulty index (p) is a measure of how easy or difficult an item is for the pupils. It is calculated by dividing the number of participants who answered the item correctly by the total number of participants (Boopathiraj & Chellamani, 2013). A high item difficulty index indicates that the item was easy for most participants, while a low index suggests that the item was challenging. A general rule for an acceptable item difficulty range is between 0.3 and 0.7. Items are deemed too difficult if $p < 0.30$, while items are considered easy when $p > 0.70$, whereby these items may require further revision or modification.

The item discrimination index (D), also referred to as the point-biserial correlation, is used to determine how well an item distinguishes between high- and low-performing pupils (Quaigrain & Arhin, 2017). It measures the correlation between participants' total scores on the test and their scores on a particular item. A high positive item discrimination index indicates that the item effectively differentiates between participants with higher and lower overall test scores, suggesting that the item is valid and contributes to the measurement of the targeted construct. Conversely, a low or negative index suggests poor discrimination, indicating that the item may not be adequately measuring the construct of interest (Boopathiraj &

Chellamani, 2013). Therefore, items with a D > 0.30 are used generally to indicate acceptable discrimination between high- and low-performing pupils, while items with D < 0.30 may require revisions or modifications (Quaigrain & Arhin, 2017). This attests to the validity of the items to be administered to the pupils in assessing their mathematics performance.

Furthermore, Cronbach's alpha coefficient, a measure of internal consistency, was employed to assess the reliability of the mathematics test. This coefficient indicates the extent to which the test items are measuring the same underlying construct (Bonett & Wright, 2015). It is computed by analyzing the intercorrelations among the items. A high Cronbach's alpha coefficient (typically > 0.70) suggests a strong internal consistency, indicating that the items in the mathematics test are reliable measures of the same construct. Conversely, a low coefficient may indicate the need for item revisions or removal to improve the overall reliability of the test.

Test Development Process

This section presents the findings of this study on the research process undertaken in structuring the developmental framework of a mathematics test designed specifically for *Orang Asli* pupils. The findings discuss each phase in detail, which were carried out sequentially during constructing the mathematics test of the *Orang Asli* pupils. A summary of the framework denoting the phases that were undertaken is presented at the end of this section.

Designing Mathematics Test Blueprint

In the initial phase of developing a mathematics test, the elements which are required for inclusion and assembly in the test blueprint were identified. Mathematics topics and key concepts based on the mathematics curriculum were initially examined and reviewed. The development of the mathematics test included test items that measure the understanding of the pupils on all the topics covered in the mathematics curriculum. Moreover, it was important to establish test purpose and specify the categories of items (i.e., mathematics computation, mathematics word problem, cognitive domains, etc.) and item format (multiple choice, true/false, short answer, constructed response, etc.) that cater to *Orang Asli* pupils.

The **test blueprint** for the mathematics test items was developed in accordance with the Level One Mathematics Curriculum and Assessment Standard. It encompassed all topics covered in the mathematics curriculum, such as whole numbers up to 100, basic operations, fractions, money, time, measurement, space, and data management. These topics were represented in the blueprint to ensure comprehensive coverage of the curriculum, and the corresponding mathematics test items were designed to evaluate pupils' performance across these areas. Additionally, the test items were categorized based on the **cognitive domains** of knowing, applying, and reasoning, as outlined in the Trends in International Mathematics and Science Study (TIMSS) framework. This categorization provided a structured approach for evaluating the hierarchical cognitive skills required in mathematics education.

The process of developing the test items was systematically guided by the TIMSS cognitive domains, which emphasize different levels of cognitive engagement (Mullis & Martin, 2017). The knowing domain was operationalized to measure pupils' ability to recall factual knowledge and understand basic mathematical principles, while the applying domain assessed their capacity to apply knowledge and mathematical procedures to solve real-world problems. The reasoning domain focused on higher-order cognitive abilities, including critical thinking, problem-solving, and logical reasoning. To ensure relevance and inclusivity, the test items were meticulously adapted to align with the cognitive demands of these domains while integrating culturally and linguistically familiar contexts. This approach enhanced engagement and understanding for

Orang Asli pupils, allowing them to relate to the test content while still adhering to the standardized framework.

By adopting this structured approach, the test development process adhered to inclusive practices that incorporated the cultural context of *Orang Asli* pupils within the national curriculum standards. This alignment was vital, as the national mathematics syllabus is based on the TIMSS framework (Tan et al., 2018) and applied uniformly across all Malaysian pupils. Although *Orang Asli* pupils bring unique cultural and epistemological perspectives to learning, aligning the test development process with the TIMSS framework was necessary to standardize assessment practices and learning outcomes.

Adapting Mathematics Test Items for Item Bank

The review process for the mathematics test items involved two mathematics experts familiar with the educational experiences, cultural backgrounds, and linguistic contexts of *Orang Asli* pupils. These experts provided valuable insights to ensure that the test items were not only aligned with the national curriculum but also culturally and linguistically appropriate. To support this task, a targeted training program was conducted for the experts. This training emphasized the importance of cultural sensitivity in assessment development, focusing on recognizing and addressing potential biases that might disadvantage *Orang Asli* pupils. The training covered strategies for adapting test items to reflect familiar contexts and language for the pupils, such as replacing abstract or unrelated examples with culturally meaningful scenarios or revising terminology to better align with the pupils' linguistic experiences. For instance, word problems involving urban settings or activities uncommon in *Orang Asli* communities were modified to include relatable situations such as rural activities or community-specific practices.

A review process was conducted to select the items for the development of the mathematics test from mathematics textbooks and sample questions from other sources such as practice books and teaching materials that are based on the national curriculum. A total of 68 draft mathematics test items were chosen from these resources, where these test items were added to the item bank for the mathematics test. The test items were adapted to suit the cultural, linguistic, and test-taking experiences of the Indigenous pupils to ensure cultural validity of the mathematics test. Mathematics experts reviewed the test items to identify cultural biases that may be embedded within the test items, such as language and terminology unfamiliar to the *Orang Asli* pupils due to different linguistic backgrounds, as well as unrelated cultural contexts and examples not typically encountered by the pupils in their daily lives. Therefore, to address this, mathematics test items that consisted of diagrams, contexts, or language that may be unfamiliar or irrelevant to the pupils were reviewed and adapted by mathematics experts to ensure that the pupils could better understand the test items and did not introduce unnecessary cognitive barriers. Through this review process, the test items were refined to align with the cognitive demands of the national curriculum while ensuring fairness and accessibility for the *Orang Asli* pupils.

Furthermore, the test items selected by the experts were aligned with the curriculum's learning standards and performance benchmarks, ensuring consistency with national educational objectives. This process helped maintain the rigor of the test items according to the mathematics standards set by the Malaysian education system, while ensuring that the test provided an accurate measure of pupils' mastery of mathematical skills and concepts. This alignment was crucial to maintain consistency in educational objectives and to ensure that the test items adhered to the content and cognitive demands expected of all primary level pupils. A scoring rubric was developed for the test items developed during this phase. As the test items developed only have a single correct answer, the items were to be scored dichotomously whereas each item will be scored either correct or incorrect. If the pupil's response matches the expected correct answer, it is considered correct, and the pupil will receive one mark for that item. On the other hand, if the pupil's response is incorrect, the pupil will receive no points or a zero score for that item.

Reviewing and Revising Mathematics Test Items for Content Validation

The focus group discussion served as the final and most critical phase in the iterative development and validation of the mathematics test items. Conducted online, this session brought together the two mathematics education experts and three mathematics teachers—both panels having previously reviewed the test items independently. The purpose of the focus group was to consolidate expert feedback, resolve outstanding disagreements, and collaboratively refine the test items to enhance their content, linguistic, and cultural validity for the *Orang Asli* pupils. The discussion was semi-structured and guided by a protocol that reflected four central validation criteria used throughout the development process: (1) alignment with the Level 1 curriculum content domains, (2) correspondence with the targeted cognitive domain (knowing, applying, or reasoning), (3) linguistic clarity and appropriateness for *Orang Asli* pupils, and (4) cultural resonance grounded in pupils' lived experiences and cultural context. The panel reviewed all 68 items collectively, focusing on how each item addressed or failed to meet these criteria based on insights from past classroom practices, test development theory, and Indigenous pedagogical considerations.

During the session, each test item was examined in relation to its intended construct and cognitive demand. Items that had previously received divergent feedback were discussed in greater detail, especially those where interpretations differed across panels. For example, mathematics education experts occasionally prioritized precision of terminology or fidelity to curriculum phrasing, while the mathematics teachers emphasized the importance of accessibility, simplicity, and classroom realities, particularly for pupils with limited exposure to formal conventional assessments or instructional language outside their everyday speech.

One recurring theme during the discussion was the role of **sentence structure** and **vocabulary** in shaping pupils' ability to understand and respond accurately to test prompts. Items written with multi-clause instructions or academic vocabulary were flagged as potentially inaccessible, even if they were mathematically sound. As the *Orang Asli* pupils typically face a mismatch between the language of assessment and their everyday language, complex sentence structure and academic vocabulary can potentially obscure their mathematical understanding which may make the test inaccessible. In such cases, consensus was reached to revise the items by simplifying sentence constructions and substituting difficult or formal terms with more direct, familiar alternatives. For instance, a test item initially phrased as "*Calculate the total quantity by aggregating the values below*" was reformulated as "*Add the numbers below*," ensuring that the linguistic demand did not obscure the mathematical task.

Another major concern raised by the teacher-practitioners was the **format** and **visual layout** of test items. It was noted that *Orang Asli* pupils are not yet familiar with most testing formats used in assessments, unlike the mainstream pupil population. These pupils might struggle with locating where to write their answers, especially in tests that lacked clear visual scaffolds. In response, the panel agreed that each item should include a standardized answer space, such as a clearly marked box or line, to guide pupils. Additionally, the use of multiple-choice questions (MCQs) was favored over open-ended formats wherever appropriate, as MCQs reduced the need for written expression and minimized performance anxiety related to writing or spelling.

In terms of cultural representation, the panel engaged in rich discussions about the authenticity and relevance of the items' context. Scenarios that included **unfamiliar names**, **urban settings**, or **tasks unrelated to the pupils' everyday experiences** (e.g., using an escalator, shopping at a mall) were critically examined. These items were revised to reflect community-specific experiences, such as using traditional foods, rivers, village events, or common trading practices as part of the mathematical scenario. In one case, an item referencing "*a factory worker's shift schedule*" was replaced with a context involving "*helping parents harvest forest fruits*," a more relatable activity for pupils in the target communities. Importantly, when disagreements persisted, such as whether a certain term would be recognized across different *Orang Asli* communities, the item was provisionally revised based on majority consensus and then re-evaluated asynchronously after the

session. This post-discussion step ensured that consensus was not forced during the live session but was based on thoughtful consideration. Items that still failed to meet the validation criteria after further revision were removed entirely from the final test bank.

Out of the initial 68 items, 36 were accepted without modification, 29 were revised based on focus group recommendations, and three were excluded due to persistent concerns over construct validity, linguistic complexity, or cultural appropriateness. From this pool, a total of 40 test items were selected to comprise the final version of the mathematics test to be administered to the *Orang Asli* pupils. Selection was based on consensus among panel members and guided by the relevance and importance of the mathematical content as stipulated in the curriculum. Priority was given to foundational knowledge and skills, particularly those in number operations and word problem-solving, which are essential for progression to higher grade levels. Ultimately, this focus group discussion functioned as a validation convergence point, where expert consensus was not merely procedural but grounded in meaningful reflection on how assessment design can include, rather than marginalize, Indigenous pupils.

Revising Mathematics Test Blueprint

Based on the comments and suggestions made by the panel of experts during the content validation process, the mathematics test items were then revised according to the respective topics and cognitive domains. During this stage, the mathematics experts selected 40 mathematics test items for the final development of the mathematics test. The 40 mathematics test items were selected based on their suitability to fairly assess the mathematical knowledge and skills of the Indigenous pupils by the mathematics experts. Subsequently, the test blueprint was redesigned to include the changes made to the selected test items based on the relevant mathematics topics and cognitive domains, as presented in Table 2.

Table 2. Revised Test Blueprint for Final Mathematics Test Items

Topics	**Cognitive Domains**		
	Knowing	**Applying**	**Reasoning**
Whole Numbers up to 100	1, 2, 3, 4, 5, 6, 7	15, 16	31
Basic Operations	8, 9, 10, 11, 12, 13	17, 18, 19, 20, 21, 22	
Fractions		23, 24, 25	
Money		26, 27	32, 33, 34
Time		28, 29, 30	
Measurement			35, 36
Space	14		37
Data Management			38, 39, 40
Total	14	16	10

Constructing Final Mathematics Test

The mathematics experts selected a total of 40 mathematics test items that were validated based on the revised test blueprint, which was then used to construct the mathematics test. For the administration of the

mathematics test to the Indigenous pupils, the mathematics experts deduced that two test items should be included for each page. The negative space between the two test items were also kept constant throughout the development of the mathematics test. A cover page was created to provide space for pupils' names, class, and gender, with the test instructions also developed for the test. This was provided to gather the necessary information about the pupils and provide instructions to the pupils on the mathematics test.

Psychometric Validation of Mathematics Test Through Pilot Test

The test was administered to 109 Grade 4 *Orang Asli* pupils from four primary schools in the states of Perak, Kelantan, and Pahang, lasting approximately one hour. **Item analysis** was subsequently carried out to determine the construct and content validity, and reliability of the mathematics test items. Therefore, the item difficulty index (p), item discrimination index (D) and Cronbach's alpha coefficient were subsequently computed.

For the item difficulty index computed for the 40 test items, 36 items were found to be within the acceptable item difficulty range ($p = 0.30 – 0.70$). Four items (Items 3, 4, 7, and 17) were found to be easy items ($p > 0.70$). On the other hand, the item discrimination index computed for the 40 test items showed that all the items were able to discriminate well between the higher performing pupils and lower-performing pupils ($D > 0.30$). This indicates that the high item discrimination index obtained for the test items in the mathematics test increases the test's validity as it can effectively measure the intended construct.

Overall, the average difficulty index obtained for the mathematics test was $p = 0.58$, while the average discrimination index obtained was $D = 0.42$. Furthermore, the Cronbach's alpha coefficient computed for the mathematics test is $\alpha = 0.76$. The findings indicate that the Cronbach's alpha value demonstrates that the test has a high internal consistency and is reliable in measuring the mathematics performance of the *Orang Asli* pupils. Table 3 presents the item difficulty index and item discrimination index values for the 40 test items developed in the mathematics test.

Based on the item analysis statistics that have been carried out for the mathematics test, the findings show that the test items developed for the test were valid and can be used to assess the pupils' mathematics performance. For test items that require revision, additional review for content and cultural validation was carried out for these items to ensure a more valid and accurate measurement of the construct intended to be measured.

Re-establishing Content and Cultural Validity

The process of **re-establishing content and cultural validity** for items that require revision and modification after item analysis crucial to ensure that the final test items can validly and accurately measure the mathematics performance of *Orang Asli* pupils. Throughout this process, collaboration and consultation with the mathematics teacher-experts and pupils' performance in the test were taken into consideration as their input, feedback and responses provide valuable insights into the cultural and educational needs of the pupils. This process is important as the rigorous reviews, revisions and modifications made by the panel of mathematics experts and mathematics teacher could help develop the final mathematics test that could accurately measure the mathematics performance of the *Orang Asli* pupils.

This item underwent simplification in the language and test design to cater to the linguistic and test-taking experience of the *Orang Asli* pupils. For instance, the number of words and sentences were reduced through the item development process. This ensured that the context of the test item could be easily understood by the pupils to allow them to solve the problem. Other than that, the answer format of the items was revised. The number of answers for the test item was reduced from three answers to two answers in the final rendition of the mathematics test. This is to mitigate any confusion that may arise due to the language and literacy barriers that may impact the pupils' performance when devising the correct conceptual knowledge

when solving this item. Table 4 presents an example of a mathematics item that underwent revisions during the phases of developing and validating the mathematics test.

Table 3. Item Difficulty and Discrimination Index for the Mathematics Test Items

Items	p	D	Alpha when deleted
1	0.39	0.62	0.66
2	0.46	0.37	0.69
3	0.92	0.35	0.71
4	0.81	0.34	0.71
5	0.66	0.36	0.70
6	0.31	0.56	0.82
7	0.71	0.40	0.69
8	0.64	0.43	0.69
9	0.67	0.46	0.83
10	0.70	0.35	0.72
11	0.46	0.56	0.67
12	0.70	0.32	0.71
13	0.48	0.42	0.68
14	0.65	0.38	0.84
15	0.54	0.35	0.84
16	0.35	0.40	0.69
17	0.71	0.55	0.67
18	0.63	0.35	0.70
19	0.51	0.45	0.83
20	0.50	0.58	0.82
21	0.64	0.37	0.83
22	0.34	0.51	0.82
23	0.33	0.33	0.74
24	0.39	0.37	0.84
25	0.30	0.30	0.60
26	0.65	0.37	0.69
27	0.37	0.35	0.83
28	0.40	0.44	0.68
29	0.34	0.54	0.82
30	0.38	0.60	0.82
31	0.48	0.50	0.82
32	0.33	0.38	0.72
33	0.48	0.39	0.83
34	0.62	0.51	0.82
35	0.70	0.53	0.82

36	0.63	0.36	0.83
37	0.33	0.30	0.71
38	0.51	0.32	0.72
39	0.48	0.47	0.83
40	0.38	0.42	0.83
Average	**0.52**	**0.42**	**0.76**

Table 4. Validation Process for Mathematics Test Item during Test Development Process

1. Mathematics Raw Test Item

RM 10

23 Mac 2022
My father gave me RM10.

23 Mac 2022
I bought a book for RM 1.

26 Mac 2022
My mother gave me RM5.

Complete the savings and expenses table by date below.

Date	Saving	Expense
23 March 2022	**RM 10**	
26 March 2022		

2. Test Adaptation by Mathematics Experts

Complete the savings and expenses table by date below.

Tajul's father gave Tajul pocket money of RM 10. Tajul has spent RM 8. How much money is left?

	Pocket Money	Expense	Balance
Father			

3. Revised Test Item from Content Validation by Mathematics Teacher-Practitioners

Ali's father gave him RM10, and his mother gave him RM5. Ali has spent RM8. How much is left? Complete the savings and expenses table below.

	Pocket Money	Expense	Balance
Father			
Mother			

4. Content and Cultural Validity after Item Analysis

Tajul has pocket money of RM 10. Tajul has spent RM 8. Complete the savings and expenditure table.

Pocket Money	Expense

The iterative modifications made to a sample test item throughout the development process reflect a systematic effort to enhance its accessibility for *Orang Asli* pupils through linguistic, cultural, and test format adaptations. Although the context of the test item underwent linguistic modification, such as simplifying the language from "Tajul's father gave him RM10, and his mother gave him RM5. Ali has spent RM8. How much is left?" to "Tajul has pocket money of RM10. Tajul has spent RM8". These changes did not alter the underlying construct being measured, which is the ability to calculate the balance of money. The linguistic adjustments were made to enhance clarity and accessibility for the *Orang Asli* pupils, ensuring that complex phrasing did not obscure their ability to demonstrate the intended mathematical skill.

Furthermore, the context was adapted to align with culturally relevant experiences, focusing on pocket money, a scenario more relatable to the daily lives of *Orang Asli* pupils. This approach ensured that the item not only measured the intended construct but also presented a context that resonated with the pupils' lived experiences, thereby fostering greater engagement and reducing the potential for cultural bias. Additionally, the test format was refined to align with the pupils' test-taking experiences, ensuring consistency with their prior learning experiences and minimizing potential confusion. Together, these modifications demonstrate a comprehensive approach to creating test items that are linguistically clear, culturally responsive, and structurally inclusive, fostering a more equitable assessment for the target population.

Finalizing Mathematics Test Blueprint and Mathematics Test

The final step in developing the mathematics test involved additional validation and approval from mathematics teachers on the finalized design and selection of the 40 test items. This stage was essential to ensure that the overall test structure, item sequence, language use, and context used in the problem statements of the test items were appropriate for *Orang Asli* pupils. Rather than modifying the mathematical content itself, the test development process focused on enhancing linguistic accessibility and cultural contextual relevance. The goal was to construct a mathematics test that accurately assessed *Orang Asli* pupils' content knowledge without introducing construct-irrelevant challenges due to unfamiliar vocabulary and language use, syntactic complexity, or abstract problem scenarios disconnected from the pupils' real-life experiences. As such, this approach emphasized *how* mathematics is presented rather than *what* mathematics is taught or tested.

To operationalize this focus, this test development process incorporated several key phases, ranging from defining the test purpose and specifications to iterative expert review and collaborative validation, which formed the procedural backbone of the developed mathematics test. These phases serve as a guide for constructing inclusive mathematics tests that accommodate the educational realities of the *Orang Asli* pupils. Table 5 presents the summary that outlines the key phases, purpose, overview of procedures, and key stakeholders carried out for each phase involved in the development and validation of the mathematics test for the Indigenous pupils.

Framework for Developing Culturally Responsive Mathematics Assessments

Drawing on the findings and summary of the phases for developing the mathematics test carried out in this study, a **generalized test development framework** was established to guide the construction of culturally responsive mathematics assessments for Indigenous pupils. This framework synthesizes eight interrelated phases, beginning with clarifying the test purpose and designing a test blueprint aligned with curriculum standards. It continues with adapting and reviewing test items for cultural and linguistic relevance, revising both item content and blueprint structure, and constructing a finalized mathematics test tailored to the learners' lived experiences. Psychometric validation and iterative revisions ensure that the assessment is not only technically sound but also contextually appropriate. The final phase involves consolidating expert feedback and teacher validation to finalize a test that reflects both instructional goals and cultural inclusivity. This process-oriented model emphasizes alignment between national curriculum

Table 5. Summary of the Test Development Process of Mathematics Test for Indigenous Pupils

Phase	Purpose	Procedural Overview	Stakeholders Involved
Designing mathematics test blueprint	Establish test structure, format and topics, aligned to curriculum and *Orang Asli* pupils' context	Identified test objectives, item types, cognitive domains (TIMSS), and curriculum topics	Mathematics education experts, curriculum guidelines, TIMSS framework
Adapting mathematics test items for item bank	Develop culturally and linguistically relevant draft items	Selected and adapted 68 items from curriculum-based sources; replaced unfamiliar contexts and vocabulary	Mathematics education experts
Reviewing and revising mathematics test items for content validation	Validate content, language, and cultural relevance	Expert and teacher panels reviewed items using 4 criteria; provided qualitative feedback; disagreements resolved through discussion or asynchronous review	Mathematics education experts and teachers
Revising mathematics test blueprint	Update blueprint to reflect validated items	Integrated validated 40 items into revised blueprint by topic and cognitive domain	Mathematics education experts and teachers
Constructing final mathematics test	Assemble test layout for administration	Formatted test pages, added pupil information cover page and instructions, and standardized answer spaces	Mathematics education experts and teachers (for feedback)
Psychometric validation of mathematics test	Establish reliability and statistical validity	Administered test to 109 Indigenous pupils: Computed item difficulty, discrimination, Cronbach's alpha	Indigenous pupils
Re-establishing content and cultural validity	Refine items flagged post-analysis for fairness and clarity	Reviewed Indigenous pupil responses; modified confusing items; simplified language, revised format (e.g., reduced options)	Mathematics education experts, teacher feedback, and Indigenous pupil response patterns
Finalizing mathematics test blueprint and mathematics test	Approve and document finalized test framework	Reviewed final items; confirmed alignment with pupil experiences, test-taking capacity, and fairness	Mathematics education experts and mathematics teachers

expectations and Indigenous learners' realities, offering practical guidance for equitable assessment design. Figure 1 illustrates this test development framework on constructing mathematics assessments for Indigenous pupils.

Test Development Process

The findings of this study emphasize the significance of a comprehensive and culturally responsive framework for developing mathematics test items tailored to the unique needs of *Orang Asli* pupils. The iterative process employed in the development of the test items underscores the value of collaboration among mathematics experts and teacher-experts familiar with the cultural context and educational experiences of *Orang Asli* pupils in developing mathematics tests that are valid, reliable, inclusive, and equitable. This discussion integrates insights on the effectiveness of the framework to address cultural and linguistic barriers, the psychometric validation of the test developed, and the potential impact of sustaining this framework in the long term to drive systemic change.

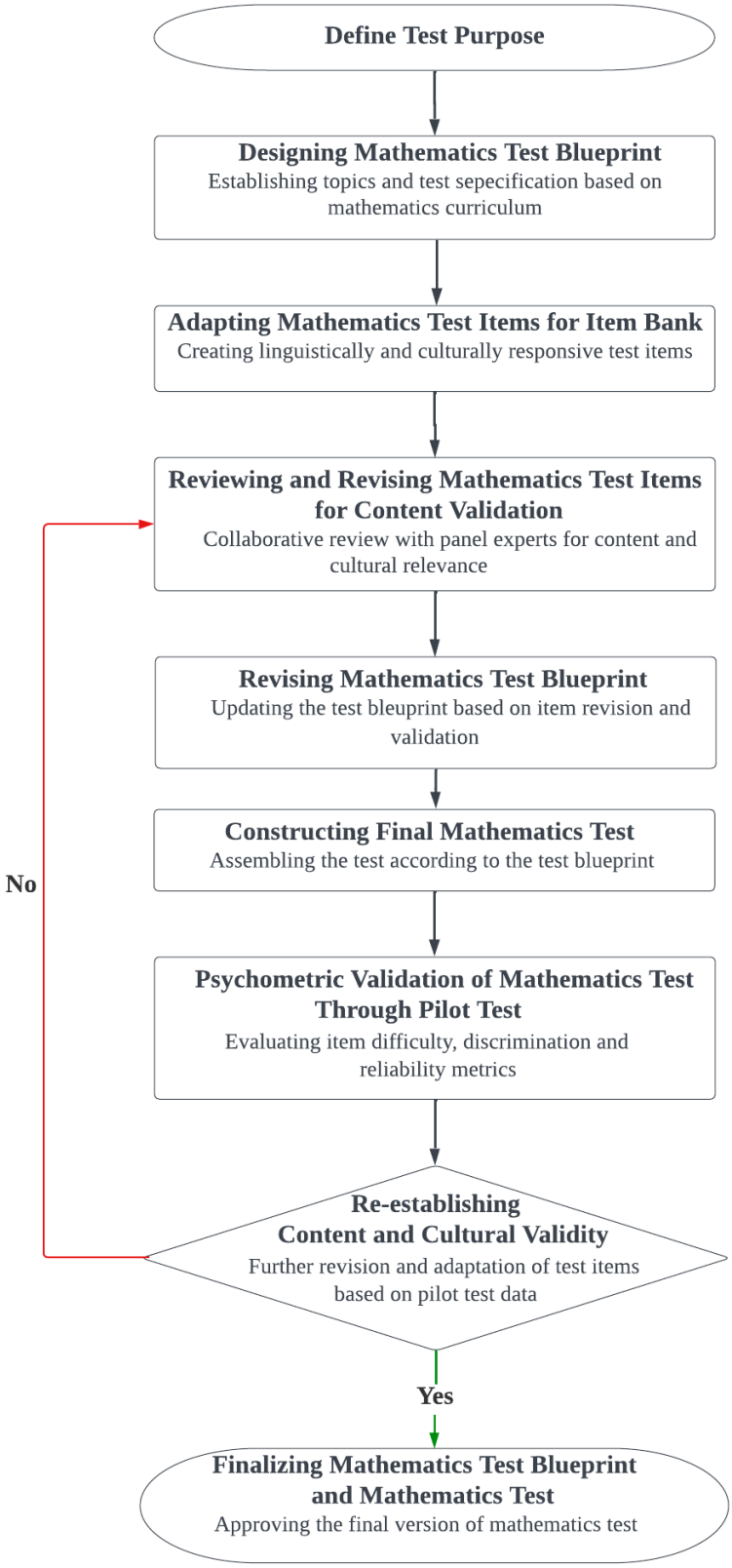
Addressing Linguistic and Cultural Barriers

This study attempted to develop a framework to mitigate the linguistic and cultural barriers that *Orang Asli* pupils face in standardized assessments. Traditional mathematics tests often reflect dominant cultural norms, language, and contexts, which inadvertently disadvantage pupils from minority communities such as the *Orang Asli*. Research indicates that culturally responsive assessments help bridge the gap between pupils' lived experiences and the educational content being assessed (Tai et al., 2022). In this study, the inclusion of problem-solving scenarios drawn from *Orang Asli* pupils' daily lives could potentially improve their engagement and comprehension when delivering tests that are culturally responsive. These efforts address the significant cultural dissonance often encountered by *Orang Asli* pupils in traditional testing contexts. Unlike mainstream pupils, who benefit from a familiar alignment between their experiences and the test scenarios, *Orang Asli* pupils may find standardized tests alienating and disconnected due to the unfamiliar language and cultural context typically used in current testing, as highlighted in past studies by Abdullah (2022), Revindran et al. (2025), and Veloo et al. (2024). The adaptations made in this study sought to address these disparities by creating a mathematics test that acknowledges and respects the pupils' unique cultural and linguistic identities.

By incorporating culturally relevant examples and simplifying linguistic structures, the test became more accessible and relatable, aligning with the principles of universal design for assessment (UDA) (Ketterlin-Geller, 2005; Nieminen, 2022). This approach directly supports the study's goal of promoting equity in assessment practices. Moreover, the framework prioritizes minimizing barriers and maximizing opportunities for all pupils to demonstrate their mathematical knowledge and skills. This aligns with the broader understanding that assessments should be inclusive and equitable, recognizing the diverse cultural and linguistic backgrounds of learners (Russel et al., 2009; Tai et al., 2022).

The findings of this study underscore the critical role of cultural context in the development of mathematics test items, emphasizing the importance of creating tests that are culturally valid and responsive to Indigenous ways of knowing, as highlighted by Aikenhead (2017), Ottmann and Jeary (2016), and Shanmugam et al. (2025). The established framework integrates cultural considerations at each stage of the test development process, ensuring that the content reflects the lived experiences, values, and knowledge systems of *Orang Asli* pupils. By involving teacher-experts who are familiar with the cultural experiences of the *Orang Asli* pupils, the framework prioritizes the inclusion of culturally relevant examples and problem-solving scenarios rooted in the daily lives of the pupils, making the test more relatable and engaging. This approach not only fosters a sense of recognition and respect for the

Figure 1. Test Development Framework for Constructing Mathematics Test



pupils' cultural identities but also addresses the historical biases of standardized assessments that often marginalize Indigenous learners (Dreher et al., 2021; Louie, 2017). Through these culturally responsive practices, the framework demonstrates its capacity to bridge the gap between Indigenous knowledge systems and formal education, ensuring that the assessments are meaningful, equitable, and inclusive.

Psychometric Properties of the Test

The quantitative findings of the study further validated the effectiveness of the developed test items. The item difficulty index and item discrimination index results indicated that the test items were appropriately challenging and capable of distinguishing between varying performance levels (Boopathiraj & Chellamani, 2013; Quaigrain & Arhin, 2017). The balance achieved through these metrics ensures that the test is neither overly simplistic nor excessively complex, thus maintaining its fairness across a diverse range of abilities. This careful calibration is crucial for providing an equitable platform that acknowledges the varying mathematical proficiencies of the *Orang Asli* pupils. Reliability analysis using Cronbach's alpha coefficient reinforced the consistency of the test. A high coefficient value affirmed that the test items collectively measured the intended construct—mathematical performance—reliably and consistently (Bonnet & Wright, 2015). This reliability indicates that the mathematics test can yield stable and reproducible results over repeated administrations, which is particularly important for tracking pupil progress over time.

The psychometric validation of the test items highlights the importance of balancing cultural relevance with quantitative rigor in the establishment of a robust framework for test development. By ensuring that the test items are appropriately challenging and consistently measure the intended construct, the framework demonstrates its capacity to create equitable assessments tailored to the needs of diverse learners, such as the *Orang Asli* pupils. This approach emphasizes the need for a development process that integrates qualitative insights, such as cultural and contextual relevance, with rigorous quantitative analysis to validate the reliability and fairness of the assessment. Ultimately, this framework provides a structured pathway for designing assessments that are not only inclusive but also capable of accurately evaluating pupils' abilities across varying skill levels, ensuring both cultural sensitivity and academic integrity.

Framework Sustainability in Indigenous Education

The sustainability of culturally responsive assessment in Indigenous education depends not only on individual efforts but on systemic support across policy, curriculum, and classroom practice. This study offers a language- and context-sensitive framework for developing mathematics tests that align with the linguistic backgrounds and lived educational experiences of the Indigenous pupils. While the study does not aim to resolve structural inequities, it contributes to an emerging body of work demonstrating how inclusive assessment design can mitigate language and cultural barriers in educational measurement. Its findings point to the importance of iterative, collaborative validation processes involving both academic experts and Indigenous-serving teachers, an approach that is vital for creating assessments that are both valid and culturally meaningful. However, for such frameworks to be sustainable, they must be embedded within broader institutional structures that prioritize Indigenous perspectives, recognize diverse learner experiences, and support ongoing teacher capacity building (Aikenhead, 2017; Dreher et al., 2021; Preston & Claypool, 2021).

Building on this, the selection of *Orang Asli* pupils who are more familiar with the formal school culture carries important implications for the test development framework. These pupils, having navigated the structures, norms, and expectations of formal education, bring a unique perspective that

bridges Indigenous cultural knowledge with mainstream educational practices, as highlighted by Nordin et al. (2020) and Renganathan (2016). This may serve as a vital reference point for designing culturally responsive tests that maintain relevance to their lived experiences while adhering to the broader academic standards. For example, test developers who are familiar with both the Indigenous cultural practices and current testing practices employed in primary schools for *Orang Asli* pupils could potentially be beneficial in creating items that are culturally grounded yet comprehensible within the formal mathematics curriculum. By centering the experiences of pupils who are more comfortable navigating the formal school culture, test developers can establish equitable testing frameworks that could gradually extend to support other groups of Indigenous pupils.

Despite meaningful strides, it is essential to acknowledge the limitations of this study. The linguistic and cultural adaptations applied to the test items were not exhaustive, and some items did not undergo significant modifications. Moreover, the absence of detailed item-by-item documentation limits the ability to evaluate the extent of cultural alignment achieved. Another limitation lies in the lack of direct engagement with students to gather their perspectives on the test tasks, such as how they interpreted and processed the information presented. While teachers' insights were instrumental in shaping the test design, the omission of student-centered approaches leaves a gap in understanding how the tasks resonate with the pupils themselves. Future studies should prioritize incorporating methods such as cognitive interviews, focus groups, or think-aloud protocols with students to capture their perspectives more directly. Such approaches would provide valuable data on how test items are perceived, highlighting potential areas for refinement and enhancing the overall robustness of the framework.

In addition to student-centered strategies, future efforts should involve deeper collaboration with *Orang Asli* communities to ensure that their values, traditions, and worldviews are meaningfully represented in assessment design. Ensuring the longevity of such frameworks requires embedding them into teacher training and professional development programs to equip educators with the skills needed to create culturally responsive assessments. Furthermore, sustained collaboration with Indigenous communities is vital to reflect their evolving cultural contexts and ensure authenticity in educational practices (Aikenhead, 2017; Meaney et al., 2021; Preston & Claypool, 2021). By addressing these limitations and exploring these directions, future work can contribute to more inclusive and effective assessment practices that better serve *Orang Asli* pupils.

While linguistic simplifications and cultural adaptations address immediate barriers, they do not fully resolve the systemic underrepresentation of Indigenous perspectives in educational assessments. As highlighted in the literature, true inclusivity requires a paradigm shift that integrates Indigenous epistemologies into the broader educational framework (Ketterlin-Geller, 2005; Nieminen, 2022). Over time, these measures can mitigate the systemic underrepresentation of Indigenous perspectives in education and promote equity. This study highlights the importance of iterative processes and ongoing dialogue with stakeholders to maintain the framework's relevance and effectiveness, contributing to a more equitable educational landscape for marginalized communities.

Conclusion

This study offers a robust framework for developing mathematics assessments that are both culturally responsive and psychometrically sound, providing a preliminary blueprint for advancing equity in education. By integrating the principles of UDA and leveraging qualitative and quantitative insights, the framework addresses the specific needs of Indigenous students while meeting rigorous academic standards. The inclusion of culturally familiar contexts and simplified language enhances the

accessibility of assessments, allowing Indigenous students to demonstrate their mathematical knowledge and skills authentically and confidently.

While the study validates the effectiveness of the developed framework for *Orang Asli* pupils, it also invites broader reflection on the role of assessments and challenges educators and policymakers to rethink the purpose and design of assessments in mathematics education. How can culturally responsive assessments be scaled to meet the diverse needs of marginalized communities worldwide? What systemic changes are necessary to ensure that Indigenous ways of knowing are authentically represented in educational practices? These questions emphasize the ongoing need for critical reflection and action in creating assessments that genuinely respect and celebrate the cultural diversity of all Indigenous pupils worldwide.

Received: 4/29/2024. **Accepted:** 8/18/2025. **Published:** 8/25/2025.

Citation: Shanmugam, S. K. S., Veloo, A., & Revindran, S. (2025). Indigenizing mathematics assessment: Developing a culturally inclusive mathematics test framework for indigenous pupils. *Practical Assessment, Research, & Evaluation*, 30(8). Available online: https://doi.org/10.7275/pare.2072

Corresponding Author: S. Kanageswari Suppiah Shanmugam, Universiti Utara Malaysia. Email: kanageswari@uum.edu.my

Statements and Declarations

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

***Financial or non-financial interests*:** No potential competing interest was reported by the authors.

Ethical approval: The research study that underpins this publication was submitted to the Education Planning and Research Division, Ministry of Education [KPM.600-3/2/3-eras(4566)] and was approved as the test items were consistent with the national mathematics syllabus and assessment standards.

References

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

Abdulrahim, N. A., & Orosco, M. J. (2020). Culturally responsive mathematics teaching: A research synthesis. *The Urban Review*, *52*(1), 1-25. https://doi.org/10.1007/s11256-019-00509-2

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

Abrams, E., Taylor, P. C., & Guo, C. J. (2013). Contextualizing culturally relevant science and mathematics teaching for indigenous learning. *International Journal of Science and Mathematics Education*, *11*, 1-21. https://doi.org/10.1007/s10763-012-9388-2

Aikenhead, G. S. (2017). Enhancing school mathematics culturally: A path of reconciliation. *Canadian Journal of Science, Mathematics and Technology Education*, *17*(2), 73-140. https://doi.org/10.1080/14926156.2017.1308043

Allen, P., & Trinick, T. (2022). Creating space for indigenous knowledge in Māori-medium mathematics classrooms. *International Journal of Qualitative Studies in Education*, 1-14. https://doi.org/10.1080/09518398.2022.2025473

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Ball, R. (2024, April 5). *Maths has no borders: Professor Rowena Ball brings Indigenous mathematics to ANU*. Australian National University. https://science.anu.edu.au/news-events/news/maths-has-no-borders-professor-rowena-ball-brings-indigenous-mathematics-anu

Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal Of Organizational Behavior*, *36*(1), 3-15. https://doi.org/10.1002/job.1960

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.

Bryant, B. R., & Rivera, D. P. (1997). Educational assessment of mathematics skills and abilities. *Journal of Learning Disabilities*, *30*(1), 57-68. https://doi.org/10.1177/002221949703000105

Dolan, R., Hall, T. E., Banerjee, M., Chun, E., & Strangman, N. (2005). Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities. *The Journal of Technology, Learning and Assessment*, *3*(7), 1-33. https://ejournals.bc.edu/index.php/jtla/article/view/1660

Dreher, A., Lindmeier, A., Feltes, P., Wang, T. Y., & Hsieh, F. J. (2021). Do cultural norms influence how teacher noticing is studied in different cultural contexts? A focus on expert norms of responding to students' mathematical thinking. *ZDM–Mathematics Education*, *53*, 165-179. https://doi.org/10.1007/s11858-020-01197-z

Dündar, S., Temel, H., & Gündüz, N. (2016). Development of a mathematical ability test: a validity and reliability study. *International Journal of Mathematical Education in Science and Technology*, *47*(7), 1061-1075. https://doi.org/10.1080/0020739X.2016.1153734

Fatimah, S., Elzamzami, A. B., & Slamet, J. (2020). Item analysis of final test for the 9th grade students of SMPN 44 Surabaya in the academic year of 2019/2020. *Journal of English Education and Literature*, *2*(1), 34-46. https://doi.org/10.51836/journeel.v2i1.81

Forsyth, H., & Evans, J. (2019). Authentic assessment for a more inclusive history. *Higher Education Research & Development*, *38*(4), 748-761. https://doi.org/10.1080/07294360.2019.1581140

Foster, N. & Piacentini, M. (2023). *Innovating assessments to measure and support complex skills*. OECD Publishing.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, *23*(1), 17-27. https://doi.org/10.1111/j.1745-3992.2004.tb00149.x

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, *12*(3), 38-47.

Han, F. (2019). Self-concept and achievement in math among Australian primary students: gender and culture issues. *Frontiers in Psychology*, *10*, 1-9. https://doi.org/10.3389/fpsyg.2019.00603

Hartell, E., & Strimel, G. J. (2019). What is it called and how does it work: examining content validity and item design of teacher-made tests. *International Journal of Technology and Design Education*, *29*(4), 781-802. https://doi.org/10.1007/s10798-018-9463-2

International Test Commission (ITC). (2017). *The ITC guidelines for translating and adapting tests (second edition)*. https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf

Lane, S., Raymond, M. R., Haladyna, T. M., & Downing, S. M. (2015). Test development process. In S. Lane, M. R. Raymond, T. M. Haladyna, & S. M. Downing (Eds.), *Handbook of test development* (pp. 3-18). Routledge.

Loewen, J., Kinshuk, Suhonen, J., & Chen, N. S. (2017). I-SLATE: Designing a culturally relevant framework for authentic learning. *Smart Learning Environments*, *4*, 1-18. https://doi.org/10.1186/s40561-017-0048-4

Louie, N. L. (2017). The culture of exclusion in mathematics education and its persistence in equity-oriented teaching. *Journal for Research in Mathematics Education*, *48*(5), 488-519. https://doi.org/10.5951/jresematheduc.48.5.0488

Jabatan Kemajuan Orang Asli (JAKAO) [Department of *Orang Asli* Development] (2022, August 16). *Tribes*. https://www.jakoa.gov.my/orang-asli/tribes/?lang=en

James, K. (2018). Universal design for learning (UDL) as a structure for culturally responsive practice. *Northwest Journal of Teacher Education*, *13*(1), 1-23. https://doi.org/10.15760/nwjte.2018.13.1.4

Jenkins, J., & Leung, C. (2019). From mythical 'standard' to standard reality: The need for alternatives to standardized English language tests. *Language Teaching*, *52*(1), 86-110. https://doi.org/10.1017/S0261444818000307

Kadonsi, K., Kenneth, M. K., & Magdalene, S. (2023). Empowering mathematical minds through indigenous pedagogies in teaching mathematics in Southern Province, Zambia. *International Journal of Research and Innovation in Social Science*, *7*(10), 2342-2370. https://dx.doi.org/10.47772/IJRISS.2023.701176

Ketterlin-Geller, L. R. (2005). Knowing what all students know: Procedures for developing universal design for assessment. *The Journal of Technology, Learning and Assessment*, *4*(2), 1-23. https://ejournals.bc.edu/index.php/jtla/article/view/1649

Klenowski, V. (2009). Australian Indigenous students: Addressing equity issues in assessment. *Teaching Education*, *20*(1), 77-93. https://doi.org/10.1080/10476210802681741

Kubiszyn, T., & Borich, G. D. (2024). *Educational testing and measurement*. Wiley & Sons.

Meaney, T., Trinick, T., & Allen, P. (2021). Ethnomathematics in education: The need for cultural symmetry. In M. Danesi (Ed), *Handbook of cognitive mathematics* (pp. 1-29). Springer.

Milne, E. (2017). Implementing Indigenous education policy directives in Ontario public schools: Experiences, challenges and successful practices. *The International Indigenous Policy Journal*, *8*(3), 1-20. https://doi.org/10.18584/iipj.2017.8.3.2

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6-20. https://doi.org/10.1111/j.1745-3992.2006.00075.x

Mislevy, R. J. (2007). Validity by design. *Educational Researcher*, *36*(8), 463-469. https://doi.org/10.3102/0013189X07311660

Mullis, I. V., & Martin, M. O. (2017). *TIMSS 2019 Assessment Frameworks*. International Association for the Evaluation of Educational Achievement.

Nelson-Barber, S., & Trumbull, E. (2007). Making assessment practices valid for Indigenous American students. *Journal of American Indian Education*, 132-147. http://www.jstor.org/stable/24398547

Nieminen, J. H. (2022, February). Universal design for assessment in mathematics. In J., Hodgen, E. Geraniou, G. Bolondi & F. Ferretti, (Eds.), *Proceedings of the twelfth congress of the European society for research in mathematics education (CERME12)* (pp. 1788-1796). Free University of Bozen-Bolzano and European Society for Research in Mathematics Education (ERME).

Nordin, R., Hassan, M. S., Fern, V. W. R., Cherley, M., & Subramaniam, L. B. (2020). Indigenous education for the Orang Asli: Legal perspectives and best practices. *Journal of Nusantara Studies*, *5*(2), 365-383. https://doi.org/10.24200/jonus.vol5iss2pp365-383

Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: Responding to issues regarding methodology, policy, and equity. *ZDM*, *50*(4), 555-570. https://doi.org/10.1007/s11858-018-0963-z

Ottmann, J., & Jeary, J. (2016). Assessment practices and aboriginal students. In S. Scott, D. E. Scott, & C. F. Webber (Eds.), *Leadership of assessment, inclusion, and learning* (pp. 327-363). Springer.

Parra, A., & Trinick, T. (2018). Multilingualism in indigenous mathematics education: an epistemic matter. *Mathematics Education Research Journal*, *30*, 233-253. https://doi.org/10.1007/s13394-017-0231-5

Preston, J. P., & Claypool, T. R. (2021). Analyzing assessment practices for Indigenous students. *Frontiers in Education*, 6, 679972. https://doi.org/10.3389/feduc.2021.679972

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, *4*(1), 1301013. https://doi.org/10.1080/2331186X.2017.1301013

Ramadhan, S., Mardapi, D., Prasetyo, Z. K., & Utomo, H. B. (2019). The development of an instrument to measure the higher order thinking skill in physics. *European Journal of Educational Research*, *8*(3), 743-751. https://doi.org/10.12973/eu-jer.8.3.743

Reid O'Connor, B., & Norton, S. (2022). Supporting indigenous primary students' success in problem-solving: Learning from Newman interviews. *Mathematics Education Research Journal*, *34*(2), 293-316. https://doi.org/10.1007/s13394-020-00345-8

Renganathan, S. (2016). Educating the *Orang Asli* children: exploring indigenous children's practices and experiences in schools. *The Journal of Educational Research*, *109*(3), 275-285. https://doi.org/10.1080/00220671.2014.945150

Revindran, S., Shanmugam, S. K. S., & Shaik-Abdullah, S. (2025). The use of recorded oral tests as test accommodations for assessing mathematics performance among Indigenous *Orang Asli* pupils. *Mathematics Education Research Journal*, 1-26. https://doi.org/10.1007/s13394-025-00524-5

Rigney, L., Garrett, R., Curry, M., & MacGill, B. (2020). Culturally responsive pedagogy and mathematics through creative and body-based learning: Urban Aboriginal schooling. *Education and Urban Society*, *52*(8), 1159-1180. https://doi.org/10.1177/0013124519896861

Rose, D. H., Robinson, K. H., Hall, T. E., Coyne, P., Jackson, R. M., Stahl, W. M., & Wilcauskas, S. L. (2018). Accurate and informative for all: Universal design for learning (UDL) and the future of assessment. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kurz (Eds.), *Handbook of accessible instruction and testing practices: Issues, innovations, and applications* (pp. 167-180). Springer.

Russel, M., Hoffmann, T., & Higgins, J. (2009). Meeting the needs of all students: A universal design approach to computer-based testing. *Innovate: Journal of Online Education, 5*(4), 1-6. https://www.learntechlib.org/p/104243/

Seipel, B., Bailey, P., & Teasdale, R. (2022). Changing gears in the assessment cycle: Preparing for gradual yet substantial shifts. In E. Meletiadou (Eds.) *Handbook of research on policies and practices for assessing inclusive teaching and learning* (pp. 157-182). IGI Global.

Shah, D. S. M., Adnan, A. H. M., Salim, M. S. A. M., Yusof, A. M., & Tahir, M. H. M. (2020). The non-formal education of Orang Asli adolescents in Malaysia: From community to university. *International Journal of Humanities Technology and Civilization*, 1-8. https://doi.org/10.15282/ijhtc.v5i1.4812

Shanmugam, S. K. S., Veloo, A., & Bin Jusoh, Y. A. (2024). The challenges of virtual learning: an exploratory study among Orang Asli Pupils from the mathematics teachers' perspectives. *Diaspora, Indigenous, and Minority Education*, *18*(3), 211-229. https://doi.org/10.1080/15595692.2023.2212823

Shanmugam, S. K. S., Veloo, A., & Revindran, S. (2025). Examining the validity of oral mathematics test accommodation in academic and native languages for *Orang Asli* pupils in Malaysia. *Applied Measurement in Education*, 1-15. https://doi.org/10.1080/08957347.2025.2480103

Simamora, R. E., & Saragih, S. (2019). Improving students' mathematical problem solving ability and self-efficacy through guided discovery learning in local culture context. *International Electronic Journal of Mathematics Education*, *14*(1), 61-72. https://doi.org/10.12973/iejme/3966

Smith, P., Frazier, P., Lee, J., & Chang, R. (2018). Incongruence between native and test administration languages: Towards equal opportunity in international literacy assessment. *International Journal of Testing*, *18*(3), 276-296. https://doi.org/10.1080/15305058.2017.1407767

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, *38*(5), 553-573. https://doi.org/10.1002/tea.1018

Stoffer, J. (2017). The importance of culturally safe assessment tools for Inuit students. *The Australian Journal of Indigenous Education*, *46*(1), 64-70. https://doi.org/10.1017/jie.2016.30

Tai, J. H. M., Dollinger, M., Ajjawi, R., Jorre de St Jorre, T., Krattli, S., McCarthy, D., & Prezioso, D. (2022). Designing assessment for inclusion: An exploration of diverse students' assessment experiences. *Assessment & Evaluation in Higher Education*, *48*(3), 403-417. https://doi.org/10.1080/02602938.2022.2082373

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

Trumbull, E., & Nelson-Barber, S. (2019). The ongoing quest for culturally-responsive assessment for Indigenous students in the US. *Frontiers in Education*, *4*(40), 1-11. https://doi.org/10.3389/feduc.2019.00040

Veloo, A., Shanmugam, S. K. S., & Revindran, S. (2024). Examining oral test accommodation in assessing Malaysian orang asli pupils' mathematics performance for computation and word problem testlets. *Malaysian Journal of Learning and Instruction*, *21*(2), 239-264. https://doi.org/10.32890/mjli2024.21.2.8

Wiese, E., & Nortvedt, G. A. (2023). Teacher assessment literacy in culturally and linguistically diverse classrooms: a Norwegian case study. *Teaching and Teacher Education*, *135*, 104357. https://doi.org/10.1016/j.tate.2023.104357

Wilkinson, L. C. (2019). Learning language and mathematics: A perspective from linguistics and education. *Linguistics and Education*, *49*, 86-95. https://doi.org/10.1016/j.linged.2018.03.005

Xu, H., & Ball, R. (2022). *Is the study of Indigenous mathematics ill-directed or beneficial?*. ArXiv. https://doi.org/10.48550/arXiv.2212.02778