EXAMINING RECORDED ORAL TEST AS A TEST ACCOMMODATION IN MEASURING ORANG ASLI PUPILS' MATHEMATICS PERFORMANCE

SUHEYSEN A/L REVINDRAN

MASTER OF EDUCATION
UNIVERSITI UTARA MALAYSIA
2025

Awang Had Salleh
Graduate School
of Arts And Sciences

Universiti Utara Malaysia

PERAKUAN KERJA TESIS / DISERTASI
(Certification of thesis / dissertation)

Kami, yang bertandatangan, memperakukan bahawa
(We, the undersigned, certify that)

SUHEYSEN REVINDRAN

calon untuk Ijazah **MASTER OF EDUCATION**
(candidate for the degree of)

telah mengemukakan tesis / disertasi yang bertajuk:
(has presented his/her thesis / dissertation of the following title):

"EXAMINING RECORDED ORAL TEST AS A TEST ACCOMMODATION IN MEASURING ORANG ASLI PUPILS' MATHEMATICS PERFORMANCE"

seperti yang tercatat di muka surat tajuk dan kulit tesis / disertasi.
(as it appears on the title page and front cover of the thesis / dissertation).

Bahawa tesis/disertasi tersebut boleh diterima dari segi bentuk serta kandungan dan meliputi bidang ilmu dengan memuaskan, sebagaimana yang ditunjukkan oleh calon dalam ujian lisan yang diadakan pada : ***21 Ogos 2024.***
That the said thesis/dissertation is acceptable in form and content and displays a satisfactory knowledge of the field of study as demonstrated by the candidate through an oral examination held on:
21 August 2024.

Pengerusi Viva: *(Chairman for VIVA)*	**Prof. Dr. Fauziah Abdul Rahim**	Tandatangan *(Signature)*
Pemeriksa Luar: *(External Examiner)*	**Assoc. Prof. Dr. Lim Hooi Lian**	Tandatangan *(Signature)*
Pemeriksa Dalam: *(Internal Examiner)*	**Dr. Marini Kasim**	Tandatangan *(Signature)*
Nama Penyelia/Penyelia-penyelia: *(Name of Supervisor/Supervisors)*	**Assoc. Prof. Dr. S. Kanageswari Suppiah Shanmugam**	Tandatangan *(Signature)*

Tarikh:
(Date) ***21 August 2024***

Permission to Use

In presenting this thesis in fulfilment of the requirements for a postgraduate degree from Universiti Utara Malaysia, I agree that the University Library may make it freely available for inspection. I further agree that permission for the copying of this thesis in any manner, in whole or in part, for scholarly purpose may be granted by my supervisor(s) or, in their absence, by the Dean of Awang Had Salleh Graduate School of Arts and Sciences. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to Universiti Utara Malaysia for any scholarly use which may be made of any material from my thesis.

Requests for permission to copy or to make other use of materials in this thesis, in whole or in part, should be addressed to:

Dean of Awang Had Salleh Graduate School of Arts and Sciences
UUM College of Arts and Sciences
Universiti Utara Malaysia
06010 Sintok
Kedah Darul Aman, Malaysia

Abstrak

Prestasi matematik murid Orang Asli lazimnya dipengaruhi oleh halangan bahasa dan literasi. Dengan memperkenalkan ujian akomodasi, ia berpotensi untuk menjadi penyelesaian bagi mengurangkan isu-isu yang dihadapi oleh murid. Kajian ini menggunakan pendekatan kaedah campuran bagi mengkaji penggunaan ujian lisan yang dirakamkan dalam bahasa pengajaran dan bahasa etnik sebagai ujian akomodasi terhadap ciri-ciri linguistik yang kompleks seperti panjang item dan kosa kata. Pendekatan kuantitatif kajian ini menilai prestasi matematik 142 murid Orang Asli Tahun 4 daripada etnik Temiar dari tujuh sekolah rendah di negeri Perak dan Kelantan. Bagi menilai keberkesanan ujian akomodasi, tiga ujian matematik yang terdiri daripada 40 item ujian telah dibangunkan, iaitu ujian bertulis dalam Bahasa Melayu, ujian lisan yang dirakamkan dalam Bahasa Melayu, dan ujian lisan yang dirakamkan dalam Bahasa Temiar. Data kuantitatif dianalisis menggunakan ujian ANOVA dan korelasi Pearson. Bagi pendekatan kualitatif, analisis kandungan dijalankan terhadap jenis kosa kata yang digunakan untuk item ujian, manakala analisis tematik terhadap temu bual separa berstruktur dengan tujuh guru matematik dijalankan untuk meneroka pandangan mereka mengenai penggunaan ujian lisan yang dirakam untuk murid Orang Asli. Dapatan kajian menunjukkan bahawa murid Orang Asli menunjukkan peningkatan prestasi yang signifikan apabila ujian lisan yang dirakam digunakan, berbanding ujian bertulis. Selain itu, antara ujian lisan yang dirakam, dapatan kajian menunjukkan bahawa murid menunjukkan prestasi yang lebih baik apabila diberikan ujian lisan dalam Bahasa Melayu, berbanding dengan Bahasa Temiar. Tiga tema utama dikenal pasti daripada input kualitatif iaitu faktor-faktor yang mempengaruhi prestasi murid Orang Asli sebelum ujian akomodasi, sokongan untuk ujian lisan yang dirakam sebagai ujian akomodasi, dan cabaran pada masa depan dalam penggunaan ujian lisan yang dirakam. Kajian ini menekankan pelaksanaan praktikal ujian lisan yang dirakam untuk mengurangkan beban kognitif yang disebabkan oleh ciri-ciri linguistik sebagai ujian akomodasi yang adil dan inklusif dalam pengujian matematik untuk murid Orang Asli di Malaysia.

Kata Kunci: Prestasi matematik, Ujian akomodasi, Ujian lisan yang dirakam, Ciri-ciri linguistik, Orang Asli

Abstract

The mathematics performance of Orang Asli pupils is commonly affected by language and literacy barriers. Introducing test accommodations may potentially serve as a solution to alleviate these issues faced by the pupils. This study employed a mixed-method approach to examine the use of recorded oral tests in the language of instruction and ethnic language as test accommodations on complex linguistic features such as item length and vocabulary. The quantitative approach of this study evaluated the mathematics performance of 142 Year 4 Orang Asli pupils from the Temiar ethnic group from seven primary schools in the states of Perak and Kelantan. To evaluate the effectiveness of the test accommodation, three mathematics tests comprising 40 test items were developed, which are a written test in Bahasa Melayu, recorded oral test in Bahasa Melayu, and recorded oral test in Bahasa Temiar. The quantitative data was analysed using ANOVA test and Pearson correlation. For the qualitative approach, content analysis was carried out on vocabulary type, while thematic analysis on semi-structured interviews with seven mathematics teachers were conducted to explore their views on the use of the recorded oral test for the Orang Asli pupils. Findings showed that the pupils performed significantly better when administered recorded oral tests, compared to the written test. Moreover, between the recorded oral tests, the findings indicated that pupils performed better when administered the test in Bahasa Melayu, compared to Bahasa Temiar. Three key themes were identified from the qualitative input, namely factors affecting pupils' performance pre-test accommodation, support for post-recorded oral test accommodation, and future challenges on the use of recorded oral test. This study underscores the practical implementation of recorded oral test to reduce cognitive load caused by linguistic features as a fair and inclusive test accommodation in mathematics testing for Orang Asli pupils in Malaysia.

Keywords: Mathematics performance, Test accommodation, Recorded oral test, Linguistic features, *Orang Asli*

Acknowledgement

I would like to express my deepest gratitude to the Almighty for granting me the strength, perseverance, and guidance throughout my academic journey, without whom this thesis would not have been possible.

To my beloved mother and family members, your unwavering love, encouragement, and support have been my pillars of strength. Your sacrifices and belief in me have been invaluable in reaching this milestone.

I extend my heartfelt appreciation to my esteemed supervisors, Assoc. Prof. Dr. Kanageswari Suppiah Shanmugam and Assoc Prof. Dr. Sarimah Shaik Abdullah, for their invaluable guidance, unwavering support, and scholarly wisdom. Their mentorship and educational support were pivotal in shaping the outcome of this research.

I am indebted to Prof. Dr. Arsaythamby Veloo for his insightful mentorship, which played a pivotal role in my growth and development as a postgraduate student. His guidance and encouragement have been instrumental in refining my academic pursuits.

Lastly, I am profoundly grateful to everyone who contributed, directly or indirectly, to this postgraduate journey. Your assistance, encouragement, and words of wisdom have been immensely valuable and deeply appreciated.

Thank you all for being a part of this significant chapter in my life and for your unwavering support throughout this academic endeavour.

Table of Contents

List of Tables

List of Figures

List of Appendices

CHAPTER ONE

INTRODUCTION

1.1 Background of Study

Mathematics has continuously been regarded as an important benchmark for the educational quality of a nation (Yolcu, 2021). Therefore, the meaningful learning of mathematics could develop the necessary mathematics skills and knowledge needed to solve everyday problems (Suarsana et al., 2019). However, in the current state of affair around the world, the mathematics score gap caused by the disparity among the different groups of pupil populations have continued to be a vexing social malaise that requires immediate attention by many educational systems (Ababneh & Kodippili, 2020; Paschall et al., 2018).

Specifically, the widening mathematics achievement gap, which has been observed between the Indigenous and mainstream pupils, continues to impede and marginalize the Indigenous pupils over the years (Berry, 2018). This notion has also been observed among the *Orang Asli* pupils, whereby in the year 2015, the Ministry of Education reported that only 61% of the Year 6 *Orang Asli* pupils had passed their national primary school examination, *Ujian Pencapaian Sekolah Rendah* (UPSR) as compared to 87% of the national average (Abdul Jalil et al., 2023). This continues to be an issue that has negatively impacted the *Orang Asli* pupils' academic progress as they move secondary schools (Abdul Halim, 2022; Ruzlan et al., 2021).

According to Arteaga and Glewwe (2019), the educational achievement of the Indigenous pupils, in general, has occasionally been affected by the lack of educational measures established by policymakers. Moreover, it has been proposed that the mismatch of the interwoven educational pedagogy between curriculum, instruction and assessment has, for the most part, been detrimental to the educational development of the Indigenous pupils, particularly as this group of pupils have seldom been considered during the developmental phases of education policies (Hollinsworth et al., 2021; McCarty, 2018; Renganathan & Kral, 2018; Zidney et al., 2020). In the seminal paper with regards to the Canadian Government's working partnership with the First Nations to reform and improve the educational quality of the Indigenous communities, Battiste (2002) stated that,

> Students in Indigenous societies around the world have, for the most part, demonstrated a distinct lack of enthusiasm for the experience of schooling in its conventional form - an aversion that is most often attributable to an alien institutional culture rather than any lack of innate intelligence, ingenuity, or problem-solving skills on the part of the students. (p. 10)

To address this issue, it has been proposed that the rich Indigenous culture, knowledge, outlook, and experiences that are deeply stemmed within these communities should also be incorporated into education policies that could promote learning and the educational development among Indigenous pupils (Hunter & Hunter, 2017; Morrison et al., 2019). In the United States, Trumbull and Nelson-Barber (2019) have highlighted the need to incorporate the unique ethnic languages of the Indigenous pupils into mathematics testing practices to uphold cultural responsiveness in mathematics assessments. The study further proposed that key cultural elements, such as the ethnic languages, should not be overlooked but rather should be included in mathematics education as a measure that could leverage language as an inclusive practice in mathematics education, as advocated by Aikenhead (2017) and Ruef et al., 2020. By leveraging language as a cultural identity, particularly in mathematics assessments, the widening mathematics achievement gap could potentially be addressed.

Therefore, by potentially employing the ethnic language of the Indigenous communities in the development of education policies, this measure could inadvertently acknowledge the "cultural repertoires and intelligences that pupils bring to the learning relationship" (Morrison et al., 2019, p. 2). As language is an important cultural element of many Indigenous communities, the incorporation of the ethnic languages during the development of culturally responsive mathematics assessments would acknowledge the needs of the Indigenous pupils to ensure fair and valid assessments of their mathematical skills, as proposed by Kükea Shultz and Englert (2021), and Trumbull and Nelson-Barber (2019).

The Indigenous pupils typically use their unique ethnic languages to communicate among members of their community during their day-to-day activities as part of their cultural heritage. However, these pupils are required to use a different language, such as the language of instruction that is typically reserved during the formal school setting, which they are least proficient in (Huang & Butler, 2020; Hudson & Angelo, 2020). It may have resulted in the potential language barrier which is caused by the discrepancy between their ethnic languages used in the social context and their low proficiency in the language of instruction used in schools (Barwell, 2020; Garcia & Solorza, 2021; Robertson & Graven, 2020). This resultant language issue may spill into mathematics learning, which could subsequently disrupt an accurate measure of the pupils' mathematics performance as they are unable to comprehend the language used in the test (Ruzlan, et al., 2021; Kanageswari et al., 2021).

In the context of this study, the Malaysian *Orang Asli* pupils experience language barriers due to the distinct languages used by them in the social context and in the formal school setting. Moreover, as the *Orang Asli* pupils have just begun their formal school education, they have yet to develop the necessary proficiency in the language of instruction as they are still reliant on their use of their respective ethnic languages in primary schools (Abdul Hakim et al., 2023; Abdul Halim, 2022; Ruzlan et al., 2021). This may have subsequently led to the reading difficulties that have affected how these pupils comprehend the written text, particularly when administered tests (Kanageswari et al., 2023; Zaleha et al., 2021) Therefore, due to the language and literacy challenges faced by the *Orang Asli* pupils, current mathematics assessments fail to provide a more accurate test score interpretation that reflects their true mathematical ability (Kanageswari et al., 2021; Arsaythamby et al., 2021).

To address the language and literacy barriers that have continued to impede the accurate assessment of the pupils' mathematical knowledge and skills, the use of test accommodations has been proposed as a testing measure aimed at reducing these barriers to create equitable testing conditions for the pupils (Abedi et al., 2020; Solano-Flores, 2014; Yang, 2020). The incorporation of test accommodations, especially in mathematics assessments, could help reduce the impact of construct irrelevant variance brought upon by the secondary dimensions, such as language (Fuchs et al., 2019; Leiss et al., 2019; Prediger et al., 2019) and reading (Cirino et al., 2018; Koponen et al., 2018), which are secondary constructs not intended to be measured in mathematics tests (Sireci, 2020).

The uniqueness of incorporating the right test accommodation can be attributed to their specific focus on addressing the language and literacy challenges faced by *Orang Asli* pupils, which distinguishes them from the standard educational requirements. By acknowledging and accommodating these distinctive challenges, the implementation of test accommodation could potentially address these issues, such as been found in other educational settings that has affected most mainstream pupils in mathematics assessments (Herman & Cook, 2019; Lee & Orgill, 2021; Schissel et al., 2021; Sireci, 2018). Therefore, the use of test accommodation may provide the necessary support for *Orang Asli* pupils to better demonstrate the mathematics skills when assessed without being impeded by the language barrier and reading demands posed by the test.

Therefore, by taking into consideration the unique educational needs of the *Orang Asli* pupils, incorporating test accommodations in mathematics assessments could be a viable measure in levelling the playing field and providing these pupils with a fair chance to demonstrate their true abilities (Abedi et al., 2020; De Backer et al., 2019; Sireci, 2020). This approach recognizes that one size does not fit all in education, particularly when it comes to the *Orang Asli* pupils with their unique educational needs and cultural backgrounds. Thus, this initiative has the potential to be a significant advancement towards a more inclusive and culturally responsive approach to mathematics testing as supported by Baidoo-Anu et al. (2020) and Shanmugam et al. (2021), which could ensure a more accurate measure of *Orang Asli* pupils' mathematics performance.

1.2 Problem Statement

Most Indigenous pupils typically encounter a range of unique challenges that have continued to affect their ability to perform well in mathematics assessments. The issues faced by the Indigenous pupils are often rooted in their cultural and language diversity that sets them apart from the mainstream pupil population, as highlighted in past studies (Preston & Claypool, 2021; Robertson & Graven, 2020; Trumbull & Nelson-Barber, 2019). The divergence in cultural and linguistic norms between the Indigenous pupils and mainstream pupil population have often led to a lack of culturally relevant mathematics educational materials and teaching approaches (Miller & Armour, 2019), which have subsequently affected their mathematics performance.

For the *Orang Asli* pupils in Malaysia, the disparity observed in their mathematics performance has been particularly concerning. The Primary School Literacy and Numeracy Programme (PLaN), which was implemented to support underachieving Year 3 pupils in Malaysian primary schools, revealed a stark contrast in mathematics performance outcomes. In 2021, 90.81% of Year 3 pupils in mainstream schools achieved at least the minimum proficiency level (*Tahap Penguasaan* 3, TP3) in mathematics. However, only 61.53% of *Orang Asli* pupils reached this benchmark for mathematics achievement in the same year (Ministry of Education, 2021; 2022).

This significant gap highlights certain issues within the education system, where teaching methods and assessment approaches employed in *Orang Asli* primary schools may not be fully inclusive or responsive to the cultural and linguistic needs of the *Orang Asli* pupils (Aidil Fitri et al., 2020). These challenges may have subtly contributed to the continued difficulties faced by these pupils, particularly in mathematics, where mathematical proficiency is crucial for future academic success.

Consequently, Indigenous pupils have struggled to connect with the mathematical content being assessed, which may lead to issues in assessing their mathematics performance (Macqueen et al., 2019; Rigney et al., 2020). This issue has also been found to affect the *Orang Asli* pupils' engagement in current mathematics teaching and learning practices, which have advertently impacted how they perform in mathematics tests (Aidil Fitri et al., 2020; Ruzlan et al., 2021; Kanageswari et al., 2023). This warrants further examination of incorporating valid mathematics testing practices that would allow for a more accurate interpretation of the pupils' mathematics test scores (Rios et al., 2020; Kanageswari et al., 2021; Vidal Rodeiro & Macinska, 2022).

One of the main issues that have been found to affect pupils' ability to perform well in mathematics tests has been attributed to the language component in mathematics (Moschkovich, 2010; Peng et al., 2020; Robertson & Graven, 2020). Language plays a pivotal role among Indigenous pupils as they typically encounter language barriers due to the discrepancy between their main ethnic language and the language of instruction used in the classroom setting (Barwell, 2020; Hudson & Angelo, 2020;

Preston & Claypool, 2021). The language mismatch experienced by the Indigenous pupils may potentially lead to the lack of development of the language of instruction, due to the pupils' reliance on their own ethnic language, which in turn may create an issue in mathematics testing that may affect how they comprehend the language used in the test (Delprato, 2021; Kanageswari et al., 2021; Reid O'Connor & Norton, 2022; Robertson & Graven, 2020).

The issue on the lack of proficiency in the language of instruction has also been observed among the *Orang Asli* pupils (Abdul Halim, 2022; Misnaton et al., 2020; Ruzlan et al., 2021). The language barrier faced by these pupils has been attributed to their lack of exposure to the language of instruction, as they are typically exposed to the language of instruction only when they begin their early formal schooling (Wong & Abdillah, 2018). Therefore, as they have yet to develop the necessary language proficiency during their early years of primary schooling, this in turn may result in the language barrier that have negatively impacted their mathematics learning, and in turn, their mathematics performance when being assessed (Abdul Hakim et al., 2023; Ruzlan et al., 2021; Kanageswari et al., 2023).

Other than that, the issue of inadequate proficiency in the language of instruction has also been found to give rise to the potential reading difficulty experienced by most pupils (McKinnon & Blair, 2019; Powell et al., 2020; Vaughn et al., 2019). To effectively tackle mathematical problems, pupils must first engage in the process of comprehending the linguistic aspects of the written mathematics test items before they can apply both conceptual and procedural mathematical knowledge (Koponen et al., 2018; Nelson & Powell, 2018; Salihu et al., 2018). When confronted with linguistically dense test items, the ability of pupils to decode and comprehend the text becomes paramount in ensuring their success in solving the test items (Leiss et al., 2019; Pongsakdi et al., 2020; Wilner & Mokhtari, 2018).

In the context of *Orang Asli* pupils, the reading difficulties they face may have a profound impact on their performance in mathematics tests (Norwaliza et al., 2016; Kanageswari et al., 2023; Aidil Fitri et al., 2020). These pupils have yet to attain the requisite reading skills necessary for comprehending the written text, partly due to their lack of language comprehension. Both language and reading ability are fundamental prerequisites for effectively engaging and solving mathematics test items (Abdul Halim, 2022; Kanageswari et al., 2021). These two-pronged challenges experienced by the *Orang Asli* pupils not only hinders their ability to decode the written text and comprehend the language used in mathematics test item, but also restricts their capacity to apply mathematical concepts and procedures effectively, as highlighted in the studies by Kanageswari et al. (2023) and Zaleha et al. (2020).

Furthermore, the complex linguistic features embedded within mathematics test items was also found to further exacerbate the difficulties most pupils' may face when exhibiting their mathematical skills during the test (Daroczy et al., 2015; Lane & Levanthal, 2015; Pongsakdi et al., 2020). Specifically, item length, or the total number of words in a test item, was found to have an impact on the mathematics performance of most mainstream pupils (Abedi, 2014; Bergqvist et al., 2018; Haag et al., 2015; Sibanda, 2017). As longer test items require more effort to comprehend the language and decode the written texts, this could impact pupils' who face language and literacy barriers that may unfairly hinder their mathematics performance (Mullis et al., 2013; Walkington et al., 2015; 2018).

The literature on the impact of item length on pupils' performance reveals divergent findings that highlight the complexity of this issue. For instance, Sibanda (2017) found that longer test items negatively impacted Grade 4 pupils' mathematics performance, likely due to the cognitive burden of processing lengthy questions. In contrast, Cruz Neri et al. (2021) reported no significant interaction between item length and performance among mainstream pupils with reading difficulties, suggesting that the challenges posed by longer items may not be as pronounced in all contexts. However, the specific impact of item length on the mathematics performance of *Orang Asli* pupils remains underexplored. The contrasting findings from previous studies raise important questions about whether the strategies to improve mathematics performance for these pupils should focus solely on linguistic factors like item length, or if other aspects such as language comprehension in general should be considered more critically.

Other than that, vocabulary is another linguistic feature embedded within mathematics test items that have also shown to impact pupils' performance (Haag et al., 2013; Powell et al., 2017; Unal, 2019). Studies have shown that pupils' performance in mathematics can significantly differ based on their understanding of general vocabulary and mathematics vocabulary used in the specific context that is intended to be measured by mathematics test items (Lee & Spratley, 2010). The intricacies of mathematics vocabulary, which often includes specialized terms and unique contextual meanings, can pose additional challenges for pupils beyond the general language used in mathematics problem statements (Abedi & Lord, 2001).

This complexity can obscure pupils' true mathematical abilities, as they may struggle with specialized mathematics vocabulary which are typically presented in the language of instruction, which may not be comprehended easily by pupils with language comprehension issues (Kopriva et al., 2007; Martiniello, 2008). As vocabulary is heavily language-dependent (Chow & Ekholm, 2015; Haag et al., 2013; Riccomini et al., 2015), this linguistic feature can be particularly daunting for pupils who already grapple with language barriers and reading difficulties. Pupils must not only decipher complex and unfamiliar terms but also grasp the meaning of these words as they relate specifically to the mathematical concepts being tested (Fuchs et al., 2021; Powell et al., 2020). This added complexity may make it harder for pupils to engage directly with the main construct intended to be measured by the mathematics test items.

Despite the recognized challenges faced by *Orang Asli* pupils in learning mathematics, there remains a notable dearth of studies that specifically investigated how language and literacy barriers, particularly those influenced by linguistic features such as vocabulary, affect their mathematics performance. While existing research has suggested that vocabulary difficulties among *Orang Asli* pupils may be associated with broader language and literacy issues that have led to their lower mathematics performance (Arsyathamby et al., 2021; Kanageswari et al., 2021; Zaleha et al., 2020), these assertions have yet to be validated, particularly in relation to general and mathematics vocabulary, with the added linguistic feature of item length, embedded within test items such as examined in this study. When item length was examined concurrently with general vocabulary and mathematics vocabulary, Haag et al. (2013) found that mathematics vocabulary was not significantly correlated to item difficulty in the mathematics assessment among mainstream pupils. When item length is controlled, an increase in the number of mathematics vocabulary became more significantly correlated to increased item difficulty, when compared to general vocabulary.

In general, the studies found that longer item length and higher number of general vocabularies found within the test item was significantly more difficult for the pupils, as also supported by Kan and Bulut (2015) and Lin et al. (2021). These studies highlighted the greater impact of the two linguistic features within mathematics test items that have been found to influence the mathematics performance of pupils during the assessment. This gap highlights the need for research that not only acknowledges the role of language proficiency and reading ability in navigating these linguistic challenges but also systematically explore how increasing word count and specific vocabulary type may affect the *Orang* Asli pupils' ability to perform well in mathematics tests.

By taking into consideration the distinct issues faced by the *Orang Asli* pupils, the potential use of test accommodation as a viable measure in mathematics tests is deemed necessary for further examination to address the language and literacy barriers faced by the pupils (Abedi et al., 2020; Schissel et al., 2021; Sireci & Faulkner-Bond, 2015). Test accommodations could offer the needed linguistic support that could remove language and reading as construct irrelevant variances in mathematics tests (Herman & Cook, 2019; Lee & Orgill, 2021). Therefore, incorporating such a measure could provide a more accurate interpretation of the *Orang Asli* pupils' mathematics test scores when assessed.

A promising strategy to address the language barriers faced by *Orang Asli* pupils during mathematics tests is to incorporate their ethnic languages into the testing process, as advocated by Matthews and López (2019) and Trumbull and Nelson-Barber (2019). This approach may enable pupils to better comprehend test items presented in their ethnic language, thus potentially enhancing their ability to demonstrate their mathematical skills in a language that they are more familiar in (Barwell, 2020; Robertson & Graven, 2020; Kanageswari et al., 2021). By incorporating the pupils' ethnic language, this measure takes into consideration the linguistic diversity that exists within educational settings that could be used in mathematics testing to help pupils to better comprehend the content and context of the mathematics test item.

Furthermore, by employing an oral test modality, this may provide an alternative measure in addressing reading difficulties faced by the pupils (Arsaythamby et al., 2021; Spiel et al., 2019; Zaleha et al., 2020). The use of orally delivered tests may enable pupils to listen and comprehend the mathematics test items whilst reducing their reliance on reading and decoding the written text (Stansfield, 2011). This may help reduce the impact of reading difficulties experienced by the *Orang Asli* pupils when administered assessments. The studies by Kanageswari et al. (2021) and Arsyathamby et al. (2021) primarily focused on how the oral delivery of mathematics tests in the ethnic language of *Orang Asli* pupils, *Bahasa Temiar*, had benefited *Orang Asli* pupils. In addition, Zaleha et al. (2020) examined the impact of administering oral mathematics tests in the language of instruction, *Bahasa Melayu*, which have also been found to improve the mathematics performance of *Orang Asli* pupils when administered.

Although these studies separately explored the advantages of a specific language delivered orally in mathematics testing, there is a need to conduct empirical comparisons between the two languages to better understand their relative effectiveness when incorporated as orally delivered test to improve mathematics performance. Moreover, as there is also a notable lack of research investigating how the oral modality of test delivery affects key linguistic features such as item length and vocabulary type, this comparison is critical in better understanding the role of language and test modality in addressing the linguistic challenges on the test performance among the *Orang Asli* pupils. By considering the issues faced by the *Orang Asli* pupils and the notable gap in current literature, this study suggests the potential use of recorded oral tests, both in the language of instruction and the ethnic language, as a possible test accommodation to provide a fairer and more reliable measure of *Orang Asli* pupils' mathematics performance.

While some studies have shown significant improvements in pupils' performance with the use of recorded oral tests (Spiel et al., 2019; Spencer et al., 2022), others have found such a measure to be ineffective compared to other means of orally delivered tests (Caviola et al., 2021; Kim, 2016). This inconsistency points to a gap in the existing literature, underscoring the need for further investigation of the recorded oral test as a viable test accommodation, particularly in the context of

Orang Asli pupils, where language barriers and reading difficulties play a critical role in affecting their mathematics performance. By addressing this gap, the present study aims to provide a more nuanced understanding of how recorded oral tests in the language of instruction and ethnic language can serve as a potential test accommodation that could ensure an unbiased and standardised testing environment (Stansfield, 2011), contributing valuable insights to the ongoing debate on the effectiveness of this measure.

Lastly, there is a dearth of research on teachers' perspectives regarding the implementation of such test accommodations for most pupil populations (Clark-Gareca, 2016; Crawford & Tindal, 2018; Wolf & Leon, 2009), especially in the context of *Orang Asli* pupils. The issues faced by the *Orang Asli* pupils in mathematics testing underscores the need for exploring the mathematics teachers' views on the use of recorded oral tests as a test accommodation strategy in mathematics testing as implementors of assessment strategies in classroom-based assessments. By exploring their unique perspectives on incorporating test accommodations, its effectiveness relies not only on their empirical validity, but also on their practical utility when incorporated with mathematics assessments. Therefore, there is a need to gain a better understanding of the perspectives of mathematics teachers who will be responsible for administering the test accommodations. This may help further establish the overall validity of the recorded oral test as a promising test accommodation which caters to the educational needs of the *Orang Asli* pupils.

1.3 Research Objectives

This study aims to examine the mathematics performance of the *Orang Asli* pupils using the recorded oral tests as a test accommodation on item length (short, moderate, and long item length) and vocabulary type (general vocabulary and mathematics vocabulary). The mathematics performance of the pupils was compared between the written test in *Bahasa Melayu* (WBM), recorded oral test in *Bahasa Melayu* (OBM) and recorded oral test in *Bahasa Temiar* (OBT). The linguistic features for both item length and vocabulary type were examined between the three mathematics tests.

In addition, this study further explores the mathematics teachers' views on the use of the recorded oral test as a possible test accommodation in assessing the mathematics performance of the *Orang Asli* pupils. This could provide the necessary evidence to further attest to the validity of using recorded oral test and to identify the factors and potential challenges for its implementation when used as a test accommodation in mathematics testing. Therefore, the objectives of this study are as follows:

1. To determine the difference in the mathematics performance of the *Orang Asli* pupils between the three mathematics tests (WBM, OBM, and OBT tests).
2. To determine the difference in the mathematics performance of the *Orang Asli* pupils by item length categories (short, moderate, and long item length) between the three mathematics tests (WBM, OBM, and OBT tests).

3. To examine the relationship between word count of vocabulary type (general vocabulary and mathematics vocabulary) and item difficulty of the test items across the three mathematics tests (WBM, OBM, and OBT tests).
4. To explore the teachers' views on the use of the recorded oral tests as test accommodations to facilitate the mathematics performance among *the Orang Asli* pupils.

1.4 Research Question

This study addresses the following research questions:

1. Is there a significant difference in the mathematics performance of the *Orang Asli* pupils between the three mathematics tests (WBM, OBM, and OBT tests)?
2. Is there a significant difference in the mathematics performance of the *Orang Asli* pupils by item length categories (short, moderate, and long item length) between the three mathematics tests (WBM, OBM, and OBT tests)?
3. What is the relationship between word count of vocabulary type (general vocabulary and mathematics vocabulary) and item difficulty of the test items across the three mathematics tests (WBM, OBM, and OBT tests)?
4. What are the teachers' views on the use of the recorded oral tests as test accommodations to facilitate the mathematics performance among *the Orang Asli* pupils?

1.5 Research Hypothesis

The null hypotheses for the research questions established for this study are as follows:

1.5.1 Research Question 1

Is there a significant difference in the mathematics performance of the *Orang Asli* pupils between the three tests (WBM, OBM, and OBT)?

H_o1: There is no statistically significant difference in the mathematics performance between the three tests (WBM, OBM, and OBT).

1.5.2 Research Question 2

Is there a significant difference in the mathematics performance of the *Orang Asli* pupils by item length (short, moderate, and long) between the three tests (WBM, OBM, and OBT)?

H_o2a: There is no statistically significant difference in the mathematics of performance for short item length test items between the three mathematics tests (WBM, OBM, and OBT tests).

H_02b: There is no statistically significant difference in the mathematics of performance for moderate item length test items between the three mathematics tests (WBM, OBM, and OBT tests).

H_o2c: There is no statistically significant difference in the mathematics of performance for long item length test items between the three mathematics tests (WBM, OBM, and OBT tests).

1.5.3 Research Question 3

What is the relationship between vocabulary type (general vocabulary and mathematics vocabulary) and item difficulty of the test items across the three mathematics tests (WBM, OBM, and OBT tests)?

H_o3a: There is no statistically significant relationship between general vocabulary and item difficulty across the three mathematics tests (WBM, OBM, and OBT tests).

H_03b: There is no statistically significant relationship between mathematics vocabulary and item difficulty across the three mathematics tests (WBM, OBM, and OBT tests).

1.6 Research Significance

The outcomes of this study bear considerable significance, as it adds to the body of literature on the validity and use of employing recorded oral test as a test accommodation to mitigate the language barriers and reading difficulties encountered by *Orang Asli* pupils in mathematics assessments. By examining the effectiveness of using test accommodation, this study highlighted the viability of using such measure to reliably assess the mathematics performance of the pupils. This research holds promise of providing valuable insights to the general educational research community on the implication of incorporating test accommodations that considers the cultural and linguistic context of culturally diverse pupils, such as the Indigenous pupils in mathematics education, as advocated by the International Test Commission (ITC) (2019).

In the broader educational landscape, this study could also serve as a strategic resource for the Malaysian Ministry of Education and education policymakers, offering a comprehensive perspective on the predicaments faced by *Orang Asli* pupils who lacks the required proficiency in the language of instruction and experience reading difficulties that have impacted their mathematics performance. The findings could provide education policymakers with empirical evidence to support the inclusion of recorded oral test as a test accommodation in mathematics testing for classroom-based assessment in *Orang Asli* primary schools throughout Malaysia, enhancing accessibility and inclusivity for *Orang Asli* pupils who face language and literacy barriers. This evidence could inform strategies for addressing linguistic challenges and improving equity in educational outcomes by tailoring assessments to better suit the linguistic needs of diverse pupil populations.

In addition to this, test developers who are responsible in developing mathematics test items that are culturally responsive to the Indigenous pupils could benefit from the findings of this study in the context of item writing. Specifically, this study underscores the necessity of embracing cultural validity in the development of mathematics assessments. Grounded in the insights by Nortvedt and Buchholtz (20180, and Trumbull and Nelson-Barber (2019), such measures undertaken may provide a fair and valid evaluation of the *Orang Asli* pupils' performance through mathematics assessments that consider their unique cultural experiences and ethnic language. This perspective emphasizes the importance of considering the language used in the development of assessments and oral test modality as means of mitigating the language and reading demands posed by standard written tests. Furthermore, this study urges test developers in Malaysia to implement such alternative testing measures and potential use of test accommodations that specifically caters to the *Orang Asli* pupils in order to enhance validity and create fair mathematics tests that could accurately assess their mathematics skills and knowledge.

Apart from that, the findings of this study would be significantly beneficial within the realm of mathematics education, particularly for classroom mathematics testing practices for teachers. This study aims to empower mathematics teachers by offering insights into the design of constructive mathematics assessments within classroom assessment practices. The proposed measure that utilises language inclusive practices in recorded oral test as a viable test accommodation are designed to help mathematics teachers to develop accessible and fair mathematics testing practices that caters to the *Orang Asli* pupils. Furthermore, as mathematics teachers acknowledged the practical utility of using the recorded oral test in the language of instruction, this approach could reduce inconsistencies in test delivery and allow for a more standardized and equitable assessment method which can be easier to implement in classroom-based assessments.

For the *Orang Asli* pupils themselves, this study holds the promise of fostering a more inclusive and culturally relevant mathematics education. By advocating for the use of their ethnic languages and oral modality in mathematics testing, the study seeks to diminish the language barriers and reading difficulties that may impede their performance. The possible incorporation of their ethnic language in recorded oral test aims to not only provide a familiar linguistic context in mathematics education, but also as a means of assessing their performance in the language they are more familiar in while providing them sufficient time to develop the necessary language proficiency and reading skills in the language of instruction. This approach, if implemented successfully, has the potential to not only improve their overall mathematics performance when assessed, but also to preserve the unique cultural identities within the current education system.

1.7 Research Limitations

This study initially assumed that the *Orang Asli* pupils were abreast with the current mathematics content knowledge that have been taught through the curriculum which have been established by the Ministry of Education. However, as the school closures were implemented during the Covid-19 pandemic, pupils in the *Orang Asli* primary schools had suffered significant learning loss (Hofer et al., 2022). The school closures for two years had negatively impacted the *Orang Asli* pupils, as the teaching and learning processes could not be conducted due to the lack of technological resources needed to carry out online classes (Noor Azlan & Noor Asyhikin, 2022). Therefore, during this period of school closures, mathematics modular learning in the form of written worksheet were used to conduct mathematics teaching and learning among *Orang Asli* pupils (Kanageswari et al., 2023)

One limitation for this study is the selection of Year 4 *Orang Asli* pupils as the primary sample. The decision was influenced by concerns that pupils in the lower grades might not have acquired foundational knowledge and experiences due to the disruptions caused by the Covid-19 pandemic-related school closures (Kanageswari et al., 2023; Lopez et al., 2022). With school closures lasting for about two years, the pupils from lower grade levels were unable to learn the new syllabus as they progressed to the next grade level.

Furthermore, advice obtained from mathematics teachers suggested the use of Year 4 pupils as the main sample for this study, rather than pupils from the lower primary levels due to their inability to read, write and count. The potential learning loss and the limited formal education during the Covid-19 period may have led to the underdevelopment of the pupils' literacy skills such as reading and writing, as highlighted by Noor Azlan and Noor Asyhikin (2022). Therefore, this study was limited to selecting Year 4 pupils as they obtained lower primary school foundational mathematical knowledge, prior to the school closures.

Based on these rational, this study was focused on evaluating the foundational mathematics knowledge and skills among *Orang Asli* pupils by adapting mathematics test items from Year 4 examinations that corresponds to the eight Year 1 mathematics topics. The topics selected were considered crucial in assessing the pupils' fundamental understanding and proficiency in mathematics. This limitation posed a notable match between the cognitive development of the Year 4 pupils who did not attend school for two years since Year 1, and subsequently the selection of topics for the mathematics tests was therefore confined to Year 1 topics that represents the basic mathematics skills and abilities that are required for Year 4 mathematics test items. The selection of these topic although posed a limitation to this study, was supported by the mathematics teachers employed for content validity during the development of the mathematics test instruments for this study.

Another notable limitation lies in the number of test items selected for the mathematics tests designed to assess the performance of the *Orang Asli* pupils. Although comprising 40 mathematics test items, categorised according to the item length categories of short, moderate and long item length, these test items that were carefully selected was deemed sufficient to capture key aspects of their mathematical abilities while minimizing potential fatigue or disengagement during the testing session. Moreover, the decision to limit the number of items for each item length category was driven by the aim to prevent cognitive strain on the *Orang Asli* pupils during test administration. Therefore, the number of test items were limited to 40 items with approximately equal number of test items across the three item length categories to ensure limited cognitive burden on the pupils.

Another significant limitation of this study was the constrained selection process for the sample, primarily influenced by the geographical distribution of the *Orang Asli* communities belonging to the *Temiar* ethnic group. The study's scope was initially confined to specific Malaysian states—Perak and Kelantan—chosen due to their concentrated population of the *Temiar* ethnic group, as reported by the *Jabatan Kemajuan Orang Asli* (JAKOA) in 2023. This geographical restriction further narrowed down the selection of districts within these states.

For example, the study focused on Sungai Siput in Perak and Gua Musang in Kelantan, as these areas had the highest concentration of settlements inhabited by the *Temiar* ethnic group. This limitation further extends to the choice of primary schools within these districts. The study exclusively targeted primary schools approved by the Ministry of Education, specifically those catering to *Orang Asli* pupils from the *Temiar* ethnic group in Sungai Siput, Perak, and Gua Musang, Kelantan.

The dialectic differences of the ethnic language were also a limitation whereby only primary schools with pupils who converse in the dialect similar to the dialect used in the development of the recorded oral test were selected. There is a clear distinction between the dialects used by the *Temiar* ethnic group in Northern Peninsular Malaysia (Kedah, Perak, and Kelantan states) and Central Peninsular Malaysia (Pahang and Selangor state) (Benjamin, 2012). The sample consisting of *Orang Asli* pupils within the districts of Sungai Siput, Perak and Gua Musang, Kelantan are situated in Northern Peninsular Malaysia. Therefore, the ethnic dialect of *Bahasa Temiar* from *Temiar* ethnic group within the Northern region of Peninsular Malaysia was utilised in the development of the recorded oral test. Therefore, the findings of this study cannot be generalised to other Indigenous ethnic groups within Peninsular Malaysia as the cultural, linguistic, educational, and social contexts of the *Temiar* ethnic group may differ significantly from those of other Indigenous groups.

1.8 Conceptual Framework

For the conceptual framework of this study, an intricate examination of three distinct independent variables was undertaken to unravel their impact on mathematics test scores among *Orang Asli* pupils. The first independent variable, mathematics tests, involved a comparative analysis of three mathematics tests, which are written test in *Bahasa Melayu* (WBM), recorded oral test in *Bahasa Melayu* (OBM), and recorded oral test in *Bahasa Temiar* (OBT). Each test was scrutinized individually, allowing for a nuanced exploration of the diverse mathematical skills and knowledge they assess.

This multifaceted approach aimed to discern any differences between these tests that might be exerted on the performance of *Orang Asli* pupils, shedding light on the potential influence of utilising test accommodations on the pupils' mathematics performance. The second independent variable, item length categories, delved into the examination of item length across the tests. Test items were categorized into short, moderate, and long items lengths, providing a comprehensive evaluation of the potential impact of item length on test performance. This categorization facilitated a systematic investigation into whether the cognitive demands associated with varying item lengths contributed significantly to differences in mathematics test scores.

Lastly, the third independent variable, vocabulary type, introduced the examination between general vocabulary and mathematics vocabulary within the context of the three mathematics tests. This dimension allowed for an investigation into the potential influence of vocabulary on mathematics test scores, discerning whether the type of vocabulary might affect the performance of *Orang Asli* pupils in mathematics tests. The conceptual framework for this study is illustrated in Figure 1.1.

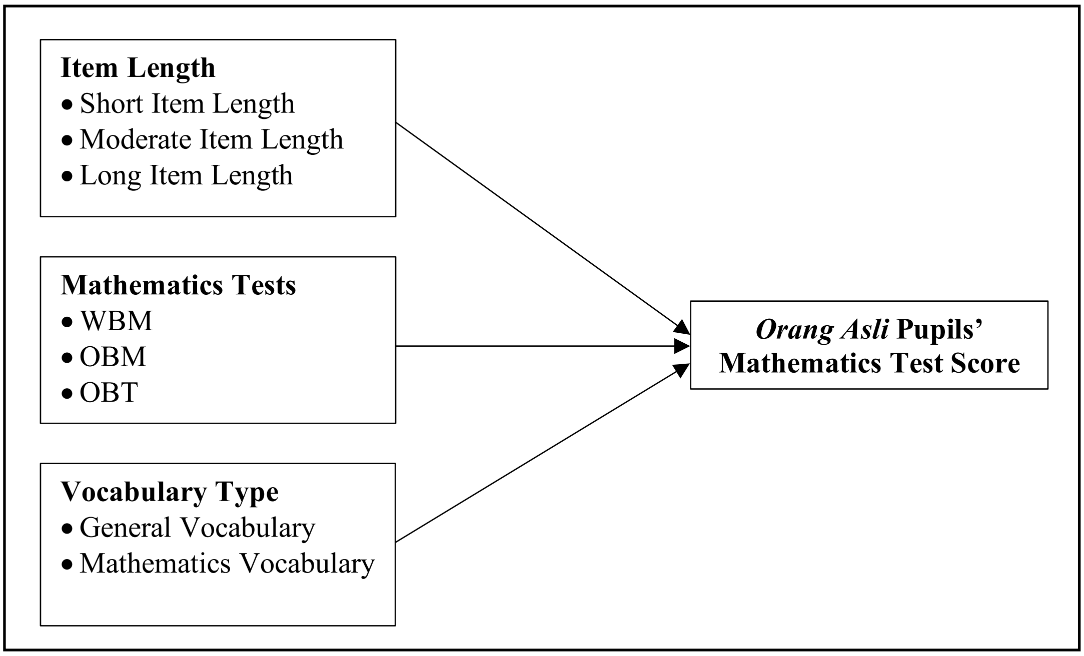


Figure 1.1 Conceptual Framework

1.9 Underlying Theory

For the present study, the Cognitive Load Theory was deemed to be a suitable theory to guide this study in examining the mathematics performance of the *Orang Asli* pupils using the recorded oral test as a test accommodation. The underlying theory guiding this study is presented in the following subsection.

1.9.1 Cognitive Load Theory

The Cognitive Load Theory (CLT) (Sweller, 1991, 1994) provides a framework that could be used to understand the impact of the language and literacy barriers with the linguistics features that may pose difficulties to the cognitive ability of the *Orang Asli* pupils during the mathematics assessment. CLT highlights that cognitive overload could potentially lead to the inability of an individual to accurately process, understand, and retain information (Puma & Tricot, 2019; Sweller, 2010). Within the theory, cognitive load involved in processing information could be categorized as: (a) intrinsic load, which represents the difficulty and complexity embedded within a task (relevant to learning and performance), (b) extraneous load, which are the irrelevant information that may or may not pertain to the task (irrelevant to learning and performance), and (c) germane or effective cognitive load, which refers to the cognitive capacity of storing information by processing and automating the knowledge gained into schemas that are processed and stored in the long-term memory (Sweller, 2010).

Intrinsic load delineates the inherent difficulty or complexity inherent in the material under study or processing information embedded within a particular task. It encompasses the level of intricacy associated with comprehending and mastering specific concepts, irrespective of the mode of presentation or instructional approach (Korbach et al., 2018; Paas et al., 2016). In the context of this study, the linguistic features of item length and vocabulary type (general vocabulary and mathematics vocabulary) are regarded as intrinsic load, demanding pupils to establish connections and comprehending the test items to deduce the appropriate mathematical procedure (Ngu et al., 2016).

Research has indicated the most pupils may encounter intrinsic load when confronted with linguistically dense test items, necessitating heightened cognitive processing to integrate various elements, including numerical ability, procedural knowledge and linguistic skills, to fully grasp and resolve mathematical problems embedded in the test items (Haag et al., 2013; Phan et al., 2017; Walkington et al., 2019). Consequently, offering language and literacy support in the form of a test accommodation has the potential to mitigate the intrinsic load arising from linguistic features such as item length and vocabulary in mathematics test items. As these features play a pivotal role in connecting interacting elements, the application of test accommodations may serve to moderate the impact of these linguistic features when assessing the mathematical performance, as highlighted by Walkington et al. (2015).

Apart from that, extraneous cognitive load is not intrinsic to the pupils' ability to exhibit their mathematical knowledge, and therefore, its reduction has been noted to be beneficial for assessing pupil's performance (Gupta & Zheng, 2020). By reducing extraneous cognitive load, more cognitive resources could be diverted into the working memory to manage the intrinsic load embedded within the problem or task (Sweller, 2016). The inability of pupils to comprehend the written test items within the mathematics test, particularly when it is not relevant to the construct being measured by the test, could be considered extraneous cognitive load which may interfere with their ability to solve the test item (Gillmor et al., 2015). The mismatch in the language used for the assessment may increase extraneous load that may limit the capacity of the pupils' working memory when they are required to understand the context of the test items in a language that they are yet to be proficient in (Roussell et al., 2017).

Additionally, the use of different test modalities, such as written versus oral tests, plays a critical role in managing extraneous cognitive load. Written tests, especially when they contain complex language, can introduce extraneous load that may hinder pupils' ability to focus on solving the mathematics problems (Gillmor et al., 2015), as more effort is required to decode and comprehend the written text. In contrast, using recorded oral test that provides an oral modality could potentially help reduce this extraneous load, allowing pupils to process the test items more efficiently by minimizing the cognitive effort required to decode the written language. This shift in modality ensures that their cognitive resources are directed towards the intrinsic demands of the mathematics tasks rather than being taxed by the test format itself.

Germane load refers to the cognitive resources dedicated to processing, constructing, and automating schema. In the context of learning, germane load is beneficial because it enhances the learner's ability to comprehend and retain information, thus facilitating deeper understanding that important for developing expertise. However, germane load is influenced by the learner's engagement and motivation, as well as the instructional design and strategies used to present information.

When instructional materials are complex or poorly designed, or when learners lack motivation or prior knowledge, germane load can be negatively affected, resulting in less effective schema construction. Germane load is more related to long-term learning and schema development, which are not the immediate focus of assessment performance. Thus, germane load is excluded from this study to allow for a more precise examination of the direct factors that impact *Orang Asli* pupils' performance on mathematics assessments, specifically in relation to language and test modality on the linguistic features of item length and vocabulary.

Despite the positive role of germane load in learning, it should be excluded from consideration in this study because it is not a primary factor for influencing pupils' mathematics performance during the administration of mathematics assessments. The focus of the study is on understanding how the language used in the mathematics test (language of instruction versus ethnic language) and test modality (written versus oral) impacts *Orang Asli* pupils' performance. Therefore, in this context, intrinsic load refers to the linguistic features of item length (short, moderate and long items) and vocabulary type (general vocabulary and mathematics vocabulary) that are embedded within mathematics test items.

On the other hand, extraneous load is the cognitive load that stems from the language use (language of instruction vs. ethnic language) and test modality (written vs recorded oral test). These components directly affect how *Orang Asli* pupils are able comprehend and solve the mathematics test items, which would subsequently affect the pupils' performance when assessed. Figure 1.2 presents the relationship between the variables established in the conceptual framework for this study with the relevant cognitive component established under the Cognitive Load Theory.

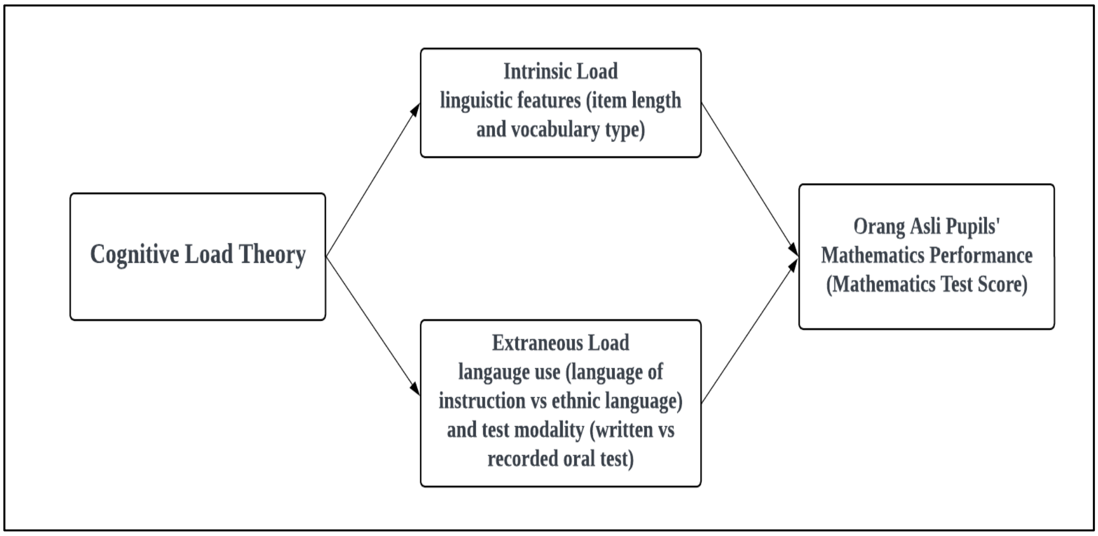


Figure 1.2 Integration of Conceptual Framework and Cognitive Load Theory

As pupils continue to acquire and comprehend mathematical concepts during classroom instruction, the implementation of test accommodation plays a vital role in reducing extraneous cognitive load. This allows pupils to allocate more of their cognitive resources toward tackling the intrinsic demands of mathematics test items (Chen et al., 2017; Paas et al., 2016). Therefore, this study proposes that recorded oral tests may provide the necessary cognitive support to overcome language barriers and reading difficulties experienced by *Orang Asli* pupils when administered mathematics assessments. By mitigating these challenges, the use of such measure could enhance their ability to engage more effectively with mathematical content and improve test performance.

1.10 Operational Definition

The definition of key terms has been operationalised in the context of this study. The operational definitions of these terms are presented in the following subsections.

1.10.1 Malaysian *Orang Asli* Pupils

The Malaysian *Orang Asli* are represented by primary school *Orang Asli* pupils from the *Temiar* ethnic group who utilise *Bahasa Temiar* as their main ethnic language at home and *Bahasa Melayu* as the language of instruction in the school setting (Ruzlan et al., 2021). The development of mathematics tests utilises *Bahasa Melayu* as the main language in the tests.

1.10.2 Year 4 *Orang Asli* Pupils

In the context of this study, the Year 4 *Orang Asli* pupils is defined as pupils who have suffered from learning loss which stemmed from the two years of school closures due to the Covid-19 pandemic, as they were unable carry out mathematics learning since they have not attended primary school since Year 1 of primary school during the Movement Control Order (MCO) established under Malaysian Government from March 2020 – December 2021 (Kanageswari et al., 2023).

1.10.3 Language of Instruction

The language of instruction refers to *Bahasa Melayu*, which is used as the formal language in the curriculum for mathematics education during primary schooling. The language is mainly used during the classroom instruction and is used as the language of assessment for content-based subjects, such as mathematics, administered among the *Orang Asli* pupils.

1.10.4 Ethnic languages

The ethnic languages of the *Orang Asli* pupils that are used for the test accommodations in this study are *Bahasa Temiar*, which is the language spoken by the *Temiar* ethnic group in Sungai Siput, Perak and Gua Musang, Kelantan. These languages are typically used by the *Orang Asli* pupils to converse among members of their ethnic group during day-to-day activities both at home and in the school-setting.

1.10.5 Test Accommodation

Test accommodation entails a type of modification made to the test or testing procedure that does not change the fundamental design of the test, or the construct intended to be measured (Sireci, 2020). For this study, test accommodation is defined as the modifications that are made to the presentation of the test items, through oral modality in the form of recorded oral test, without altering the constructs intended to be measured by the test.

1.10.6 Recorded Oral Test

Recorded oral test is defined as audio recorded test items that are played to test takers alongside written test items, which would allow test takers to access the test items through audio presentation, rather than relying on the written test items (Stansfield, 2011). In this study, the recorded oral tests are administered by playing the audio recording of the mathematics test items in *Bahasa Melayu* and *Bahasa Temiar*, which will be accompanied with the written test items in *Bahasa Melayu* during the test administration.

1.10.7 Written Test in *Bahasa Melayu* (WBM)

The written test in *Bahasa Melayu* (WBM) refers to the mathematics test developed in the language of instruction which is comprised of 40 mathematics test items that were administered to the *Orang Asli* pupils. The WBM test represents the standard unaccommodated mathematics test that is typically administered to the pupils for mathematics assessments.

1.10.8 Recorded Oral Test in *Bahasa Melayu* (OBM)

The recorded oral test in *Bahasa Melayu* (OBM) refers to the mathematics test comprising the audio recordings of the test items in the language of instruction, whereby each test item was read aloud, and audio recorded in *Bahasa Melayu* using the WBM test as a script. The OBM test represents the test accommodation examined in this study which was played to the *Orang Asli* pupils during test administration.

1.10.9 Recorded Oral Test in *Bahasa Temiar* (OBT)

The recorded oral test in *Bahasa Temiar* (OBT) refers to the mathematics test comprising the audio recordings of the test items in the ethnic language, whereby each test item was first transadapted to ensure cultural and linguistic equivalence to the *Orang Asli* pupils' cultural and language background. The test items were then read aloud, and audio recorded in *Bahasa Temiar* whilst ensuring total word count was similar with the WBM and OBM tests. The OBT test represents the test accommodation examined in this study, which was played to the *Orang Asli* pupils during test administration.

1.10.10 Item Length

Item length is defined by the total number of words found within the mathematics test items (Bergqvist et al., 2018; White, 2012). In the context of this study, item length examined consists of three distinct categories, namely short, moderate, and long item length.

1.10.10.1 Short Item Length

Short item length test item is defined as a mathematics test item that contains total number of words of less than five words (White, 2012).

1.10.10.2 Moderate Item Length

Moderate item length test item is defined as a mathematics test item that contains total number of words between five to ten words in a one simple sentence.

1.10.10.3 Long Item Length

Long item length test item is defined as a mathematics test item that contain total number of words of more than 10 words in two or more simple sentences (White, 2012).

1.10.11 Mathematics Vocabulary

Mathematics vocabulary refers to words or terms that are primarily used in mathematics and have one or a specific meaning within the context of mathematics (Baumann & Graves, 2010).

1.10.12 General Vocabulary

General vocabulary refers to generalised words or terms that are used across different content areas whereby the meaning changes depending upon the specific context of which it is used (Baumann & Graves, 2010).

1.10.13 Mathematics Performance

Mathematics performance is defined as the total test score of 40 marks that could be obtained from the 40 test items developed for the mathematics tests.

1.11 Chapter Summary

This chapter provided the introduction to this study and discussed the gap in the literature that warrants further research as discussed in the problem statement. In addition, the research significance and limitations of the study were also presented. The research objective and questions were identified to lay the groundwork that would pave the direction of this study. Lastly, the operational definitions that conceptualises this study were presented to establish the parameters of conducting the research for this study.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

The purpose of this chapter is to provide a comprehensive review of literature that informs the research objectives established in this study. The chapter begins by providing a background of *Orang Asli* in Malaysia in section *2.2*. To address the first research objective, the chapter provides a discussion on the influence of language and reading difficulty on mathematics performance which are presented in section *2.3*, *2.4* and *2.5*, while a review of the literature on test accommodation, particularly on the use of oral test and recorded oral test as test accommodations are presented in sections *2.6* and *2.7*. Additionally, the test development process involving test translation and adaptation, and ethnic language assessment as further implemented as test accommodation for *Orang Asli* pupils were further discussed in sections *2.8*, *2.9* and *2.10*.

Next, the second and third research objectives are addressed by reviewing studies on the linguistic features in mathematics tests, particularly with respect to item length and vocabulary type (general vocabulary and mathematics vocabulary), which are presented in section *2.11*. Finally, the chapter addresses the fourth research objective with a review of teachers' views on test accommodation that broadly discusses the factors affecting *Orang Asli* pupils' mathematics performance and challenges encountered by teachers for implementing test accommodations, as presented in section *2.12*. Each of these sections are aligned with the aim and objectives of this study, providing the necessary background to support the research inquiry and identifying gaps in the existing body of knowledge that the study seeks to address.

2.2 Background of *Orang Asli* Community

The Indigenous community or '*Orang Asli*' people is the oldest occupants found in Peninsular Malaysia. However, these communities continue to remain a minority among the Malaysian population today (Noraini et al., 2018). The *Orang Asli* communities are subdivided into three main tribes, namely *Negrito*, *Senoi* and *Proto-Malay* which consist of a total of eighteen ethnic groups (JAKOA, 2021). The *Negrito* tribe is comprised of six ethnic groups: *Kensiu*, *Kintaq*, *Jahai*, *Mendriq*, *Bateq*, and *Lanoh*. The *Senoi* tribe is comprised of six ethnic groups: *Temiar*, *Semai*, *Jah Hut*, *Che Wong*, *Semoq Beri*, and *Mah Meri*. Lastly, the Proto-Malay tribe is comprised of six ethnic groups: *Temuan*, *Jakun*, *Semelai*, *Orang Kuala*, *Orang Seletar*, and *Orang Kanaq*. Figure 2.1 shows the distribution of the *Orang Asli* ethnic groups across Peninsular Malaysia.

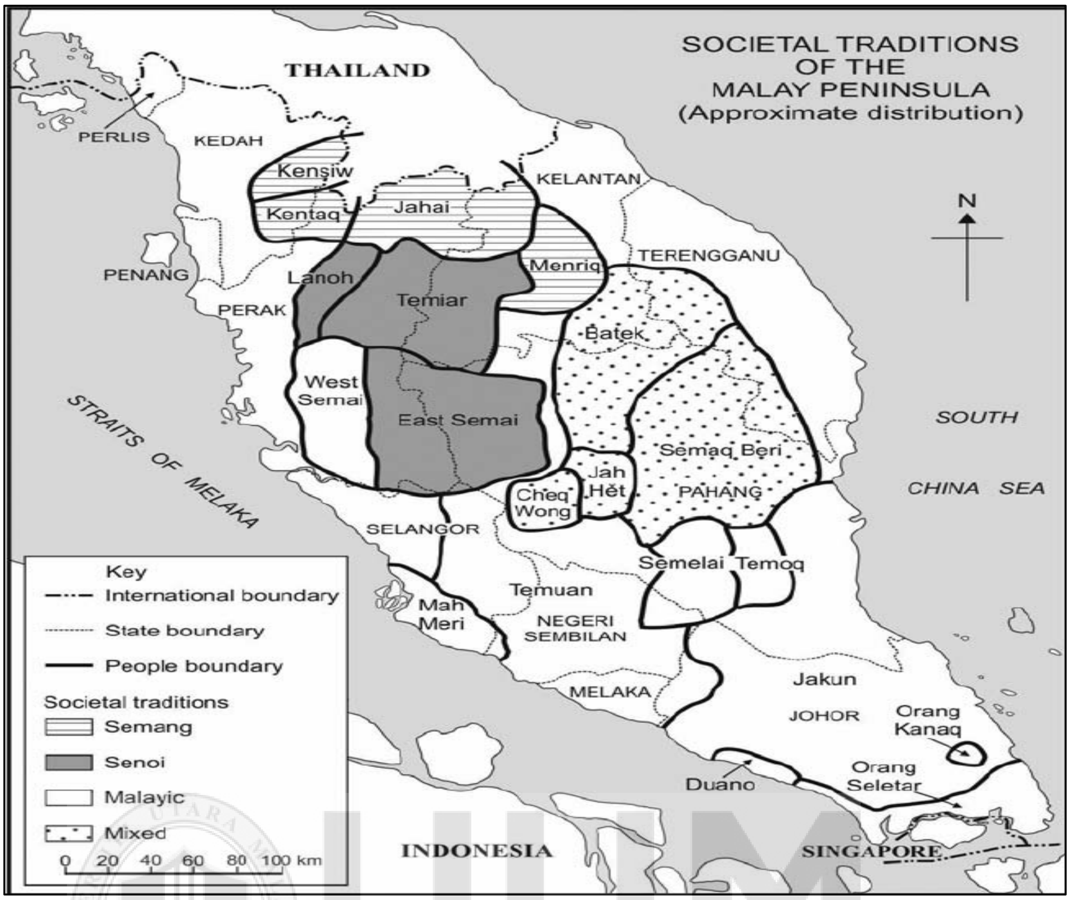


Figure 2.1 Distribution of *Orang Asli* Ethnic Groups (Benjamin, 2012).

The Indigenous *Orang Asli* communities maintain their distinct cultural beliefs and languages, fostering deep bonds among members of their ethnic group (Benjamin, 2012). Despite *Bahasa Melayu* being the national language and the language of instruction for mathematics assessments, the *Orang Asli* pupils persist in using their ethnic languages for everyday communication, even as they commence formal education (Ruzlan et al., 2021). Consequently, there's a need to integrate and safeguard these ethnic languages within the schools' mathematics curriculum to better serve the *Orang Asli* pupils.

Advocates like Trumbull and Nelson-Barber (2019) proposed integrating Indigenous languages into mathematics assessments, aiming to enhance cultural responsiveness for the Indigenous pupils. This approach, as noted by Reid O'Connor and Norton (2022), could help Indigenous pupils to better demonstrate their mathematics knowledge and skills. Moreover, the incorporation of their ethnic languages could establish cultural validity when developing mathematics tests, as supported by Kükea Shultz & Englert (2021) and Ruef et al. (2020), and particularly among *Orang Asli* pupils (Kanageswari et al., 2021).

Among these, the *Temiar* group stands out within the *Senoi* tribe due to its size, cultural preservation, and language. Their sizable population allows for more robust research samples and potentially more statistically significant findings. Their commitment to heritage and language preservation makes them an ideal case for exploring recorded oral test as a possible test accommodation in mathematics assessments. Acknowledging that the *Temiar* ethnic group might not fully represent all Indigenous ethnicities, incorporating their ethnic language in mathematics assessment presents a promising path towards a more inclusive testing practice that is applicable to the wider *Orang Asli* communities (Abdul Hakim et al., 2023; Arsaythamby et al., 2021; Kanageswari et al., 2021). Therefore, examining the use of the recorded oral test that incorporates the language of instruction and ethnic language of the pupils may be crucial, not only for language preservation but also for ensuring equitable and fair assessments when evaluating their mathematics performance.

2.3 Language Influence on Mathematics

The influence of language has been found to be an important component in most content-based assessments, particularly in mathematics assessments (Chow & Wehby, 2018; Erath et al., 2021; Planas, 2018). The language component is as significant during the administration of mathematics assessments, as it is during mathematics teaching and learning in classrooms (Hornburg et al., 2018). Studies have posited that mathematics performance is language-dependent, whereby language comprehension is vital to ensure that pupils are able to understand the content and context delivered through the language used in mathematics assessments (Martin & Fuchs, 2019; Planas et al., 2018), particularly among Indigenous pupils (Barwell, 2020; Miller & Armour, 2021; Robertson & Graven, 2020).

Overall, numerous studies have underscored the influence of language proficiency on mainstream pupils' mathematics performance (Chow & Ekholm, 2019; Delprato, 2021; Peng et al., 2020; Powell et al., 2020). For instance, a meta-analysis of 344 studies by Peng et al. (2020) on the relationship of language and mathematics highlighted that both language proficiency and mathematics skills are developed concurrently. Similarly, Delprato (2021) emphasized the pivotal role of language proficiency in pupils' performance during mathematics assessments. Therefore, proficiency in the language of instruction would significantly aid pupils in grasping the contextual intricacies of mathematical problems presented in test items, as further supported by Powell et al. (2020) and Prediger et al. (2019).

Language proficiency among Indigenous pupils has also been identified as a critical factor that have influenced their performance in mathematics (Robertson & Graven, 2020; Trumbull & Nelson-Barber, 2019). The mismatch between the language of instruction employed in mathematics tests and their ethnic language has given rise to significant language barriers. The disparity between these languages impedes the Indigenous pupils' comprehension of mathematical concepts embedded within the test items (Macqueen et al., 2019; Preston & Claypool, 2021). This in turn may have led to their underperformance in most mathematics assessments as they may have yet to develop the required proficiency in the language of instruction.

The issue of language proficiency is particularly evident among the *Orang Asli* pupils in Malaysia, where the use of *Bahasa Melayu* as the language of instruction, poses a barrier to understanding mathematical concepts during mathematics assessments (Abdul Hakim et al., 2023). This issue has also been further highlighted in the studies by Ruzlan et al. (2021) and Kanageswari et al. (2023), whereby these studies have posited that language barriers faced by the pupils are caused by the mismatch between the language of instruction and their ethnic languages. This have subsequently led to difficulties in conducting mathematics teaching and learning processes in classrooms, which have affected the *Orang Asli* pupils' ability to perform well when administered mathematics tests in the language of instruction (Kanageswari et al., 2021; Arsaythamby et al., 2021).

Hence, for the *Orang Asli* pupils, the close relationship between language and mathematics is evident, and indicates that the proficiency in the language of instruction is vital in allowing pupils to exhibit their mathematical knowledge and skills. However, the language barrier that is faced by these pupils have resulted in mathematics test scores that do reflect their true mathematical ability (Kanageswari et al., 2021; Zaleha et al., 2020). Therefore, by providing a form of language support using test accommodation is vital to ensure that construct irrelevant variance caused by the language barrier do not impede the assessment of pupils' mathematics performance (Abedi et al., 2020; Rios et al., 2020; Stone & Cook, 2018).

2.4 Reading Difficulties on Mathematics Performance in Global Context

The relationship between reading ability and mathematics performance has been extensively documented in various studies, particularly among mainstream pupil populations. Research consistently showed that pupils with lower reading proficiency tended to achieve lower test scores in mathematics assessments compared to their peers with stronger reading skills (Nortvedt et al., 2016; Purpura et al., 2019; Salihu & Räsänen, 2018). This trend has consistently shown the potential reading difficulties experienced by pupils who have yet to develop the required literacy skills which have been found to directly influence their ability to perform well in mathematics assessments, largely due to challenges in comprehending and decoding written test items (Cirino et al., 2018; Gillmor et al., 2015; Haag et al., 2015).

Studies such as those by Hakkarainen et al. (2013) have found that pupils facing reading difficulties tend to perform poorly in mathematics tests. This is primarily because these pupils tended to struggle with understanding the language of the test items, which often include complex word problems, and the lack of the requisite reading ability (Purpura et al., 2019). Without the ability to decode and comprehend these test items, pupils are unable to apply their mathematical knowledge effectively. Koponen et al. (2018) and Salihu et al. (2018) further corroborate these findings by indicating that decoding abilities are essential for both reading and mathematics performance. Their research highlighted that reading difficulties, caused by an inability to decode written text, are significant predictors of low mathematics performance. Furthermore, these reading difficulties are not only present in early primary education, but continues to impact pupils as they progress to higher grade levels (Salihu et al., 2018).

Furthermore, as studies by Ding and Homer (2020) and Nelson and Powell (2018) also reported that pupils with lower reading proficiency tend to struggle more in mathematics assessments, these studies emphasized the interplay between reading comprehension and mathematics performance and suggested that interventions aimed at improving reading skills could improve mathematics performance. Given this interdependence, test accommodation was viewed as a suitable assessment strategy that could provide the needed linguistic support, such as reading the test aloud or simplifying the language used in test items to mitigate the impact of reading difficulties on mathematics performance (Gandhi et al., 2018; Giusto & Ehri, 2019; Walkington et al., 2018). This approach could offer pupils the necessary support to access the mathematical content being assessed, reducing the cognitive load associated with reading difficulties when decoding the written text.

2.5 Reading Difficulties on Mathematics Performance among *Orang Asli* Pupils in Malaysian Context

While the issue of reading difficulties impacting mathematics performance is well-documented among the mainstream pupil populations, it is important to recognise that these challenges are not only unique to this group of pupils. Indigenous pupils, who often face additional cultural and linguistic barriers, may also experience similar reading difficulties (Chamberlain & Medina, 2020; Miller, 2018; Rigney et al., 2020). This gap is particularly evident in the Malaysian context, where studies on *Orang Asli* pupils have highlighted the unique challenges they face, especially in terms of reading and its impact on mathematics performance. Research conducted among *Orang Asli* pupils has shed light on how reading difficulties affect their ability to perform well in mathematics assessments, revealing a pressing need to establishing measures that could mitigate this issue (Kangeswari et al., 2023; Zaleha et al., 2020).

Ruzlan et al. (2021) explored the perspectives of mathematics teachers in *Orang Asli* primary schools and found that reading difficulties significantly affect the mathematics teaching and learning process, subsequently influencing pupil performance. Teachers reported that pupils often struggle to understand mathematical problems because of their limited reading skills, which hinders their ability to engage with the mathematics content. This finding is also supported by Kanageswari et al. (2023), who noted that reading difficulties are prevalent among *Orang Asli* pupils and continue to contribute to their lower scores in mathematics assessments. The study by Khan et al. (2013) further emphasizes the issue of reading difficulties among *Orang Asli* pupils, specifically focusing on the *Semai* ethnic group. The research found that the reading ability of these pupils is relatively poor, which corroborates with other findings that suggest a strong link between reading difficulties and poor academic performance.

These studies highlight the interconnectedness of reading and mathematics skills and suggest that addressing reading difficulties among *Orang Asli* pupils could lead to better outcomes in mathematics performance. The challenges faced by *Orang Asli* pupils mirror those observed in the broader research contexts, where inadequate literacy support leads to lower test scores in mathematics. This issue underscores the need for enhanced literacy support and assessment strategies that cater to the specific needs of *Orang Asli* pupils (Kanageswari et al., 2021; Zaleha et al., 2020). Given that mainstream pupils also experience similar issues, implementing test accommodations that provide linguistic support could be a beneficial strategy. This measure could help mitigate the reading difficulties experienced by *Orang Asli* pupils and provide them with the necessary support to achieve better mathematics performance outcomes.

2.6 Test Accommodation

Ercikan et al. (2015) have posited that the there is a significant relationship that could be observed between mathematics and language. It was further suggested that this relationship have resulted in the widening achievement gap between pupils with language difficulties and their mainstream counterparts. Furthermore, studies have also found that the reading ability is also another construct that have affected the validity of mathematics assessments as it is another construct irrelevant variance that are present when assessing mathematical ability (Gilmour et al., 2019; Theens, 2019).

Due to these complexities that stem from the lack of language proficiency used in mathematics assessments and reading difficulties faced by the *Orang Asli* pupils (Ruzlan et al., 2021; Kanageswari et al., 2023), these issues may have led to the pupils' inability to perform well during the mathematics assessments. The resultant test scores obtained from these assessments do not reflect the pupils' true mathematical ability. Hence, incorporating test accommodations in mathematics assessments have been advocated to ensure fairness, and to obtain a more valid interpretations of pupil's test scores (De Backer et al., 2019; Wilner & Mokhtari, 2018).

According to Sireci (2020), test accommodation refers to a type of modification made to the test or testing procedure that does not alter the design of the test, or the constructs embedded within test items that are intended to be measured. In addition, test accommodation is also defined as a measure used in assessments to eliminate construct irrelevant variance in the test, which may otherwise affect the test-takers' ability to display their knowledge and skills on the intended constructs (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014).

In general, test accommodations have been viewed in relation to the pupils' characteristics (disability/non-disability) and the method of assessments (accommodated/standard conditions). Ideally, the test accommodation employed would function in a way that it is aligned to the interaction hypothesis (Haag et al., 2015; Philips, 1994; Sireci et al., 2005) or differential boost hypothesis (Fuchs et al., 2000; Lindstrom, 2010; Peltier & Harrison, 2018).

Interaction hypothesis is based on the relationship that test accommodation could be used to improve test scores of pupils who need it, but not for the pupils who do not need the test accommodation, whereby these pupils would obtain similar test scores with or without the test accommodation (Sireci et al., 2005). On the other hand, the differential boost hypothesis has a more realistic take on the validity of assessments whereby it relaxes the guiding principles of the interaction hypothesis on test accommodations that are deemed invalid if test scores of pupils who do not need the accommodations also increased relative to the pupils who need it (Peltier & Harrison, 2018).

The differential boost hypothesis acknowledges that both pupils with and without language and reading difficulties will in fact benefit from an accommodated test, however, an effective and valid test accommodation is depicted when the test scores of the pupils who need the accommodation are greater than the test scores of the pupils who do not need it (Fuchs et al., 2000). The differential boost hypothesis allows the incorporation of valid and effective test accommodations for mathematics assessment to ensure that pupils could exhibit their content knowledge and mathematical ability without being impeded by the language difficulties that are not intended to be measured in mathematics assessments (Fuchs et al., 2000; Lindstrom, 2010).

The differential boost hypothesis holds relevance when considering test accommodations for *Orang Asli* pupils to overcome the issue of language and literacy barriers. The use of test accommodations that are tailored to their specific needs can notably enhance test scores for pupils facing substantial linguistic challenges. For these pupils, such interventions can serve as a vital bridge, compensating for linguistic-related disadvantages and enabling a fairer assessment of their true knowledge. In contrast, pupils who do not face significant language barriers may not experience the same degree of improvement with the use of such measures since their performance is less hindered by linguistic constraints.

Harrison et al. (2013) highlighted four main criteria that are needed to determine if the test accommodation administered is a valid measure. These four criteria are: 1) change to the typical standard condition practices, 2) no modifications to the academic standards, 3) ability to mitigate the effects of the pupil's disability to facilitate equitable access to the standard curriculum, and 4) provide differential boost for pupils with disabilities. These criteria emphasised the importance of maintaining the validity of the assessment. This is to ensure that any test accommodation selected and implemented does not compromise the essential components of the test, allowing fair and accurate measurement of pupils' abilities.

To ensure that valid test accommodations are selected to measure the pupils' mathematical knowledge, it is important to take into consideration of the distinctive language needs that addresses the linguistic complexities caused by language and reading in mathematics assessments (Abedi et al., 2020; Gilmour et al., 2019). Thus, Rivera et al. (2006) had created a taxonomical framework to determine the type of test accommodations that could be implemented for pupils who require specific test accommodation in the United States, and had classified test accommodations into two distinct categories, which are direct linguistic support and indirect linguistic support.

Direct linguistic support includes the use of the pupils' home or ethnic language (test translation, bilingual test) and English language accommodation or the language of instruction used in the assessments (linguistic modifications, glossary, read aloud), or the combination of the two. On the other hand, indirect linguistic support is the adjustments made to the assessment conditions that enables the pupils to process the language components embedded within the test items without any alteration to the test itself. This may include providing extra time for reading and comprehension or change in the venue of examination (Rivera et al., 2006).

The framework allows the selection of the right test accommodation used in mathematics assessments that is tailored to the needs and characteristics of the pupils intended to be assessed (Gilmour et al., 2019; Sireci et al., 2018). Based on the issues faced by the *Orang Asli* pupils, using test accommodations that could provide direct linguistic support may be warranted to address the language and literacy barriers faced by these pupils when administered mathematics assessments. As the issues faced by the pupils are embedded within the test items, direct linguistic support could aid comprehension of the test items to ensure that they are able to better comprehend the language and written text, in order to exhibit their mathematical knowledge.

2.7 Types of Test Accommodations

When determining the appropriate test accommodation, it is important to consider the pupils' specific learning needs and behaviours (Lai & Berkeley, 2012). Zabala (2005) highlighted the need to gather the relevant information from the administration of formative assessments, individual diagnostic tests, observational data and annual assessments to choose the right test accommodation for the pupils who need it. Moreover, this notion was further supported by Abedi et al. (2020) and Sireci (2020), where the right test accommodation ensure that pupils have an equal opportunity to demonstrate their knowledge and skills by levelling the playing field without being hindered by their limitation or the educational challenges they face. Therefore, the selection of the right test accommodation had been broadly categorized into four main categories: timing or schedule accommodations, setting accommodations, presentation accommodations and response accommodations (Florida Department of Education [FDOE], 2022).

Based on the guidelines established under the FDOE, the timing or schedule accommodations refer to the time and how it is allocated, managed, and scheduled for the pupils during the assessments. The setting accommodations are the changes made to either the location or the environmental conditions of the assessment site. Additionally, the presentation accommodations are the administration of specialized formats of the assessment that is presented to the pupils if they are unable to hear, read or see the assessment. Lastly, response accommodations are accommodations that aid the pupils to respond to the constructs of the assessment by speaking, writing, drawing, or through other ways of expression. Table 2.1 lists some examples of the test accommodations from the four types of test accommodation category that have been approved for the Florida state assessments.

Table 2.1

Categories and Examples of Test Accommodations

Test Accommodation Categories	Examples
Timing/Scheduling Accommodations	• extended time • breaks • schedule adjustments
Setting Accommodations	• physical access • accessibility workstation • preferential seating • special lighting • acoustical treatment
Presentation Accommodations	• restructured print • braille • large print • digital text (e-text) • read aloud by teacher or text to speech • oral test • recorded oral test • graphic enhanced or symbolated text • captioned videos/images • tactile graphics • manipulatives

Table 2.1 (Cont.)

Test Accommodation Categories	Examples
Response Accommodations	• scribe • word processor • word prediction software • braillewriter • voice recorders • voice recognition software • sign language • cued speech • augmented and alternative communication (AAC)

Within the context of this study, to address the issues of the lack of proficiency in the language of instruction and the reading difficulties faced by the *Orang Asli* pupils, the category of test accommodation that was deemed most suited to the needs of the *Orang Asli* pupils was the presentation accommodations. More specifically, the orally delivered test (oral test) or recorded oral test are deemed appropriate as valid test accommodations that could provide a form of direct linguistic support in mathematics assessments for the *Orang Asli* pupils. Therefore, this study considers the delivery of the mathematics test items in the language of instruction and ethnic languages as recorded oral tests that warrants further examination.

2.7.1 Oral Test as a Test Accommodation

The oral test is defined as test items or instruction which are presented orally to the pupils when the test is being administered (Buzick & Stone, 2014; Crawford & Tindal, 2018). The oral test is typically utilised as a form of test accommodation to address the language and literacy barriers faced by pupils during assessments (Cho et al., 2020; Wickstrom et al., 2020). This measure provides the needed language support that could help pupils to better comprehend both the written text and language embedded within the test items (Penner, 2016; Rogers et al., 2019). Various studies have been carried out to determine the effectiveness of implementing oral test to the mainstream pupils (Gandhi et al., 2018; Giusto & Ehri, 2019), and among Indigenous pupils (Zaleha et al., 2020; Arsaythamby et al., 2021).

Among the mainstream pupil population, Bolt and Thurlow (2007) highlighted that pupil encountering reading difficulties displayed significantly improved performance in oral mathematics tests compared to written ones. It was highlighted that the oral test may have reduced the reading demands posed by the mathematics test items, allowing pupils to better comprehend the test items easily without the need to decode the written text. This was further supported in the studies by Gandhi et al. (2018) and Giusto and Ehri (2019) on the effective use of the oral test to address the reading difficulties faced by most mainstream pupils. These studies examined different forms of the oral tests, such as partial oral test where question stems and answers are orally delivered, and full oral test which included the test instruction. Although the use of oral tests used in these studies as test accommodations were not implemented in mathematics assessments, nevertheless, the findings of these studies highlighted the effectiveness of the oral test as test accommodation in addressing the literacy needs of the pupils.

With respect to the *Orang Asli* pupils, Zaleha et al. (2020) highlighted that the *Orang Asli* pupils' mathematics performance had improved when administered the oral mathematics test, compared to the written test. The study found that the *Orang Asli* pupils were able to comprehend the test items easily which were presented orally, which subsequently allowed them to exhibit their mathematics knowledge and skills during the mathematics test. This is similar to the findings by Arsaythamby et al. (2021), whereby the study found that there was a significant difference in the mathematics performance of the *Orang Asli* pupils between the oral test and written test. The findings of these studies for both mainstream, and *Orang Asli* pupils, underscore the effectiveness of employing the oral test as a viable measure to more accurately assess the pupils' mathematics performance.

2.7.2 Recorded Oral Test as a Test Accommodation

Stansfield (2011) defined recorded oral tests as tests where test instructions and test items are audio recorded and are presented audially to the pupils. The recorded oral test is used a standardized testing method that ensures that the audio quality of the test items presented are clear and consistent. By employing the recorded oral test, this measure aims to eliminate variations in audio quality that could affect the fairness and accuracy of the assessment (Stansfield, 2011; Vidal Rodeiro & Macinska, 2022). While most studies have been conducted in examining the use of the recorded oral test particularly among the mainstream pupil population (Kim, 2016; Lopez et al., 2019; Spiel et al., 2019), the findings from these studies could be leveraged to offer crucial insights into its use among the Indigenous pupils, particularly the *Orang Asli* pupils.

The use of the recorded oral test has served as a valuable test accommodation aimed at mitigating reading difficulties experienced by pupils (McMahon et al., 2016; Spiel et al., 2019). For instance, McMahon et al. (2016) investigated the use of the recorded oral test among 47 Grade 6 mainstream pupils who experience reading difficulties in a science test. The study found that the pupils had obtained the highest mean test score when administered the recorded oral test, compared to the unaccommodated test. The study attributed this improved performance to the effectiveness of the recorded oral test as a measure in reducing the reading demands posed by the test items.

Furthermore, this notion is further supported by Spiel et al. (2019) whereby this study examined the use of recorded oral test as a test accommodation among pupils with reading difficulties in a reading test. The study found that the performance of the pupils with reading difficulties had significantly improved when the recorded oral test was administered, compared to the pupils without reading difficulties. Although these studies did not examine the use of the recorded oral test in mathematics assessments, the findings point to the effectiveness of the recorded oral test in reducing the reading demands posed by the test items as pupils are able to comprehend the test items easily without the need to decode the written text, as suggested by Spencer et al. (2022)

However, despite the perceived effectiveness of recorded oral tests, certain studies suggest that its effectiveness might be compromised when compared to the human-delivered oral tests (Caviola et al., 2021; Kim, 2016; Li, 2014). For instance, Li (2014) found that recorded oral tests did not boost the mathematics test scores for most mainstream pupils who experience language and literacy barriers. Instead, the study found that the pupils had obtained improved test scores when the oral test was delivered by a human proctor. This finding is also supported in the study by Kim (2016), where young mainstream pupils were able to perform better in a reading test when administered a human-delivered oral test, compared to the recorded oral test. These studies highlight that the use of human proctors may have benefitted the pupils as they may have provided hints to the answers through conversational cues from the written text. This may have resulted in higher test scores obtained by the pupils.

Nonetheless, this was an issue highlighted by Stansfield (2011) as human proctors may also present the test items according to the social and communicative cues during the assessments. This may allow the pupils to grasp important clues to solve the problem embedded within the test item. By standardizing the oral test using audio recordings, it was proposed that the recorded oral test as a test accommodation could uphold the validity of assessing pupils' mathematical abilities, free from external factors that might unfairly sway their performance. Therefore, it becomes imperative to conduct a comprehensive examination on the use of the recorded oral test as a test accommodation specifically for *Orang Asli* pupils to ensure an equitable and unbiased testing environment that enables them to genuinely exhibit their mathematical knowledge.

2.8 Test Development

Tests or assessments are important tools that are used to monitor and obtain information on the teaching and learning processes, the pupils' academic progress and improvements in the learning outcomes (Pandra et al., 2017). As the test is an important measuring tool, its development should be of good quality and adheres to the highest validity when used as the basis for interpretation of the construct intended to be measured (Clark & Watson, 2019; Haladyna et al., 2002). In relation to mathematics, the test development process is an important step that ensures that the test could validly measure the true mathematics performance of the pupils who took the test.

To ensure test validity, studies have over the years proposed test development frameworks and guidelines that could be employed. The processes adhere to strict protocol in ensuring that the test validly measures what is intended. For instance, Hambleton et al. (2004) developed eight steps in the test development process, which are: developing the test specifications, (2) item pool, (3) pilot study, (4) compilation of norms reference (for norm-referenced test), (5) criterion specification for pass score (for criterion-referenced test), (6) test of reliability, (7) item validity, and (8) development of final test. Other than that, Ramadhan et al. (2019) proposed a nine-step process for test development, which consists of developing the test specification, writing the test, understanding the test, pilot study, item analysis, test revision, reconstruct the test, test administration, and interpreting the test results. This process provided a more comprehensive framework in developing valid tests that measures its intended construct.

However, a more comprehensive and recent iteration of the test development process have been proposed by Lane et al. (2015), which comprised of a twelve-step process. This process could be extended to all types of tests, regardless of the mode of the test (traditional paper-and-pencil or computer-based tests) and test format (selected response, constructed response and performance tasks). These twelve steps are overall plan, content definition, test specification, item development, test design and assembly, test production, test administration, scoring test responses, passing scores, reporting test results, item banking, and test technical report. Overall, the test development processes proposed by these studies provide the necessary guidelines that could be used to develop more effective tests that could maximise the improvements to test validity.

One of the key stages in the test development process is item development, whereby test validity and accessibility is observed by employing best practices (Rodriguez & Albano, 2011). Nonetheless, most of the attention on item writing guidelines have pertained to selected response items, or multiple choice. For instance, Haladyna et al. (2002) proposed a taxonomy consisting of 43 guidelines as the basis for item writing, these guidelines provide a template for general item writing, procedures, item content, formatting and style, writing the stems and writing the options.

Therefore, in the context of test validity, most of these guidelines provide clarity, efficiency in word usage and sentence structure, and good writing practices that should provide improved accessibility to the test takers. In relation to the *Orang Asli* pupils, the test development process proposed by Ramadhan et al. (2019) was used as the main framework in for this study. This ensured that the mathematics tests developed could address the construct irrelevant variance brought upon by both language and reading difficulties that might unfairly impact the mathematics performance of the *Orang Asli* pupils. Hence, through the test development framework employed in this study, valid and fair mathematics tests were developed that specifically cater to the needs of the *Orang Asli* pupils when assessing their true mathematical skills and knowledge.

2.9 Test Translation and Test Adaptation

There has been a growing need for multilingual versions of assessments in recent years (Hernández et al., 2020), whereby it has become crucial when assessing individuals from multicultural and multilingual backgrounds (Duarte & Rossier, 2008). Two terms have been sometimes mutually used in the field of research are test translation and test adaptation (Hambleton et al., 2004; Muniz et al., 2013). According to Ercikan and Por (2020), test translation refers to an instrument developed in a language with a focus of linguistic equivalence to the instrument created in another language. On the other hand, test adaptation is the development of a different language version of the instrument that undergoes a broader process of ensuring cultural and linguistic equivalence to the original instrument.

For cross-cultural comparison, test developers typically opt for test adaptation whereby score comparability could be made between the two different versions of the test without jeopardizing validity (Ercikan et al., 2015). Furthermore, it is more feasible to translate and adapt the test that is validated in another culture, than to create a new test for the same measure (Hambleton & Lee, 2013). Language and culture influence how pupils respond to the test which subsequently affects score comparability (Ercikan & Por, 2020). For instance, in a study by Ercikan et al. (2014), it was found that the pupils' cultural experience outside of school had significantly affected how they understand and respond to items in a test.

The difference in the language used at home and schools could potentially lead to pupils' inability to perform well in the test due to the unfamiliarity for the conditions of the test. Moreover, the complex relationship between the pupils' background and the construct intended to be measured in the assessment could significantly impact the way pupils respond to the test (Thomson, 2018). Hence, to ensure a more valid test adaptation process, the International Test Commission (ITC) had developed guidelines that could ensure well adapted test are developed to a multicultural and language diverse group of test takers (Hambleton, 1996; Hambleton, 2005; ITC, 2018). A total of 18 guidelines were developed and presented in six sections: 1) Precondition, 2) Test Development, 3) Confirmation, 4) Administration, 5) Score scale and interpretation, and 6) Documentation (Hambleton, 2005; ICT, 2018).

2.9.1 Ethnic Language Assessment

Translation of content-based assessments into the pupils' main home or ethnic languages, such as in mathematics, has been found to be beneficial for pupils who are not native speakers of the language used when developing assessments (De Angelis, 2014; Turkan & Oliveri, 2014). Ercikan et al. (2012) had defined translated assessment as the conversion of the language that is expressed from the source language to the target language in a translated version of the assessment without altering the content, or the intended construct. The use of translated assessment is in line with the view that pupils who are second language learners and language minority have the right to be assessed in their respective native or ethnic languages (McCarty, 2021).

In a study by Trumbull and Nelson-Barber (2019), the researchers highlighted the importance of incorporating the ethnic languages of the Indigenous pupils to ensure cultural validity. It was further suggested that efforts should be geared towards the use of the ethnic languages as an effective measure that could establish culturally responsive assessments that acknowledges the rich culture and heritage of the pupils. Therefore, the use of translated assessments as a viable test accommodation meets this demand, whereby the culture and languages of the various ethnic groups should be taken into consideration when developing mathematics assessments that acknowledges the cultural validity of the pupils and allows true construct abilities to be measured in the assessment (Schissel et al., 2021).

Translation of an assessment could be carried out in two formats, which are written translation and oral translation. Written translations as a test accommodation encompasses the use of a bilingual glossary, dual language bilingual test, assessment direction that are translated into the ethnic language and a fully translated written assessment (Solano-Flores, 2012). On the other hand, oral translations could be administered as a suitable test accommodation if pupils have reading difficulties that makes it difficult for them to comprehend the text of the assessments (Lopez, 2020). This can be carried out by employing a bilingual speaker to read aloud the test items in target language, or pre-recorded audio recording of the translated material that can be played to the pupils (International Test Commission (ITC), 2019).

Stansfield (2011) examined the use of oral translations as a test accommodation in the ethnic language of the pupils to highlight the need of addressing the pupils' unfamiliarity to the language of instruction. The researcher made clear distinctions between sight translation and recorded oral translation, whereby the latter was defined as an audio recording that is made of a scripted translation of the assessment items and/or directions in the target language of the pupils, without altering the construct of the assessment. The main difference between the sight translation and recorded oral translation is the method of delivery, whereby the recorded oral translation ensures a standardized delivery of the test accommodation which removes discrepancies caused extraneous factors that may be caused from sight translation, such as intonation and conversational cues (Robinson, 2010; Stansfield, 2011).

In studies that were conducted to investigate the different modes of delivery and uses of translations that were employed as a form of test accommodation in content-based assessments, the effectiveness of using translated assessment, particularly through the oral delivery accommodation of the translated assessment requires further empirical research (Turkan & Oliveri, 2014; Yang, 2020). Further research is pivotal to determine the extent to which oral delivery of translated assessments can genuinely serve as a dependable accommodation method for ensuring equitable evaluation among pupils from diverse linguistic and cultural backgrounds in the

In conclusion, the *Orang Asli* pupils represent a marginalised group of pupils within Malaysian education context facing language discrepancies between their ethnic languages and the language of instruction used in mathematics testing (Abdul Hakim et al., 2023; Hema et al., 2021; Kanageswari et al., 2023). Their unique cultural and linguistic backgrounds necessitate a customized testing approach that surpasses mere evaluation of mathematical skills. By integrating their ethnic language into mathematics assessments, this comprehensive method not only fosters more inclusive testing practices but also offers an avenue for *Orang Asli* pupils to engage with and excel in mathematics education on a more profound level.

2.10 Test Accommodations Among Indigenous Pupils

The mathematics performance of the Indigenous pupils continues to lag their mainstream peers due to factors such as the lack of knowledge and interest in the notion of schooling and assessment (Norwaliza et al., 2017; Tanius et al., 2020). Moreover, it has been suggested that the language discrepancy between the ethnic languages of the Indigenous pupils and the language of instruction used during the teaching and learning process, and in assessments have negatively affected the mathematics performance of these pupils during assessments (Meaney et al., 2017; Van Ryzin & Vincent, 2017).

As many Indigenous pupils might not be proficient in the language used in most standardized testing, past studies have proposed that the language barriers posed by the lack of proficiency in the language of instruction may have resulted in their lower mathematics performance (Preston & Claypool, 2021; Trumbull & Nelson-Barber, 2019). For the Indigenous pupils, one such recommendation for test accommodations has typically been the incorporation of the pupils' ethnic language in mathematics assessments. Therefore, to cater to the linguistic needs of the pupils, there is a need to employ language inclusive measures that could potentially mitigate this language barrier in mathematics tests.

Over the years, various studies have attempted to address this issue using test accommodations for mathematics assessments, which caters to the language needs of the *Orang Asli* pupils (Zaleha et al., 2020; Kanageswari et al., 2021; Arsaythamby et al., 2021). For instance, Zaleha et al. (2020) examined both written and oral mathematics tests in *Bahasa Melayu* on the pupils' mathematics numeracy skills. The study revealed higher test scores obtained in the oral test compared to written tests. The findings of this study highlights that although current consensus indicate that these pupils lack proficiency in the language of instruction, the use oral test in *Bahasa Melayu* did not impede their ability to perform better compared to the written test.

Other than that, Kanageswari et al. (2021) examined a trilingual (*Bahasa Melayu*, English, and *Bahasa Temiar*) and written mathematics tests among Year 5 *Orang Asli* pupils. Through Rasch analysis, the study found comparability between modes, advocating the validity of oral delivery in the pupils' ethnic language. Results highlighted improved performance in the trilingual test, which emphasised the crucial role of the ethnic language in aiding comprehension of mathematics items. Similarly, Arsaythamby et al. (2021) delved into the use of the ethnic language in an oral test that was supplement to the written test in *Bahasa Melayu* for Year 5 *Orang Asli* pupils.

Consistent with prior studies, this study found that the pupils had also performed better when the ethnic language was incorporated in the testing practice. Collectively, these studies advocate for incorporating oral language assessments, either in the language of instruction or the pupils' ethnic languages, when evaluating

Orang Asli pupils' mathematics performance. They propose oral tests with recorded delivery as a valid measure, recognizing that language accommodations are crucial for these pupils to bridge the academic achievement gap with their mainstream peers. These findings challenge traditional assessment norms and highlight the importance of linguistic inclusivity in evaluating academic performance among marginalized groups. Addressing language barriers through tailored assessment approaches could contribute significantly to ensuring equitable educational opportunities for *Orang Asli* pupils.

2.11 Linguistic Features in Mathematic Assessment

To perform well in mathematics assessments, the relationship between both language and reading ability are intertwined (Nagy & Townsend, 2012), whereby the pupils' inability to interpret and process the linguistic features of the test items had subsequently led to their lower mathematics performance (Fang, 2016; Plath & Leiss, 2018). Bergqvist et al. (2018) posited that the tests items developed in mathematics assessments are in fact a measure of both language comprehension and reading ability as these are important components needed to solve the mathematics problem. Therefore, the role of language in mathematics is characterised by the linguistic features of the test items, and its association to reading in mathematics (Cruz Neri et al., 2022).

There are basic assumptions that underlie the role of linguistic features within mathematics test items. For instance, Peng et al. (2020) proposed that linguistic features act to convey information and gather knowledge through the different cognitive domains. On the other hand, Walkington et al. (2018) claim that the pupils' ability to comprehend the complex linguistic features are based on the test takers characteristics, such as their reading ability, and how the test takers perceive the presentation of mathematics items in the assessment. Various studies have shown that certain linguistic features may affect pupils' interpretation of a test item, which could result in the wrong answer (Noble et al., 2014; Sibanda, 2017). Hence, language and reading are underlying factors that makes it challenging for pupils to comprehend items with complex linguistic features (Haag et al., 2015; Riccomini et al., 2015; Sibanda, 2017; Wolf & Leon, 2009).

Generally, for any content-based assessments, a comprehensive study by Kachchaf et al. (2016) identified more than sixty linguistic features that could affect pupils' performance. The study categorised linguistic features typically found in assessments into three distinct levels. The first level is the word-level, which includes frequency of words or vocabulary, general or academic vocabulary, technical terms, context-specific terms, and low frequency technical term. The next level is the sentence-level, which includes subordinate and adverbial clause, subordinate clause, relative clause and the number of verbs and nouns found in the item stem and answer options. Lastly, the final level is the item-level, which includes forced comparison items and reference back items.

Given the complexity of language and reading in content-based assessments, it is essential to examine how specific linguistic features, such as item length encompassing number of words in a sentence or the frequency of words and vocabulary, and vocabulary type which includes general vocabulary and mathematics vocabulary, may influence pupils' understanding of mathematics test items (Haag et al., 2013; Kan & Bulut, 2015; Lin et al., 2021). These linguistic features are commonly associated with language comprehension and reading difficulties, which may hinder pupils' ability to accurately interpret and solve mathematics test items. Therefore, investigating these linguistic features is crucial for understanding how they may ultimately affect pupils' performance in mathematics assessments.

2.11.1 Item Length

Item length is a linguistic feature at the sentence level that is defined as the total word count of a test item (Bergqvist et al., 2018; White, 2012). In mathematics tests, the item length of a test item has been found to affect the mathematics performance of most mainstream pupils, whereby longer test items have been attributed to increased difficulties when pupils are administered mathematics tests. (Haag et al., 2015; Sibanda, 2017; Walkington et al., 2018). As longer test items require adequate proficiency in the language and reading ability that are needed to both decode and comprehend the long test item, deficiency in any of these abilities may impact the pupils' performance when administered mathematics assessments (Chow & Ekholm, 2019; Cruz Neri et al., 2022; Wilkinson, 2019)

Nonetheless, the lack of specific research dedicated to examining the influence of item length on Indigenous pupils, particularly *Orang Asli* pupils, poses a gap in understanding the impact of item length on their mathematics performance. However, by extrapolating from the findings and insights garnered from broader studies within this review, parallels can be drawn to illuminate potential challenges faced by the mainstream pupil population in general. This may help shed light on impact of item length in mathematics assessments that might similarly affect the *Orang Asli* pupils.

In the studies conducted among mainstream pupils, Shaftel (2006) have posited that long mathematics items with higher word count had significantly impacted the pupils' performance. It was deduced that this may have been attributed to the difficulty faced by the pupils to comprehend the longer test items. The increased length of test items frequently correlates to the complexities faced by pupils who experience language barriers or reading difficulties, as proposed by Gillmor et al. (2015) and Haag et al. (2015). This issue often imposes challenges in comprehension due to the linguistic nuances that may impact the pupils' ability to decode and solve the problem (Walkington et al., 2018).

For instance, Sibanda (2017) examined the linguistic difficulties of the mathematics test items in the 2013 Annual National Assessment (ANA) mathematics assessment among Grade 4 mainstream pupils. The study found that longer test items with more than 10 words had negatively affected pupils' mathematics performance on the assessment. It was further highlighted that the low performance caused by the long items may have affected the pupils who might have encountered obstacles in effectively processing and comprehending long test items, which can subsequently impede their accurate interpretation of the questions or tasks at hand, as indicated in the studies by Haag et al. (2015) and Walkington et al. (2015).

Conversely, the study by Cruz Neri et al. (2021) examined the interaction effect of item length and reading comprehension ability among 2051 German pupils in Grade

10 and 11 on a science assessment, which controlled for cognitive ability, ethnic language used at home, gender, and school grade. The study found that for pupils with reading difficulties, there was no interaction effect of item length on the performance of the pupils. It implies that the difficulty posed by longer test items did not exacerbate their challenges in comprehension beyond what was already observed with shorter items (Härtig et al., 2015). In the context of this current study, these findings may suggest that efforts to improve the mathematics performance of the *Orang Asli* pupils may also need to consider other factors, such as their language comprehension of the language of instruction, rather than merely shortening or simplifying the length of test items.

Therefore, to address the influence of item length on the mathematics performance, studies have suggested the use of test accommodations to provide the needed support in addressing the complex linguistic features of test items in assessments (Abedi et al., 2020; Clark-Gareca, 2016; Yang, 2020). For instance, Haag et al. (2015) and Walkington et al. (2015) had implemented linguistic modifications, such as shorter test items and simple language, as a form of test accommodation in addressing the item length of test items in a mathematics assessment. These studies found that the use of test accommodation addressing the item length feature had significantly improved the pupils' performance. Therefore, implementing test accommodations that address the unique language and literacy related barriers faced by the *Orang Asli* pupils could ensure a fairer evaluation of their mathematical knowledge and skills.

2.11.2 General Vocabulary and Mathematics Vocabulary

Another important component of linguistic features that have affected the mathematics performance of pupils is the vocabulary used in mathematics tests items (Nagy & Townsend, 2012). Baumann and Graves (2010) defined vocabulary into two main components, which are academic vocabulary and general vocabulary. Academic vocabulary is defined as vocabulary that is domain specific and are specific to concepts and procedures used in mathematics. According to Thompson and Rubenstein (2000), the vocabulary found specifically in mathematics with words such as 'perpendicular' and 'some', and phrases such as 'twice as many' are typically referred to as academic vocabulary. On the other hand, general vocabulary refers to vocabulary that have more than one meaning across different content areas (Baumann & Graves, 2010). Examples of general vocabulary typically used 'side' (*side* of a triangle or *side* of the road) or 'foot' (*foot* in inches or *foot* of the bed) (Riccomini et al., 2015).

Studies have suggested that proficiency in both mathematics vocabulary and general vocabulary is needed for pupils to grasp the context of mathematics test items (Haag et al., 2013; Powell et al., 2017; Schuth et al., 2017). Mastery of mathematics vocabulary is pivotal as it encompasses specialized terms essential for understanding and effectively communicating mathematical concepts. Without a strong command of mathematics vocabulary specific to mathematical principles, most pupils may encounter difficulties in comprehending complex mathematical problems, interpreting mathematical notations, or articulating their mathematical reasoning coherently (Powell & Nelson, 2017; Virgana & Lapasau, 2019). Conversely, general vocabulary proficiency, while crucial in everyday communication, might not suffice for grasping the intricacies of mathematical language or specialized vocabulary typically required in formal mathematical contexts (Riccomini et al., 2015).

Studies have been carried out to examine the impact of both general and mathematics vocabulary on the mathematics performance of most mainstream pupils in primary schools (Forsyth & Powell, 2017; Powell et al., 2020). Nonetheless, there has been a dearth of empirical research that have been conducted on the influence of the vocabulary types among Indigenous pupils (Jorgensen, 2018; Rigney et al., 2020). The findings from the corresponding studies on the mainstream pupils' mathematics performance in relation to the vocabulary could serve as a valuable foundation for further exploring the impact of general and mathematics vocabulary among the *Orang Asli* pupils.

In the study by Powell and Nelson (2017), it was found that the mathematics performance of Grade 1 mainstream pupils is significantly correlated to the pupils' proficiency in mathematics vocabulary. The study proposed that the development of mathematics vocabulary had enabled the pupils to better grasp the mathematical concepts more easily. This is in line with the findings by Forsyth and Powell (2017) whereby it was found that the pupils' proficiency in mathematics vocabulary was able to mitigate the potential reading difficulties experienced by them during mathematic assessments. Therefore, by having a firmer grasp of the mathematics vocabulary used in the test items, this may help improve their mathematics performance.

This notion was also observed in the study by Unal (2019) whereby the study examined the relationship between general vocabulary, mathematics vocabulary and mathematics computation among 34 mainstream primary school pupils. The findings concluded that mathematics vocabulary aids in the development of both general vocabulary and mathematics computation, indicating that mathematics proficiency is therefore mediated by the development of mathematics vocabulary. This further highlights the importance of mathematics vocabulary not only in the development of mathematical proficiency, but also in their later years of primary schooling as general vocabulary development begins to become less important in affecting the pupils' mathematics performance. This notion has also been found to be in line with other studies, namely by Powell et al. (2020), Purpura et al. (2019) and Schuth et al. (2017).

In linguistically dense and complex mathematics test items, Peng and Lin (2019) highlighted that mathematics vocabulary was a significant contributor to mathematics proficiency and performance among Grade 4 Chinese pupils. In particular, the study found that mathematics vocabulary that was specific to the topics of measurement and geometry was correlated to the pupils' mathematics performance, particularly when such vocabulary use coincides with words and terms used in their everyday setting such as 'four side shapes' (which represent a quadrilateral) or numbers such as '11' (which is termed *ten-one*, rather than *eleven*).

With respect to the impact of vocabulary within longer test items, Haag et al. (2015), Powell and Nelson (2017), and Virgana and Lapasau (2019) found that mathematics test items that are typically longer may contained more technical terms that are specific to the mathematics topics being assessed. It was posited that the additional mathematics vocabulary may negatively impact the pupils' performance when administered these tests, particularly if the pupils are not familiar, or lack the necessary conceptual knowledge of the meanings for those vocabulary or terms. Therefore, for most pupils' who know and understand the specialized mathematics vocabulary, this may have led to their better mathematics performance in the topics that contained these mathematics vocabulary (Haag et al., 2015; Unal, 2019).

Regarding the impact of mathematics vocabulary on the academic performance of Indigenous pupils, Roberts (1998) emphasized the significance of the Indigenous language in conveying mathematical concepts within their communities. This language has developed its distinct vocabulary and grammatical framework for discussing mathematical ideas, posing a challenge for Indigenous pupils in aligning with Western mathematical principles (Meaney & Evans, 2013). Expanding on this, Jorgensen's qualitative case study in 2018 delved into the perspectives of educators and support staff within schools. The study underscored mathematics vocabulary as a pivotal concern affecting the academic achievements of Indigenous pupils in Australia. Notably, it highlighted the pivotal role of language when imparting mathematical concepts in an unfamiliar language to the pupils.

Despite these critical observations, there remains a notable dearth of empirical research addressing the influence of vocabulary specifically among Indigenous pupils in mathematics (Meaney & Trinick, 2020; Moschkovich, 2020), particularly within the context of mathematics assessments. This absence of empirical studies signifies an essential gap in comprehending the intricate interplay between language, vocabulary type, and mathematics performance of the Indigenous pupils. Therefore, in the context of this study, the influence of vocabulary, particularly with respect to general vocabulary and mathematics vocabulary, underscores the pressing need for comprehensive studies to inform more effective assessment strategies such as the use of test accommodations that are tailored to their unique linguistic and cultural backgrounds, particularly for *Orang Asli* pupils who grapple with language and literacy barriers.

2.12 Teachers' Views on Employing Test Accommodation

This section discusses teachers' perspectives on the use of test accommodations, which is organized into two subsections; factors affective mathematics performance among *Orang Asli* pupils, and challenges of implementing test accommodations among teachers. The review of the literature aims to explore how these factors and challenges shape the effectiveness and practicality of using test accommodations for *Orang Asli* pupils in mathematics assessments.

2.12.1 Factors Affecting Mathematics Performance among *Orang Asli* Pupils

The mathematics performance of *Orang Asli* pupils is influenced by a myriad of interconnected factors that have continued to affect their mathematics performance over the years (Abdul Halim, 2022; Ruzlan et al., 2021; Kanageswari et al., 2023). For instance, Ruzlan et al. (2021) explored the perspectives of mathematics teachers on the possible factors that affected mathematics teaching and learning among *Orang Asli* pupils. The study found that factors such as pupils' attitude and disengagement, and the lack of parental involvement in the pupils' educational progress had made it difficult for teachers to teach effectively during classroom mathematics teaching.

These findings are also in line with Aidil Fitri et al. (2020), Mohd Azizul et al. (2020) and Mohammad Safwat et al. (2020), whereby these studies further highlighted the need for teacher to be more proactive in stimulating engagement of *Orang Asli* pupils in classrooms. Although these studies were conducted in classroom of other subjects, the *Orang Asli* pupils' attitude and their parental involvement in the education of their children were some common factors across these studies, which further proposed that this may have led to the possible decline in the academic performance of the pupils (Hamidah et al., 2017; Sumathi, 2016).

Furthermore, Kanageswari et al. (2023) carried out a qualitative study among five mathematics teachers, who identified several key challenges that impacted their mathematics teaching practices in *Orang Asli* classrooms. The teachers emphasized the lack of technological resources as a major barrier, making it difficult to incorporate innovative instructional methods into their lessons. Additionally, the teachers pointed to language and literacy challenges, particularly the language barrier between the pupils' ethnic language and the language of instruction and the possible reading difficulties that are faced by the majority of *Orang Asli* pupils in the primary schools.

These issues were viewed as significant obstacle when conducting classroom teaching and learning processes, in line with the study by Salim et al. (2022). The pupils' limited proficiency in the language of instruction often made it difficult for them to fully grasp mathematical concepts, which contributed to their lower performance in mathematics assessments. These challenges were echoed in the findings of Aidil Fitri et al. (2020) and Ruzlan et al. (2021), whereby the combination of technological constraints and language difficulties may have adversely affected the performance of *Orang Asli* pupils when they were administered mathematics tests.

In conclusion, the educational landscape for *Orang Asli* pupils is profoundly influenced by the language and literacy barriers, pupils' engagement in classroom mathematics teaching and parental support in learning, all of which have significantly shaped their mathematical proficiency and, consequently, their performance on mathematics tests (Abdul Halim, 2022; Sumathi, 2016). Understanding these multifaceted factors that impact *Orang Asli* pupils' mathematics performance when administered assessments is crucial in devising appropriate testing strategies that could accurately evaluate their mathematical abilities. By acknowledging these factors, mathematics teachers and educators could implement more accessible and inclusive assessment methods that may accurately assess their pupils' mathematics performance.

2.12.2 Challenges of Implementing Test Accommodations Among Teachers

Implementing test accommodations in classroom assessments is essential for providing equitable educational opportunities to all pupils. These accommodations are designed to ensure that assessments measure a pupil's knowledge and skills to obtain a more valid interpretation of their test scores (De Backer et al., 2019; Wilner

& Mokhtari, 2018). However, implementing test accommodations in classroom based assessments effectively are fraught with various challenges.

One of the primary challenges in implementing test accommodations in classroom assessments is determining which pupils should receive accommodations and what specific accommodations they should be provided (Fletcher et al., 2018; Rios et al., 2020). For instance, Zhang and Rivera (2021) pointed out that teachers often struggle to identify the right type of test accommodation for their pupils due to the lack of clear, standardized criteria. This uncertainty can lead to inconsistencies in how test accommodations are provided, potentially disadvantaging some pupils while giving others inappropriate levels of support, as teachers who may find it difficult to select and implement the right test accommodations for their pupils (Clark-Gareca, 2016; Lovett & Harrison, 2021).

Other than that, the role of teachers in the implementation of test accommodations is crucial, yet it is also a significant source of challenge. Wickstrom et al. (2020) and Yang (2020) noted that many teachers lack the training and resources necessary to implement test accommodations effectively. This lack of preparation can lead to improper facilitation, where the selected test accommodation may either be inadequately provided, or inconsistently applied (Lovett & Harrison, 2021). This may pose an issue whereby the test accommodation implemented may not adequately address the pupils' needs, undermining the effectiveness of these supports and potentially impacting the pupils' academic performance (Xu & Kuti, 2024; Zhang & Rivera, 2021).

Another critical challenge in implementing test accommodations is ensuring that the test accommodation do not compromise the validity and reliability of the assessments (Kūkea Shultz & Englert, 2021; Yang, 2020). Radmehr and Vos (2020) emphasised that creating reliable and valid assessments while accommodating diverse needs is a complex task. Teachers must balance the need to provide support with the need to maintain the integrity of the assessment to ensure that the assessment administered is still measuring what it intends to measure (Reynolds et al., 2021). These concerns highlight the need for careful consideration and ongoing research to develop assessment practices that are both fair and valid for all pupils.

Despite extensive research on the challenges of implementing test accommodations in various educational contexts, there remains a significant gap in the literature regarding the specific challenges associated with implementing test accommodations for *Orang Asli* pupils, particularly on the use of recorded oral tests. The unique linguistic and cultural characteristics of *Orang Asli* pupils require tailored accommodation strategies that address their specific needs. However, current studies have primarily focused on mainstream populations, often neglecting the perspectives of teachers working in *Orang Asli* primary schools. There is a pressing need to explore these teachers' views to better understand the potential challenges and practical difficulties that they may encounter when implementing recorded oral test as a possible test accommodation for mathematics assessments.

2.13 Chapter Summary

The literature review conducted for this study examined the key areas relevant to the research objectives. First, it explored the mathematics performance of *Orang*

Asli pupils, highlighting the impact of language and literacy barriers, particularly how language proficiency and reading difficulties affect pupils' performance in mathematics assessments. In addressing the test accommodation aspect, this chapter also examined the use of oral test and recorded oral test, along with the test development processes involving translation, adaptation, and ethnic language assessments.

The review further explored the linguistic features of mathematics tests, such as item length and vocabulary type, and how these features embedded within mathematics test items influence pupils' mathematics performance when mathematics assessments are administered. Finally, the literature highlighted teachers' perspectives on test accommodation, including the factors impacting *Orang Asli* pupils' mathematics performance and the challenges teachers face in implementing accommodations. Through this review, gaps in the existing research were identified, supporting the need for further examination into how recorded oral test can be implemented as a viable test accommodation for *Orang Asli* pupils for fair and valid mathematics assessments.

CHAPTER THREE

RESEARCH METHODOLOGY

3.1 Introduction

This chapter presents a comprehensive overview of the research methodology employed to investigate the research objectives and questions established in this study. The research design, population, sampling technique, sample and participants, instruments, pilot study, procedure and data analysis are discussed. This chapter provides insight into the systematic process employed to carry out the research process for this study.

3.2 Research Design

In this study, a concurrent triangulation mixed methods design was employed, where both quantitative and qualitative data were collected simultaneously to provide a comprehensive and integrated understanding of the research problem (Creswell & Creswell, 2018). The quantitative component involved administering three mathematics tests to *Orang Asli* pupils—a written test in the language of instruction as the standard assessment, recorded oral test in the language of instruction, and recorded oral test in the pupils' ethnic language. Concurrently, qualitative data were gathered through interviews with mathematics teachers to explore their perspectives on the use of recorded oral test as a potential test accommodation in mathematics testing.

For the quantitative approach, this study utilised the random equivalent group design as the main research approach for this study (Gunawan et al., 2021; Wang et al., 2020). This design allowed the researcher to administer different test forms to the pupils whereby the randomisation of the pupils into groups would provide comparable test scores on the construct measured by the test (Kim & Walker, 2021). One of the assumptions on the use of this design is that the differential performance observed between the groups was attributed to test form difficulty of the mathematics assessment administered to the groups of *Orang Asli* pupils.

This allowed the researcher to better examine the impact of the test accommodation in both the language of instruction and ethnic language of the *Orang Asli* pupils, when compared to the written mathematics test. Moreover, this design minimises the time to complete the test, while eliminating the order effect which may stem from the potential bias or influence that the sequence or order of presenting different test forms can have on the pupils' responses (Kolen & Brennan, 2014). Based on preliminary interviews conducted with the mathematics teachers in *Orang Asli* primary schools, these pupils shared similar key characteristics (e.g. lack language proficiency in *Bahasa Melayu*, preference towards ethnic language, similar mathematics skills and knowledge) based on current classroom teaching and learning practices and previous classroom assessments. Therefore, the groups are considered equivalent and that the differences observed in the outcomes were due to the tests administered, while ensuring item difficulty and person ability are not compromised (Kolen & Brennan, 2014).

Conversely, for the qualitative approach, the case study design was employed (Creswell & Poth, 2018). The case study is a qualitative research design that involves an in-depth and detailed examination of a particular case to gain a comprehensive understanding of a phenomenon, event, individual, group, or situation (Asenahabi, 2019). Case studies are particularly useful when exploring complex real-life scenarios to understand the intricacies of a particular subject within its natural setting.

The nested case study design within this mixed method approach allowed for the validation of the quantitative approach, through the process of triangulation, using the qualitative input obtained from the participants to further support the quantitative findings (Guetterman & Fetters, 2018). By utilising the concurrent triangulation mixed method design, both quantitative and qualitative data were analysed separately, but the results were compared and integrated during the interpretation phase. This approach allowed for the triangulation of findings, ensuring that the insights from the qualitative data validated and enriched the quantitative results. Ultimately, this provided a more robust and holistic understanding of the use of recorded oral test as a viable test accommodation in classroom mathematics assessments for the *Orang Asli* pupils.

3.3 Population

The main population for this study involved Year 4 *Orang Asli* pupils from *Orang Asli* primary schools that have been designated by the Malaysian Ministry of Education. The population is comprised of *Orang Asli* pupils from the *Temiar* ethnic group within the districts of Sungai Siput, Perak and Gua Musang, Kelantan. The selection of the sample from these districts and states are to ensure that the language and its respective dialect used in the development of the recorded oral test in the ethnic language are similar and uniform when administered to these pupils in their respective primary schools.

On the other hand, the population for the qualitative approach for this study comprised of mathematics teachers currently teaching in *Orang Asli* primary schools within the districts of Sungai Siput, Perak and Gua Musang, Kelantan. These teachers represented the population based on their direct involvement in classroom mathematics instruction and their experience with teaching and assessing *Orang Asli* pupils in primary schools. Table 3.1 presents the population of the Year 4 *Orang Asli* pupils and mathematics teachers in the districts of Sungai Siput, Perak and Gua Musang, Kelantan.

Table 3.1

Population of Orang Asli Pupils and Mathematics Teachers in Sungai Siput, Perak and Gua Musang, Kelantan

State	District	School	Number of Pupils	Number of Mathematics Teachers
Perak	Sungai Siput	A	34	2
		B	11	3
		C	31	2
		D	64	3
		E	42	2
		F	63	2
		G	24	2
		H	24	3
		I	18	1

Table 3.1 (Cont.)

State	District	School	Number of Pupils	Number of Mathematics Teachers
Kelantan	Gua Musang	P	21	1
		Q	111	3
		R	43	2
		S	43	1
		T	44	2
	Total		**573**	**29**

3.4 Sample and Participants

This section presents the sampling techniques used to select the samples for the quantitative approach and participants for the qualitative approach for this study. In this section, the sampling techniques employed in both the quantitative and qualitative approaches are discussed in the following subsections.

3.4.1 Quantitative Sample

For the selection of the final sample that would be used for the administration of the mathematics tests, this study employed the multistage cluster sampling. This sampling method is typically used for geographically isolated populations, whereby the main population is subdivided into smaller clusters or groups whereby samples are drawn from the groups at each stage (Sedgwick, 2015; Creswell & Creswell, 2018). At each stage of the sampling, the cluster or groups selected are smaller than the previous group until the final sample is obtained. The selection of the sample involved the selection of the Malaysian states, followed by the districts, schools, and classroom.

In the first stage of the multistage cluster sampling method, the *Temiar* ethnic group of the *Orang Asli* was first chosen, As the *Temiar* community are widely scattered across the Northern and Central region of Peninsular Malaysia, the first stage of the multistage cluster sampling was carried out for the selection of Malaysian states that is comprised of the largest population of the *Temiar* ethnic community (JAKOA, 2023). Therefore, for the *Temiar* ethnic group, the states of Perak and Kelantan were selected as these states had the highest population of the *Orang Asli* community from this ethnic group.

After the selection of the states of Perak and Kelantan, the second stage of the cluster sampling procedure was carried out to select the districts within these states, which contained the highest number of *Orang Asli* peoples. Within the states selected, the districts of Sungai Siput, Perak and Gua Musang, Kelantan were selected as these districts had the largest number of *Orang Asli* peoples of the *Temiar* ethnic group (Government of Malaysia, 2023). Therefore, the two districts were selected as the second cluster.

From the districts selected, the third stage of the multistage cluster sampling was carried out for the selection of the primary schools. During this stage, it was ensured that only *Orang Asli* primary schools designated by the Ministry of Education and are comprised of pupils from the *Temiar* ethnic group within the districts of Sungai Siput, Perak and Gua Musang, Kelantan were selected. This stage of the sampling was important to ensure that the *Orang Asli* primary schools selected as clusters or groups were externally homogenous as it represented the *Orang Asli* pupils in general, but internally heterogenous whereby the pupils within these primary schools originate specifically from the *Temiar* ethnic group. This represented the third subsequent cluster.

There was a total of nine primary schools within the district of Sungai Siput, Perak, and five primary schools within the district of Gua Musang, Kelantan. However, certain schools had to be excluded from this study due to the post-COVID recovery measures had further impacted accessibility, as the ongoing restrictions at the time of data collection made travelling and in-person data collection more challenging. Furthermore, the schools' remote locations also pose significant challenges for data collection as some schools were located in areas with very poor road conditions, making travel dangerous. Some primary schools could only be reached using four-wheel drive vehicles, which presented serious safety concerns for the researcher. As a result, access to some schools for carrying out data collection was severely restricted. Given these limitations, only primary schools that could be reached safely and to conduct the data collection efficiently were considered for the study. Consequently, only five primary schools from the district of Sungai Siput, Perak and two primary schools from the district of Gua Musang, Kelantan were selected.

After the selection of the primary schools, the fourth and final stage of the multistage cluster sampling method was employed to select the classrooms within these primary schools. For most schools, the Year 4 *Orang Asli* pupils from the first two classrooms were selected as the final sample of this study. However, the number of *Orang Asli* pupils within the two classrooms selected was also dependent on their attendance at the time of data collection in the respective schools. Moreover, based on the recommendations made by the mathematics teachers during the preliminary interviews, only Year 4 O*rang Asli* pupils were selected. The selection of the Year 4 primary level was deemed appropriate and suitable to assess their foundational mathematics skills and knowledge gained during their lower primary school years. This selection was further substantiated by the notion that this primary level had also attended Year 1 primary level at primary schools prior to the school closures during the Covid-19 pandemic, whereby Year 1 to Year 3 pupils had to be excluded due to the potential learning loss suffered during the period of school closures at the time (Kanageswari et al., 2023).

From these schools, a total of 176 Year 4 *Orang Asli* pupils were selected to represent the main sample for this study. Following this, pupils were randomly assigned into three equivalent groups using simple random sampling, and the mathematics tests were administered using a spiral administration method (Kolen & Brennan, 2014). This approach involved distributing the different test forms in a rotating sequence across the three groups to ensure that no one group consistently received an easier or harder version of the test.

This method prevented any single group from facing a test that was disproportionately more difficult or easier than the others, thereby ensuring equivalence across the groups (Livingston, 2014). Furthermore, because the random assignment of the test forms across pupils accounted for variability in person ability, each group's performance reflected a true measure of their ability, rather than being influenced by differences in test difficulty (Alordiah & Oji, 2024). Thus, the spiral administration of the three test forms ensured equivalence across the groups, which allowed for fair comparison between the groups without skewing the results due to discrepancies in test form difficulty. Table 3.2 presents the sample of this study within each primary school, respectively.

Table 3.2

Sample of Orang Asli Pupils

State	District	School	Number of Pupils
Perak	Sungai Siput	A	28
		B	13
		C	23
		D	37
		E	25
Kelantan	Gua Musang	P	26
		Q	24
	Total		176

3.4.2 Qualitative Participants

For the qualitative approach employed for this study, purposive sampling was carried out to select the mathematics teachers to participate in the interview sessions (Creswell & Poth, 2018; Patton, 2005). The selection of participants for this study was based on their willingness to engage in in-depth interviews regarding their perspective on the use of test accommodation for *Orang Asli* pupils. An important component in the selection of participants was their readiness and availability during data collection. Therefore, the selection of participants for this study was limited to their willingness to participant and their availability during the data collection process which were necessary to facilitate scheduling and conducting the interviews effectively.

The selection of the research participants does not necessarily require representativeness of the broader population, as suggested by Marshall and Rossman (2016). Instead, by employing purposive sampling, this allows for in-depth information to be gathered from a smaller number of participants. The focus was on obtaining rich, detailed perspectives that capture the nuanced experiences of the participants that are directly involved in the phenomena or issue being studied (Braun & Clark, 2013). By prioritizing participants' willingness to share their experiences, the study aimed to delve deeply into specific, context-rich insights that could inform the development and effectiveness of test accommodations that does not compromise the final data.

To answer the fourth research question established in this study, a single encompassing criterion was established to select the participants. The criterion for selecting the participants is as follow: primary school teachers with at least three years' experience of teaching mathematics in *Orang Asli* primary schools within the district of Sungai Siput, Perak. A total of seven mathematics teachers from four *Orang Asli* primary schools within the district of Sungai Siput, Perak were selected to participate in this study. During the data collection period, there were no mathematics teachers available in schools selected within the district of Gua Musang, Kelantan, which precluded their participation in this study. The profile of the participants selected for this study are presented in Table 3.3.

Table 3.3

Participants Profile for Semi-Structured Interviews

Participants	School	Profile
P1	E	• Head of Mathematics Department. • Has seven years of experience teaching mathematics to *Orang Asli* pupils. • Previous taught Islamic studies before transitioning to mathematics. • Graduate from *Kolej Islam Selangor*.
P2	E	• Native *Orang Asli* ethnic member of the Semai ethnic group. • Began teaching career in 2019 as a Bahasa Melayu and mathematics teacher for Level One primary *Orang Asli* pupils. • Field of specialization is in special education. • Graduated from IPG Tengku Ampuan Afzan, Kuala Lipis.
P3	A	• Mathematics teacher with 24 years of teaching experience. • Have been teaching mathematics for 14 years. • Currently teaching mathematics to Level Two *Orang Asli* pupils
P4	A	• Has been a mathematics teacher for 14 years. • Taught mathematics to Year 1 to Year 5 *Orang Asli* pupils for five years. • Currently hold the post as Head teacher for mathematics department.

Table 3.3 (Cont.)

Participants	School	Profile
P5	D	• Began teaching mathematics to Year 2 and Year 3 *Orang Asli* pupils in 2016. • Four years of experience teaching mathematics to *Orang Asli* pupils. • Currently teaching mathematics to Year 2, Year 3, Year 4 and Year 6 *Orang Asli* pupils.
P6	B	• Employed as a CEFR certified English teacher for teaching *Orang Asli* pupils. • Began teaching mathematics for Year 1 to Year 3 pupils since 2017. • Teaching reading and language literacy as part of language programme in the primary school.
P7	B	• Began teaching mathematics since 2021 in *Orang Asli* primary school for Year 1 to Year 5 pupils. • One of three mathematics teachers who is responsible for mathematics education under *Kelas Dewasa Ibu Bapa Orang Asli dan Penan* (KEDAP), which is a programme to teach language and mathematics skills to parents of *Orang Asli* pupils.

The researcher acknowledges that there is a difference in the unit of analysis employed in this study. However, it is imperative that the perspectives of the mathematics teachers are considered to attest to the validity of using the recorded oral test as a test accommodation among *Orang Asli* pupils. Moreover, as these teachers are implementors of current mathematics assessment practises who better understand the needs of the *Orang Asli* pupils, their view on its implementation is warranted to determine its feasibility as a promising test accommodation that can be incorporated in current mathematics testing practices.

3.5 Instrument

For both quantitative and qualitative approaches employed in this study, this section delves into the distinct instruments utilised for both approaches to answer the research questions established in this study. The following subsection presents the instruments utilised for both quantitative and qualitative approaches.

3.5.1 Mathematics Tests

For the quantitative approach, three mathematics tests were developed sequentially as the main instruments for this study, namely the written test in *Bahasa Melayu* (WBM), recorded oral test in *Bahasa Melayu* (OBM) and recorded oral test in *Bahasa Temiar* (OBT). Based on the proposed framework by Ramadhan et al. (2019), the test development process for the three mathematics tests were carried out as discussed in detail in the following sub-subsections. A total of three types of experts were involved in the development of the mathematics tests, namely mathematics experts, mathematics researchers, and mathematics teacher-experts. Table 3.4 provides a summary of the roles undertaken by these experts for this study.

Table 3.4

Roles of Experts in the Development of Mathematics Tests

Experts	**Role**
Mathematics Experts (ME)	To examine the suitability and difficulty of the test items to Year 4 *Orang Asli* pupils, and to establish content validity.
Mathematics Researchers (MR)	To establish content validity of test items.
Mathematics Teacher Experts (MTE)	To transadapt test items to *Bahasa Melayu* and *Bahasa Temiar* for linguistic, cultural and semantic equivalence, and subsequent audio recording of the test items.

3.5.1.1 Written Test in *Bahasa Melayu* (WBM)

During the initial phases of developing the written test in *Bahasa Melayu* (WBM), the selection of mathematics test items was made from the Year 4 syllabus that were aligned with the eight Year 1 topics outlined in the *Dokumen Standard Kurikulum dan Pentaksiran* (DSKP) or Standard-based Curriculum and Assessment (Malaysia Ministry of Education, 2017). This included eight mathematics topics, namely whole numbers up to 100, basic operations, fractions, money, time, measurement, space, and data management.

The rational for the selection of the mathematics test items that corresponded to the Year 1 topics was based on the advice given by the mathematics experts due to the potential learning loss that may have resulted from the two years of school closures during the Covid-19 pandemic (Noor Azlan & Noor Asyhikin, 2022; Kanageswari et al., 2023). Moreover, the mathematics experts highlighted that the learning loss suffered by the *Orang Asli* pupils may have affected the learning of mathematical knowledge and proficiency during this period and therefore proposed on selecting the foundational mathematics topics of Year 1 within the current Year 4 syllabus for developing the mathematics test items.

This justification was also endorsed by the mathematics researchers. In addition, due to the lack of language and literacy development among the *Orang Asli* pupils, they also advised on language simplicity and reduced test item difficulty in accordance with the pupils' language, literacy and cognitive ability. A total of 40 mathematics test items were developed and adapted for the Year 4 *Orang Asli* pupils for the written mathematics test. The written mathematics test blueprint for the 40 test items is presented in Appendix A.

The test items were then classified based on the respective item length categories, which are short, moderate and long item. Short item length refers to mathematics test items that contain total number of words that are less than five words, while moderate item length refers to test items that contain total number of words between five to ten words in one simple sentence. Lastly, long item length refers to test items that contain total number of words of more than 10 words in two or more simple sentences.

Additionally, for the categorisation of the vocabulary of the mathematics test items for analysis, the researcher first distinguished between general vocabulary and mathematics vocabulary for each test item. This process involved referring to the *Glossary for mathematics primary school dual language programme* by Mian et al. (2021) to classify the terms based on the two distinctions, general vocabulary and mathematics vocabulary. This reference book was used as it had integrated the glossaries and keywords from the *Kurikulum Standard Sekolah Rendah* (KSSR) mathematics textbooks used by all primary level schools in Malaysia.

General vocabulary included words that are broadly applicable across different subjects and do not carry specialized meanings within mathematics (Baumann & Graves, 2010). Conversely, mathematics vocabulary encompasses terms that are unique to the discipline or have specialized meanings in mathematical contexts (Baumann & Graves, 2010). After categorizing the vocabulary type according to general vocabulary and mathematics vocabulary, the researcher quantified each category by counting the frequency and tallied the number of general and mathematics vocabulary for each test item. This allowed for a quantitative comparison of the total word count of the two vocabulary types. Table 3.5 presents the comprehensive categorisation of the 40 mathematics test items according to item length category, and vocabulary type and total word count for both general vocabulary and mathematics vocabulary.

Table 3.5

Test Item Categorisation According to Item Length Category and Vocabulary Type

Item	Item Length Category	Vocabulary Type			
		Mathematics Vocabulary	Number of Words	General Vocabulary	Number of Words
1	Moderate	bilangan; tidak sama banyak	2	bulatkan, gambar; yang; menunjukkan; cawan	6
2	Moderate	bilangan; sama banyak	2	bulatkan; gambar; yang; menunjukkan; pensel	5
3	Moderate	*berapakah; bilangan*	2	*gula-gula; yang; terdapat; dalam; gambar*	5
4	Short	*jumlah*	1	bulatkan; *guli; yang; betul*	4
5	Moderate	*nombor*	1	*tuliskan; yang; tertinggal; dalam; kotak*	6
6	Moderate	*jumlah*	1	kirakan; *guli; yang; terdapat; dalam; gambar*	6
7	Short	*jumlah*	1	*tuliskan; bola; dalam; kotak*	4
8	Short	No vocabulary	0	No vocabulary	0
9	Short	No vocabulary	0	No vocabulary	0
10	Short	No vocabulary	0	No vocabulary	0
11	Short	No vocabulary	0	No vocabulary	0
12	Short	*pecahan*	1	*tuliskan; bagi; bahagian; hitam*	4
13	Short	*jumlah; duit syiling*	2	*tuliskan*	1

Table 3.5 (Cont.)

Item	Item Length Category	Vocabulary Type			
		Mathematics Vocabulary	Number of Words	General Vocabulary	Number of Words
14	Short	*jumlah; wang*	1	*tuliskan*	2
15	Short	*waktu*	1	*tuliskan; pada; jam*	3
16	Short	*bilangan; sisi*	2	*nyatakan*	1
17	Short	*nombor*	1	*nyatakan; yang; paling; kecil*	4
18	Moderate	*dibundarkan; puluh yang terdekat*	2	*nyatakan; kotak; yang; betul; kepada*	5
19	Moderate	*simbol*	1	*bulatkan; yang; betul; dalam; kotak*	5
20	Long	*bilangan; paling sedikit*	4	*bulatkan; kumpulan; yang; mempunyai*	6
21	Long	*nombor; simbol; tambah; sama dengan; lima; tujuh; dua belas*	8	*tuliskan; dan*	2
22	Short	*setengah*	1	*bulatkan; gambar; yang; menunjukkan*	4
23	Long	*jumlah; wang*	2	*Ali; mempunyai; di bawah; nyatakan; yang; ada; pada; seperti*	8
24	Moderate	*waktu*	1	*Ali; bersarapan, pagi; tuliskan; sedang*	5

Table 3.5 (Cont.)

Item	Item Length Category	Vocabulary Type			
		Mathematics Vocabulary	Number of Words	General Vocabulary	Number of Words
25	Short	*bentuk*	1	*namakan; di; bawah*	3
26	Moderate	kurang; lebih	2	*bulatkan; jawapan; yang; betul; daripada*	5
27	Moderate	*sepuluh-sepuluh*	1	*bulatkan; siapa; yang; membilang*	5
28	Long	*jumlah; enam; dua*	3	berapakah; *emak; bapa; kek; memberi; ada; yang; pada; Suzana; dan*	10
29	Moderate	*nombor; bulatan*	2	*tuliskan; yang; betul; dalam*	4
30	Long	*baki*	2	*Ali; ada; biji; kuih; memberi; kepada; adiknya; yang; ada; pada*	10
31	Long	*jumlah*	1	*di; dalam; botol; ada; guli; hitam; putih; tuliskan; terdapat; yang*	10
32	Long	*jumlah*	1	*ada; biji; telur; itik; yang; ada; pada; mereka; Bob; Liyana; berapakah*	11
33	Moderate	*hitamkan; bulatan; pecahan*	3	*di; bawah; untuk*	3

Table 3.5 (Cont.)

Item	Item Length Category	Vocabulary Type			
		Mathematics Vocabulary	Number of Words	General Vocabulary	Number of Words
34	Long	*wang saku; sebanyak*	2	*mendapat; lengkapkan; jadual; di; bawah; membelanjakan*	8
35	Moderate	*berapakah, belajakan*	2	*belon; buku; membeli; Suzana*	4
36	Long	No vocabulary	0	*bulatkan; hari; Nadia; bertugas; jadual; kelas; Isnin; Selasa*	11
37	Short	*lebih; berat*	2	*bulatkan; siapa; yang; Tajul; John*	3
38	Moderate	*lebih banyak*	2	*bulatkan; bekas; yang; mempunyai; air*	5
39	Long	No vocabulary	0	*cita-cita; yang; paling; diminati; askar; bomba; polis; guru*	10
40	Long	*susunan; bilangan; tertib menaik*	4	*bulatkan; buku; dalam; bagi; Tajul; Arsay; Rosie*	7
Short Items				15	
Moderate Items				14	
Long Items				11	
Total				**40**	

To ensure that the mathematics test items were appropriately suited to the pupils' abilities, a thorough validation process was conducted where each test item was reviewed by three mathematics experts (ME) and two mathematics researchers (MR) who were employed to take part in this study. These experts were deemed necessary for the validation process due to their specialised knowledge in mathematics education and assessment, and their vast experience in teaching *Orang Asli* pupils in primary school level. The profile of the mathematics experts and the mathematics researchers are presented in Table 3.6.

Table 3.6

Mathematics Experts' and Mathematics Researchers' Profile

Expert	Profile
ME-A	• Assistant headmaster for academic affairs. • Mathematics teacher with 29 years of teaching experience with four years in *Orang Asli* primary school. • Graduated with a Bachelor of Education (B Ed.) in mathematics.
ME-B	• Head of the school mathematics department • Mathematics teacher with 17 years of teaching experiences. • Currently teaching mathematics to Year 2, 4, 5 and 6 *Orang Asli* pupils • Bachelor of Education (B Ed.) in mathematics.
ME-C	• Mathematics teacher with ten years of teaching experience. • Bachelor of Education (B Ed.) in special education and Malay language (*Bahasa Melayu*).

Table 3.6 (Cont.)

Expert	Profile
MR-A	• A mathematics educator with almost 20 years of teaching experience • Serving as a senior lecturer in a public university. • Expertise in the field of test development, psychometrics, and evaluation. • Specialisation in educational measurement on reliability and validity of cross-cultural testing. • Has six years of experience in mathematics test development for Indigenous *Orang Asli* pupils.
MR-B	• A Mathematics educator with 30 years of teaching experience. • Serving as a professor in a public university. • Expertise in the field of psychometrics, testing and evaluation and mathematics education. • Has 20 years' experience in mathematics test development.

The ME experts, together with the MR experts, evaluated the test items for content validity, ensuring that the test items were aligned to the curriculum and covered the relevant mathematical concepts. In addition to this, the ME experts used their professional judgment to assess the difficulty level of each test item. They considered factors such as the cognitive demands of the tasks, the complexity of the mathematical vocabulary used, and the alignment of the items with the pupils' grade level and learning progression in primary school level.

To further ensure that the difficulty of the test items was appropriate for the pupils, the ME experts referenced the pupils' previous performance in classroom-based assessments, or *Pentaksiran Bilik Darjah* (PBD) and curriculum guidelines to gauge the suitability of the items' difficulty level, as the experts are experienced mathematics item developers for *Orang Asli* pupils. Through this expert validation, it was ensured that the mathematics test items would neither be too challenging nor too simplistic, where the difficulty of the test items were appropriately matched to the Year 4 *Orang Asli* pupils' abilities.

During the validation process, the experts also meticulously reviewed the distinctions between general vocabulary and mathematics vocabulary categorized by the researcher to ensure the accuracy of the categorization for each test item, according to the distinctions made by Baumann and Graves (2010). The experts assessed whether the words and terms were appropriately classified based on their usage and relevance within the mathematical context. Through a consensus process, the ME and MR experts validated that the categorization were aligned with the content of the mathematics assessment based on the reference text on the glossary of the primary school textbooks and the linguistic ability of the Year 4 *Orang Asli* pupils, as exhibited in Table 3.5.

Additionally, the validation process was also carried out to ensure that the test items developed and adapted were linguistically and culturally applicable to the *Orang Asli* pupils. During this process, it was agreed that the test items were to be scored dichotomously whereby one mark was given for the correct answer, while zero marks was given for the incorrect answer. The total test score represents the number of correct responses from the 40 test items in the mathematics test. Overall, all 40 test items were adapted to a short-answer format, following a consensus reached during the validation process to ensure that the test items would be more easily understood by the *Orang Asli* pupils without being overly complex or ambiguous (refer Appendix A). The test items were deemed appropriate to evaluate the pupils' fundamental mathematical skills and basic mathematics concepts typically taught in the lower primary level. The 40 validated mathematics test items formed the written test in *Bahasa Melayu* (WBM), serving as the main template used to construct the recorded oral tests in *Bahasa Melayu* (OBM) and recorded oral test in *Bahasa Temiar* (OBT), respectively.

3.5.1.2 Recorded Oral Test in *Bahasa Melayu* (OBM)

For the development of the recorded oral test in *Bahasa Melayu* (OBM), the WBM test comprising of the 40 mathematics test items were used as the written script to audio record the test items in *Bahasa Melayu*. The test consists of the same mathematics test items developed, where each test item was read aloud and simultaneously audio recorded by a mathematics teacher expert (MTE). The MTE employed in this study is an *Orang Asli* teacher from the *Temiar* ethnic group who is fluent in both *Bahasa Melayu* and *Bahasa Temiar*. This mathematics teacher expert was employed for the construction of both the OBM and OBT tests that were utilised in this study. The profile of the MTE is presented in Table 3.7.

Table 3.7

Mathematics Teacher Expert Profile for OBM and OBT Tests

Expert	Teacher's Background
Mathematics Teacher Expert (MTE)	• Is a native of the *Temiar* ethnic group from Gua Musang, Kelantan. • Special rehabilitation teacher (*guru pemulihan khas*) for lower primary pupils. • Has four years teaching experience in Orang Asli primary school. • Native speaker of both *Bahasa Melayu* and *Bahasa Temiar*.

All 40 mathematics test items were audio recorded individually and were saved as separate files. The audio recordings of these mathematics test items were compiled as an audio test, which formed the OBM test.

3.5.1.3 Recorded Oral Test in *Bahasa Temiar* (OBT)

For the development of the recorded oral test in *Bahasa Temiar* (OBT), the WBM test was also used as the main template in its construction. The mathematics test items developed in the WBM test were reviewed by the same MTE who assisted in the recording and development of the OBM test as the expert was also fluent in *Bahasa Temiar*. The profile of the MTE was presented in Table 3.5. The review process of the mathematics test items was carried in the presence of the researcher to ensure that the test items were appropriately trans adapted into *Bahasa Temiar* by reviewing the selection of words used in the WBM and OBM tests.

Each mathematics test items were first trans adapted, whereby the test items underwent an adaptation process followed by the translation into *Bahasa Temiar*. This was to ensure language (linguistic), meaning (semantic) and cultural equivalence during the entire translation process by using the same number of words and equivalent terms when translating the test items from *Bahasa Melayu* to *Bahasa Temiar*. For example, the language equivalent for the term "*berlorek*" was not present in *Bahasa Temiar*. Therefore, the term "*berlorek*" was adapted to the term "*bahagian hitam*" as this was the only language equivalent term in *Bahasa Temiar*.

Nonetheless, the test items were adapted to ensure that the total number of words were equivalent across the three mathematics tests, whilst ensuring that the test items are measuring the same construct across the tests. Once the mathematics test items had been adapted and translated, each test item were individually read aloud in *Bahasa Temiar* and were simultaneously audio recorded with the assistance of the MTE. All 40 test items were audio recorded individually and were saved as separate audio files. The audio recording of these mathematics test items was then compiled as an audio test, which formed the OBT test. The recordings were further validated with the expert to ensure test equivalence with the mathematics tests developed in *Bahasa Melayu* (WBM and OBM tests).

3.5.2 Semi-Structured Interview Questions

The semi-structured interview questions were developed for the qualitative approach of this study to gauge the mathematics teachers' view on the use of the recorded oral test as a test accommodation. Merriam and Tisdell (2016) posited that interviews are used to gather valuable information from the perspectives of the participants on the behaviours, reaction, emotions, and their perception of the world around them. Therefore, to gather this valuable data, this study developed semi-structured interview questions according to the research question of this study to gather insight on the perspectives of the participants.

Therefore, to gather quality data from the participants, open-ended questions that are neutral and clear were developed to allow participants to fully express their opinion and experiences of the phenomenon being studies (DeJonckheere & Vaughn, 2019). In the present study, an initial interview guide was developed, which included ten open-ended questions that were developed for the semi-structured interview sessions with the participants. The following interview questions are as follows:

1. How are mathematics assessments carried out for the pupils?
2. What are the factors that affect the pupils' performance during assessments?
3. How does language affect pupils' performance?
4. How does reading ability/reading difficulties affect pupils' performance?
5. What measures have you taken to address these issues during classroom instruction?
6. How does oral presentation of the test fit into the current assessment practices for the *Orang Asli* pupils?
7. How can the oral presentation of the test be applied during mathematics assessments?
8. What are the challenges faced by the pupils if the oral presented test were to be implemented?

9. What are the challenges faced by the teachers if the oral presented test were to be implemented?
10. What are the challenges faced by the school administration if the oral presented test were to be implemented?

3.6 Pilot Study

Pilot study is a stage where the sampling strategies, procedures and other components of the research are tested in preparation for the final data collection (Srinivasan et al., 2017). Specifically, the pilot study is an important phase of the research that could help improve and validate the research instruments (Brooks et al., 2016; Pearson et al., 2020). This process allowed the researcher to identify test items from the mathematics tests which may require further review, and to take the necessary steps to re-validate the instruments before the final data collection process.

Therefore, to determine the validity of the mathematics test items developed for this study, it is important that the instruments are pilot tested to determine its validity and reliability for its use in the actual data collection (Pearson et al., 2020). As this study employed a mixed-method research design, the instruments that have been both developed for the quantitative and qualitative approaches had undergone pilot testing to determine the respective validity for the final data collection (Malmqvist et al., 2019; Williams-McBean, 2019). Therefore, the following sections will present the pilot testing of the mathematics tests and the semi-structured interview questions developed for this study.

3.6.1. Pilot Testing Mathematics Tests

The pilot study was carried out in three *Orang Asli* primary schools: two schools in the district of Sungai Siput, Perak and one school in the district of Gua Musang, Kelantan. The three mathematics tests (WBM, OBM, and OBT tests) were administered to 125 Year 4 *Orang Asli* pupils. Out of the 125 pupils selected, 66 pupils (53%) are male, while 59 pupils (47%) are female. All the sample of the *Orang Asli* pupils selected were from the *Temiar* ethnic group. The *Orang Asli* primary schools and the number of Year 4 *Orang Asli* pupils who participated during the pilot study are presented in Table 3.8.

Table 3.8

Sample of Pupils Selected for Pilot Study

State	District	School	Number of Pupils
Perak	Sungai Siput	I	27
		J	39
Kelantan	Gua Musang	K	59
		Total	125

Before carrying out this research, permission was obtained from the Ministry of Education through the *Bahagian Perancangan dan Penyelidikan Dasar Pendidikan*, or Educational Planning and Policy Research Division (EPRD). The letter of EPRD approval to carry out this study is presented in Appendix B. Subsequently, through the sampling procedure undertaken to identify the sample, the states and districts which had fit the necessary requirements for its inclusion as the sample in this study was then selected. Therefore, permission and approval to carry out this research were obtained from the State Education Departments for the states of Perak and Kelantan, as presented in Appendix C and Appendix D, respectively, before contacting the *Orang Asli* primary schools within the districts of Sungai Siput, Perak and Gua Musang, Kelantan. Introductions were made to seek permission and approval to carry out this research.

The mathematics tests were administered by the researcher to the *Orang Asli* pupils to gather data in determining the validity and internal reliability of the mathematics tests. In each primary school selected, the three mathematics tests were spirally administrated to the Year 4 *Orang Asli* pupils, whereby the WBM, OBM, and OBT tests were administered to the pupils who have been randomly assigned to three groups. The first group was administered the WBM test, while the second and third groups were administered the OBM and OBT tests respectively.

For the groups administered the OBM and OBT tests, a copy of the WBM test was also administered to the pupils as a reference and answer booklet to each test item. The tests were conducted concurrently with the help of some schoolteachers present at the time. The administration of the three mathematics tests lasted approximately one hours. During this process, the duration it took to complete the mathematics tests, field notes and pupils' responses were noted for further investigation and as additional support in determining the reliability and validity of the test items within the mathematics tests during item analysis.

3.6.2 Item Analysis

Item analysis is a measure used to investigate the responses of the pupils to the individual test items to gauge the characteristics of the test items and the quality of the assessment (Ryser, 2018). The item analysis is mainly used to identify test items that are suitable for the pupils when measuring the intended construct and to determine the psychometric properties of the test items (Quaigrain & Arhin, 2017). Item analysis therefore provides the statistical information necessary to pinpoint test items that are deemed inappropriate for the *Orang Asli* pupils to ensure that the content and construct validity are preserved. Therefore, in this study, item difficulty index, item discrimination index and reliability coefficient, through Cronbach's Alpha, were computed to determine the suitability of the mathematics test items developed to gauge the *Orang Asli* pupils' mathematics performance.

The item difficulty index refers to "the percentage of the total number of correct responses to the test items" (Quaigrain &Arhin, 2017, p. 5). The item difficulty index is denoted in the formula $p = \frac{R}{T}$, where p is the item difficulty index, R is the number of pupils that answered the item correctly and T is the total number of pupils that took part in the assessment, according to classical test theory (CTT). Therefore, rather than identifying the intrinsic difficulty of the item, the index considers the relative frequency of the pupils that were able to answer the item correctly (Thorndike et al., 1991).

Another measure that is also calculated is the item discrimination index, which is the degree to which the items distinguish between the pupils of differing skills and content knowledge of the construct measured in the assessment (Boopathiraj & Chellamani, 2013). The item discrimination index is obtained through the point biserial correlation which is denoted as the formula D1 = $\frac{UG-LG}{n}$, where *UG* is the upper 27% of pupils who produced the correct response, *LG* is the lower 27% of pupils who produced the correct responses and *n* is the number of pupils in the largest of the two groups (Quaigrain & Arhin, 2017). The discrimination index could either be positive or negative. A positive index indicates that high performing pupils of the whole test scores higher on that item, while a negative index indicate that the low performing pupils of the whole test scored higher on that item.

The final measure that was computed was the reliability of the mathematics assessment. The items were examined using the reliability analysis of Cronbach's Alpha. Within the context of this study, the Cronbach's Alpha coefficient was computed to determine the internal consistency and reliability of the four mathematics test booklets used during the pilot study in the *Orang Asli* primary schools. The coefficient of Cronbach's Alpha range between 0 to 1, where the closer the alpha coefficient is to 1, the higher the reliability of the test booklets to be administered for final data collection.

However, Taber (2018) suggested that higher coefficient values for Cronbach's Alpha does not offer much evidence on the reliability of an instrument and proposed that it may be undesirable when developing content-based assessments. Therefore, it was proposed the coefficient value of 0.70 is suitable to gauge the reliability or internal consistency of the mathematics instrument. This notion is in line with Cortina (1993) where the coefficient value of 0.70 is acceptable to determine the internal consistent of the instrument. Nonetheless, the threshold of the acceptable Cronbach's Alpha coefficient values can only be used as guidelines, where lower values should not be considered as indicating unsatisfactory reliability of the instrument (Plummer & Ozcelik, 2015).

3.6.2.1 Written Test in *Bahasa Melayu*

Item analysis was carried out on the 40 mathematics test items developed for the WBM test. Table 3.9 presents the item difficulty and item discrimination indexes for the 40 mathematics test items in the WBM tests.

Table 3.9

Item Difficulty and Item Discrimination Index for WBM Test

Item	p	D	Alpha if Item Deleted	SD
1	0.47	0.74	0.58	0.51
2	0.38	0.66	0.59	0.49
3	0.91	0.15	0.66	0.30
4	0.75	0.05	0.67	0.44
5	0.94	0.29	0.65	0.25
6	0.72	0.26	0.65	0.46
7	0.78	0.52	0.62	0.42
8	0.78	-0.48	0.72	0.42
9	0.44	0.55	0.61	0.50
10	0.81	0.02	0.67	0.40
11	0.44	0.47	0.62	0.50
12	0.09	0.27	0.65	0.30
13	0.66	0.51	0.62	0.48
14	0.88	0.15	0.66	0.34
15	0.44	-0.20	0.71	0.50
16	0.75	0.41	0.63	0.44
17	0.41	0.56	0.61	0.50
18	0.34	-0.35	0.72	0.48

Table 3.9 (Cont.)

Item	p	D	Alpha if Item Deleted	SD
19	0.56	0.33	0.68	0.47
20	0.31	0.33	0.64	0.47
21	0.34	0.57	0.85	0.48
22	0.84	0.21	0.86	0.37
23	0.75	0.55	0.85	0.44
24	0.53	0.07	0.87	0.51
25	0.47	0.80	0.83	0.51
26	0.50	0.53	0.85	0.51
27	0.78	0.50	0.85	0.42
28	0.16	0.41	0.85	0.37
29	0.31	-0.04	0.87	0.47
30	0.45	0.31	0.86	0.48
31	0.22	0.20	0.85	0.42
32	0.28	0.74	0.86	0.46
33	0.50	0.69	0.85	0.51
34	0.41	0.50	0.87	0.50
35	0.53	0.53	0.83	0.51
36	0.63	0.58	0.85	0.49
37	0.78	0.49	0.85	0.42
38	0.88	0.31	0.85	0.34
39	0.44	0.70	0.87	0.50
40	0.06	0.24	0.85	0.25

Based on Table 3.9, the test items were within acceptable item difficulty ranges. With respect to item discrimination, Items 4 and 10 had very low item discrimination index of 0.05 and 0.02 respectively, which suggest that these items do not discriminate well between the good and poor performing pupils when administered the test. Most of the test items displayed positive item discrimination index, however, Items 8, 15 and 18, and 29 displayed negative item discrimination indexes of -0.48, -0.20, -0.35 and -0.04 respectively. This indicates that for these items, poor performing pupils were able to answer the item correctly compared to good performing pupils. To determine the internal consistency of the WBM test, Cronbach's Alpha was computed. Table 3.10 presents the Cronbach's Alpha value obtained for WBM test.

Table 3.10

Reliability Test for WBM Test

Test	Cronbach's Alpha
WBM Test	0.76

Based on Table 3.10, the Cronbach's Alpha for the WBM test consisting of the 40 mathematics test items is 0.76. This indicated that the WBM test is found to be reliable (Taber, 2018) and was deemed suitable for administration to the *Orang Asli* pupils.

3.6.2.2 Recorded Oral Test in *Bahasa Melayu*

The 40 mathematics test items developed for the OBM test was also analysed using item analysis. Table 3.11 presents the item difficulty and item discrimination indexes for the 40 mathematics test items in the OBM tests.

Table 3.11

Item Difficulty and Item Discrimination Index for OBM Test

Item	p	D	Alpha if Item Deleted	SD
1	0.56	0.66	0.71	0.50
2	0.35	0.50	0.73	0.49
3	0.94	0.27	0.75	0.24
4	0.74	0.52	0.73	0.45
5	0.94	0.27	0.75	0.24

Table 3.11 (Cont.)

Item	p	D	Alpha if Item Deleted	SD
6	0.88	0.16	0.76	0.33
7	0.67	0.31	0.75	0.48
8	0.76	0.49	0.73	0.43
9	0.38	0.40	0.74	0.49
10	0.68	-0.08	0.78	0.47
11	0.56	0.37	0.74	0.50
12	0.35	-0.01	0.77	0.49
13	0.85	0.24	0.75	0.36
14	0.82	0.29	0.75	0.39
15	0.24	0.05	0.77	0.43
16	0.88	0.36	0.74	0.33
17	0.12	0.31	0.75	0.33
18	0.68	0.33	0.75	0.47
19	0.18	0.50	0.73	0.39
20	0.65	0.60	0.72	0.49
21	0.26	0.66	0.82	0.45
22	0.74	0.58	0.82	0.45
23	0.97	-0.04	0.84	0.17
24	0.59	0.66	0.82	0.50
25	0.71	0.55	0.82	0.46
26	0.53	0.68	0.82	0.51
27	0.87	0.27	0.86	0.39
28	0.50	0.85	0.81	0.51
29	0.41	0.45	0.83	0.50
30	0.15	0.11	0.86	0.49
31	0.29	0.51	0.83	0.46
32	0.59	0.19	0.84	0.50
33	0.35	-0.10	0.86	0.49
34	0.71	0.66	0.82	0.46
35	0.47	-0.06	0.86	0.51
36	0.50	0.75	0.81	0.51
37	0.82	0.15	0.84	0.39
38	0.95	0.21	0.73	0.37
39	0.32	0.42	0.83	0.47
40	0.21	0.54	0.83	0.41

Based on Table 3.11, both item difficulty and item discrimination index are within suitable values except for Items 10, 12, 15, 23, 33, and 35, which was found to have negative discrimination index values. These values indicate that the items could not discriminate between poor and good performing pupils, and therefore may require revision for better discrimination. Although these values are relatively low, the analysis suggest that poor performing pupils were able to perform better in these items compared to good performing pupils. However, as these items have relatively high item difficulty index, this may suggest that almost all pupils administered the OBM booklet found these items easy to solve, therefore, these items were found to be easy for both poor and good pupils. Subsequently, the internal consistency of the OBM test was also carried out by computing the Cronbach's Alpha. Table 3.12 presents the Cronbach's Alpha value obtained for OBM test.

Table 3.12

Reliability Test for OBM Test

Test	Cronbach's Alpha
OBM Test	0.80

With respect to internal consistency for the OBM test, the Cronbach's Alpha obtained for the OBM test consisting of the 40 mathematics test items is 0.80. As the alpha values obtained are higher than the threshold value of 0.70 (Taber, 2018), the OBM test was deemed reliable and suitable for administration to the *Orang Asli* pupils.

3.6.2.3 Recorded Oral Test in *Bahasa Temiar*

For the 40 mathematics test items developed for the OBT test, item analysis was also used to examine the validity of the test items. Table 3.13 presents the item difficulty and item discrimination index for the 40 mathematics test items in the OBT test.

Table 3.13

Item Difficulty and Item Discrimination Index for OBT Test

Item	p	D	Alpha if Item Deleted	SD
1	0.60	0.44	0.66	0.49
2	0.53	0.29	0.68	0.50
3	0.98	0.54	0.67	0.15
4	0.65	-0.03	0.71	0.48
5	0.89	0.21	0.67	0.41
6	0.63	0.49	0.65	0.49
7	0.95	0.59	0.66	0.21
8	0.86	0.42	0.66	0.35
9	0.63	0.17	0.69	0.49
10	0.86	0.37	0.67	0.35
11	0.35	0.53	0.64	0.48
12	0.26	0.19	0.69	0.44
13	0.56	0.20	0.69	0.50
14	0.95	0.19	0.68	0.21
15	0.02	0.11	0.69	0.15
16	0.88	0.36	0.67	0.32
17	0.60	0.03	0.71	0.49
18	0.77	0.22	0.68	0.43
19	0.05	-0.10	0.70	0.21
20	0.74	0.46	0.65	0.44
21	0.26	0.43	0.73	0.44
22	0.81	-0.06	0.77	0.39
23	0.93	0.20	0.75	0.26
24	0.63	0.00	0.77	0.49
25	0.72	0.76	0.70	0.45
26	0.53	0.50	0.73	0.50
27	0.91	0.39	0.74	0.29
28	0.09	0.27	0.75	0.29
29	0.37	-0.07	0.78	0.49
30	0.19	0.11	0.71	0.28
31	0.23	0.54	0.73	0.43

Table 3.13 (Cont.)

Item	p	D	Alpha if Item Deleted	SD
32	0.35	0.35	0.74	0.48
33	0.72	0.21	0.75	0.45
34	0.79	0.46	0.73	0.41
35	0.42	0.30	0.75	0.50
36	0.60	0.63	0.71	0.49
37	0.98	0.20	0.75	0.15
38	0.93	0.11	0.76	0.26
39	0.74	0.43	0.73	0.44
40	0.28	0.55	0.72	0.45

Based on Table 3.13, all items had recorded acceptable item difficulty and item discrimination index values except for Items 4, 17, 19, 22, and 29. Although the item discrimination index for these items were relatively low, the results of the analysis suggested that these items were unable to discriminate between the good and poor performing pupils. One possible explanation to the negative item discrimination values obtained for these test items is that the translated terms might not have an equivalent meaning or might be interpreted differently, leading to a possible confusion among test-takers. Subsequently, the Cronbach's Alpha was computed to determine the internal consistency of the OBT test. Table 3.14 presents the Cronbach's Alpha value for the OBT test.

Table 3.14

Reliability Test for OBT Test

Test	Cronbach's Alpha
OBT Test	0.72

With respect to internal consistency for the OBT test, the Cronbach's Alpha value obtained from the item analysis for the 40 mathematics test items is 0.72. As this value is higher than the threshold value of 0.70 (Taber, 2018), this shows that the OBT test is a reliable instrument and deemed suitable for administration to the *Orang Asli* pupils.

3.6.2.4 Revised Mathematics Tests

Based on the findings obtained from the item analysis carried out on the three mathematics tests (WBM, OBM, and OBT tests) during the pilot testing, a general concession on each of the 40 mathematics test items was carried out by reviewing and re-adapting the test items with two mathematics researchers (MR) that were previously employed during the content validation process. This process was undertaken with the same experts to re-evaluate and re-establish content validity to ensure that the mathematics test items developed in this study assesses the mathematics construct intended to be measured. The decisions and rationales for each of the 40 mathematics test items were based on the results of the item analysis, as presented in Table 3.15.

Table 3.15

Summary of Decisions on Test Items Post-Item Analysis

Item	Decision	Rational
1	Retained	Revision made to format of answering the question.
2	Retained	Revision made to format for answering the question.
3	Removed	As in measures the same content with Item 4, consensus for removal

Table 3.15 (Cont.)

Item	Decision	Rational
4	Retained but requires revision	Low discrimination in WBM test. Negative discrimination in OBT.
5	Retained	No revisions made. Item difficulty and discrimination index were in acceptable ranges.
6	Retained	Revision made to format for answering the question.
7	Retained	Revision made to format for answering the question.
8	Retained	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
9	Retained	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
10	Retained	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
11	Retained	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
12	Revised	Content remained the same. Item format was revised to fill in the box. (Previously; circle the answer)
13	Retained	Revision made to wording and format of answering.
14	Removed/Revised	Measures same content as Item 13, consensus for removal. However, still retained for now.
15	Removed/Revised	Very low discrimination and difficulty index in OBM and OBT tests. Negative discrimination in WBM test. Bad item writing.
16	Accepted with revision made	Issue with equivalent word use for native language. Change "*Nyatakan*" to "*Tuliskan*"
17	Accepted with revision made	Issue with equivalent word use for native language. Change "*Nyatakan*" to "*Tuliskan*"
18	Accepted with revision made	Issue with equivalent word use for native language. Change "*Nyatakan*" to "*Tuliskan*"

Table 3.15 (Cont.)

Item	Decision	Rational
19	Revised	Very low discrimination and difficulty index in WBM and OBM tests. Negative discrimination in OBT test.
20	Retained	More clear instruction provided. Need revision for recording.
21	Retained	Revision made to format of answering the item.
22	Retained with revision	Negative discrimination in OBT but an easy item. Overall item difficulty and discrimination indexes are good for WBM and OBM tests.
23	Revised	Low item difficulty index in WBT and OBM tests. Issue with the word "*Padankan*". Revised to fill in the box.
24	Retained	Acceptable ranges for difficulty index and discrimination for OBM, but very low discrimination in WBM and OBT tests. May need revision. Probably simplify item format to *box* + 3 = 9.
25	Retained	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
26	Retained	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
27	Retained	Language was simplified for better understanding.
28	Retained	Language was simplified for better understanding.
29	Revised	Negative discrimination in OBT test, while low value was obtained for WBM test. Item difficulty index is within acceptable ranges from all tests. The item may need revision. (Revision already made to "circle the correct fraction"
30	Revised	No data for WBM, OBM and OBT tests (extreme measures). Will retain but requires revision.

Table 3.15 (Cont.)

Item	Decision	Rational
31	Retained but revised	Below acceptable range for both item difficulty and discrimination index for WBM and OBM tests. Acceptable range for both item discrimination and item difficulty in OBT test which may need revision.
32	Revised	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
33	Retained	Acceptable in WBM and OBT test, but very low discrimination in OBM. The item may need revision.
34	Retained	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
35	Retained	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
36	Retained	No revisions required. Item difficulty and discrimination index were in acceptable ranges.
37	Retained	Acceptable ranges in WBM test, but very low discrimination index in OBM and OBT. Item was retained.
38	Retained	Acceptable ranges in WBM test, however negative item discrimination in OBM test. Item difficulty index is high (too easy). Item can be removed.
39	Retained	No revisions required.
40	Retained	Although very low difficulty index in WBM and OBM tests, item discrimination index is within acceptable ranges.

The mathematics test items with negative discrimination index were reviewed and modified to further uphold linguistic and cultural equivalence. For instance, items which contained the words '*nyatakan*' was revised to '*tuliskan*', whereby the latter word had a linguistic equivalent in *Bahasa Temiar*. Furthermore, the test items were modified for linguistic simplification to ensure that the *Orang Asli* pupils would be better able to understand the test items better.

The decisions on the validity of the test items developed for the WBM test was conducted to ensure that the items and the overall formatting of the mathematics test were better suited for the *Orang Asli* pupils, as these group of pupils were unfamiliar to the standard assessment format that is typically used by mainstream primary school pupils. The test items were revised for the WBM test, however the final test blueprint remained intact and unaltered. The final test blueprint is as exhibited in Appendix A as no changes were made to the topics, subtopics and item length, where only answering format and phrasing of certain words and terms were carried out after the pilot study. The revised and final WBM test used as the main instrument for this study is presented in Appendix E.

The final revised WBM test was then used as the main template for the development of the recorded oral test in *Bahasa Melayu* (OBM) and *Bahasa Temiar* (OBT). The same mathematics teacher expert who participated in the translation and audio recordings for the initial 40 mathematics test items during the development of the recorded oral tests were re-employed to develop the audio tests for both OBM and OBT tests, respectively. The same procedures stipulated in Section 3.5.1.2 to Section 3.5.1.3 were carried out to develop the OBM and OBT tests, respectively. The total word count of the test items was ensured to be equivalent across the three mathematics tests to ensure a fair comparison of the differences in performance between the tests.

3.6.3 Semi-Structured Interview Question

To validate the semi-structured interview questions developed for the qualitative approach, purposive sampling was employed to select two research methodology experts (RM). This process was employed to enhance the potency of the interview questions, ensuring their relevance in accomplishing this study's objective (Castillo-Montoya, 2016; Yeong et al., 2018). Table 3.16 presents the profile of the two research methodology experts employed for validating the interview questions.

Table 3.16

Experts Profile for Validating Interview Questions

Experts	Profile
RM-1	A mathematics educator with almost 20 years of teaching experience Serving as a senior lecturer in a public university. Expertise in the field of test development, psychometrics, and evaluation. Specialisation in educational measurement on reliability and validity of cross-cultural testing.
RM-2	Serving as a senior lecturer in a public university. Expertise in the field of educational psychology in academic literacy and alternative practices. Aligned mainly in interpretive (qualitative) research approaches.

For the validation of the ten semi-structured interview questions developed, a meeting was conducted between the two research methodology experts and researcher to discuss the suitability of the interview questions in relation to the fourth research question established in this study. Following the initial review with the experts, the interview questions were reviewed to ensure that the questions were in line with the objective of this study. The meeting with the experts lasted approximately 30 minutes. During the meeting, each interview question was discussed pertaining to its language, wording, relevance and open-endedness. Table 3.17 presents the comments made to the ten interview questions.

Table 3.17

Original Questions, Comments and Action Taken for Interview Questions

Questions	Comments	Action
1. How are mathematics assessments carried out for the pupils?	This question is relevant to the research question. This is relevant to explore the practices implemented by the teachers to assess the pupils.	Retain
2. What are the factors that affect the pupils' performance during assessments?	This question is relevant to the research question. This question allows the researcher to explore current issues that may affect pupils' performance in mathematics tests.	Retain
3. How does language affect pupils' performance?	Language is already an issue faced by the *Orang Asli* pupils. This does not require teachers to clarify. Question not relevant to research question.	Remove
4. How does reading ability/reading difficulties affect pupils' performance?	The question is not relevant to this study's research questions. This is already examined in other studies/literature. Question not relevant	Remove

Table 3.17 (Cont.)

	Questions	Comments	Action
5.	What measures have you taken to address these issues during classroom instruction?	This question is relevant to the research question. This question would allow the researcher to understand the ways teachers mitigate the issues faced by the *Orang Asli* pupils during assessment.	Retain
6.	How does the oral delivery of the test fit into the current assessment practices for the *Orang Asli* pupils?	It is important to explore the context of the recorded oral delivery test accommodation examine in this study, but teachers' views on its implementation in current practices would be better.	Remove
7.	How can the oral delivery of the test be applied during mathematics assessments?	This question is relevant to the research question. This is to explore the way the teachers may try to implement the oral delivery of the assessment in the context of the *Orang Asli* pupils.	Retain
8.	What are the challenges faced by the pupils if the orally delivered test were to be implemented?	This question is relevant to the research question. This question allows the researcher to explore the issues of implementing this form of test accommodation to the pupils.	Retain
9.	What are the challenges faced by the teachers if the orally delivered test were to be implemented?	This question is relevant to the research question. This question explores the issues that would arise due to the test accommodation from the teachers' perspective.	Retain

Table 3.17 (Cont.)

	Questions	Comments	Action
10.	What are the challenges faced by the school administration if the orally delivered test were to be implemented?	This question is relevant to the research question. This question explores the issues that would arise due to the test accommodation from the school administrators and staff's perspective.	Retain

Based on the review by the experts on the initial ten interview questions presented in Table 3.17, three questions were agreed to be removed as it was not relevant in answering the research question for the qualitative approach. The remaining seven interview questions were further modified and revised between the researcher and two RM experts to unsure its suitability in exploring the mathematics teachers' perspectives. The seven revised interview questions were then deemed appropriate to gather the necessary data from the participants during the final data collection process. The interview protocol containing these interview questions is presented in Appendix F. Table 3.18 lists the seven revised and validated open-ended interview questions in both English and *Bahasa Melayu*.

Table 3.18

Interview Questions in English and Bahasa Melayu

Interview Questions (English)	**Interview Question (*Bahasa Melayu*)**
1. How are mathematics assessments implemented for *Orang Asli* pupils?	1. *Bagaimanakah pentaksiran matematik dilaksanakan untuk murid Orang Asli?*
2. What are the factors that affect the *Orang Asli* pupils' performance during mathematics assessments?	2. *Apakah faktor-faktor yang mempengaruhi prestasi murid Orang Asli semasa pentaksiran matematik?*
3. What steps have been taken to address these issues when conducting mathematics assessments or activities in the classroom?	3. *Apakah langkah-langkah yang telah diambil untuk menangani isu-isu ini semasa menjalankan pentaksiran atau aktiviti matematik di dalam kelas?*
4. What is the teachers' view on the administration of recorded oral tests in *Bahasa Melayu* and *Bahasa Temiar* to be implemented during mathematics assessments?	4. *Apakah pandangan guru terhadap pentadbiran ujian lisan yang direkodkan dalam Bahasa Melayu dan Bahasa Temiar yang akan dilaksanakan semasa pentaksiran matematik?*
5. What are the teachers' views on the use of the recorded oral test to fairly assess mathematics performance for *Orang Asli* pupils?	5. *Apakah pandangan guru tentang penggunaan ujian lisan yang direkodkan untuk menilai prestasi matematik secara adil bagi murid Orang Asli?*

Table 3.18 (Cont.)

Interview Questions (English)	Interview Question (*Bahasa Melayu*)
6. What challenges will *Orang Asli* pupils may experience if the recorded oral test was to be implemented during mathematics assessments?	6. *Apakah cabaran yang mungkin dialami oleh murid Orang Asli sekiranya ujian lisan yang direkodkan dilaksanakan semasa pentaksiran matematik?*
7. What challenges will the teachers and school administrators experience if the recorded oral test was to be implemented during mathematics assessments?	7. *Apakah cabaran yang akan dialami oleh guru dan pentadbir sekolah sekiranya ujian lisan yang direkodkan dilaksanakan semasa pentaksiran matematik?*

In the series of seven interview questions, the initial three interview questions may not directly address the developed research question. Nonetheless, these interview questions play a crucial role in illuminating the challenges and context surrounding the educational issues and needs of the *Orang Asli* pupils. By exploring these aspects, the interview aims to identify key factors influencing the potential adoption of recorded oral test as a test accommodation. This broader perspective is essential for comprehensively understanding the needs and circumstances of the pupils in question. The remaining four questions are specifically designed to capture the perspectives of mathematics teachers regarding the implementation of recorded oral test. These inquiries delve into the practical aspects and potential challenges associated with integrating such test accommodation into mathematics assessments.

3.7 Procedure

The data collection procedures for both quantitative and qualitative approach are presented in this section. For the quantitative approach, the procedures presented are for the administration of the three mathematics tests, namely WBM, OBM and OBT tests. On the other hand, the data collection process for the qualitative approach was presented on the semi-structured interviews with the participants.

3.7.1 Mathematics Tests

The *Orang Asli* primary schools selected were contacted via phone calls to seek approval to conduct this research. A brief background of the research, accompanied with the letters from EPRD and the State Education Departments from the states of Perak and Kelantan were provided to the schools. After receiving necessary approval from the schools, appointments with the schools were made on the appropriate date and time that is convenient to carry out the administration of the mathematics tests.

On the day of the final data collection, a brief introduction between the researcher and the school administrators on the research background and samples required for the administration of the mathematics tests was conducted. To maintain test equating, this study employed the random equivalent group design, whereby the mathematics tests were administered through a spiralling process. According to Livingston (2004), this design equates the relationship between the different groups of pupils that are administered different versions of the test form, thereby allowing generalization to the target population. Furthermore, by administering the tests through the spiralling process, the pupils that are administered the test would be divided equally between the different test forms to balance out the order effect (Boyer, 2018).

The process of the random equivalent group design as outlined, was employed to ensure that the groups established in this study are comparable between each other. This process involved the random assignment of the sample of *Orang Asli* pupils who were randomly allocated into three different groups. The groups were ensured to approximately contain the same number of pupils and were spirally administered the three mathematics tests. In this study, a total of 156 *Orang Asli* pupils were administered the mathematics tests in the seven primary schools within the districts of Sungai Siput, Perak and Gua Musang, Kelantan.

In this study, the WBM test was administered to all three groups, however two out of the three groups were administered the OBM and OBT tests respectively as audio files. The audio files comprising of the audio recording of the test items in *Bahasa Melayu* for the OBM test, and *Bahasa Temiar* for the OBT tests were played during the test administration process. The WBM test booklet was used not only as a guide for the pupils to refer to, but also as the answering sheet for the pupils to write their answers in. The OBM and OBT tests were administered through audio recordings that were played for each item separately by the researcher. The audio recording for each item were played twice, and five minutes was allocated for the pupils to answer the test item before proceeding to the next test item in sequence as developed in the WBM test.

Although a maximum of five minutes was allocated for each test item, in practice, there was a degree of leniency in administering the test, as the researcher proceeded to the next question once the pupils had finished responding. This often occurred within 1-2 minutes, or even less for some test items, depending on the complexity of the test items and the pupils' response time. The pupils were permitted to request for the audio recording to be replayed within the provided time interval allocated for answering the test item, however in most cases, the time required by the pupils to the replay of the audio recording was also considerably shorter. This allowed for a more efficient progression through the test. As a result, the test administration did not exceed two hours.

3.7.2 Semi-Structured Interview

During the pilot study, short interview sessions were conducted between the headmasters and school administrators on the issues faced by the various stakeholders within the *Orang Asli* primary schools when carrying out mathematics teaching and learning for the *Orang Asli* pupils. Subsequently, this had led to the need to interview the mathematics teachers to gather their views on the use of the recorded oral test as a test accommodation to mitigate the issues faced by the pupils and facilitate their mathematics performance.

Initial contact was established through purposive sampling, targeting participants who possess relevant experiences and knowledge pertinent to the study's focus (Creswell & Poth, 2018). Therefore, to gather the necessary data for the qualitative approach employed in this study, permission letters explaining the research's purpose, objectives, and the voluntary nature of participation were sent to the school administrators of four *Orang Asli* primary schools within the district of Sungai Siput, Perak. This was carried out to obtain the headmasters' permission to interview the mathematics teachers within those primary schools. Once permission was granted, the mathematics teachers' contact details were obtained from the respective school offices. This was followed by personalized invitations, either via email or direct contact, providing potential participants with an opportunity to deliberate and make an informed decision about their involvement in this study.

A total of 12 mathematics teachers from the four primary schools were contacted by the researcher, however only seven mathematics teachers agreed to participant in this study. The teachers who agreed to attend the interviews were subsequently used as the main participants for this study. Upon obtaining informed consent, semi-structured interviews were conducted as the primary mode of data collection. The validated interview protocol ensured the exploration of key themes while allowing flexibility for participants to expound on their unique experiences and perspectives (Kallio et al., 2016). The semi-structured nature of these interviews facilitated a balance between maintaining a coherent structure and allowing for organic conversation, thereby eliciting rich and nuanced data (Adeoye-Olatunde & Olenik, 2021; Ruslin et al., 2022)

During the interview sessions, the researcher acted as the human instrument to obtain the relevant data from the participants (Creswell & Poth, 2018). The researcher was able to allocate more time and provide follow-up questions where it was deemed necessary to extract rich information from the participants. To ensure accuracy and detail in capturing the participants' input, all interviews were meticulously audio-recorded with the participants' explicit consent. An audio recording tool was used to capture verbal nuances, tone, and emotions embedded within the dialogue. This allowed for a comprehensive examination during the transcription phase, minimizing the loss of valuable information and maintaining the integrity of the participants' narratives.

The transcription process involved careful and systematic verbatim conversion of the audio recordings into written text. To uphold confidentiality, each participant was assigned a pseudonym, and any identifying information was redacted from the transcripts. The transcripts were cross-checked by the participants for accuracy and completeness, ensuring fidelity to the original dialogue and maintaining the integrity of the participants' voices. Member checking was employed as a methodological strategy to enhance rigor and credibility of the data (Candela, 2019; Motulsky, 2021). Participants were provided with an opportunity to review excerpts of their own transcripts, affording them the chance to validate the accuracy and contextual relevance of their contributions through instant messaging platforms, such as WhatsApp.

3.8 Data Analysis

The study employed a mixed method approach encompassing both quantitative and qualitative data analyses techniques to address this study research questions. The data analysis is presented according to the quantitative and qualitative approaches in following subsections.

3.8.1 Statistical Tests: One-Way ANOVA Test and Pearson Correlation

For research questions one and two, the study employed descriptive statistics and one-way ANOVA test (Heavy, 2019; Polit & Beck, 2018). For research question one, the one-way ANOVA test was used as a tool to assess the differences in mathematics performance among *Orang Asli* pupils across the three mathematics tests (WBM, OBM, and OBT), treating test types as independent variables and overall mathematics scores as dependent variables. Meanwhile, research question two utilized data stratification by item length categories to examine the mathematics performance variations attributed to differing item length categories within each mathematics test. This stratification resulted in subsets of test scores based on the short, moderate and long item length, which were examined across WBM, OBM, and OBT tests.

For research question one, only one independent variable was examined which is the mathematics tests (WBM, OBM, and OBT tests), while for research question two, the independent variable examined was item length (short, moderate and long items) across the three mathematics tests, respectively. On the other hand, for both research questions, the dependent variable was the total mathematics test scores obtained for the respective mathematics tests and for each item length category respectively. Therefore, for research question one and two, the one-way ANOVA test was utilised to examine the differences in the mathematics performance of the pupils between the three mathematics tests, and between the three item length categories across each test.

Before conducting the ANOVA test, the essential assumptions to carrying out the one-way ANOVA test were validated (Emerson, 2022; Judd et al., 2017). These assumptions include.

a. Independence of Data: Each observation within the groups is independent of every other observation. The groups are not influenced by or related to each other.
b. Normality Test: The assumption that the population for which the samples are drawn are normally distributed. H_o: The data is normally distributed.
c. Homogeneity of Variance: The assumption that the variances of the groups being examined are equal. H_o: The variances are equal across the groups.

Independence of data was ensured via careful data collection procedures using random sampling during test administration. Normality tests using the Shapiro-Wilks test were conducted to affirm that the data between the variables followed a normal distribution (Orcan, 2020). Subsequently, the Levene's test assessed homogeneity of variance among the three mathematics tests to determine if the variance across the groups is approximately equal (Mishra et al., 2019; Parra-Frutos, 2013). If the results of the Levene's test is not significant, this justifies the use of the one-way ANOVA to examine if there is a significant difference in the mathematics performance of the pupils across WBM, OBM and OBT tests.

Conversely, a significant result from the Levene's test necessitated the Brown-Forsythe test to examine if there is a significant difference in the pupils' mathematics performance. The Brown-Forsythe test is employed when the assumption of homogeneity of variance is violated in the context of a one-way

ANOVA. Instead of relying on the equality of variances among groups, the Brown-Forsythe test uses the medians and absolute deviations from the medians to assess differences between group variances (Parra-Frutos, 2016). This makes it a robust alternative in situations where the variances are not equal, providing a more reliable analysis of group differences. Post hoc analysis was also carried out to determine the specific groups that differed from each other to determine if there are significant difference in the group pairs between the three mathematics tests (Williams & Abdi, 2010).

Research question three sought to explore relationships between general vocabulary and mathematics vocabulary, with item difficulty. Based on the CTT used for item analysis, the item difficulty was computed for the 40 mathematics test items across the three mathematics tests (WBM, OBM, and OBT tests). The test items were then categorised based on the item difficulty range, which are easy ($p > 0.70$), moderate ($0.30 < p < 0.70$), and hard ($p < 0.30$), as suggested by Bichi (2016), and Quaigrain and Arhin (2017)

Pearson correlation and scatter plot diagrams was employed to assess if there is a relationship between vocabulary type and item difficulty across WBM, OBM, and OBT tests. The Pearson correlation coefficient provideed a numerical measure of the strength and direction of the linear relationship between two variables, while a scatter plot diagram visually represented this relationship. These techniques allowed for a more intuitive understanding and identification of patterns within the data (Schober et al., 2018), with respect to vocabulary and item difficulty. Both methods complement each other in analysing the relationship between these variables.

3.8.2 Content Analysis and Thematic Analysis

Additionally, content analysis was also carried out to answer research question three. Content analysis involves systematically examining and interpreting the qualitative data, such as text, images, or other media to identify patterns or themes within the data (Krippendorff, 2018; Vaismoradi & Snelgrove, 2019). Content analysis was used to identify the types of vocabulary that showed improved performance when administered the recorded oral tests, compared to written test based on the item difficulty.

The process began with coding, where segments of data were systematically assigned descriptive or interpretive labels that capture their essence. Through an iterative process, codes are grouped into categories, and overarching themes and subtheme emerge from these categories. The themes and subtheme represent broader patterns or concepts that encapsulate the underlying meaning or content within the data (Lindgren et al., 2020). The themes and subthemes were identified to examine the relationship between the type of vocabulary and its impact on the difficulty of test items that were found easier when administered the recorded oral tests, compared to the written test.

Lastly, for research question four, thematic analysis was employed to analyse the qualitative data from the participants semi-structured interviews (Corbin & Strause, 2015; Creswell & Poth, 2018). Through the process of coding and identification of recurring ideas, major themes and subthemes emerged. This contributed to a deeper understanding of the perspectives of the participants, based on the research question established in this study.

To ensure the consistency and reliability of the coding process, an external audit of the qualitative data was carried out as part of the data validation strategy (Herber et al., 2020; White et al., 2012). In this study, the mathematics researcher (MR-A), who possesses expertise in qualitative research methodology and had participated in the content validation process of the mathematics test items, was re-employed to independently review and audit a subset of the coded data. The MR-A was not involved in the initial data collection or analysis, which provided an objective perspective on the coding scheme.

The MR-A evaluated the alignment of the codes with the data and provided feedback on the appropriateness and accuracy of the assigned codes. This process ensured that the coding scheme was not only consistent but also reflective of the underlying themes in the participants' responses (Akkerman et al., 2008). This mitigated potential biases that contributed to the trustworthiness of the qualitative analysis, providing additional rigor and credibility to the findings.

Based on Creswell and Creswell (2018), the qualitative findings were the triangulated with quantitative results to provide a comprehensive perspective of the use of recorded oral tests as a test accommodation for *Orang Asli* pupils. This approach allowed for the interpretation of data, where the qualitative insights from the teacher interviews were aligned with the quantitative performance data of the *Orang Asli* pupils from the mathematics tests. Table 3.19 provides a summary of the extensive data analysis procedures employed to answer the research questions which integrates quantitative and qualitative methodologies to yield a comprehensive understanding of the study's objectives.

Table 3.19

Summary of Data Analysis Based on the Research Questions

Research Question	Data	Data Analysis
1. Is there a significant difference in the mathematics performance of the *Orang Asli* pupils between the three mathematics tests (WBM, OBM, and OBT tests)?	40 mathematics test items in each mathematics tests (WBM, OBM, and OBT tests)	One-way ANOVA Independent variables: WBM, OBM, and OBT tests. Dependent variable: overall mathematics test scores obtained from the 40 mathematics test items.
2. Is there a significant difference in the mathematics performance of the *Orang Asli* pupils by item length categories (short, moderate, and long item length) between the three mathematics tests (WBM, OBM, and OBT tests)?	40 mathematics test items in each mathematics tests (WBM, OBM, and OBT tests)	One-way ANOVA Independent variable: Short, moderate and long item length. Dependent variable: overall mathematics test scores obtained from the 40 mathematics test items for WBM, OBM, and OBT tests respectively.
3. What is the relationship between vocabulary type (general vocabulary and mathematics vocabulary) and item difficulty of the test items across the three mathematics tests (WBM, OBM, and OBT test)?	40 mathematics test items in each mathematics tests (WBM, OBM, and OBT tests)	Pearson correlation and scatter plot diagram, and content analysis.
4. What are the teachers' views on the use of the recorded oral test as a test accommodation to facilitate the mathematics performance among the *Orang Asli* pupils?	Interview protocol (semi-structured interview questions)	Thematic analysis

3.9 Chapter Summary

This study employed the explanatory mixed method research design that examined the use of the recorded oral test as a test accommodation for the *Orang Asli* pupils. This chapter discussed the population and sample, instrumentation, pilot study, procedures, and data analysis that were employed for this research study for both quantitative and qualitative approaches. In conclusion, the meticulous selection and application of the chosen research methodology carried out laid the foundation for a comprehensive and robust investigation into the core aspects of this study that answers the established research questions.

CHAPTER FOUR

FINDINGS

4.1 Introduction

This chapter presents the comprehensive findings for both quantitative and qualitative data analyses. The quantitative approach involved examining the mathematics performance of *Orang Asli* pupils between the WBM, OBM, and OBT tests, and examining their performance between the item length categories across the three mathematics tests. Additionally, the relationship between total word count for general vocabulary and mathematics vocabulary with the item difficulty were examined. Qualitative content analysis was then carried out to identify the types of vocabulary that was found to be easier by the pupils when administered the recorded oral tests, compared to written test. Lastly, the qualitative input of the semi-structured interviews with the participant was analysed to shed light on their perspectives regarding the use of recorded oral tests as accommodations. The findings presented in this chapter are aimed at answering the research objectives established in this study.

4.2 Response Rate of Participants

During the test administration for the final data collection, a total of 156 Year 4 *Orang Asli* pupils of the *Temiar* ethnic group from seven primary schools in the districts of Sungai Siput, Perak and Gua Musang, Kelantan were administered the mathematics tests. All *Orang Asli* pupils that were selected as the sample had responded in the mathematics tests that were administered by researcher for this study, yielding a response rate of 100%. Nonetheless, after meticulous data cleaning, 142 out of 156 pupils represented the final sample that were analysed for this study. The final sample of 142 pupils, constituting 91% of the initial sample pool formed the basis for drawing meaningful conclusions and insights from the data analysis conducted for this study. Table 4.1 presents the number of responses and profile of the samples according to the respective schools and gender for the three tests (WBM, OBM, and OBT tests).

Table 4.1

Responses of Orang Asli Pupils for Mathematics Tests

State	**School**	**WBM**		**OBM**		**OBT**		**Total**
		Male	Female	Male	Female	Male	Female	
Perak	A	5	3	3	5	4	3	23
	B	2	2	1	2	2	2	11
	C	1	2	3	4	3	3	16
	D	5	5	6	5	6	5	32
	E	3	3	3	4	3	4	20
Kelantan	J	5	4	3	3	4	4	23
	N	3	2	4	3	3	2	17
Total		24	21	23	26	25	23	142
		45		49		48		

4.3 Assumptions Testing for One-Way ANOVA Test

The one-way ANOVA test was carried out to answer research question one and two. For research question one, the independent variables were the three mathematics tests, namely the WBM, OBM, and OBT test, which represented the three groups that were tested. The dependent variable was the mathematics test scores obtained from the total of 40 marks (40 test items) that represented their overall mathematics performance. On the other hand, for research question two, the independent variable that will be tested was the item length categories which consisted of three groups, namely short, moderate and long item length for WBM, OBM, and OBT tests, respectively. The dependent variable was the average mathematics test scores obtained from the pupils for each item length category in each test.

Before proceeding to carry out the one-way ANOVA test to address research question one and research question two, it was imperative to ensure that the assumptions for ANOVA test were met to ensure the reliability and validity of the results (Emerson, 2022). Firstly, the assumption for the independence of the data was ensured. Secondly, test for normality was computed to confirm that the data distribution for each mathematics test group was approximately normally distributed. Lastly, the assumption for the homogeneity of variance was examined to ensure the variability or variance within each group is approximately equal. These assumptions were therefore examined and are presented in the subsequent sections.

4.3.1 Independence of Data

For research question one, the data related to the independent variable were collected based on the mathematics tests (WBM, OBM, and OBT tests), which represented the three groups that were examined. The study design and procedure were thoroughly examined to ensure that the data collected for each group were independent of each other.

For research question two, the independent variables are the item length categories, which are short, moderate and long item lengths categorised from the 40 mathematics test items that have been developed in this study. The three item length categories represents the three groups that were examined using the ANOVA test between the three-mathematics test (WBM, OBM, and OBT tests). The three groups of item length category were therefore treated as independent variables for each mathematics test.

4.3.2 Normality Test

The normality of the data collected for both research question one and two were computed using the Shapiro-Wilk test. The Shapiro-Wilk test was used to assess whether the groups of independent variables are normally distributed within a given population (Alizadeh Noughabi, 2016). This was done to determine if parametric test or non-parametric test were to be used to determine if there are significant difference between the groups. The Shapiro-Wilk test was computed to determine if the three groups of mathematics tests (WBM, OBM, and OBT tests) had met the assumption for normality. Table 4.2 presents the results of the Shapiro-Wilk test for WBM, OBM, and OBT tests.

Table 4.2

Shapiro-Wilk Test for WBM, OBM, and OBT Tests

Test	Statistics	df	Sig.
WBM	0.95	45	0.07
OBM	0.97	49	0.14
OBT	0.97	48	0.43

Based on the Shapiro-Wilk test performed for the three groups of mathematics tests, the results show that the WBM test (W = 0.95, p = 0.07), OBM test (W = 0.96, p = 0.14), and OBT test (W = 0.97, p = 0.43) are not significant. The three tests failed to reject the null hypothesis, which indicates that the data collected for the three mathematics tests are normally distributed. Therefore, the three groups represented the three mathematics tests, which also contained the mathematics test items classified according to the respective item length categories (short, moderate, and long item lengths). These groups were assumed to be approximately normally distributed.

4.3.3 Homogeneity of Variance

For research question one, the Levene's test was used to test the assumption of the homogeneity of variances between the groups of independent variables (Mishra et al., 2019). This test was used to determine if the variances within the three groups of the mathematics tests (WBM, OBM, and OBT tests) and item length categories (short, moderate, and long item lengths) were approximately equal. This is to ensure that the assumption underlying the ANOVA test was met to ensure a more accurate interpretation of the data (Ntumi, 2021). Table 4.3 presents the results of the Levene's test for the three groups of mathematics tests, WBM, OBM and OBT tests.

Table 4.3

Levene's Test between WBM, OBM, and OBT Tests

Total Sum	Levene Statistic	df1	df2	Sig.
Based on Mean	8.016	2	122	0.001

Based on the results of the Levene's test, the assumption for homogeneity of variance for the three groups of mathematics tests was significant, [F (1, 122) = 8.02, p = 0.001]. Therefore, the variances between the three groups are not equal. As equal variances were not assumed, the Brown-Forsythe test was therefore utilised (Cavus & Yazici, 2020; Wang et al., 2017). This test assumes unequal variance between the three groups and was used to answer research question one to determine if there are significant differences in the mathematics performance of the *Orang Asli* pupils between the WBM, OBM and OBT tests.

For research question two, the Levene's test was also computed for the three groups of item length categories for short, moderate and long item length across the WBM, OBM, and OBT tests respectively. Table 4.4 presents the results of the Levene's test between the three groups of short, moderate and long item length across the three mathematics tests.

Table 4.4

Levene's Test between Short, Moderate and Long Item Length

Test	**Item Length**	**Levene's Statistic**	**df1**	**df2**	**Sig.**
Based on Mean	Short	10.22	2	121	0.001
	Moderate	2.07	2	121	0.013
	Long	2.20	2	121	0.008

Based on the results of the Levene's test, the assumption for homogeneity of variance for short item length [F (2, 121) = 10.22, p = 0.001], moderate item length [F (2, 121) = 2.07, p = 0.013], and long item length [F (2, 121) = 2.20, p = 0.008] were significant. Therefore, the variance between the three groups of item length categories across the three mathematics tests were not equal. As unequal variances were not assumed, the Brown-Forsythe test was therefore used to carry out the ANOVA test (Cavus & Yazici, 2020; Wang et al., 2017). This test assumes unequal variances between the three groups and was used to answer research question two to determine if there are significant differences in the mathematics performance of the *Orang Asli* pupils between the item length categories for short, moderate and long item lengths across WBM, OBM, and OBT tests.

4.4 Research Question One

RQ1: Is there a significant difference in the mathematics performance of the *Orang Asli* pupils between the three mathematics tests (WBM, OBM, and OBT tests)?

Prior to conducting the Brown-Forsythe test, the descriptive statistics for the mathematics performance was compared based on the total mathematics test scores obtained from the 40 test items in the WBM, OBM, and OBT tests. The descriptive statistics was carried out to examine and compare the average test scores obtained by the pupils across the three mathematics tests. Table 4.5 presents the results of the descriptive statistics for the mathematics test scores obtained for WBM, OBM, and OBT tests.

Table 4.5

Descriptive Statistics for WBM, OBM, and OBT Tests

Test	N	Mean (M)	Std. Dev (SD)	Std. Error
OBM	45	28.35	4.93	0.78
OBT	49	27.79	5.70	0.87
WBM	48	23.64	7.78	1.20

Based on the results from the descriptive statistics, the *Orang Asli* pupils had obtained the highest average test scores when administered the OBM test (M = 28.35, SD = 4.93), followed by the OBT test (M = 27.79, SD = 5.70). The *Orang Asli* pupils had performed the lowest when administered the WBM test (M = 23.64, SD = 7.78). Overall, the results showed that the *Orang Asli* pupils who were administered the OBM and OBT tests had higher mathematics performance, compared to those administered WBM test.

Subsequently, the Brown-Forsythe test was carried out to examine if there was a significant difference in the mathematics performance of the *Orang Asli* pupils between the mathematics tests (WBM, OBM, and OBT tests). This test was used as the selected ANOVA test in this study as equal variances between the groups was not assumed. Table 4.6 presents the results of the Brown-Forsythe test on the mathematics performance of the pupils between WBM, OBM, and OBT tests.

Table 4.6

Brown-Forsythe Test between WBM, OBM, and OBT Tests

ANOVA Test	Statistics	df1	df2	Sig.
Brown-Forsythe	7.05	2	106.42	.001

The result of the Brown-Forsythe test was statistically significant, [F (2, 106) = 7.05, p = .001]. Therefore, the results show that there is a significant difference in the mathematics performance of the *Orang Asli* pupils between WBM, OBM, and OBT tests. To determine if there are significant differences between the specific groups of the mathematics tests administered, the post hoc test using the Games-Howell post-hoc analysis for pair-wise multiple comparisons was used. The Games-Howell post hoc test was computed as equal variances between the three groups of mathematics tests was not assumed. Table 4.7 presents the results of the Games-Howell post-hoc analysis between the WBM, OBM, and OBT tests.

Table 4.7

Games-Howell Post Hoc Analysis between WBM, OBM and OBT Tests

Tests	Tests	Mean Difference	Std. Error	Sig.
OBM	WBM	4.71	1.43	.002
OBT	WBM	4.15	1.48	.022
OBM	OBT	0.56	1.17	.883

The results show that there was a significant difference between the OBM test and WBM test (p = .002), and between OBT test and WBM test (p = .022). On the other hand, the results showed that there was no significant difference between the OBM test and OBT test ($p = .883$). From the Games-Howell post hoc analysis, there is a difference in the mathematics performance when administered the OBM and OBT tests, compared to the WBM test. However, there is no differences in the mathematics performance of the *Orang Asli* pupils between the OBM and OBT tests.

4.5 Research Question Two

RQ2: Is there a significant difference in the mathematics performance of the *Orang Asli* pupils by item length categories (short, moderate, and long item lengths) between the three mathematics tests (WBM, OBM, and OBT tests)?

The 40 mathematics test items were categorised based on the respective item length categories, which are short, moderate, and long item length. Each item length categories were compared respectively across the WBM, OBM, and OBT tests. By analysing the differences in the mathematics performance between the item length categories across the three mathematics tests, this study examined how varying item lengths of the mathematics test items could potentially influence the *Orang Asli* pupils' mathematics performance across the three tests.

The subsequent sub-sections present the results of the descriptive statistics and Brown-Forsythe tests of the mathematics performance of the pupils based on the short, moderate, and long item lengths. Post-hoc test was also carried out to further examine the difference between the groups of item length categories, whereby multiple pair-wise comparison between the groups were conducted to examine the differences between the specific groups of mathematics tests according to short, moderate, and long item lengths categories.

4.5.1 Short Item Length

From the 40 mathematics test items developed for the mathematics tests, a total of 15 test items were categorised as short item length. In the context of this study, short item length is defined as mathematics test items that contain total number of words of less than five. Descriptive statistics was computed to examine the *Orang Asli* pupils' average test scores obtained on short item length across WBM, OBM, and OBT tests. Table 4.8 presents the results of the descriptive statistics for short item length test items across WBM, OBM, and OBT tests.

Table 4.8

Descriptive Statistic for Short Item Length

Tests	N	Mean (M)	Std. Dev (SD)
OBT	45	10.19	2.02
OBM	49	10.03	1.72
WBM	48	9.24	3.30

The results from the descriptive statistics showed that the *Orang Asli* pupils had obtained the highest mean test scores for the mathematics test items categorised under short item length when administered the OBT test (M = 10.19, SD = 2.02), followed by the OBM test (M = 10.13, SD = 1.61), and lastly, the WBM test (M = 9.24, SD = 3.30). For short item length, this study found that the *Orang Asli* pupils were able to perform better when they were administered the OBT and OBM tests, compared to the WBM test. In order to determine if there is a difference in the mathematics performance of the pupils for short item length test items between the three mathematics tests, the Brown-Forsythe test was computed between the three tests. Table 4.9 presents the results of the Brown-Forsythe test for short item length between the WBM, OBM and OBT tests.

Table 4.9

Brown-Forsythe Tests for Short Item Length

Item Length	Statistics	df1	df2	Sig.
Short	1.82	2	91.41	.172

The result obtained from the Brown-Forsythe test was not statistically significant, [F (2, 91) = 1.82, p = .172]. This shows that the result had failed to reject the null hypothesis, indicating that there is no significant difference in the mathematics performance of the *Orang Asli* pupils for short item length between the WBM, OBM, and OBT tests. Therefore, based on the results of the Brown-Forsythe test, it can be concluded that the *Orang Asli* pupils' mathematics performance for test items categorised as short item length are similar across WBM, OBM, and OBT tests, regardless of test modality or language used for the test.

4.5.2 Moderate Item Length

A total of 14 mathematics test items were categorised as moderate item length. In the context of this study, moderate item length is defined as mathematics test items that contain total number of words between five to ten words in one simple sentence. The descriptive statistics was calculated to examine the average test scores obtained by the *Orang Asli* pupils for moderate item length across WBM, OBM, and OBT tests. Table 4.10 presents the results of the descriptive statistics for moderate item length test items between the WBM, OBM, and OBT tests.

Table 4.10

Descriptive Statistic for Moderate Item Length

Tests	N	Mean (M)	Std. Dev (SD)
OBT	45	10.60	2.00
OBM	49	10.53	2.21
WBM	48	8.76	2.43

The results from the descriptive statistics showed that the *Orang Asli* pupils had obtained the highest average test scores for test items categorised under moderate item length when administered the OBT test (M = 10.60, SD = 2.00), followed by the OBM test (M = 10.53, SD = 2.21) and WBM test (M = 8.76, SD = 2.43). Therefore, this study found that the *Orang Asli* pupils had performed better on moderate item length test items on the OBT and OBM tests, compared to WBM test.

In order to determine if there is a difference in the mathematics performance of the *Orang Asli* pupils for the mathematics test items that are categorised as moderate item length between the three mathematics tests, the Brown-Forsythe test was performed. Table 4.11 presents the results of the Brown-Forsythe test on the mathematics performance of the pupils for moderate item length test items between WBM, OBM and OBT tests.

Table 4.11

Brown-Forsythe Tests for Moderate Item Length

Item Length	Statistics	df1	df2	Sig.
Moderate	9.23	2	118.52	.001

The result obtained from the Brown-Forsythe test was statistically significant, [F (2, 119) = 9.23, p = .001]. Therefore, the result of the test rejected the null hypothesis, whereby it was found that there is a statistically significant difference in the mathematics performance of the *Orang Asli* pupils for test items categorised as moderate item length between WBM, OBM, and OBT tests. Subsequently, to further examine this difference between the groups of mathematics tests, the Games-Howell post hoc test was carried out for group pairwise comparison between WBM, OBM, and OBT tests for moderate item length. Table 4.12 presents the results of the Games-Howell post-hoc test carried out between the WBM, OBM, and OBT tests for moderate item length test items.

Table 4.12

Games-Howell Post Hoc Test for Moderate Item Length

Tests	Tests	Mean Diff.	Sig.
OBT	WBM	1.84	.001
OBM	WBM	1.76	.001
OBT	OBM	0.08	.981

Based on the results of the Games-Howell post hoc test, the study found that there was a statistically significant difference between the OBT and WBM tests (p = .001), and between OBM and WBM tests (p = .001). Conversely, the study found that there is no statistically significant difference in the mathematics performance of the pupils between the OBM and OBT tests (p = .981). Overall, this study found that there is a difference in the mathematics performance of the pupils when administered the OBT and OBM tests, when compared to the WBM test. However, this study also found that there is no difference in the mathematics performance of the pupils when administered either the OBM or OBT tests.

4.5.3 Long Item Length

From the 40 mathematics test items, a total of 11 mathematics test items were categorised as long item length. In the context of this study, long item length is defined as mathematics test items that contain total number of words of more than 10 words in two or more simple sentences. The descriptive statistics was also calculated to examine the average test scores obtained by the *Orang Asli* pupils for long item length test items across WBM, OBM, and OBT tests. Table 4.13 presents the results of the descriptive statistics for long item length test items between WBM, OBM, and OBT tests.

Table 4.13

Descriptive Statistic for Long Item Length

Tests	N	Mean (M)	Std. Dev (SD)
OBM	45	7.80	2.07
OBT	49	7.00	2.32
WBM	48	5.64	2.89

The results from the descriptive statistics showed that the *Orang Asli* pupils had obtained the highest average test scores for test items categorised under long item length when administered the OBM test (M = 7.80, SD = 2.07) followed by the OBT test (M = 7.00, SD = 2.32), and lastly the WBM test (M = 5.64, SD = 2.89). This study also found that the *Orang Asli* pupils were able to perform better in the OBM and OBT tests, compared to the WBM test for long item length test items. Subsequently, the Brown-Forsythe test was performed to determine if there is a difference in the mathematics performance of the *Orang Asli* pupils for the mathematics test items that are categorised as long item length between the three mathematics tests. Table 4.14 presents the results of the Brown-Forsythe test for long item length test items between the WBM, OBM and OBT tests.

Table 4.14

Brown-Forsythe Tests for Long Item Length

Item Length	Statistics	df1	df2	Sig.
Long	8.19	2	113.68	.001

The result obtained from the Brown-Forsythe test was statistically significant, [F (2, 114) = 8.19, p = .001]. The result of the test rejected the null hypothesis, whereby this study found that there is a significant difference in the mathematics performance of the *Orang Asli* pupils for test items categorised as long item length between the three mathematics tests. Subsequently, the Games-Howell post hoc test was carried out between the WBM, OBM, and OBT tests for long item length. This test was carried out to examine the difference via pairwise multiple comparison across the three groups of mathematics tests. Table 4.15 presents the results of the Games-

Howell post-hoc test carried out between the WBM, OBM, and OBT tests for long item length test items.

Table 4.15

Games-Howell Post Hoc Test for Long Item Length

Tests	Tests	Mean Diff.	Sig.
OBM	WBM	2.16	.001
OBT	WBM	1.36	.041
OBM	OBT	0.80	.233

Based on the results of the Games-Howell post hoc test, the study found that there is a statistically significant difference between the OBM and WBM tests (p = .001), and between OBT and WBM tests (p = .041). This study also found that there is no statistically significant difference in the mathematics performance for long item length test items between the OBM and OBT tests (p = .233). Therefore, for long item length test items, this study found that there is a difference in the mathematics performance of the *Orang Asli* pupils when administered the OBM and OBT tests, when compared to the WBM test. However, similarly to moderate item length test items, there was no difference in the mathematics performance of the pupils between the OBM and OBT tests.

4.5.4 Summary of Findings for Research Question Two

To sum up the results obtained from the data analysis carried out to address research question two, this study found that for all item length categories, the *Orang Asli* pupils had obtained the highest mathematics test scores and performed better when administered the recorded oral tests in either the language of instruction or ethnic language, compared to the written test. Specifically, the study found that there were no differences in the mathematics performance of the pupils between the three mathematics tests for short item length test items. Therefore, for shorter test items, the pupils were able to perform equally in all three tests.

On the other hand, for moderate and long test items, the study found that there were differences in the mathematics performance of the pupils between the two recorded oral tests, compared to the written test. However, there was no difference in the mathematics performance of the pupils between the two recorded oral tests. This shows that the *Orang Asli* pupils were able to perform better when the recorded oral tests were administered to the pupils, compared to the administration of the written test, particularly for moderate and long test items.

4.6 Research Question Three

RQ3: What is the relationship between vocabulary type (general vocabulary and mathematics vocabulary) and item difficulty of the test items across the three mathematics tests (WBM, OBM, and OBT tests)?

For the vocabulary type for the mathematics test items, general vocabulary and mathematics vocabulary were categorised using the *Glossary for mathematics primary school dual language programme* by Mian et al. (2021), which is aligned with the *Kurikulum Standard Sekolah Rendah* (KSSR) mathematics textbooks used in Malaysian primary schools. The findings to answer research question three were presented in two main parts in the following subsections. Firstly, the frequency of both types of vocabulary for each test item were quantified to examine the relationship between the total word count of general vocabulary and total word count of mathematics vocabulary with the item difficulty for each test item respectively. The data for the test items categorised as general vocabulary and mathematics vocabulary were analysed using scatterplot diagram and Pearson correlation. Then, qualitative content analysis was carried out to identify emerging patterns on the types of vocabulary for test items which exhibited improved item difficulty when administered the recorded oral tests, compared to written test.

4.6.1 Total Word Count of Vocabulary Type on Item Difficulty

The relationship between the total word count for general vocabulary and mathematics vocabulary with the item difficulty of the 40 mathematics test items across WBM, OBM, and OBT tests were subsequently examined. The total number of general and mathematics vocabulary words for each test item were quantified and juxtaposed with the respective item difficulties across the three mathematics tests. The study further highlights that the total word count for both general vocabulary and mathematics vocabulary of the test items across the three mathematics tests were equal.

By employing scatter plot diagrams and Pearson correlation coefficients, the study examined the relationship between the total word count of vocabulary types (general and mathematics vocabulary) with item difficulty index for the 40 mathematics test items. The scatter plot diagram provided graphical representations which facilitated the visual examination of data point distribution that was used to identify trends or cluster while the Pearson correlation coefficient, ranging from -1 to +1, was computed to gauge the strength and direction of the relationship. This statistical measure enabled an assessment of the linear relationship between continuous variables, elucidating perfect negative correlation (-1), perfect positive correlation (+1), or no linear correlation (0) (Akoglu, 2018; Cohen et al., 2009).

4.6.1.1 Relationship of General Vocabulary and Item Difficulty.

The total word count for general vocabulary was quantified and recorded for the 40 mathematics test items developed for the three mathematics tests. The total word count for general vocabulary of each test item was the same across WBM, OBM, and OBT tests. Table 4.16 presented the total word count for general vocabulary with the corresponding item difficulty for each test item across WBM, OBM and OBT tests.

Table 4.16

Total Word Count of General Vocabulary and Item Difficulty

Item	Item Difficulty (p)			General Vocabulary
	WBM	OBM	OBT	
1	0.52	0.97	0.86	6
2	0.76	0.69	0.84	5
3	0.86	0.95	0.98	5
4	0.90	0.82	0.77	4
5	1.00	0.95	1.00	6
6	0.71	0.79	0.86	6
7	0.81	0.77	0.84	4
8	0.90	0.90	0.86	0
9	0.52	0.72	0.67	0
10	0.79	0.69	0.84	0
11	0.55	0.77	0.81	0
12	0.48	0.03	0.12	4
13	0.57	0.90	0.88	1

Table 4.16 (Cont.)

Item	Item Difficulty (p)			General Vocabulary
	WBM	OBM	OBT	
14	0.57	0.28	0.37	2
15	0.40	0.67	0.60	3
16	0.56	0.85	0.86	1
17	0.81	0.97	0.98	4
18	0.60	0.72	0.35	5
19	0.69	0.64	0.77	5
20	0.69	0.92	0.93	4
21	0.79	0.85	0.77	2
22	0.21	0.79	0.72	4
23	0.55	0.36	0.37	8
24	0.57	0.90	0.53	5
25	0.07	0.00	0.00	3
26	0.60	0.46	0.70	5
27	0.36	0.85	0.79	4
28	0.50	0.72	0.63	10
29	0.26	0.38	0.70	4
30	0.33	0.85	0.51	10
31	0.43	0.87	0.86	10
32	0.24	0.51	0.49	11
33	0.67	1.00	0.77	3
34	0.52	0.72	0.70	6
35	0.33	0.38	0.51	4
36	0.62	0.72	0.65	8
37	0.90	0.97	0.86	3
38	0.86	1.00	0.95	4
39	0.64	0.87	0.67	9
40	0.33	0.51	0.42	7

Based on Table 4.16, the relationship between total word count for general vocabulary and item difficulty of the test items across the three mathematics tests were then examined using scatterplot diagram and Pearson correlation coefficient. The findings are presented in the following sub-subsections.

4.*6.1.1.1. **Correlation Analysis using Scatterplot Diagram.*** A scatter plot diagram was employed as a visual tool to examine the relationship between the two variables. Figure 4.1 depicts the scatter plot diagram and line of best fit between the total word count of general vocabulary and item difficulty across WBM, OBM, and OBT tests.

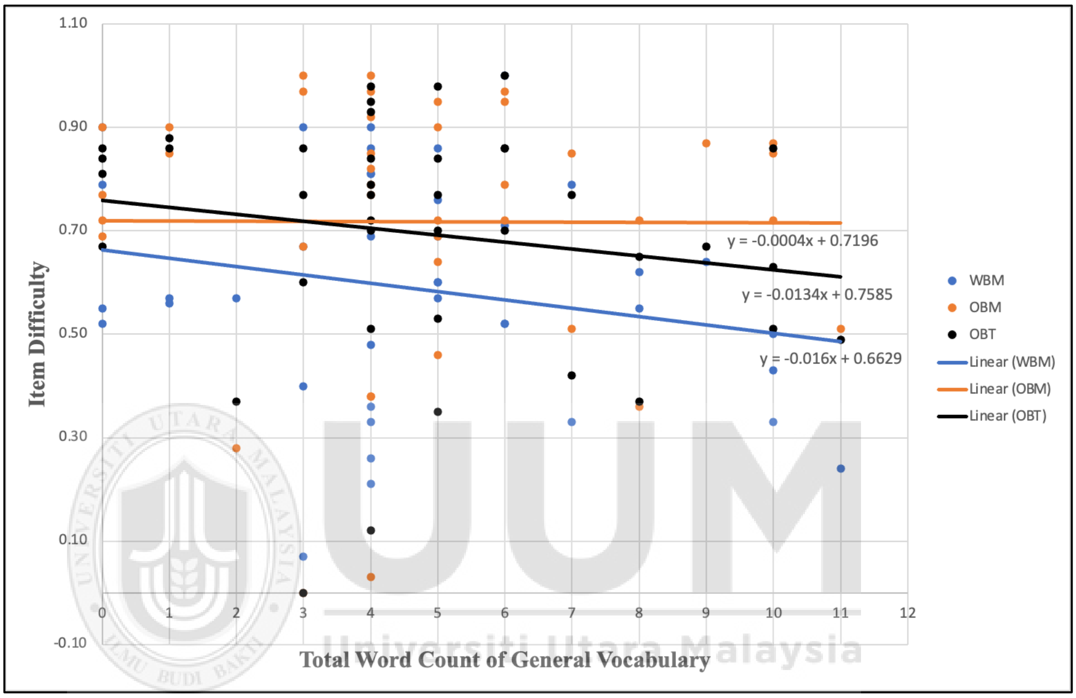


Figure 4.1 Scatter Plot of General Vocabulary and Item Difficulty

The scatter plot diagram in Figure 4.1 offers valuable insights into the relationship between total word count of general vocabulary and item difficulty across three mathematics tests: WBM, OBM, and OBT. The y-intercept of the lines of best fit serves as a baseline for item difficulty when no general vocabulary is present. Notably, the y-intercepts suggest that at this baseline level, the OBT test had the highest item difficulty, followed by OBM and WBM tests, implying that without general vocabulary, pupils found the OBT test easiest while the WBM test was the hardest.

The gradients of these lines depict a negative linear relationship, indicating that as the total word count of general vocabulary increases, item difficulty decreases. This implies that *Orang Asli* pupils perceived the test items to be more challenging as the number of general vocabulary words increased. Specifically, the steeper negative gradient for the WBM test compared to OBM and OBT tests suggests a stronger inverse relationship between general vocabulary word count and item difficulty in the WBM test.

By examining specific word count ranges, when vocabulary count was zero to three words, the OBT test was deemed easiest. However, as the word count increased from three to 11 words, the OBM test became comparatively easier. Overall, with rising general vocabulary word count, the OBT test was found to be easier than the OBM test, particularly for shorter test items (< 5 words). Conversely, for moderate to longer test items, the OBM test was found to be easier over the OBT test in terms based on item difficulty values.

4.6.1.1.2 Correlation Analysis using Pearson Correlation. The study conducted Pearson correlation analysis to explore the association between total word count in general vocabulary and item difficulty across three math tests: WBM, OBM, and OBT. The findings revealed nonsignificant correlations between total word count for general vocabulary and item difficulty: WBM [$r(38) = -0.25$, $p = .117$], OBT [$r(38) = -0.18$, $p = .261$], and OBM [$r(38) = -0.03$, $p = .864$]. Consequently, no statistical support emerged for a substantial relationship between these variables, suggesting external factors beyond the study's scope might influence their connection.

However, the correlation coefficients among the tests indicated a stronger inverse relationship between general vocabulary word count and item difficulty in WBM, followed by OBT, and least in OBM. Hence, an increase in general vocabulary words led to harder items for Orang Asli pupils in WBM compared to OBM and OBT. Furthermore, the weaker inverse relations in OBM and OBT tests suggested a smaller difficulty decrease with increasing general vocabulary words. This implied Orang Asli pupils found OBM and OBT items comparatively easier as vocabulary words increased. In summary, the study suggests *Orang Asli* pupils perceived OBM and OBT tests easier than WBM as general vocabulary word count rose.

4.6.1.2 Relationship of Mathematics Vocabulary and Item Difficulty

To examine the relationship between mathematics vocabulary and item difficulty, the total word count for mathematics vocabulary was also counted and recorded for the 40 mathematics test items across WBM, OBM, and OBT tests. The total word count for the mathematics vocabulary for the test items were the same across the three mathematics tests. Table 4.17 presents the total word count for mathematics vocabulary and the item difficulty for each test item across WBM, OBM and OBT tests.

Table 4.17

Total Number of Word Count of Mathematics Vocabulary and Item Difficulty

Item	Item Difficulty (p)			Mathematic Vocabulary
	WBM	OBM	OBT	
1	0.52	0.97	0.86	2
2	0.76	0.69	0.84	2
3	0.86	0.95	0.98	2

Table 4.17 (Cont.)

Item	Item Difficulty (p)			Mathematic Vocabulary
	WBM	OBM	OBT	
4	0.90	0.82	0.77	1
5	1.00	0.95	1.00	1
6	0.71	0.79	0.86	1
7	0.81	0.77	0.84	1
8	0.90	0.90	0.86	0
9	0.52	0.72	0.67	0
10	0.79	0.69	0.84	0
11	0.55	0.77	0.81	0
12	0.48	0.03	0.12	1
13	0.57	0.90	0.88	2
14	0.57	0.28	0.37	1
15	0.40	0.67	0.60	1
16	0.56	0.85	0.86	2
17	0.81	0.97	0.98	1
18	0.60	0.72	0.35	2
19	0.69	0.64	0.77	1
20	0.69	0.92	0.93	2
21	0.79	0.85	0.77	7
22	0.21	0.79	0.72	1
23	0.55	0.36	0.37	2
24	0.57	0.90	0.53	1
25	0.07	0.00	0.00	1
26	0.60	0.46	0.70	2
27	0.36	0.85	0.79	1
28	0.50	0.72	0.63	3
29	0.26	0.38	0.70	2
30	0.33	0.85	0.51	2
31	0.43	0.87	0.86	1
32	0.24	0.51	0.49	1
33	0.67	1.00	0.77	3
34	0.52	0.72	0.70	3
35	0.33	0.38	0.51	2
36	0.62	0.72	0.65	0
37	0.90	0.97	0.86	2
38	0.86	1.00	0.95	1
39	0.64	0.87	0.67	0
40	0.33	0.51	0.42	4

The relationship between total word count for mathematics vocabulary and the respective item difficulty obtained for WBM, OBM, and OBT tests were examined. This was carried out by employing scatterplot diagram and Pearson correlation are presented in the following sub-subsections.

4.6.1.2.1 Correlation Analysis using Scatterplot Diagram. A scatter plot diagram was also employed as a visual tool to examine the relationship between the total word count of mathematics vocabulary and item difficulty between the three mathematics tests. Figure 4.2 depicts the scatter plot diagram and line of best fit between the two variables across the WBM, OBM, and OBT tests.

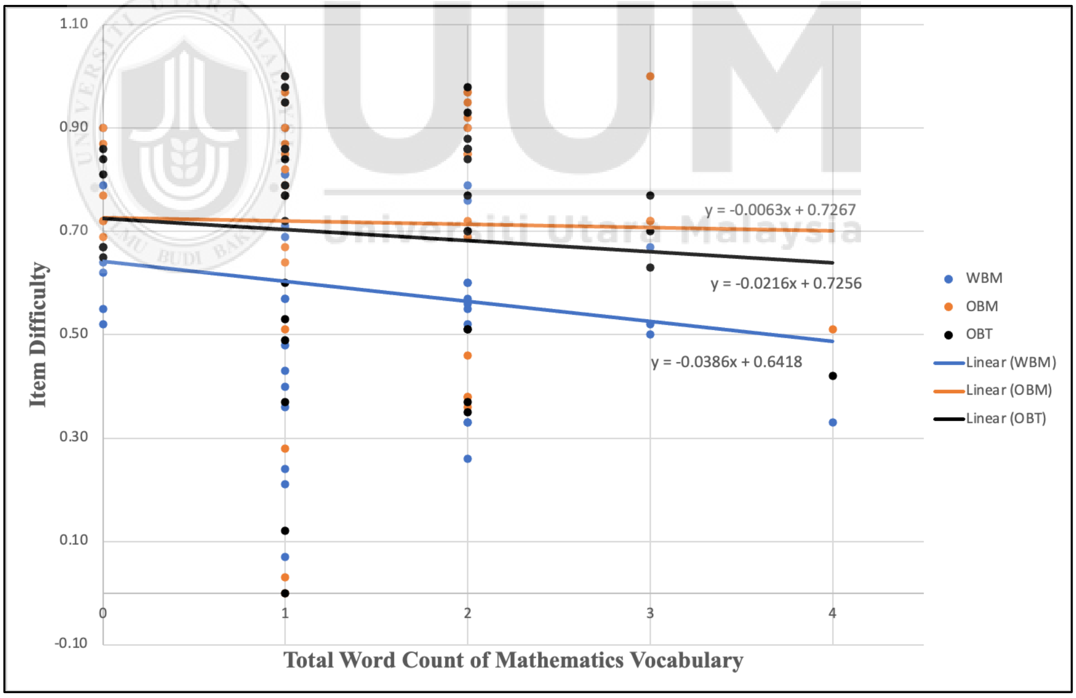


Figure 4.2 Scatter Plot of Mathematics Vocabulary and Item Difficulty

The analysis of Figure 4.2 presents crucial insights into the association between total mathematics vocabulary and item difficulty in WBM, OBM, and OBT tests. The y-intercepts established baseline difficulty levels in the absence of mathematics vocabulary, demonstrating that both OBM and OBT tests were relatively easier than WBM when devoid of mathematics-related terms. This highlights the significant impact of vocabulary presence on item difficulty perception among *Orang Asli* pupils.

Furthermore, the negative gradients depicted a consistent decrease in item difficulty values as the total word count of mathematics vocabulary increased across all tests. This implies that as the number of mathematics vocabulary increased, the *Orang Asli* pupils found the test item to be more difficult. Overall, the pupils found the OBM test to be the easiest, followed by OBT test and WBM test as the total word count of mathematics vocabulary increased within each test item.

4.6.1.2.2 Correlation Analysis using Pearson Correlation. The Pearson correlation coefficient was also performed to examine the relationship between the total word count of mathematics vocabulary and item difficulty between WBM, OBM, and OBT tests. For the three mathematics tests, the results of the Pearson correlation indicated that the relationship between the total word count for mathematics vocabulary and item difficulty were not significant for WBM test, [$r(38) = -0.03$, $p = .871$], OBT test [$r(38) = -0.03$, $p = .850$], and OBM test [$r(38) = -0.04$, $p = .825$]. Based on the results, there is no statistical evidence to support a significant relationship between total word count of mathematics vocabulary and item difficulty.

Nonetheless, by comparing the Pearson correlation coefficient values obtained between the three mathematics tests in the context of this study, the study found that the relationship between the total word count for mathematics vocabulary and item difficulty for WBM, OBM, and OBT tests were negligible negative correlation between the two variables. Therefore, this study found that there may not be a significant relationship between the total word count of mathematics vocabulary and item difficulty of the test items across the WBM, OBM, and OBT tests. In other words, the total word count for mathematical vocabulary did not have a substantial impact on the difficulty level of the test items.

In conclusion, based on the statistical analysis carried out using the scatterplot diagram and Pearson correlation coefficient, this study found that although there is an overall negative correlation between the total word count of general vocabulary and mathematics vocabulary with item difficulty, the relationship between these variables is not statistically significant with respect to the general population. For both vocabulary type between the three mathematics tests (WBM, OBM, and OBT tests), as the total word count of general vocabulary and mathematics vocabulary increases, the item difficulty decreases.

This indicates that the *Orang Asli* pupils had found the test items harder when total word count of both vocabulary types increases across the three mathematics tests. Nonetheless, the study also found that the magnitude of the decrease in item difficulty was lower when administered the OBM and OBT tests, compared to the WBM test. This study found that, between WBM, OBM and OBT, the pupils found the test items with increasing total word count of both vocabulary types easier when administered the OBM test followed by the OBT test, as opposed to the WBM test.

4.6.2 Qualitative Content Analysis of Vocabulary Type and Item Difficulty

As both general vocabulary and mathematics vocabulary were similar across the three mathematics tests, qualitative content analysis was carried out to explore the types of general vocabulary and mathematics vocabulary embedded within the mathematics test items based on the item difficulty across the three mathematics tests. Subsequently, based on the content analysis, the development of themes and subtheme were carried out to identify the patterns on the types of vocabulary which improved item difficulty levels when administered the recorded oral tests.

The following subsection presents firstly the summary of the content analysis carried out based on the item length categories for short, moderate and long item length test items. Based on the content analysis, the next subsection presented the corresponding themes and subthemes that corresponded to the test item which showed improved item difficulty when administered the OBM and OBT tests, compared to WBM test.

4.6.2.1 Summary of Content Analysis of Vocabulary Type

The detailed analysis of the content analysis conducted on the 40 mathematics test items on both general vocabulary and mathematics vocabulary according to item length categories is presented in Appendix G. To sum up, the content analysis carried out on the mathematics test items yielded compelling insights. This study found that the item length categories, comprising of short, moderate, and long item lengths, did not influence item difficulty of the test items among the *Orang Asli* pupils. This assumption was made as the pupils' performance on the test items developed in this study varied regardless of the total word count of the test items.

Nonetheless, this study found that vocabulary types, which include both general vocabulary and mathematics vocabulary, used for the test items have a more significant role in influencing the item difficulty. It was evident from the analysis across the three mathematics tests (WBM, OBM, and OBT tests) that certain vocabulary types, such as specific words, contextual terms and terminologies posed a greater challenge to the *Orang Asli* pupils. Therefore, the findings of this study highlighted that there is a clear relationship between vocabulary type, comprising of general vocabulary and mathematics vocabulary, on the item difficulty for this sample of *Orang Asli* pupils.

Based on item difficulty index computed for each test item, the study found that a total of 23 mathematics test items were found to be easier for the *Orang Asli* pupils when administered the OBM and OBT tests, compared to the WBM test. The remaining 17 test items, including two items (Item 8 and 10) which were purely arithmetic items with no vocabulary, were found to have approximately similar item difficulty range across the three tests. The 17 test items did not merit further examination on the relationship between item difficulty and vocabulary types, as these test items were within similar item difficulty ranges across the three mathematics tests. As such, this study focused on the 23 test items, including Item 9 and 11 which had no vocabulary, which were found to be easier when administered the recorded oral test in either language. These test items were further analysed using content analysis to identify the types and characteristics of the vocabulary that was found to be easier by the pupils when delivered orally in the mathematics assessment.

Based on 23 test items, the process of coding was carried out, which was then used to explore the patterns related to the types of vocabulary used in the test items which led to the improved mathematics performance observed among the *Orang Asli* pupils. The analysis of the codes revealed distinct themes and corresponding subthemes that corresponded to the vocabulary types within the 23 mathematics test items. Table 4.18 presents the themes and subthemes of the types of vocabulary which showed improved item difficulty when administered the OBM and OBT tests.

Table 4.18

Themes and Subthemes of Vocabulary Type for Items Exhibiting Improved Item Difficulty for OBM and OBT Tests

Themes	Subthemes	Items
General Vocabulary	General vocabulary typically used in mathematics learning	20, 27, 34
Mathematics Vocabulary	Mathematics vocabulary used in real-world context	1, 13, 20, 22, 24, 29, 31, 32, 34
	Mathematics vocabulary accompanied with diagram	12, 14, 23, 25
	Mathematics vocabulary specific to mathematics content	16, 18, 27, 30, 33
No Vocabulary	No words used and included numerals only.	9, 11

From the thematic analysis carried out, this study found three emergent themes from the thematic analysis of the test items that are related to vocabulary type, namely general vocabulary, mathematics vocabulary and no vocabulary. The first theme that emerged from the data is general vocabulary, which consisted of one subtheme, namely general vocabulary typically used in mathematics learning. Three items (Items 20, 27 and 34) were found to have general vocabulary that are usually used in the classroom during mathematics teaching and learning. The study found that test items that contained familiar general vocabulary that are typically used during mathematics teaching and learning was found to make the test item easier for the Orang Asli pupils when administered the OBM and OBT test. For instance, the word '*kumpulan*' and '*membilang*' for Items 20 and 27 respectively are typically used in the classroom setting.

The second theme that emerged from the data is mathematics vocabulary. Three subthemes were identified under this theme, which are mathematics vocabulary used in real-world context, mathematics vocabulary accompanied with diagram, and mathematics vocabulary specific to mathematics content. For the first subtheme of mathematics vocabulary used in real-word context, there were a total nine test items (Item 1, 13, 20, 22, 24, 29, 31, 32 and 34) that corresponded to this first subtheme. Vocabulary such as '*tidak sama banyak*' (not the same amount), '*duit syling'* (coin), '*waktu*' (time) and '*wang saku*' (pocket money) are some examples of mathematics vocabulary that are culturally relevant and relatable to the *Orang Asli* pupils encountered in their everyday lives and has real-word application. For example, Item 13 relates to the topic of money, whereby the term '*duit syling*' (coin) is a mathematical term that is used in dealing with everyday money transactions.

Therefore, this study shows that the pupils found the test items to be easier when using mathematics vocabulary that are aligned or encountered in their daily lives.

Other than that, four test items (Item 12, 14, 23, and 25) were related to mathematics vocabulary that were accompanied by a diagram. This study found that the *Orang Asli* pupils were able to answer these test items easily in the OBM and OBT tests as these test items contained additional information presented in the diagram that was associated with mathematics vocabulary found within the test items. For example, the term '*pecahan*' (fraction) in Item 12 was accompanied with a diagram that contained a single shaded region of a square that was divided equally by four. Item 25 contained a diagram of a cube, where the pupils were required to solve the test item by writing the shape (*bentuk*). The findings illustrate that pupils found the test items easier when the mathematics vocabulary was closely related to the diagram.

Additionally, this study also found that five test items (Item 16.18, 27, 30 and 33) which pertained to the third subtheme, which is mathematics vocabulary that are specific to the mathematics content. These mathematics vocabularies are closely aligned with the curriculum content and mathematics topics that are taught during mathematics teaching and learning. For instance, Item 16 contains the mathematics term '*sisi*' (sides), which relates to the number of sides of a shape. This term bears a different meaning outside of the classroom context which is unrelated to the mathematics content. On the other hand, Item 27 contained the mathematics term '*sepuluh-sepuluh'* (tens) which is not typically used by the pupils outside of mathematics teaching. This mathematics vocabulary is typically used when teaching the mathematics topics of 'whole numbers up to 100' during mathematics lessons.

Lastly, for the theme of no vocabulary, one subtheme emerged from the data, which is no words used and included numerals only. For this subtheme, only two items (Item 9 and 11) were found to not contain both mathematics vocabulary and general vocabulary, but instead contained numerals. This study found that despite the lack of vocabulary within these test items, the *Orang Asli* pupils had found the test items easier in the OBM and OBT tests, compared to the WBM test. Therefore, this study also found that the OBM and OBT tests were easier for the *Orang Asli* pupils on computational mathematics test items that do not contain any words or terminology that may influence language proficiency or reading ability of the pupils.

Hence, this study conclusively established a noteworthy relationship between the vocabulary type for general vocabulary and mathematics vocabulary on the item difficulty obtained for the mathematics test items across WBM, OBM, and OBT tests. Furthermore, the study also further highlights the association between the total word count of general vocabulary and mathematics vocabulary with the item difficulty, and types of vocabulary that were found to be easier when administered the OBM and OBT tests to the *Orang Asli* pupils, compared to the WBM test. These results effectively answer research question three, shedding light on the crucial role played by vocabulary type in influencing item difficulty of the mathematics test items among the *Orang Asli* pupils.

4.7 Research Question Four

RQ4: What are the teachers' views on the use of the recorded oral test as a test accommodation to facilitate the mathematics performance among *Orang Asli* pupils?

To answer the fourth research question developed for this study, the qualitative research approach which utilised the use of semi-structured interview questions was carried out. Thematic analysis of the semi-structured interview data from the seven mathematics teachers interviewed for this study revealed three overarching themes, which are: factors affecting pupils' performance pre-test accommodation, support for post-recorded oral test accommodation and future challenges on the use of recorded oral test. The themes, subthemes and responses from the participants are presented in the following section.

4.7.1 Factors impacting pupils' performance pre-test accommodation

The first main theme that emerged from the data is the 'factors impacting pupils' performance pre-test accommodation'. Under this main theme, four subthemes were identified, which are pupils' lack of interest in learning, parental involvement, inadequate language comprehension and reading difficulties.

The first sub-theme identified was the lack of interest when learning mathematics. For instance, participant P3 expressed that "*It is their attitude. What I mean, is that their thoughts on education… is that they care less*". One participant stated that the *Orang Asli* pupils had also lack motivation when answering mathematics questions during the assessments. Participant P6 stated that "*when we conduct assessments, the pupils do not have a desire to answer the questions. So, as teachers, we have to guide them... become their guide. We have to help motivate the pupils to answer the questions*". In addition, Participant P6 stated that "*they do not want to take risks... or they do not want to try*". Conversely, to promote interest when learning mathematics, Participant P2 asserted that "*for them to be motivated, we (the* *teachers) must do group activities... then they will start to show interest and be motivated*".

The lack of parental involvement in the pupils' mathematics education was the second subtheme that was identified from the data. This was also reoccurring subtheme whereby most participants highlighted that the *Orang Asli* pupils seldom receive any help with respect to their education from their parents. Participant P2 highlighted that "*the parents do not provide any support to their children... because when we give them (pupils) homework, the parents do not help them. Sometimes, the parents do not know how to read. So, how will they be able to help?*". This notion also echoes the response by Participant P3 whereby the participant stated that "*their parents' level of education too is low. So, they do not know how to teach their children. They don't know. So, the pupils are only able to learn at school*".

Besides pupils' lack of interest in learning and lack of parental involvement, the next subtheme that emerged is inadequate language comprehension. The participants suggested that despite the *Orang Asli* pupils can converse in *Bahasa Melayu*, the pupils still have difficulties in comprehending mathematics questions and test items when presented in the language of instruction. For instance, Participant P5 stated that "*their main language... their first language is Bahasa Temiar. Their second language is Bahasa Melayu. So, when we give them questions, they have difficulties in understanding the question... Sometimes, the teachers will have to explain what the question requires... even though they know how to speak in Bahasa Melayu*". Moreover, Participant P6 noted that both the language of instruction and ethnic language have similarities, however, the pupils are unable to comprehend the language of instruction, as similar terminologies may have different contextual meaning in their ethnic language. Participant P6 asserted that "*language proficiency of the pupils... actually, they can understand. However, on questions that involve connecting relationships… there are similar terminologies used in both Bahasa Temiar and Bahasa Melayu but are used in different context*".

The last subtheme that was identified is reading difficulties. Participant P5 stated that "*sometimes, they can read. But when it comes to understanding the question, it takes time for them to process the questions. Sometimes they may misinterpret the context of the question. This may lead them to answer the question wrongly*". Additionally, Participant P7 asserted that "*for mathematics, there are pupils who can read but do not understand the question... Meaning, they can read the question, but they still could not answer. When asked, "Why can't you answer (the question)?" ... They answer, "I don't understand". This shows that they can only read the question*".

Lastly, there is also a clear distinction on the issue between language proficiency and reading difficulties faced by the pupils. There seems to be a clear divide on what may be the main cause in affecting the mathematics performance of the pupils. Nonetheless, from the data, reading difficulties may play a bigger part in affecting their performance when administered mathematics tests, as resonated by most participants. On one such occasion when asked about the possible factor that may have affected the pupils' mathematics performance between the two issues, Participant P3 sincerely asserted that "*This is not a language problem, but it is a* *reading problem*", as uttered in *Bahasa Melayu "bukan masalah bahasa, tetapi masalah membaca".*

4.7.2 Support for post-recorded oral test accommodation

The second overarching theme was support for 'post-recorded oral test accommodation'. Two subthemes were identified from the responses that highlights this main theme, which are preference towards oral delivery and promotion of text comprehension. For the first subtheme, preference towards oral delivery, which suggested that most participants agreed that the use of the recorded oral test would be beneficial and helpful for the *Orang Asli* pupils. For instance, Participant P1 stated that "*Yes... If it is an oral test, where we read aloud, this will help*". Furthermore, Participant P3 asserted that "*the method of use (recorded oral delivery) ... I also think it is good... orally in Malay or orally in Temiar. Students will be able to listen and answer*".

Other than that, the second subtheme was also identified from the mathematics teachers' responses when asked about the use of the recorded oral test. Most participants proposed that the support for the use of the test accommodation stems for its use to promote text comprehension when the mathematics test is administered. For example, Participant P3 asserted that "*when we give mathematics questions... we understand that not all pupils have mastered reading. They cannot read well. So, we have to help the pupils by reading out loud the question to help them understand the question*". This notion resonated with Participant P5, whereby it was stated that "*the pupils will understand better. They can understand easily when we conveyed orally. It doesn't matter if it is in Bahasa Temiar or Bahasa*

Melayu. But, as long as I have been here, when questions are delivered orally, they can understand easily and can answer any question".

4.7.3 Future challenges on the use of recorded oral test

The final theme that emerged from the data is 'future challenges on the use of recorded oral test'. During the interview sessions, the participants proposed certain issues and challenges that may arise if the recorded oral test were to be administered to the *Orang Asli* pupils. From the data, two subthemes were identified which are pupils' challenges and teachers' challenges.

For the first subtheme, pupils' challenges, one of the potential challenges that the *Orang Asli* pupils may face is that the use of the ethnic language for the test accommodation may inhibit their development of proficiency in the language of instruction. For instance, when asked on the potential challenges that the pupils may face when using the ethnic language in the recorded oral test, Participant P1 simply stated that "*their (proficiency) Bahasa Melayu will be weak*".

Other than that, Participant P4 highlighted that "*they (pupils) need to be proficient in Bahasa Melayu... We may solve the problem in mathematics, but it may lead to problems in other subjects like Bahasa Melayu*". Participant P6 stated further that "*even though Bahasa Temiar could be introduced (as a test accommodation) in mathematics... I believe that the oral test in Bahasa Melayu needs to be emphasized*". This issue was also echoed by Participant P4, whereby Participant P4 had highlighted that "*the pupils may develop problems in other subjects such as science and Bahasa Melayu, which requires pupils to build sentences and write* *essays... it may solve the problem in mathematics but there will be problems with other subjects*".

The second subtheme is teachers' challenges in implementing recorded oral test accommodation. One of the major challenges that the participants had highlighted was on the issues for implementation of the recorded oral test for the *Orang Asli* pupils. For instance, Participant P5 posited that "*in terms of implementation... Because this is an oral test, it requires more time... we have to read (questions) one by one… In terms of the benefits, it's really good... because the pupils will be able to understands easily. But in terms of its implementation... in terms of its practicality, we have to investigate that...*".

Moreover, this notion was also supported by Participant P4 who further added that "*there may be problems in terms of duration (of the test) … and its implementation. The way it is implemented and the duration of the test*". Other than that, some participant voiced concern over the development of the recorded oral test in the ethnic language, whereby Participant P6 stated that "*to find a teacher who can speak Bahasa Temiar... it is very hard*". This view was also shared by Participant P7 who asserted that "*to find a teacher from the Temiar ethnic group... there isn't any*".

4.8 Chapter Summary

This chapter presented quantitative and qualitative findings based on the research questions established in this study. This chapter presented the quantitative findings that provided statistical evidence to support the use of the recorded oral tests in either the language of instruction or ethnic language as possible and viable test accommodation to assess the mathematics performance of the *Orang Asli* pupils. Additionally, the qualitative data offered valuable insights into the use of the recorded oral tests that could potentially be incorporated into current mathematics testing practice that could cater to the linguistic and educational needs of the *Orang Asli* pupils.

CHAPTER FIVE

DISCUSSION AND IMPLICATION

5.1 Introduction

This study employed a mixed method approach, whereby the quantitative approach was aimed at examining the use of the recorded oral tests in both the language of instruction (*Bahasa Melayu*) and ethnic language (*Bahasa Temiar*) as a test accommodation on test items categorised as short, moderate and long item length, and between general vocabulary and mathematics vocabulary among 142 Year 4 *Orang Asli* pupils. On the other hand, the qualitative approach of this study explored the perspectives of seven mathematics teachers on the use of the recorded oral test to facilitate the mathematics performance of the *Orang Asli* pupils during classroom mathematics testing. The following research questions were developed to guide this study:

RQ 1: Is there a significant difference in the mathematics performance of the *Orang Asli* pupils between the three mathematics tests (WBM, OBM, and OBT tests)?

RQ 2: Is there a significant difference in the mathematics performance of the *Orang Asli* pupils by item length categories (short, moderate, and long) between the three mathematics tests (WBM, OBM, and OBT tests)?

RQ 3: What is the relationship between vocabulary type (general vocabulary and mathematics vocabulary) and item difficulty of the test items across the three mathematics tests (WBM, OBM, and OBT tests)?

RQ 4: What are the teachers' views on the use of the recorded oral test as a test accommodation to facilitate the mathematics performance among *Orang Asli* pupils?

For the quantitative approach of this study, the mathematics performance of the *Orang Asli* pupils was examined between three mathematics tests (WBM, OBM, and OBT tests). This study also investigated the impact between the mathematics tests on the item length categories; short, moderate and long item length test items, and vocabulary; general vocabulary and mathematics vocabulary. Subsequently, content analysis was carried out to explore the types of vocabulary within the test items that were found to show improved mathematics performance when administered the recorded oral tests, compared to the written mathematics test.

For the qualitative approach undertaken in this study, thematic analysis was conducted on the semi – structured interviews with the mathematics teachers to identify the themes and subthemes that pertains to their views on the use of the recorded oral test in mathematics assessments for the *Orang Asli* pupils. This chapter provides a summary of the research findings from the data analyses carried out, discussion of the findings, research implications and recommendation for future research.

5.2 Summary of Findings

Through rigorous investigation and data analysis carried out, the findings were pertinent and effective in addressing the research questions for this study. To address research question one, the Brown-Forsythe test and Games-Howell post hoc test was carried out on the average mathematics test scores obtained by the *Orang Asli* pupils between the three mathematics tests (WBM, OBM, and OBT tests). Overall, the study found that the *Orang Asli* pupils had obtained the highest average test scores when administered the OBM test, followed by OBT and WBM tests. The findings showed that the *Orang Asli* pupils performed better in the OBM and OBT tests, compared to the WBM test.

Furthermore, this study also found that there was a significant difference in the mathematics test scores between the three mathematics tests. Further analysis using the Games-Howell post-hoc test found that there was a significant difference between the OBM and WBM tests, and OBT and WBM tests. However, the study found that there was no significant difference between the OBM and OBT tests. Therefore, the findings indicated that there was a difference in the mathematics performance of the *Orang Asli* pupils between the recorded oral tests and the standard written test. Conversely, there was no difference in their performance between the recorded oral tests in either language.

To address research question two, the mathematics performance of the *Orang Asli* pupils was also examined across WBM, OBM, and OBT tests for the three item length categories, namely short, moderate and long item length. For short item length, a total of 15 test items were examined. The study found that, for short item length test items, the pupils performed better when administered the recorded oral test in both the languages, compared to the written test. However, the result of the Brown-Forsythe test indicated that there were no significant differences for short item length test items across the three tests. This showed that the pupils were able to perform equally well in all three mathematics tests for short test items.

A total of 14 test items categorised as moderate item length test items were analysed. The study found that the *Orang Asli* pupils had obtained the highest average test scores when administered the OBM test, followed by the OBT test and WBM test respectively. The results of the Brown-Forsythe test indicated that there was a significant difference between the three mathematics tests for moderate item length test items. Subsequently, the results of the Games-Howell post hoc test showed that there was a significant difference in the pupils' mathematics performance between the OBM and WBM tests, and OBT and WBM tests. However, there was no difference in their performance between OBM and OBT tests.

Lastly, a total of 11 test item categorised as long item length were examined. The study also found that the pupils had obtained the highest average test scores when administered the OBM test, followed by the OBT test and WBM test respectively. Overall, for long item length, the pupils performed better when administered the recorded oral test in both languages, compared to the written test. Subsequently, the Brown-Forsythe test indicated that there was a significant difference in the mathematics performance of the pupils between the three mathematics tests. The results of the Games-Howell post-hoc test showed that there was a significant difference in the mathematics performance between the OBM and WBM tests, and OBT and WBM tests. However, the study also found no significant difference in the pupils' mathematics performance between OBM and OBT tests for long item length test items.

In summary, the findings showed that the *Orang Asli* pupils were able to perform equally between the three mathematics tests for test items categorised as short item length. However, for moderate and long item length test items, there was a difference in the mathematics performance of the pupils between both the recorded oral tests and written test. Overall, with respect to the linguistic feature of item length, the findings indicated that test modality (oral vs written) and length of test item, rather than language choice was more effective in enabling the pupils to better exhibit their mathematical skills, as evidenced by their higher mathematics performance on the recorded oral tests, compared to the written test.

For research question three, the item difficulty for the 40 mathematics test items across the three mathematics tests were computed before examining the relationship between vocabulary type and item difficulty. The findings showed that the *Orang Asli* pupils found a higher number of mathematics test item to be easier when administered the OBM and OBT tests, compared to the WBM test. On the other hand, the pupils found more test items to be moderately difficulty when administered the WBM test. Two data analyses techniques were used, which are (a) using statistical analysis which included scatterplot diagram and Pearson correlation to examine the relationship between total word count of general vocabulary and mathematics vocabulary with item difficulty, and (b) qualitative content analysis to identify patterns of the type of vocabulary within the test items that showed improved item difficulty when administered the OBM and OBT tests, compared to the WBM test.

The findings from the scatterplot diagram showed that there was a negative relationship observed from the line of best fit between the total word count of general vocabulary and mathematics vocabulary, with item difficulty respectively. The findings showed that as the overall total word count increased for both vocabulary types, the *Orang Asli* pupils found the mathematics test items to be much easier when administered the recorded oral test in the language of instruction, followed by the recorded oral test in the ethnic language. For increasing total word count of both vocabulary types, the pupils found the test items to be more difficult in the written test. Therefore, as the number of words for general vocabulary and mathematics vocabulary increased within the test items, the lower the item difficulty or the more difficulty the test items were to the *Orang Asli* pupils.

The Pearson correlation was also carried out to determine the significance of the relationship between the two variables. The correlation analysis was in line with the scatterplot diagram whereby it was found that there was a negative weak correlation between total word count of general vocabulary and mathematics vocabulary with the item difficulty respectively. With respect to the correlation between the three mathematics tests, there was a relatively stronger negative correlation between both the total word count of general vocabulary and total word count of mathematics vocabulary with item difficulty in the WBM test, followed by the OBT test and OBM test. The test items were found to be easier when administered the OBM and OBT test, compared to the WBM test for increasing total word count of both vocabulary types. Nonetheless, this study found that the relationship between both general vocabulary and mathematics vocabulary with item difficulty were not significant across the three mathematics tests when Pearson correlation was carried out.

Subsequently, the content analysis was carried out to explore the patterns of the type of vocabulary and its relationship with item difficulty for the 40 mathematics tests items. Therefore, content analysis was employed to examine item difficulty range of the test items based on the item length categories for short, moderate and long item length across the three mathematics tests. Overall, this study found that there was no association between item length categories and vocabulary type as there were no patterns observed between the two variables. Nonetheless, a comparison of the test item across the three mathematics tests showed that a total of 23 test items which comprised of both general vocabulary and mathematics vocabulary were found to be easier by the *Orang Asli* pupils when they were administered the OBM and OBT tests, compared to the WBM test.

Based on the 23 test items, three themes emerged from the patterns identified from the qualitative content analysis, which are general vocabulary, mathematics vocabulary and no vocabulary. For the theme of general vocabulary, one subtheme was identified, which is general vocabulary typically used in mathematics learning. The second theme was mathematics vocabulary and consisted of three subthemes, namely mathematics vocabulary used in real-world context, mathematics vocabulary accompanied with diagram and mathematics vocabulary specific to mathematics content. The last theme, which is no vocabulary, contained one subtheme which is no words or terminology and included numerals only.

For research question four, thematic analysis was carried out on the transcribed data to identify the themes and subthemes based on the responses from participants during the semi structured interviews. A total of three main themes emerged from the data. The first theme that emerged from the data was ‘factors affecting pupils’ performance prior to test accommodation’ which included four subthemes, namely pupils’ lack of interest in learning, parental involvement, inadequate language comprehension and reading difficulties. Other than that, the second theme that emerged was ‘support for post-recorded oral test accommodation’. This theme highlighted two subthemes, which are preference towards oral delivery and promotion of text comprehension. Lastly, the final theme identified was ‘future challenges on the use of recorded oral test’ and consisted of two subthemes which are pupils’ challenges and teachers’ challenges. These themes and subthemes highlighted the perspective of the mathematics teachers on the use of the recorded oral test as a viable test accommodation for the *Orang Asli* pupils.

5.3 Discussion of Findings

The findings from the quantitative approach undertaken in this study provide statistical evidence for incorporating the test accommodations to address the language barriers and reading difficulties faced by the *Orang Asli* pupils. Furthermore, the findings may also shed light on the impact of the test accommodation on the item length categories (short, moderate and long item length) and vocabulary type (general vocabulary and mathematics vocabulary) of test items. On the other hand, the findings from the qualitative approach provide additional converging evidence which attest to the potential utility of the recorded oral test as a promising test accommodation for the *Orang Asli* pupils. The findings for each research question formulated in this study were discussed in the subsequent subsections.

5.3.1 Research Question One

Research question one was established to determine if there was a significant difference in the mathematics performance of the *Orang Asli* pupils when administered the recorded oral tests in both the language of instruction and ethnic language. This study found that the pupils had performed higher when administered the OBM and OBT test, as opposed to their lower performance in the WBM test. Although there was significant difference between both recorded oral tests and written test, it was found that there were no differences in the performance of the pupils between the recorded oral tests in both the language of instruction and ethnic language. This finding may suggest that the recorded oral test could provide the needed linguistic support that could potentially alleviate the language and reading demands posed by the test when administered to the *Orang Asli* pupils, as supported by Arsaythamby et al. (2021), Kanageswari et al. (2021) and Zaleha et al. (2020).

This finding supports the results of past studies, whereby it was proposed that proficiency in the language of instruction plays a major role in influencing mathematics performance of the general pupil population (Chow & Ekholm, 2019; Martin & Fuchs, 2019; Peng et al., 2020). Nonetheless, this notion contradicts past studies carried out on *Orang Asli* pupils (Abdul Halim Abdullah, 2022; Ruzlan et al., 2021, Kanageswari et al., 2023), whereby these studies proposed that the lack of proficiency in the language of instruction is a factor that have negatively impacted the mathematics performance of the pupils. This is not the case as the findings in this study represents a unique contribution to the field, suggesting that, at least for *Orang Asli* pupils, language may not necessarily be a predominant factor in influencing their mathematical performance, particularly as they progress to higher grade levels as the sample of this study are Year 4 pupils.

Furthermore, it may appear that the improved the mathematics performance was due to the different test modality used in this study, whereby the oral modality of the mathematics test was found to benefit the pupils compared to the written test, as highlighted in the studies by Gandhi et al. (2018), and Giusto and Ehri (2019). The better performance observed in the oral modality of the mathematics tests, supplemented by the recorded oral tests, may suggest that the reading demands posed by the test items were potentially reduced when the test items were delivered orally to the pupils. This finding is supported by past studies which highlighted the use of the recorded oral test as a valid measure of addressing the reading demands posed by the test items (McMahon et al., 2016; Spiel et al., 2019).

This intriguing finding considers that reading difficulties and text comprehension could be pivotal factors in determining the *Orang Asli* pupils' mathematical success. The findings highlight that pupils were able to perform better when administered the recorded oral tests in both languages, compared to written tests. Therefore, a possible explanation is that the orally delivered test may have mitigated the reading demands posed by the mathematics test items, which may have subsequently allowed the pupils to perform better. These findings highlight two important issues.

Firstly, the *Orang Asli* pupils may be facing reading difficulties that have negatively impacted their ability to perform well in mathematics assessments, as highlighted in past studies (Ruzlan et al., 2021; Kanageswari et al., 2023; Zaleha et al., 2020). In essence, it underscores the crucial role of reading ability as a supplementary skill that is required for the pupils to excel in mathematics assessments. This notion is in line with studies by Koponen et al. (2018) and Purpura et al. (2019), which proposed that reading difficulties experienced by pupils may affect their ability to fully comprehend the mathematics test items, subsequently leading to lower mathematics performance.

Secondly, when pupils face reading difficulties, such as struggling to decode the written text or comprehend sentences effectively as proposed by Cirino et al. (2018) and Salihu et al. (2018), these pupils may require linguistic support in the form of test accommodation. These findings highlight the effectiveness of incorporating test accommodations in mathematics testing that caters to the *Orang Asli* pupils to help mitigate the reading difficulties experienced by them, as highlighted in past studies among mainstream pupil population who may experience similar literacy issues (Kim, 2016; Spencer et al., 2022; Spiel et al., 2019). The findings of this study are aligned with notion that the use of the recorded oral test could therefore provide the necessary literacy support needed in removing reading as an access skill when measuring the *Orang Asli* pupils' mathematics performance.

The findings also align with the study by Zaleha et al. (2020) whereby it was found that the oral delivered test was effective in improving the mathematics performance of the *Orang Asli* pupils by addressing the language and reading component embedded within the mathematics test items. The orally delivered test, using the recorded oral test, was able to minimise construct irrelevant variance brought upon by both language and reading as secondary dimension that were unintended to be measure (Buono & Jang, 2021; Kan et al., 2019). This allowed the *Orang Asli* pupils to better exhibit their mathematical knowledge and skills when being assessed.

Additionally, the incorporation of the ethnic language in the test accommodation has been found to be beneficial in allowing the *Orang Asli* pupils to exhibit their mathematical knowledge and skills. This is in line with the findings by Arsaythamby et al. (2021) and Kanageswari et al. (2021) whereby it was found that *Orang Asli* pupils were able to perform better when the ethnic language was also incorporated, compared to the standard written mathematics assessment in the language of instruction. The effectiveness of incorporating the ethnic language in mathematics testing points to the efforts in creating culturally responsive and accessible tests that acknowledges the cultural and language nuances of the *Orang Asli* pupils, subsequently allowing them to use the language they are more proficient in when assessing their mathematics performance, as advocated by Kūkea Shultz and Englert (2021), and Lopez (2020).

Although past studies have posited that language proficiency in the language of instruction may impact mathematics performance, particularly among early school-going pupils (Abdul Halim, 2022; Lovett & Nelson, 2021; Misnaton Rabahi et al., 2020), the use of the ethnic language in mathematics may still be beneficial and warrants consideration when developing mathematics tests for the *Orang Asli* pupils. This additionally corroborates the recommendation by McCarty (2021), Trumbull and Nelson-Barber (2019) and Turkan and Oliveri (2016) on the incorporation of the ethnic languages of the pupils to ensure culturally responsive mathematics tests are developed to validly assess the mathematics performance of the pupils. The inclusion of the ethnic language in the mathematics test in this study proved to be beneficial in allowing pupils to exhibit their mathematical knowledge and skills.

Lastly, the findings are also aligned to the framework of the Cognitive Load Theory which shed light on the impact of the test accommodation as language and literacy support on the pupils' mathematics performance. As the language barriers faced by the pupils in assessments may lead to increased extraneous cognitive load as highlighted by Gillmor et al. (2015) and Roussell et al. (2017), the test accommodation proposed in this study could potentially help reduce the cognitive burden posed by this language mismatch. This is in line with studies by Haag et al. (2015) and Walkington et al. (2018), whereby the approach of using test accommodation that provides direct linguistic support offers a crucial avenue for reducing on the cognitive resources faced by the *Orang Asli* pupils.

In conclusion, the findings obtained to answer this research question establishes three key points which attest to the use of the recorded oral tests in both the language of instruction and ethnic language for the *Orang Asli* pupils. Firstly, proficiency in the language of instruction plays a pivotal role in enhancing mathematics performance of the pupils. Secondly, the use of the oral test modality underscores reading difficulty as a more prominent factor in influencing mathematics performance, rather than language choice use in the development of the mathematics test. Lastly, the integration of the ethnic language continues to yield notable benefits for the *Orang Asli* pupils in primary schools. By aligning with these broader findings, incorporating the recorded oral tests tailored to the linguistic needs of *Orang Asli* pupils emerges as a promising strategy.

5.3.2 Research Question Two

Research question two was established to examine the impact of the item length categories when assessing mathematics performance of the *Orang Asli* pupils. For all categories of item length (short, moderate & long item length), it was highlighted that the *Orang Asli* pupils had obtained a higher mathematics performance when administered the recorded oral test in both the language of instruction and ethnic language, compared to the written test. For short items, the study found that there is no difference in the mathematics performance between the three tests, however for moderate and long items, there is a significant difference in the performance of the *Orang Asli* pupils. The better performance observed in the recorded oral tests, regardless of the language used in the mathematics tests may indicate that reading ability has a more prominent role in ensuring that the pupils are able to comprehend the text easily (Cruz Neri et al., 2021; Walkington, 2017).

The findings from this study regarding the relationship between item length and pupil’s mathematics performance are aligned with previous research. The findings corroborate with several studies which have highlighted the challenges posed longer test items, particularly among pupils who face language and literacy barriers when assessed in mathematics (Bergqvist et al., 2018; Haag et al., 2015; Sibanda, 2017). The issue of language comprehension and reading difficulties may have resulted in the *Orang Asli* pupils' low mathematics performance when administered the written tests, compared to those administered the recorded oral tests., and that the pupils found it difficult to decode and comprehend the context of the question, especially when dealing with the longer test items, as suggested by Walkington et al. (2018) This may be attributed to the reading difficulties faced by these pupils, as proposed by Ruzlan et al. (2021) and Kanageswari et al. (2023) which could manifest as a struggle to effectively decipher the words and phrases used in the test item when dealing with written mathematics tests.

Moreover, the cognitive demands associated with processing longer and more complex linguistic structures may have played a role in affecting the mathematics performance of the *Orang Asli* pupils. As the test items become lengthier, the cognitive demand placed on the *Orang Asli* pupils may make them struggle to fully grasp the content and requirements of the test items due to the higher number of words. The complexity of lengthy items, with intricate sentence structures or numerous details, can impede effective information extraction, as highlighted in studies by Cruz Neri et al. (2021) and Walkington (2015).

As a result, these pupils may have found it challenging to formulate an accurate response, even if they possess the necessary mathematical knowledge and skills (Sibanda, 2017). However, as there is a dearth of literature that focusses exclusively on the impact of reading difficulties faced by *Orang Asli* pupils in mathematics testing, the findings of this study may suggest that the use of the recorded oral test may have, to certain degree, mitigated the possible reading difficulties experienced by the pupils, as supported by Kanageswari et al. (2021) and Zaleha et al. (2020). By

By incorporating an oral test modality for mathematics tests, this measure may potentially reduce the decoding demands posed by the moderate and long test items that could further impact comprehension. This is in line with studies by Gandhi et al. (2018) and Giusto and Ehri (2019) among most mainstream pupils who experiences reading difficulties in mathematics test, whereby these studies found that test accommodation that are delivered orally were able to help improve the performance of pupils who have low decoding ability, especially when dealing with long worded mathematics test items. Therefore, using test accommodation, through recorded oral test, could potentially be used to help pupils navigate test items that poses substantial reading demands typically encountered in standard the written test (Cirino et al., 2018; Koponen et al., 2018; Sireci, 2020)

5.3.3 Research Question Three

To answer this research question, item difficulty index was computed for the 40 mathematics test items to determine the relationship between the vocabulary type within the test items with item difficulty. Overall, the findings of this study provide compelling evidence that the use of the recorded oral tests in the language of instruction and ethnic language were found to be much easier among the *Orang Asli* pupils, compared to the written test. The higher item difficulty observed for the written test could be indicative of the complexities posed by the vocabulary, particularly mathematics vocabulary, that may negatively impact pupils' performance in mathematics.

This notion is in line with previous studies whereby mathematics involves not only the comprehension of mathematical concepts and problem-solving techniques but also the interpretation of specialised vocabulary that can be daunting for pupils who may already be struggling with language-related issues (Powell & Nelson, 2017, Riccomini et al., 2015; Unal, 2019). By presenting mathematics test items in an auditory format, the *Orang Asli* pupils are not as heavily reliant on the written vocabulary, instead, the pupils could focus more on the spoken language which may be more accessible to the pupils when vocabulary is delivered orally. Therefore, the findings may highlight the effectiveness of the recorded oral tests as a measure that may help reduce the language demands posed by both general vocabulary and mathematics vocabulary of the test items.

Moreover, this study found that there is a negative association between total word count of vocabulary type and item difficulty, whereby pupils found the test items to be more difficult as the number of words for both general vocabulary and mathematics vocabulary increased. The impact of an increasing number of words for both general vocabulary and mathematics vocabulary on the *Orang Asli* may be best understood through the lens of Cognitive Load Theory. When the pupils encounter numerous unfamiliar words in test items, it significantly increases their cognitive load, which can have detrimental effects on their ability to solve mathematical problems effectively, as proposed in the studies by Gillmor et al. (2005) and Walkington et al. (2019).

In this context, the increase in the number of words for both vocabulary type contributes to extraneous cognitive load, which is often considered detrimental to learning and problem-solving. This notion corroborates with the findings by Haag et al. (2013) whereby this study proposed that the increase in the item difficulty of mathematics test items by the German language minority pupils were potentially due to the longer item length, rather than vocabulary type. Although complex vocabulary may exist within the test items, the length of the test items was found to impart cognitive burden on the pupils solving the mathematics tests. The higher word count for both vocabulary type was subsequently mitigated using the recorded oral tests.

On the other hand, the findings could also be explained with respect to the use unfamiliar words for both vocabulary types. Taken into context that the *Orang Asli* pupils typically converse in the ethnic language, some vocabulary may not be familiar to them or used in their everyday context. Therefore, when the pupils are administered test items loaded with unfamiliar words, their cognitive resources are diverted towards decoding and understanding the language rather than applying mathematical principles, as highlighted by Moussa-Inaty et al. (2018). This diversion of cognitive resources leads to cognitive overload, making it difficult for pupils to focus on the core mathematical content of the test. As a result, they may struggle to apply mathematical principles correctly, even if they possess the requisite knowledge.

The better mathematics performance observed when administered the recorded oral tests points to some form of linguistic support awarded by the test accommodation to help the *Orang Asli* pupils to better comprehend the test items within increasing linguistic complexities. This is line with the findings with studies by Walkington

(2017), and Wolf and Leon (2009) whereby the impact of complex linguistic features in mathematics test items could be mitigated with the use of test accommodation. In this case, the support provided by the recorded oral tests as a test accommodation may have reduced the impact of unfamiliar mathematics vocabulary that may have impeded the pupils' comprehension and ability to employ the correct solution strategies.

Another interesting finding of this study is the higher performance of the *Orang Asli* pupils on items that incorporate mathematics vocabulary within real-world contexts when administered the recorded oral tests. One such explanation is that the improved performance observed by the pupils in the recorded oral tests may suggest that the pupils are able to solve the test item if it is within their own cultural context. This may enhance their understanding of mathematical concepts and vocabulary, as they could visualize and conceptualize them within real-world contexts that they are familiar with, in line with past studies (Ercikan and Por, 2020; Rigney et al., 2020), and among the *Orang Asli* pupils (Kanageswari et al., 2021).

When mathematics vocabulary is presented in real-world contexts, the pupils may be able to make connections between the abstract concepts and their everyday experiences, facilitating comprehension and problem-solving abilities. This finding aligns with studies emphasizing the importance of contextualizing mathematics test items based on *Orang Asli* pupils' experiences and language use both at home and school setting to promote deeper understanding as pupils would be able to better understand and respond to the test items (Ercikan & Por, 2020; Thomson, 2018). However, it is important to note that not all the findings obtained for each test item favoured the recorded oral tests. Some test items exhibited lower performance in the recorded oral test compared to the written test. One possible explanation is the lack of familiarity with orally delivered mathematics vocabulary.

For the test items that were within similar item difficulty ranges across the three test, a possible explanation would be that the *Orang Asli* pupils may be accustomed to reading and writing certain mathematical vocabulary or terms rather than hearing and using them in oral communication for certain topics taught during mathematics teaching and learning. Complex mathematics vocabulary can be challenging to understand and articulate correctly in an oral format or could be translated incorrectly, particularly if the pupils are not exposed to the mathematics vocabulary in their everyday conversations or do not have a linguistic equivalent word or term in their ethnic language (Solano-Flores et al., 2012). This essentially points to the role of language proficiency that is required during early formal schooling in developing familiarity with the vocabulary used within the context of mathematics, as proposed by Meaney and Trinick (2020), and Moschkovich (2020).

Moreover, the findings also show that the lower performance obtained on the items when presented with the recorded oral test had also contained diagrams. Therefore, this may suggest that when the recorded oral tests are employed, the *Orang Asli* pupils are required to process both visual and auditory stimuli simultaneously, which had subsequently increased their cognitive load when attempting to solve the test items. This is in line with the study by Ernest et al. (2018) whereby for low performing pupils, the inclusion of both visual and auditory stimuli was found to have increased the cognitive among pupils when attempting to solve the mathematics test items. Therefore, in the context of this study, it can be assumed that by adding additional stimuli in the mathematics test items, this may have inadvertently diverted cognitive resources of the *Orang Asli* pupils leading to increased intrinsic cognitive load. This may have subsequently led to the lower performance when examined.

5.3.4 Research Question Four

For research question four, the findings of the qualitative approach were discussed for the three main themes that emerged from the data. The subsequent subsections provide the discussion for the main themes, namely factors affecting pupils' performance prior to test accommodation, support for post-recorded oral test accommodation and future challenges on the use of recorded oral tests.

5.3.4.1 Factors impacting pupils' performance pre-test accommodation

The study found that one of the factors that can impact mathematics performance is the pupils' attitude themselves towards mathematics instruction. This is in line with studies by Mohd Azizul et al. (2020) and Mohammad Safwat (2020), whereby pupils' attitudes were identified as a key determinant of academic performance. These studies emphasize that how pupils perceive and approach mathematics can significantly influence their engagement and understanding of the subject, ultimately affecting their overall mathematics performance. Therefore, as *Orang Asli* pupils lack interest when learning mathematics, teachers are therefore needed to ensure that teaching strategies implemented throughout the mathematics teaching and learning process could arouse interest in learning that would help them perform well in mathematics tests.

One such measure that was recommended was to promote more group activity participation among the pupils. Activities such as group work classroom mathematics teaching and learning could promote collaboration, which may foster a positive attitude towards learning. Other than that, the collaboration between teachers and parents may create a strong support system, where teachers can provide guidance and personalised mathematics instructions while parents can reinforce the learning at home. As this study also found that parental involvement in the pupils' education is a key factor in affecting pupils' performance, this aligns with findings from Sumathi (2016), who emphasized that active parental engagement, such as providing academic support and fostering a conducive learning environment at home, significantly contributes to improving pupils' academic outcomes.When teachers, parents and pupils work together, there would be a higher likelihood of identifying and addressing learning challenges early.

The findings also underscore that the key determinant of success for the *Orang Asli* pupils in their mathematics performance is not primarily rooted in mathematical ability, but rather hinges significantly on their language proficiency and reading ability. It is evident that the *Orang Asli* pupils continue to encounter formidable language barriers, which impede their ability to excel in mathematics, as highlighted by Ruzlan et al. (2021) and Salim et al. (2022). This lingering challenge stems from their ongoing struggle to attain the necessary linguistic competence in the language of instruction, as they lack the opportunities to developed proficiency in the language as the use of their ethnic language remains dominant in conducting their day-to-day activities. Therefore, this language barrier has consistently impacted the *Orang Asli* pupils' mathematics performance.

These findings align closely with prior research by Macqueen et al. (2019), Preston and Claypool (2021) and Robertson and Graven (2020), which emphasized the pivotal role of language proficiency in shaping Indigenous pupils' mathematical capabilities. In essence, these perspectives collectively assert that the crux of the issue lies in language competency rather than innate mathematical aptitude. To facilitate improved mathematical performance among *Orang Asli* pupils, it is therefore imperative to prioritize and invest in measures that is aimed as accommodating their language and reading abilities. This shift in focus toward language development is crucial for enabling these pupils to manifest their genuine mathematical potential, ultimately highlighting that the root of their mathematics challenges lies in linguistic barriers that urgently require resolution.

Moreover, it is also vital that pupils can comprehend the language used in the test. Therefore, as these pupils are more adept in their respective ethnic language, the incorporation of the ethnic language as a form of test accommodation could be beneficial in improving the *Orang Asli* pupils' mathematics performance. As evidenced from the findings from this study, the pupils were able to perform significantly better in the test when the ethnic language was incorporated, as compared to their performance in the written test. This finding underscores the potential benefits of using the ethnic language as a tool to address linguistic challenges faced by these pupils, particularly in comprehending mathematics test items, as further advocated by Trumbull and Nelson-Barber (2019).

By considering the inclusion of either the language of instruction or ethnic language to supplement mathematics assessments, the recorded oral test as a valid measure deserves special attention in mathematics testing among *Orang Asli* pupils. The test accommodation may provide an equal and fair opportunity for *Orang Asli* pupils to demonstrate their knowledge and understanding that could better tap into their listening skills and understanding, which may be more developed in both languages. Other than it offers a form of inclusion, this test accommodation may allow these pupils to showcase their abilities without being hindered by language barriers that written tests might present. Therefore, this could potentially offer a more accurate reflection of their mathematical skills and knowledge when assessed.

5.3.4.2 Support for post-recorded oral test accommodation

This study also highlights the preference towards the recorded oral test in both the language of instruction and ethnic language as a means of addressing the language barriers and reading difficulties faced by the pupils. Pupils will be able to exhibit their true mathematical knowledge when the mathematics test is supplemented with an oral form. This is in line with the study by De Backer et al (2019) whereby it was proposed that to mitigate the issues faced by pupils who lack proficiency in the language of instruction, teachers agreed that the use of test accommodation for this group of pupils may be a valid measure in assessing their mathematics performance.

Furthermore, as evidenced from the findings of this study, the use of the recorded oral test could promote reading comprehension. As the findings suggest that although pupils are able to read, their ability to comprehend may be nominal during the test. This notion corroborates the findings of the studies by McMohan et al. (2016) and Spiel et al. (2019) where the oral modality of the test may enable pupils to better comprehend the context of the mathematics test items and subsequently help pupils to perform significantly better when administered the mathematics test, as opposed to the written test. Hence, with the inclusion of the recorded oral test as a test accommodation, the *Orang Asli* pupils may be able to comprehend the mathematics items better without the need to decode the written text presented in the test.

5.3.4.3 Future challenges on the use of recorded oral test

This study further highlights that the main challenge that may be faced by the mathematics teachers on the use of the recorded oral test is the difficulty of implementing the test accommodation for the *Orang Asli* pupils. In line with the study by Wickstrom et al. (2020), the implementation of test accommodation may require prerequisite skills, knowledge and experiences on the part of the teachers to develop and carry out the test accommodation in mathematics tests. This is consistent with past studies on the lack of available resources, logistics and expertise that is required when incorporating test accommodation in most content-based assessments, as highlighted by Clark- Gareca (2016), Lovett and Harrison (2021), and Radmehr and Vos (2020). Without targeted training or adequate support, the practical utility of employing the recorded oral test may be limited, potentially affecting the accuracy and fairness of the assessments among the *Orang Asli* pupils.

The challenge in ensuring the validity and reliability of mathematics assessments for *Orang Asli* pupils are further complicated by the use of test accommodations such as recorded oral tests in the ethnic language. While the use of the test accommodation is aimed to support pupils, there is a concern that over-reliance on the ethnic language may hinder their development of proficiency in the language of instruction, potentially affecting long-term academic performance in other subjects that uses the language of instruction. This issue has also been observed among teachers in other studies (Radmehr & Vos, 2020; Yang, 2020), where the implementation of test accommodations was carefully managed to ensure that it supported the pupils' needs without compromising the reliability of the assessment.

Despite this, the quantitative findings of this study indicated no significant differences in pupils' mathematics performance between the recorded oral tests administered in either language. In fact, pupils performed better when tested in the language of instruction, suggesting that it may serve as a more reliable and valid measure of their mathematical abilities. This also indicates that using the recorded oral test in the language of instruction may ease implementation for teachers and support the development of language proficiency, while maintaining the validity of the mathematics assessment, as advocated by Reynold et al. (2021).

5.3.5 Triangulation of Quantitative and Qualitative Findings

The convergence of quantitative and qualitative findings in this study on the use of the recorded oral test as a test accommodation for assessing the mathematics performance of *Orang Asli* pupils presents a comprehensive perspective. The quantitative data highlighted that the recorded oral tests in both language of instruction and ethnic language were effective test accommodations for the Orang Asli pupils. Nonetheless, the pupils appeared to perform better when administered the recorded oral test in the language of instruction. This suggests a potential advantage in using the language of instruction for mathematics assessments. On the other hand, the incorporation of the ethnic language remains crucial, as it reflects cultural relevance and inclusivity in mathematics education, in line with studies by Kükea Shultz and Englert (2021), and Ruef et al. (2020).

The qualitative findings from the participants provided deeper insights into the effectiveness of recorded oral tests as possible test accommodations for the *Orang Asli* pupils. The mathematics teachers perceived this test accommodation as a valuable tool for providing language and literacy support to the pupils. They emphasized that using the recorded oral tests could enhance pupils' comprehension and understanding of mathematics test items. Moreover, the teachers expressed a preference for the language of instruction over the ethnic language due to several challenges, including their non-native proficiency in the ethnic language and the additional time and effort required for its implementation.

The convergence between quantitative and qualitative findings underscores the validity of using recorded oral tests as a promising test accommodation for assessing the mathematics performance of *Orang Asli* pupils. The quantitative data's indication of higher performance in the language of instruction aligns with qualitative insights from the mathematics teachers, who view the use of the language of instruction as more practical and feasible, due to their familiarity and ease in test development. This correspondence indicates a consensus on the efficacy of the language of instruction in facilitating better performance among the pupils during mathematics assessments.

Furthermore, the qualitative findings substantiate the quantitative results by elucidating the mechanisms through which recorded oral tests aid Orang Asli pupils.

Teachers highlighted the role of recorded oral tests in providing necessary language and literacy support, thereby addressing potential barriers to understanding mathematics test items. This alignment strengthens the argument on the use of the recorded oral tests as a valid test accommodation, emphasizing its potential to bridge language gaps and improve text comprehension among the *Orang Asli* pupils. However, it's essential to acknowledge the nuanced perspectives revealed in the qualitative data. While the preference for the language of instruction is evident due to practical considerations, the importance of the ethnic language should not be underestimated. Cultural sensitivity and inclusivity are paramount in mathematics assessments, as highlighted in Trumbull and Nelson-Barber (2019). The incorporation of the ethnic language remains crucial for providing equitable opportunities for all pupils, whilst preserving cultural identity and fostering a supportive learning environment.

In conclusion, the triangulation of quantitative and qualitative findings supports the use of recorded oral tests as a valid test accommodation for assessing the mathematics performance of *Orang Asli* pupils. The alignment between higher performance in the language of instruction and teachers' preference towards the use of the language of instruction underscores its practicality and efficacy as a promising measure in the mathematics testing. However, this should not overshadow the significance of the ethnic language in ensuring cultural relevance and inclusivity in education. Balancing both languages as a possible language support in the form of test accommodations can better cater to the diverse needs of *Orang Asli* pupils to promote fair and accessible mathematics assessments.

5.4 Implication

The *Orang Asli* pupils' unique challenges brought upon by the language barriers caused by the mismatch between the language of instruction and ethnic language, and reading difficulties should be addressed to obtain a more accurate interpretation of their mathematics test scores. Previous studies have proposed that the use of test accommodation was deemed suitable and necessary in providing the needed language support to mitigate the impact of these issues during the mathematics tests. In response to the current literature and based on the findings of this study, several implications namely theoretical, methodological, empirical and practical implications were identified.

5.4.1 Theoretical Implication

The use of recorded oral test as a test accommodation to address both the language barriers and reading difficulties faced by *Orang Asli* pupils has significant theoretical implications, particularly in light of Cognitive Load Theory. This theory focuses on the limitations of working memory and how cognitive demands posed by the test items could impact performance during the mathematics test. By gaining a better understanding on cognitive load theory, the right test accommodation could be employed to alleviate the burden on pupils' working memory which would allow them to focus more effectively on mathematical problem-solving rather than on the language component of the test items, as proposed by Haag et al. (2013) and Walkington et al. (2019).

Firstly, the language barrier and reading difficulties experienced by *Orang Asli* pupils can impose a high cognitive load during mathematics tests. When pupils have limited proficiency in the language of instruction, they must dedicate considerable cognitive resources to decoding written instructions, comprehending mathematical terminology, and understanding the problem within their experiences and cognitive resources. This linguistic processing consumes valuable working memory capacity, leaving less cognitive resources available for mathematical reasoning and problem-solving. As a result, the *Orang Asli* pupils may struggle to demonstrate their true mathematical abilities and knowledge due to the overwhelming cognitive demands imposed by the language barrier.

By supplementing the recorded oral tests as a test accommodation with mathematics test, this measure may help reduce the cognitive load associated with the language barriers and reading difficulties experienced by the pupils. Instead of relying solely on written instructions, pupils could listen to the test items, allowing them to engage their auditory processing system rather than relying solely on reading skills. By presenting the test items in an oral format, the recorded oral test taps into a different modality, enabling pupils to process information using their auditory and verbal working memory capacities. This redistribution of cognitive resources allows pupils to allocate more mental effort to the actual mathematical problem-solving rather than language comprehension. As a result, the pupils can better demonstrate their true mathematical abilities as the accommodation aligns with their strengths and reduces the cognitive burden associated with the language barrier.

Furthermore, the use of recorded oral tests could be used as a means of managing extraneous cognitive load. By providing the recorded oral tests, this extraneous cognitive load is minimized as pupils would no longer need to engage in the complex task of reading and comprehending written text, particularly with long text or unfamiliar mathematics vocabulary, as highlighted by Gillmor et al. (2015) and Walkington et al. (2015) on providing language and linguistic support in mathematics testing. Additionally, the extraneous cognitive load could also be reduced by incorporating the pupils' ethnic language to help them to better comprehend the test items in a language that they are more familiar in. This reduction in extraneous cognitive load frees up working memory capacity, allowing pupils to direct their cognitive resources towards essential aspects of the task, such as understanding the mathematical problem and formulating a solution strategy.

Moreover, this test accommodation can also address intrinsic cognitive load, which is inherent to the complexity of the mathematical task itself. By reducing the extraneous cognitive load associated with both the mismatch in the language and reading difficulties, more cognitive resources become available for pupils to engage with the intrinsic cognitive load of the mathematical problem, as highlighted by Haag et al. (2013) and Phan et al. (2017). This enables deeper processing of mathematical concepts and promotes a more efficient use of working memory. As a result, the *Orang Asli* pupils can focus on the inherent complexity of mathematical tasks without being overwhelmed by the additional cognitive demands imposed by the language barrier.

In conclusion, the theoretical implications of using recorded oral tests as a test accommodation for mathematics exams to address the language barrier and reading difficulties faced by *Orang Asli* pupils are significant within the framework of the Cognitive Load Theory. By reducing extraneous cognitive load and managing working memory capacity, this test accommodation in the language of instruction or ethnic language would allow for the allocation of more cognitive resources towards mathematical problem-solving and reasoning (Korbach et al., 2018; Ngu et al., 2016). As a result, the *Orang Asli* pupils can better demonstrate their mathematical abilities, fostering a more valid interpretation of their mathematics test scores rather than being hindered by the language barriers or reading difficulties faced by them during the mathematics tests.

5.4.2 Methodological Implication

In carrying out this study, the research conducted provided several methodological implications. Firstly, by incorporating both quantitative and qualitative data collection methods in this study on the use of recorded oral tests as test accommodations for the *Orang Asli* pupils, the findings yielded a holistic understanding of assessing the utility of accommodating to the language and literacy needs of the pupils. The quantitative data, gathered through the administration of written mathematics tests and recorded oral tests in both language of instruction and ethnic language, allowed for the measurement of mathematics performance and the effectiveness of recorded oral tests as a promising test accommodation. On the other hand, the qualitative data collected unveiled valuable perspectives, shedding light on the practical challenges and benefits of implementing the test accommodation to the pupils. Together, the quantitative and qualitative approaches provided a comprehensive and nuanced understanding of the complex issue of establishing inclusive mathematics assessment practices.

Other than that, the methodological implication of this study stems from the construction of the recorded oral tests in both the language of instruction and ethnic language were found to be pivotal for developing valid assessment tools to measure the mathematics performance of the *Orang Asli* pupils. This was carried out by ensuring that construct, content and cultural validity were established during the entirety of the test development process undertaken in this study.

The development of the mathematics tests entailed a careful selection of test items that accurately measure the targeted construct of mathematical knowledge and skills from the relevant topics within the curriculum. Additionally, the development of the recorded oral test in the ethnic language involved not only linguistic adaptation and translation, but effort was also made to ensure that the test content was culturally sensitive and aligned to the pupils' cultural contexts, values, and experiences. Therefore, the resulting mathematics tests served as valuable and valid tools for measuring the pupils' mathematics performance.

5.4.3 Empirical Implication

The empirical implications arising from the findings of this study shed light on the effectiveness of the recorded oral test as a valid and promising test accommodation that caters to the needs of the *Orang Asli* pupils. The empirical evidence obtained from this study show that the *Orang Asli* pupils performed better when administered the recorded oral test in the language of instruction, followed by the recorded oral test in their ethnic language, and lastly, in the written test in the language of instruction. These findings shed light on the significance of oral proficiency in the language of instruction and suggest potential benefits of incorporating the ethnic language in educational settings.

The findings challenge the traditional notion that proficiency in the language of instruction is an important determinant to mathematics proficiency and increased mathematics performance. While previous studies have indicated that a lack of proficiency in the language of instruction hampers mathematics performance among *Orang Asli* pupils (Abdul Hakim et al., 2023; Ruzlan et al., 2021), the current empirical findings suggest that the choice of language, whether it be the language of instruction or the ethnic language, does not significantly impact mathematics performance. This implies that language preference does not have a direct bearing on the ability to comprehend and excel in mathematics, at least among this sample of *Orang Asli* pupils at Year 4 primary level.

Furthermore, the improved performance of the pupils when administered the recorded oral tests highlights the role of the oral language skills in mathematics education, particularly in dealing with reading difficulties that may impact pupils' when decoding and comprehending the written text as indicated in past studies (Gandhi et al., 2018; Giusto & Ehri, 2019). As the findings demonstrate that the pupils were able to excel when supplemented the recorded oral tests in both the language of instruction and their ethnic language suggests that oral proficiency plays a crucial role in their academic achievements. This finding underscores the importance of providing opportunities for pupils to develop their oral language skills, regardless of the language they use.

This study uncovered a noteworthy disparity whereby complex linguistic features such as moderate item length, long item length and mathematics vocabulary suggest that the implication of this finding is twofold. Firstly, it highlights the possibility that items with a greater number of words and unfamiliar mathematics vocabulary may place a greater cognitive demand on the *Orang Asli* pupils which may potentially affect their performance differently (Haag et al., 2013; Walkington et al., 2015). Secondly, the findings also imply that selecting and developing test items that contain simple language and familiar linguistic features ensures a valid mathematics test is developed, which could assess the *Orang Asli* pupils' true mathematical knowledge and skills.

5.4.4 Practical Implication

The practical implications of using recorded oral tests in either the language of instruction or the ethnic language as a test accommodation for mathematics tests can be significant in addressing the language barriers and reading difficulties faced by the pupils. Based on findings of this study, the recorded oral test could be a practical measure in providing inclusive and equitable mathematics tests which caters to the needs of *Orang Asli* pupils, as previously observed among mainstream pupils (Lopez et al., 2019; McMahon et al., 2016; Spiel et al., 2019). Essentially, this could provide a more accurate interpretation of their mathematics test scores, thereby allowing for the assessment of their true mathematical knowledge and abilities.

One of the main practical implications from the findings of this study is that the use of recorded oral tests in either the language of instruction and ethnic language can provide a practical solution for *Orang Asli* pupils who face language barriers and reading difficulties. Mathematics tests often require a deep understanding of problem-solving techniques and concepts, which can be hindered by limited proficiency in the language of instruction. By providing recorded oral tests in the language of instruction, educators can assess the mathematical knowledge and skills of *Orang Asli* pupils without solely relying on reading and written language abilities. This test accommodation could allow pupils to demonstrate their mathematical competence and comprehension through orally delivered mathematics test items, as supported by Kanageswari et al. (2021) and Zaleha et al. (2020). Such a measure could provide a fair and inclusive assessment method that takes into account their unique linguistic challenges.

Moreover, the ease of implementation associated with recorded oral tests in the language of instruction can enhance their practicality as a test accommodation for mathematics assessments, compared to the use of the ethnic language. As the findings of the study show that there are no differences in the pupils' performance on the recorded oral tests for both languages, the mathematics teachers' preferences on the recorded oral test in the language of instruction highlight its convenience and compatibility with existing assessment practices. Implementing an oral test in the language of instruction requires minimal adjustments to the assessment framework and reduces the need for additional resources or training, particular during the trans adaptation process of the mathematics test items to ensure cultural and linguistic equivalence to *Bahasa Temiar* used by the *Orang Asli* pupils.

This practical advantage can lead to a more valid and reliable assessment process, enabling teachers to focus on evaluating pupils' mathematical understanding rather than grappling with the logistics of developing the test accommodation in the ethnic language to accommodate language barriers, where such challenges have been found in past studies to have hindered the implementation of test accommodation for pupils (Lovett & Harrison, 2021; Wickstrom et al., 2020; Yang, 2020). Moreover, the ease of implementation may encourage more educators to adopt this approach, leading to greater consistency and standardization in assessing *Orang Asli* pupils' mathematical abilities across different classrooms, schools, and ethnic groups across Malaysia.

Nonetheless, despite the practical implications and advantages of using recorded oral tests in the language of instruction for mathematics assessments, it is crucial to acknowledge the potential limitations and considerations. One important aspect to consider is the preservation of cultural identity and the significance of the ethnic language for *Orang Asli* pupils. While the use of the language of instruction may facilitate assessment and provide equal access to educational opportunities, it is essential to ensure that the ethnic language is not marginalized or neglected, as advocated by Trumbull and Nelson-Barber (2019). Therefore, efforts should be made to integrate and promote the use of the ethnic language during mathematics teaching and learning that promotes a deeper understanding of the mathematics knowledge and concepts taught to the pupils.

In conclusion, the practical implications of employing recorded oral test as a promising and valid test accommodation in mathematics assessments for *Orang Asli* pupils are significant. The ease of implementation associated with this approach further enhances its practicality, ensuring that mathematics teachers can focus on assessing mathematical understanding rather than be burdened by logistical challenges that may disrupt its implementation to the larger *Orang Asli* pupil population. However, it is crucial to strike a balance between accommodating language barriers and preserving cultural identity by incorporating the ethnic language in other educational aspects, such as during the classroom teaching to ensure that pupils could effectively learn and comprehend the language in mathematics education. By considering these practical implications, educators can create a more inclusive and effective assessment environment for the *Orang Asli* pupils.

5.5 Recommendation for Future Studies

Future studies are recommended to further attest to the use of the recorded oral tests as a valid and potential test accommodation for the *Orang Asli* pupils. One future study that could be carried is to expand the study sample to include a larger and more diverse group of *Orang Asli* pupils from different ethnic backgrounds. This could provide a broader representation and increase the generalizability of the findings to include all *Orang Asli* pupils across Malaysia, and Indigenous pupils. Additionally, a longitudinal study may be conducted to assess the long-term impact of the recorded oral tests on the mathematics performance across the grade levels of the pupils. By tracking their progress in mathematics when supplemented with the recorded oral test as a test accommodation over an extended period, this may provide valuable insights into the sustainability and effectiveness of this test accommodation.

Some qualitative recommendations that could be carried out is the observation of the use of the test accommodation in the classroom setting. The implementation of the recorded oral test could be observed in actual classroom settings to gain a better understanding on how the test could be integrated by the mathematics teachers during formative assessments. These observations will provide insights into the practical challenges faced by the teachers and help identify potential areas for improvement in the implementation of the test accommodation for the pupils.

Other than that, the role of parents and community members could be investigated on the use of the recorded oral test as a test accommodation for the *Orang Asli* pupils. By employing qualitative research methods such as surveys or interviews, the findings could shed light on the level of parental awareness, community engagement, and attitudes toward the use of the test accommodation, thus providing a more holistic understanding of its impact on long-term mathematics performance of the pupils.

5.6 Summary

This study was aimed at examining the mathematics performance of the *Orang Asli* pupils using the recorded oral tests in the language of instruction and ethnic language as a test accommodation on item length and vocabulary. The study was carried out using both quantitative and qualitative approaches on 142 Year 4 *Orang Asli* pupils from the *Temiar* ethnic group and eight mathematics teachers from *Orang Asli* primary schools respectively. Overall, the study found that the pupils were able to perform better when administered the recorded oral tests, compared to the written test.

Moreover, it was found that the choice of language used for the recorded oral tests is not an important component in affecting mathematics performance among the pupils, but rather it was the modality of the test that played a more significant role in improving mathematics performance. With respect to item length and vocabulary, this study also found that the recorded oral tests were able to mitigate the impact of complex linguistic features such as moderate and long item length and mathematics vocabulary found in the test items to ensure pupils were able to better solve the mathematics problem.

The qualitative findings also support the quantitative findings obtained from this study, whereby the mathematics teachers do support the use of the recorded oral test as a promising test accommodation to help mitigate the language and literacy barriers that have negatively impacted the pupils' mathematics performance. Moreover, as the *Orang Asli* pupils were found to perform the better when the recorded oral test in the language of instruction was administered, compared to the use of the ethnic language This provides further support on the findings from the qualitative data which suggested the incorporation of the language of instruction, rather than the ethnic language for ease of implementing the test accommodation with current assessment practices used in the primary schools today. Hence, this study concludes that both quantitative and qualitative findings support the use of the recorded oral test as a test accommodation that caters to the needs of the *Orang Asli* pupils in validly assessing their mathematics performance.

The findings and conclusions drawn from this study may not be applicable or generalizable to non-Indigenous pupils whose mother tongue/ethnic language is not

Bahasa Melayu, such as the Malaysian Chinese and Indian pupils. As mainstream pupils could still converse in *Bahasa Melayu* and are not segregated due to their remote location as opposed to the *Orang Asli* pupils, this study specifically focuses on *Orang Asli* primary schools whereby the primary language used by the *Orang Asli* pupils outside of their educational institutions is their ethnic language. Therefore, the outcomes of this study should be cautiously interpreted and not extrapolated to educational settings where the primary language and cultural context significantly differ from those observed within the *Orang Asli* primary school context.

REFERENCES

Ababneh, E.G., & Kodippili, A. (2020). Investigating the association of some variables with mathematics achievement gap between rural and urban Jordanian students. *Journal of Education and Practice, 11*(21), 147-158.

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

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

Abedi, J., & Lord, C. (2001). The language factor in mathematics tests. *Applied Measurement in Education*, *14*(3), 219-234.

Abedi, J. (2014). English language learners with disabilities: Classification, assessment, and accommodation issues. *Journal of Applied Testing Technology*, *10*(2), 1-30. https://doi.org/10.1207/S15324818AME1403_2

Abedi, J., Zhang, Y., Rowe, S. E., & Lee, H. (2020). Examining effectiveness and validity of accommodations for English language learners in mathematics: An evidence-based computer accommodation decision system. *Educational Measurement: Issues and Practice*, *39*(4), 41-52. https://doi.org/10.1111/emip.12328

Adeoye-Olatunde, O. A., & Olenik, N. L. (2021). Research and scholarly methods: Semi-structured interviews. *Journal of the American College of Clinical Pharmacy*, *4*(10), 1358-1367. https://doi.org/10.1002/jac5.1441

Aidil Fitri Sawalludin, Min, C. L. J., & Mohamad Izzuan Mohd Ishar. (2020). The struggle of Orang Asli in education: quality of education. *Malaysian Journal of Social Sciences and Humanities*, *5*(1), 46-51. https://doi.org/10.47405/mjssh.v5i1.346

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

Akkerman, S., Admiraal, W., Brekelmans, M., & Oost, H. (2008). Auditing quality of research in social sciences. *Quality & Quantity*, *42*, 257-274. https://doi.org/10.1007/s11135-006-9044-4

Alordiah, C., & Oji, J. (2024). Test equating in educational assessment: a comprehensive framework for promoting fairness, validity, and cross-cultural equity. *Asian Journal of Assessment in Teaching and Learning*, *14*(1), 70-84. https://doi.org/10.37134/ajatel.vol14.1.7.2024

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

Arsaythamby Veloo, Kanageswari S. Suppiah Shanmugam., Ruzlan Md-Ali., Yus'aiman Jusoh Yusoff, & Rosna Awang-Hashim. (2021). Grade five Indigenous (Orang Asli) pupil's achievement in bilingual versions of mathematics test. *Journal of Language and Linguistic Studies*, *17*(4), 1863-1872.

Arteaga, I., & Glewwe, P. (2019). Do community factors matter? An analysis of the achievement gap between indigenous and non-indigenous children in Peru. *International Journal of Educational Development*, *65*, 80-91. https://doi.org/10.1016/j.ijedudev.2017.08.003

Asenahabi, B. M. (2019). Basics of research design: A guide to selecting appropriate research design. *International Journal of Contemporary Applied Researches*, *6*(5), 76-89.

Barwell, R., Wessel, L., & Parra, A. (2019). Language diversity and mathematics education: New developments. *Research in Mathematics Education*, *21*(2), 113-118. https://doi.org/10.1080/14794802.2019.1638824

Barwell, R. (2020). Learning mathematics in a second language: Language positive and language neutral classrooms. *Journal for Research in Mathematics Education*, *51*(2), 150-178. https://doi.org/10.5951/jresematheduc-2020-0018

Battiste, M. (2002). *Indigenous knowledge and pedagogy in First Nations education: A literature review with recommendations* (pp. 1-69). Ottawa: National Working Group on Education.

Baumann, J. F., & Graves, M. F. (2010). What is academic vocabulary?. *Journal of Adolescent & Adult Literacy*, *54*(1), 412.

Benjamin, G. (2012, November). Why you should study Aslian languages. In *Proceedings of the Second International Conference on Linguistics, Literature and Culture (ICLLIC2012) Penang, Malaysia* (Vol. 8).

Bergqvist, E., Theens, F., & Österholm, M. (2018). The role of linguistic features when reading and solving mathematics tasks in different languages. *The Journal of Mathematical Behavior*, *51*, 41-55. https://doi.org/10.1016/j.jmathb.2018.06.009

Berry, R. Q. (2018). Disrupting policies and reforms in mathematics education to address the needs of marginalized learners. In *Toward equity and social justice in mathematics education* (pp. 3-20). Springer.

Betts, A., Thai, K. P., Jacobs, D., & Li, L. (2020). Math readiness: Early identification of preschool children least ready to benefit from formal math instruction in school. In *The IAFOR International Conference on Education–Hawaii 2020 Official Conference Proceedings. Paper Presented at the IAFOR International Conference on Education, Hawaii, USA. The International Academic Forum, Japan*.

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

Bishop, M., & Vass, G. (2021). Talking about culturally responsive approaches to education: teacher professional learning, Indigenous learners and the politics of schooling. *The Australian Journal of Indigenous Education*, *50*(2), 340-347. https://doi.org/10.1017/jie.2020.30

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

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.

Braun, V., & Clarke, V. (2013). *Successful qualitative research: A practical guide for beginners*. Sage Publications.

Brooks, J., Reed, D. M., & Savage, B. (2016, June). Taking off with a pilot: The importance of testing research instruments. In *ECRM2016-Proceedings of the 15th European Conference on Research Methodology for Business Management": ECRM2016. Academic Conferences and publishing limited*(pp. 51-59).

Brown, L. (2019). Indigenous young people, disadvantage and the violence of settler colonial education policy and curriculum. *Journal of Sociology*, *55*(1), 54-71.

Buzick, H., & Stone, E. (2014). A meta-analysis of research on the read aloud accommodation. *Educational Measurement: Issues and Practice*, *33*(3), 17-30. https://doi.org/10.1111/emip.12040

Candela, A. G. (2019). Exploring the function of member checking. *The Qualitative Report*, *24*(3), 619-628. https://nsuworks.nova.edu/tqr/vol24/iss3/14

Castillo-Montoya, M. (2016). Preparing for interview research: The interview protocol refinement framework. *Qualitative Report*, *21*(5), 811-831. http://nsuworks.nova.edu/tqr/vol21/iss5/2

Caviola, S., Visentin, C., Borella, E., Mammarella, I., & Prodi, N. (2021). Out of the noise: Effects of sound environment on maths performance in middle-school students. *Journal of Environmental Psychology*, *73*, 101552. https://doi.org/10.1016/j.jenvp.2021.101552

Chamberlain, M. C., & Medina, E. J. (2020). A case of being the same? Australia and New Zealand's reading in focus. *Australian Journal of Education*, *64*(3), 243-263. https://doi.org/10.1177/0004944120953235

Chen, O., Castro-Alonso, J. C., Paas, F., & Sweller, J. (2018). Undesirable difficulty effects in the learning of high-element interactivity materials. *Frontiers in Psychology*, *9*(1483). 1-7. https://doi.org/10.3389/fpsyg.2018.01483

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

Chow, J. C., & Wehby, J. H. (2018). Associations between language and problem behavior: A systematic review and correlational meta-analysis. *Educational Psychology Review*, *30*(1), 61-82. https://doi.org/10.1007/s10648-016-9385-z

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

Cirino, P. T., Child, A. E., & Macdonald, K. T. (2018). Longitudinal predictors of the overlap between reading and math skills. *Contemporary Educational Psychology*, *54*, 99-111.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological assessment*, *31*(12), 1412-1427. https://psycnet.apa.org/doi/10.1037/pas0000626

Clark-Gareca, B. (2016). Classroom assessment and English Language Learners: Teachers' accommodations implementation on routine math and science tests. *Teaching and Teacher Education*, *54*, 139-148.

Corbin, J., & Strauss, A. (2015). *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage Publications.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology,* *78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Crawford, L., & Tindal, G. (2018). Effects of a read-aloud modification on a standardized reading test. In *Large-scale testing of students with disabilities* (pp. 89-106). Routledge.

Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry and research design: Choosing among five approaches*. SAGE Publication.

Creswell, J. W., & Creswell, J. D. (2018). Research design: Qualitative, quantitative and mixed method approaches (5th ed.). SAGE Publication.

Cruz Neri, N., Guill, K., & Retelsdorf, J. (2021). Language in science performance: do good readers perform better? *European Journal of Psychology of Education*, *36*(1), 45-61. https://doi.org/10.1007/s10212-019-00453-5

Cruz Neri, N., & Retelsdorf, J. (2022). The role of linguistic features in science and math comprehension and performance: A systematic review and desiderata for future research. *Educational Research Review*, 100460. https://doi.org/10.1016/j.edurev.2022.100460

Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H. C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in psychology*, *6*(348). 1-13. https://doi.org/10.3389/fpsyg.2015.00348

De Angelis, G. (2014). A multilingual approach to analysing standardized test results: Immigrant primary school children and the role of languages spoken in a bi-/multilingual community. *Intercultural Education*, *25*(1), 14-28. https://doi.org/10.1080/14675986.2014.883167

De Backer, F., Slembrouck, S., & Van Avermaet, P. (2019). Assessment accommodations for multilingual learners: Pupils' perceptions of fairness. *Journal of Multilingual and Multicultural Development*, *40*(9), 833-846.

DeJonckheere, M., & Vaughn, L. M. (2019). Semistructured interviewing in primary care research: a balance of relationship and rigour. *Family Medicine and Community Health*, *7*(2). e000057.

Delprato, M. (2021). Indigenous learning gaps and home language instruction: New evidence from PISA-D. *International Journal of Educational Research*, *109*, 101800. https://doi.org/10.1016/j.ijer.2021.101800

Ding, H., & Homer, M. (2020). Interpreting mathematics performance in PISA: Taking account of reading performance. *International Journal of Educational Research*, *102*, 101566. https://doi.org/10.1016/j.ijer.2020.101566

Duarte, M. E., & Rossier, J. (2008). Testing and assessment in an international context: Cross-and multi-cultural issues. In J. A. Athanasou & R. Van Ana Hernández, María Dolores Hidalgo, Ronald K. Hambleton, and J. Gómez-Benito (Eds.), *International handbook of career guidance* (pp. 489- 510). Springer.

Emerson, R. W. (2022). ANOVA assumptions. *Journal of Visual Impairment & Blindness*, *116*(4), 585-586. https://doi.org/10.1177/0145482X221124187

Erath, K., Ingram, J., Moschkovich, J., & Prediger, S. (2021). Designing and enacting instruction that enhances language for mathematics learning: A review of the state of development and research. *ZDM–Mathematics Education*, *53*, 245-262. https://doi.org/10.1007/s11858-020-01213-2

Ercikan, K., Roth, W-M., Simon, M., Sandilands, D., & Lyons-Thomas, J. (2014). Inconsistencies in DIF detection for sub-groups in heterogeneous language groups. *Applied Measurement in Education, 27*(4), 273–285.

Ercikan, K., Chen, M. Y., Lyons-Thomas, J., Goodrich, S., Sandilands, D., Roth, W. M., & Simon, M. (2015). Reading proficiency and comparability of mathematics and science scores for students from English and non-English backgrounds: An international perspective. *International Journal of Testing*, *15*(2), 153-175.

Ercikan, K, & Por, H. H. (2020). Comparability in multilingual and multicultural assessment contexts. In A. I. Berman, E. H. Haertal, & J. W. Pellegrino (Eds.), *Comparability of large-scale assessment: issues and recommendation* (pp. 205-220). National Academy of Education.

Fang, Z. (2016). Text complexity in the US Common Core State Standards: A linguistic critique. *Australian Journal of Language and Literacy,39*(3), 195-206.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2018). *Learning disabilities: From identification to intervention*. Guilford Publications.

Forsyth, S. R., & Powell, S. R. (2017). Differences in the mathematics-vocabulary knowledge of fifth-grade students with and without learning difficulties. *Learning Disabilities Research & Practice*, *32*(4), 231-245.

Fuchs, L. S., Fuchs, D., Eaton, S. B., Hamlett, C. L., & Karns, K. M. (2000). Supplementing teacher judgments of mathematics test accommodations with objective data sources. *School Psychology Review*, *29*(1), 65-85.

Fuchs, L. S., Fuchs, D., Seethaler, P. M., Cutting, L. E., & Mancilla-Martinez, J. (2019). Connections between reading comprehension and word-problem solving via oral language comprehension: Implications for comorbid learning disabilities. *New directions for child and adolescent development*, *2019*(165), 73-90.

Fuchs, L. S., Seethaler, P. M., Sterba, S. K., Craddock, C., Fuchs, D., Compton, D. L., Geary, D. C., & Changas, P. (2021). Closing the word-problem achievement gap in first grade: Schema-based word-problem intervention with embedded language comprehension instruction. *Journal of Educational Psychology, 113*(1), 86–103. https://doi.org/10.1037/edu0000467

Gandhi, A. G., Ogut, B., Stein, L., Bzura, R., & Danielson, L. (2018). Enhancing accessibility for students with decoding difficulties on large-scale reading assessments. *Journal of learning disabilities*, *51*(6), 540-551. https://doi.org/10.1177%2F0022219417714774

García, O., & Solorza, C. R. (2021). Academic language and the minoritization of US bilingual Latinx students. *Language and Education*, *35*(6), 505-521.

Giusto, M., & Ehri, L. C. (2019). Effectiveness of a partial read-aloud test accommodation to assess reading comprehension in students with a reading disability. *Journal of learning disabilities*, *52*(3), 259-270. https://doi.org/10.1177%2F0022219418789377

Gillmor, S. C., Poggio, J., & Embretson, S. (2015). Effects of reducing the cognitive load of mathematics test items on student performance. *Numeracy: Advancing Education in Quantitative Literacy, 8*(1), 1-18.

Gilmour, A. F., Fuchs, D., & Wehby, J. H. (2019). Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities. *Exceptional Children*, *85*(3), 329-346.

Government of Malaysia. (2023, June 24). Demography: The indigenous peoples. https://data.gov.my/dashboard/orang-asli

Gunawan, M. A., Retnawati, H., & Kartowagiran, B. (2021). Vertical Equating Accuracy Using Kernel Method. *Technium: Romanian Journal of Applied Sciences and Technology*, *3*(7), 110–120.

Gupta, U., & Zheng, R. Z. (2020). Cognitive load in solving mathematics problems: validating the role of motivation and the interaction among prior knowledge, worked examples, and task difficulty. *European Journal of STEM Education*, *5*(1), 1-14.

Guetterman, T. C., & Fetters, M. D. (2018). Two methodological approaches to the integration of mixed methods and case study designs: A systematic review. *American Behavioral Scientist*, *62*(7), 900-918. https://doi.org/10.1177/0002764218772641

Haag, N., Heppt, B., Stanat, P., Kuhl, P., & Pant, H. A. (2013). Second language learners' performance in mathematics: Disentangling the effects of academic language features. *Learning and Instruction*, *28*, 24-34.

Haag, N., Heppt, B., Roppelt, A., & Stanat, P. (2015). Linguistic simplification of mathematics items: effects for language minority students in Germany. *European Journal of Psychology of Education*, *30*(2), 145-167. https://doi.org/10.1007/s10212-014-0233-6

Hakkarainen, A., Holopainen, L., & Savolainen, H. (2013). Mathematical and reading difficulties as predictors of school achievement and transition to secondary education. *Scandinavian Journal of Educational Research*, *57*(5), 488-506. https://doi.org/10.1080/00313831.2012.696207

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied measurement in education*, *15*(3), 309-333. https://doi.org/10.1207/S15324818AME1503_5

Hambleton, R. K. (1996). Adaptación de tests para su uso en diferentes idiomas y culturas: fuentes de error, posibles soluciones y directrices prácticas [Adapting tests for use in different languages and cultures: Sources of error, possible solutions, and practical guidelines]. In J. Muñiz (Coord.), Psicometría (pp. 207-238). Madrid.

Hambleton, R. K., Merenda, P. F., & Spielberger, C. D. (2004). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 15-50). Psychology Press.

Hambleton, R.K. (2005). Issues, designs and technical guidelines for adapting tests into multiple languages and cultures. In R.K. Hambleton, P.F. Merenda & S.D. Spielberger (Eds.), Adapting educational and psychological tests for cross-cultural assessment (pp. 3-38). Lawrence Erlbaum Associates.

Hambleton, R. K., & Lee, M. K. (2013). Methods for translating and adapting tests to increase cross-language validity. In D. H. Saklofske, C. R. Reynolds, V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 172-181). Oxford University Press.

Hamidah Yusof, Norasibah Abdul Jalil, Yin, K. Y., Mahaliza Mansor, and Maryam Mahdinezhad (2017). Parental commitment in leading learning of the Orang Asli students. *International Journal of Academic Research in Business and Social Sciences*, *7*(4), 816-827.

Harrison, J., Bunford, N., Evans, S. W., & Owens, J. S. (2013). Educational accommodations for students with behavioral challenges: A systematic review of the literature. *Review of Educational Research, 83*(4), 551–597.

Härtig, H., Heitmann, P., & Retelsdorf, J. (2015). Analyse der Aufgaben zur Evaluation der Bildungsstandards in Physik-Differenzierung von schriftsprachlichen Fähigkeiten und Fachlichkeit [Analyses of the tasks for evaluating the educational standards in physics - differentiation between written language proficiency and content knowledge]. *Zeitschrift für Erziehungswissenschaft, 18,* 763–779. https://doi.org/10.1007/s11618-015-0646-2.

Heavy, E. (2019). *Statistics for Nursing: A practical guide* (3rd ed.). Jones and Bartlett Learning.

Hema Letchamanan, Nur Surayyah Madhubala Abdullah, & Kamal Solhaimi Fadzil. (2021). Language education for orang asli children in Malaysia. *Pertanika Journal of Social Sciences & Humanities*, *29*(3), 443-458.

Herber, O. R., Bradbury-Jones, C., Böling, S., Combes, S., Hirt, J., Koop, Y., Nyhagen, R., Veldhuizen, J. D., & Taylor, J. (2020). What feedback do reviewers give when reviewing qualitative manuscripts? A focused mapping review and synthesis. *BMC medical research methodology*, *20*, 1-15. https://doi.org/10.1186/s12874-020-01005-y

Herman, J., & Cook, L. (2019). Fairness in classroom assessment. In S. M. Brookhart & J.H McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 243-264). Routledge.

Hernández, A., Hidalgo, M. D., Hambleton, R. K., & Gómez Benito, J. (2020). International test commission guidelines for test adaptation: A criterion checklist. *Psicothema, 32*(3), 390-398.

Hofer, S. I., Reinhold, F., & Koch, M. (2022). Students home alone—profiles of internal and external conditions associated with mathematics learning from home. *European Journal of Psychology of Education*, 1-34.

Hollinsworth, D., Raciti, M., & Carter, J. (2021). Indigenous students' identities in Australian higher education: Found, denied, and reinforced. *Race Ethnicity and Education*, *24*(1), 112-131. https://doi.org/10.1080/13613324.2020.1753681

Hornburg, C. B., Schmitt, S. A., & Purpura, D. J. (2018). Relations between preschoolers' mathematical language understanding and specific numeracy skills. *Journal of experimental child psychology*, *176*, 84-100. https://doi.org/10.1016/j.jecp.2018.07.005

Huang, B. H., & Butler, Y. G. (2020). Validity considerations for assessing language proficiency in young language minority students. *Language Assessment Quarterly*, *17*(5), 461-466. https://doi.org/10.1080/15434303.2020.1826486

Hudson, C., & Angelo, D. (2020). Teacher views on the implementation of English language proficiency scales for young Indigenous learners of standard English. *Language Assessment Quarterly*, *17*(5), 491-518. https://doi.org/10.1080/15434303.2020.1826489

Hunter, R., & Hunter, J. (2017). Maintaining a cultural identity while constructing a mathematical disposition as a Pāsifika learner. In E. A. McKinley & L.T. Smith (Eds.), *Handbook of Indigenous education*, 1-19. Springer.

International Test Commission (ICT). (2018). Guidelines for translating and adapting tests (second edition). *International Journal of Testing, 18*(2), 101-134.

International Test Commission (ITC). (2019). ITC guidelines for the large-scale assessment of linguistically and culturally diverse populations. *International Journal of Testing*, *19*(4), 301-336.

Jabatan Kemajuan Orang Asli (JAKOA). (2023). *Taburan ethnic orang asli mengikut ethnic/subethnic mengikut negeri* [Ethnic distribution of orang asli by ethnic/subethnic by state]. https://www.jakoa.gov.my/orang-asli/taburan-etnik-orang-asli-mengikut-etnik-sub-etnik-mengikut-negeri/

Jorgensen, R. (2018). Building the mathematical capital of marginalised learners of mathematics. *ZDM*, *50*(6), 987-998. https://doi.org/10.1007/s11858-018-0966-

Judd, C. M., Westfall, J., & Kenny, D. A. (2017). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology*, *68*, 601-625. https://doi.org/10.1146/annurev-psych-122414-033702

Kachchaf, R., Noble, T., Rosebery, A., O'Connor, C., Warren, B., & Wang, Y. (2016). A closer look at linguistic complexity: Pinpointing individual linguistic features of science multiple-choice items associated with English language learner performance. *Bilingual Research Journal*, *39*(2), 152-166. https://doi.org/10.1080/15235882.2016.1169455

Kallio, H., Pietilä, A. M., Johnson, M., & Kangasniemi, M. (2016). Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. *Journal of Advanced Nursing*, *72*(12), 2954-2965. https://doi.org/10.1111/jan.13031

Kan, A., & Bulut, O. (2015). Examining the language factor in mathematics assessments. *Education Research and Perspectives*, *42*, 582-606. https://search.informit.org/doi/10.3316/ielapa.136354527893104

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

Kanageswari, S. Suppiah Shanmugam, Arsaythamby Veloo, & Yus'aiman Jusoh (2023). The challenges of virtual learning: an exploratory study among Orang Asli Pupils from the mathematics teachers' perspectives. *Diaspora, Indigenous, and Minority Education*, 1-19. https://doi.org/10.1080/15595692.2023.2212823

Khan, R., Mustapha, R., Jantan, R., Nachiappan, S., Masran, M. N., & Shukor, A. A. (2013). Reading ability among semai (Orang asli) school children in Malaysia. In *ICERI2013 Proceedings* (pp. 3884-3888). IATED.

Kim, Y. S. G. (2016). Do live versus audio-recorded narrative stimuli influence young children's narrative comprehension and retell quality? *Language, speech, and hearing services in schools*, *47*(1), 77-86.

Kim, S., & Walker, M. (2021). Assessing Mode Effects of At-Home Testing Without a Randomized Trial. *ETS Research Report Series*, *2021*(1), 1-21. https://doi.org/10.1002/ets2.12323

Kolen, M. J., & Brennan, R. L. (2014). Nonequivalent groups: Linear methods. In M. J. Kolen & R. L. Brennan (Eds.), *Test equating, scaling, and linking* (pp. 103-142). Springer, New York, NY.

Koponen, T., Aro, M., Poikkeus, A. M., Niemi, P., Lerkkanen, M. K., Ahonen, T., & Nurmi, J. E. (2018). Comorbid fluency difficulties in reading and math: Longitudinal stability across early grades. *Exceptional Children*, *84*(3), 298-311.

Kopriva, R. J., Emick, J. E., Hipolito-Delgado, C. P., & Cameron, C. A. (2007). Do proper accommodation assignments make a difference? Examining the impact of improved decision making on scores for English language learners. *Educational Measurement: Issues and Practice*, *26*(3), 11-20. https://doi.org/10.1111/j.1745-3992.2007.00097.x

Korbach, A., Brünken, R., & Park, B. (2018). Differentiating different types of cognitive load: A comparison of different measures. *Educational Psychology Review*, *30*(2), 503-529. https://doi.org/10.1007/s10648-017-9404-8

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology*. Sage Publications.

Kūkea Shultz, P., & Englert, K. (2021). Cultural validity as foundational to assessment development: An indigenous example. *Frontiers in Education*, *6*, 701973. https://doi.org/10.3389/feduc.2021.701973

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

Lane, S., & Leventhal, B. (2015). Psychometric challenges in assessing English language learners and students with disabilities. *Review of Research in Education*, *39*(1), 165-214. https://doi.org/10.3102%2F0091732X14556073

Lee, C. D., & Spratley, A. (2010). *Reading in the disciplines: the challenges of adolescent literacy: Final report from carnegie corporation of new york's council on advancing adolescent literacy*. Carnegie Corporation of New York. https://files.eric.ed.gov/fulltext/ED535297.pdf

Lee, E. N., & Orgill, M. (2021). Toward equitable assessment of english language learners in general chemistry: Identifying supportive features in assessment items. *Journal of Chemical Education*, *99*(1), 35-48. https://doi.org/10.1021/acs.jchemed.1c00370

Leiss, D., Plath, J., & Schwippert, K. (2019). Language and mathematics-key factors influencing the comprehension process in reality-based tasks. *Mathematical Thinking and Learning*, *21*(2), 131-153. https://doi.org/10.1080/10986065.2019.1570835

Li, H. (2014). The effects of read-aloud accommodations for students with and without disabilities: A meta-analysis. *Educational Measurement: Issues and Practice*, *33*(3), 3-16. https://doi.org/10.1111/emip.12027

Lin, C. S., Lu, Y. L., & Lien, C. J. (2021). Association between test item's length, difficulty, and students' perceptions: Machine learning in schools' term examinations. *Universal Journal of Educational Research*, *9*(6), 1323-1332. https://doi.org/10.13189/UJER.2021.090622

Lindgren, B. M., Lundman, B., & Graneheim, U. H. (2020). Abstraction and interpretation during the qualitative content analysis process. *International Journal of Nursing Studies*, *108*, 103632. https://doi.org/10.1016/j.ijnurstu.2020.103632

Lindstrom, J. H. (2010). Mathematics assessment accommodations: Implications of differential boost for students with learning disabilities. *Intervention in School and Clinic*, *46*(1), 5-12. https://doi.org/10.1177%2F1053451210369517

Livingston, S. A. (2014). *Equating test scores (without IRT)*. Educational testing service.

Lopez, A. A., Guzman-Orth, D., & Turkan, S. (2019). Exploring the use of translanguaging to measure the mathematics knowledge of emergent bilingual students. *Translation and Translanguaging in Multilingual Contexts*, *5*(2), 143-164.

Lopez, A. A. (2020). Examining how Spanish-speaking English language learners use their linguistic resources and language modes in a dual language mathematics assessment task. *Journal of Latinos and Education*, 1-14. https://doi.org/10.1080/15348431.2020.1731693

Lopez, J., Salim, S. S., Zaremohzzabieh, Z., & Ahrari, S. (2022). The role, experience, and challenges to headmasters of indigenous primary schools amid covid-19 in Malaysia. *Asian Journal of University Education (AJUE)*, *18*(1), 232-243.

Lovett, B. J., & Harrison, A. G. (2021). De-implementing inappropriate accommodations practices. *Canadian Journal of School Psychology*, *36*(2), 115-126. https://doi.org/10.1177/0829573520972556

Macqueen, S., Knoch, U., Wigglesworth, G., Nordlinger, R., Singer, R., McNamara, T., & Brickle, R. (2019). The impact of national standardized literacy and numeracy testing on children and teaching staff in remote Australian Indigenous communities. *Language Testing*, *36*(2), 265-287. https://doi.org/10.1177%2F0265532218775758

Malaysian Curriculum Development Division. (2017). *Standards-based curriculum and assessment (DSKP): Mathematics Year 1*. Ministry of Education: Curriculum Development Division.

Malmqvist, J., Hellberg, K., Möllås, G., Rose, R., & Shevlin, M. (2019). Conducting the pilot study: A neglected part of the research process? Methodological findings supporting the importance of piloting in qualitative research studies. *International Journal of Qualitative Methods*, *18*, https://doi.org/10.1177/1609406919878341

Marshall, C. & Rossman, G. (2016) *Designing qualitative research (6th Ed).* Sage Publishing.

Martin, B. N., & Fuchs, L. S. (2019). The mathematical performance of at-risk first graders as a function of limited English proficiency status. *Learning Disability Quarterly*, *42*(4), 244-251. https://doi.org/10.1177%2F0731948719827489

Martiniello, M. (2008). Language and the performance of English-language learners in math word problems. *Harvard Educational Review*, *78*(2), 333-368. https://doi.org/10.17763/haer.78.2.70783570r1111t32

Matthews, J. S., & López, F. (2019). Speaking their language: The role of cultural content integration and heritage language for academic achievement among Latino children. *Contemporary Educational Psychology*, *57*, 72-86.

McCarty, T. L. (2018). Twelfth annual brown lecture in education research: so that any child may succeed: indigenous pathways toward justice and the promise of brown. *Educational Researcher*, *47*(5), 271-283.

McCarty, T. L. (2021). The holistic benefits of education for Indigenous language revitalisation and reclamation (ELR2). *Journal of Multilingual and Multicultural Development*, *42*(10), 927-940. https://doi.org/10.1080/01434632.2020.1827647

McKinnon, R. D., & Blair, C. (2019). Bidirectional relations among executive function, teacher–child relationships, and early reading and math achievement: A cross-lagged panel analysis. *Early Childhood Research Quarterly*, *46*, 152-165.

McMahon, D., Wright, R., Cihak, D. F., Moore, T. C., & Lamb, R. (2016). Podcasts on mobile devices as a read-aloud testing accommodation in middle school science assessment. *Journal of Science Education and Technology, 25*(2), 263-273.

Meaney, T., & Evans, D. (2013). What is the responsibility of mathematics education to the Indigenous students that it serves? *Educational Studies in Mathematics*, *82*(3), 481-496. https://doi.org/10.1007/s10649-012-9439-1

Meaney, T., Trinick, T., & Fairhall, U. (2017). Language choice and ethnomathematics in the Pacific: transforming education. *Journal of Mathematics and Culture*, *11*(3), 112-132.

Meaney, T., Trinick, T. (2020). Indigenous Students in Mathematics Education. In S. Lerman (Eds), *Encyclopedia of mathematics education* (pp. 369-373). Springer. https://doi.org/10.1007/978-3-030-15789-0_76

Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative research: A guide to design and implementation (4th ed*). Jossey- Bass.

Miller, J. F. (2018). *Do you read me?: Learning difficulties, dyslexia and the denial of meaning*. Routledge.

Miller, J., & Armour, D. (2021). Supporting successful outcomes in mathematics for aboriginal and Torres Strait Islander students: a systematic review. *Asia-Pacific Journal of Teacher Education*, *49*(1), 61-77.

Ministry of Education Malaysia (MOE). (2022). *List of primary and secondary schools June 2022*. https://www.moe.gov.my/muat-turun/laporan-dan-statistik/senarai-sekolah/5341-senarai-sekolah-rendah-dan-menengah-jun-2022/file

Ministry of Education Malaysia (MOE). (2022). Malaysia education blueprint (2013-2025): Annual report 2021. https://www.moe.gov.my/storage/files/shares/Dasar/PPPM/MEB%20Annual%20Report%202021.pdf

Ministry of Education (MOE). (2023). Malaysia education blueprint (2013-2025): Annual report 2022. https://www.moe.gov.my/storage/files/shares/Dasar/PPPM/MEB%20Annual%20Report%202022.pdf

Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, *22*(1), 67-72.

Misnaton Rabahi, Hamidah Yusof, Marinah Awang, & Arif Jawaid. (2020). Strategic considerations for kindergardens in leading learning of orang asli students. *International Journal of Academic Research in Business and Social Sciences*, *10*(7), 559-570.

Mohd Azizul Sulaiman, Nor Farhana Mohd Azmi, Arifi Ridzuan, Azwan Shah Aminuddin, and Norhana Aini Saini. (2020). Differences in teacher's perspective on academic engagement among students of Orang Asli and non-Orang Asli students. *Gading Journal for Social Sciences*, *23*(1), 13-23.

Mohamad Safwat Ashahri Mohd Salim, Airil Haimi Mohd Adnan, Shah, Dianna Suzieanna Mohamad Shah, Mohd Haniff Mohd Tahir, & Ahmad Muhyiddin Yusof. (2020). The Orang Asli in Malaysian formal education: Orang Asli teachers' sentiments and observations. *International Journal of Humanities Technology and Civilization*, *1*, 95-108.

Moschkovich, J. N. (2010). Language(s) and learning mathematics: Resources, challenges, and issues for research. In J. N. Moschkovich (Ed.), *Language and mathematics education: Multiple perspectives and directions for research* (pp. 1–28). IAP Information Age Publishing.

Moschkovich, J. N. (2020). Bilingual/multilingual issues in learning mathematics. In S. Lerman (Eds.), *Encyclopedia of mathematics education* (pp.75-79), Springer.

Morrison, A., Rigney, L. I., Hattam, R., & Diplock, A. (2019). *Toward an Australian culturally responsive pedagogy: A narrative review of the literature*. University of South Australia.

Motulsky, S. L. (2021). Is member checking the gold standard of quality in qualitative research? *Qualitative Psychology, 8*(3), 389–406. https://doi.org/10.1037/qup0000215

Mullis, I. V., Martin, M. O., & Foy, P. (2013). The impact of reading ability on TIMSS mathematics and science achievement at the fourth grade: An analysis by item reading demands. In M. O. Martin & I. V. S. Mullis (Eds.), *TIMSS and PIRLS 2011: Relationships among reading, mathematics, and science achievement at the fourth grade: Implications for early learning* (pp. 67–108). TIMSS & PIRLS International Study Center, Boston College and IEA.

Muñiz, J., Elosua, P., & Hambleton, R. K. (2013). Directrices para la traducción y adaptación de los tests: segunda edición [Guidelines for test translation and adaptation: Second edition]. *Psicothema, 25*, 151- 157.

Nagy, W., & Townsend, D. (2012). Words as tools: Learning academic vocabulary as language acquisition. *Reading Research Quarterly*, *47*(1), 91-108. https://doi.org/10.1002/RRQ.011

Nelson, G., & Powell, S. R. (2018). Computation error analysis: Students with mathematics difficulty compared to typically achieving students. *Assessment for Effective Intervention*, *43*(3), 144-156. https://doi.org/10.1177/1534508417745627

Ngu, B. H., Phan, H. P., Hong, K. S., & Usop, H. (2016). Reducing intrinsic cognitive load in percentage change problems: The equation approach. *Learning and Individual Differences*, *51*, 81-90. https://doi.org/10.1016/j.lindif.2016.08.029

Noble, T., Rosebery, A., Suarez, C., Warren, B., & O'Connor, M. C. (2014). Science assessments and English language learners: Validity evidence based on response processes. *Applied Measurement in Education*, *27*(4), 248-260.

Noor Azlan Mohd Noor, & Noor Asyhikin Abd Razak. (2022). Overcoming learning challenges during the COVID-19 pandemic: Traditional knowledge as an educational alternative for the Orang Asli. *Islam and Civilisational Renewal Journal*, *13*(1), 31-50.

Noraini Mohd Shah, Ridzwan Che'Rus, Ramlee Mustapha, Mohd Azlan Mohammad Hussain, & Norwaliza Abdul Wahab. (2018). The Orang Asli profile in peninsular Malaysia: background & challenges. *International Journal of Academic Research in Business and Social Sciences*, *8*(7), 1157-1164.

Nortvedt, G. A., & Buchholtz, N. (2018). Assessment in mathematics education: Responding to issues regarding methodology, policy, and equity. *ZDM*, *50*(4), 555-570.

Norwaliza Abdul Wahab, Ridzuan Jaafar, Ramlee Mustapha, Arasinah Kamis, & Norhayati Mohd Affandi. (2017). The effectiveness of likes method in improving reading skills of orang asli students. *Asian Social Science*, *13*(6), 1-74. https://doi.org/10.5539/ass.v13n6p74

Orcan, F. (2020). Parametric or non-parametric: Skewness to test normality for mean comparison. *International Journal of Assessment Tools in Education*, *7*(2), 255-265. https://doi.org/10.21449/ijate.656077

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2016). Cognitive load measurement as a means to advance cognitive load theory. In F. Paas, A. Renkl, & J. Sweller (Ed.), *Educational psychologist* (pp. 63-71). Routledge.

Pandra, V., & Mardapi, D. (2017). Development of mathematics achievement test for third grade students at elementary school in Indonesia. *International Electronic Journal of Mathematics Education*, *12*(3), 769-776. https://doi.org/10.29333/iejme/647

Parra-Frutos, I. (2013). Testing homogeneity of variances with unequal sample sizes. *Computational statistics*, *28*, 1269-1297. https://doi.org/10.1007/s00180-012-0353-x

Parra-Frutos, I. (2016). Preliminary tests when comparing means. *Computational statistics*, *31*, 1607-1631. https://doi.org/10.1007/s00180-016-0656-4

Paschall, K. W., Gershoff, E. T., & Kuhfeld, M. (2018). A two decade examination of historical race/ethnicity disparities in academic achievement by poverty status. *Journal of Youth and Adolescence*, *47*(6), 1164-1177. https://doi.org/10.1007/s10964-017-0800-7

Pearson, N., Naylor, P. J., Ashe, M. C., Fernandez, M., Yoong, S. L., & Wolfenden, L. (2020). Guidance for conducting feasibility and pilot studies for implementation trials. *Pilot and Feasibility Studies*, *6*, 1-12. https://doi.org/10.1186/s40814-020-00634-w

Peltier, C., & Harrison, J. R. (2018). Selecting accommodations for mathematics assessments: Legal and practical considerations. *Preventing School Failure: Alternative Education for Children and Youth*, *62*(4), 300-310. https://doi.org/10.1080/1045988X.2018.1443425

Peng, P., & Lin, X. (2019). The relation between mathematics vocabulary and mathematics performance among fourth graders. *Learning and Individual Differences*, *69*, 11-21.

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

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

Phan, H. P., Ngu, B. H., & Yeung, A. S. (2017). Achieving optimal best: Instructional efficiency and the use of cognitive load theory in mathematical problem solving. *Educational Psychology Review*, *29*(4), 667-692.

Phillips, S. E. (1994). High stakes testing accommodations: Validity versus disabled rights. *Applied measurement in education*, *7*(2), 93-120.

Planas, N., Morgan, C. R., & Schütte, M. (2018). Mathematics education and language: Lessons and directions from two decades of research. In T. Dreyfus, M. Artigue, D. Potari, S. Prediger, & K. Ruthven (Eds.), *Developing research in mathematics education: Twenty years of communication, cooperation and collaboration in Europe***.** Routledge.

Plath, J., & Leiss, D. (2018). The impact of linguistic complexity on the solution of mathematical modelling tasks. *ZDM*, *50*(1), 159-171. https://doi.org/10.1007/s11858-017-0897-x

Plummer, J. D., & Ozcelik, T. A. (2015). Preservice teachers developing coherent inquiry investigations in elementary astronomy. *Science Education*, *99*(5), 932-957.

Polit, D. F. & Beck, C.T. (2018). *Essentials of nursing research: Appraising evidence for nursing practice,* (9th ed.). Wolters Kluwer.

Pongsakdi, N., Kajamies, A., Veermans, K., Lertola, K., Vauras, M., & Lehtinen, E. (2020). What makes mathematical word problem solving challenging? Exploring the roles of word problem characteristics, text comprehension, and arithmetic skills. *ZDM*, *52*(1), 33-44. https://doi.org/10.1007/s11858-019-01118-9

Powell, S. R., & Nelson, G. (2017). An investigation of the mathematics-vocabulary knowledge of first-grade students. *The Elementary School Journal*, *117*(4), 664-686.

Powell, S. R., Berry, K. A., & Tran, L. M. (2020). Performance differences on a measure of mathematics vocabulary for English learners and non-English learners with and without mathematics difficulty. *Reading & Writing Quarterly*, *36*(2), 124-141.

Prediger, S., Erath, K., & Opitz, E. M. (2019). The language dimension of mathematical difficulties. In A. Fritz, V. G. Haase & P. Räsänen (Ed.), *International handbook of mathematical learning difficulties* (pp. 437-455). Springer.

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

Puma, S., & Tricot, A. (2019). Cognitive load theory and working memory models: Comings and goings. In S. Tindall-Ford, S. Agostinho, & J. Sweller (Ed.), *Advances in Cognitive Load Theory* (pp. 41-52). Routledge.

Purpura, D. J., Litkowski, E. C., & Knopik, V. (2019). Mathematics and reading develop together in young children: practical and policy considerations. *Policy Insights from the Behavioral and Brain Sciences*, *6*(1), 12-20.

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

Radmehr, F. & Vos, P. (2020). Issues and challenges of 21st century assessment in mathematics education. In L. Leite, E. Oldham, A. S. Afonso, F. Viseu, L. Dourado, H. Martinho (Eds.), *Science and mathematics education for 21st century citizens: Challenges and ways forward* (pp. 437–462). Nova Science Publishers.

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.

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

Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). *Mastering Modern Psychological Testing: Theory and Methods*. Springer.

Ruef, J. L., Jacob, M. M., Walker, G. K., & Beavert, V. R. (2020). Why indigenous languages matter for mathematics education: a case study of Ichishkíin. *Educational Studies in Mathematics*, *104*, 313-332. https://doi.org/10.1007/s10649-020-09957-0

Riccomini, P. J., Smith, G. W., Hughes, E. M., & Fries, K. M. (2015). The language of mathematics: The importance of teaching and learning mathematical vocabulary. *Reading & Writing Quarterly*, *31*(3), 235-252.

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

Rios, J. A., Ihlenfeldt, S. D., & Chavez, C. (2020). Are accommodations for English learners on state accountability assessments evidence-based? A multistudy systematic review and meta-analysis. *Educational Measurement: Issues and Practice*, *39*(4), 65-75. https://doi.org/10.1111/emip.12337

Rivera, C., Collum, E., Shafer Willner, L., & Sia, J. K., Jr. (2006). An analysis of state assessment policies addressing the accommodation of English language learners. In C. Rivera & E. Collum (Eds.), *A national review of state assessment policy and practice for English language learners* (pp. 1–173). Erlbaum.

Roberts, T. (1998). Mathematical registers in Aboriginal languages. *For the Learning of Mathematics*, *18*(1), 10-16.

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

Robinson, J. P. (2010). The effects of test translation on young English learners' mathematics performance. *Educational Researcher*, *39*(8), 582-590.

Rodriguez, M. C., & Albano, A. D. (2011). Item writing for test developers, researchers, and teachers. *Examinations Research*, *27*, 85-94.

Rogers, C. M., Lazarus, S. S., Thurlow, M. L., & Liu, K. K. (2020). A summary of the research on the effects of K-12 test accommodations: 2017. NCEO Report 418. *National Center on Educational Outcomes*.

Roussel, S., Joulia, D., Tricot, A., & Sweller, J. (2017). Learning subject content through a foreign language should not ignore human cognitive architecture: A cognitive load theory approach. *Learning and Instruction*, *52*, 69-79. https://doi.org/10.1016/j.learninstruc.2017.04.007

Ruef, J. L., Jacob, M. M., Walker, G. K., & Beavert, V. R. (2020). Why indigenous languages matter for mathematics education: A case study of Ichishkíin. *Educational Studies in Mathematics*, *104*, 313-332. https://doi.org/10.1007/s10649-020-09957-0

Ruslin, R., Mashuri, S., Rasak, M. S. A., Alhabsyi, F., & Syam, H. (2022). Semi-structured interview: A methodological reflection on the development of a qualitative research instrument in educational studies. *IOSR Journal of Research & Method in Education*, *12*(1), 22-29.

Ruzlan Md-Ali, Arsaythmaby Veloo, Kanageswari S. Suppiah Shanmugam., Yus'aiman Jusoh, & Rosna Awang Hashim. (2021). The issues and challenges of mathematics teaching and learning in Malaysia" orang asli" primary schools from teachers' perspectives. *Malaysian Journal of Learning and Instruction*, *18*(2), 129-160.

Ryser, G. R. (2018). Qualitative and quantitative approaches to assessment. In S. K. Johnsen (Eds.) *Identifying gifted students* (pp. 33-57). Routledge.

Salihu, L., Aro, M., & Räsänen, P. (2018). Children with learning difficulties in mathematics: Relating mathematics skills and reading comprehension. *Issues in Educational Research*, *28*(4), 1024-1038.

Salihu, L., & Räsänen, P. (2018). Mathematics skills of Kosovar primary school children: A special view on children with mathematical learning difficulties. *International Electronic Journal of Elementary Education*, *10*(4), 421-430.

Schissel, J. L., De Korne, H., & López-Gopar, M. (2021). Grappling with translanguaging for teaching and assessment in culturally and linguistically diverse contexts: teacher perspectives from Oaxaca, Mexico. *International Journal of Bilingual Education and Bilingualism*, *24*(3), 340-356. https://doi.org/10.1080/13670050.2018.1463965

Schuth, E., Köhne, J., & Weinert, S. (2017). The influence of academic vocabulary knowledge on school performance. *Learning and Instruction*, *49*, 157-165. https://doi.org/10.1016/j.learninstruc.2017.01.005

Sedgwick, P. (2015). Multistage sampling. *BMJ*, *351*. https://doi.org/10.1136/bmj.h4155

Shaftel, J., Belton-Kocher, E., Glasnapp, D., & Poggio, J. (2006). The impact of language characteristics in mathematics test items on the performance of English language learners and students with disabilities. *Educational Assessment*, *11*(2), 105-126.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763-1768. https://doi.org/10.1213/ANE.0000000000002864

Sibanda, L. (2017). Grade 4 learners' linguistic difficulties in solving mathematical assessments. *African Journal of Research in Mathematics, Science and Technology Education*, *21*(1), 86-96.

Sireci, S. G., Scarpati, S. E., & Li, S. (2005). Test accommodations for students with disabilities: An analysis of the interaction hypothesis. *Review of Educational Research*, *75*(4), 457-490. https://doi.org/10.3102%2F00346543075004457

Sireci, S. G., & Faulkner-Bond, M. (2015). Promoting validity in the assessment of English learners. *Review of Research in Education*, *39*(1), 215-252. https://doi.org/10.3102/0091732X14557003

Sireci, S. G., Banda, E., & Wells, C. S. (2018). Promoting valid assessment of students with disabilities and English learners. In *Handbook of accessible instruction and testing practices* (pp. 231-246). Springer.

Sireci, S. G. (2020). Standardization and UNDERSTANDardization in educational assessment. *Educational Measurement: Issues and Practice*, *39*(3), 100-105. https://doi.org/10.1111/emip.12377

Solano-Flores, G. (2012). Translation Accommodations Framework for Testing English Language Learners in Mathematics. *Smarter Balanced Assessment Consortium*.

Solano-Flores, G. (2014). Probabilistic approaches to examining linguistic features of test items and their effect on the performance of English language learners. *Applied Measurement in Education*, *27*(4), 236-247. https://doi.org/10.1080/08957347.2014.944308

Spencer, M., Fuchs, L. S., Geary, D. C., & Fuchs, D. (2022). Connections between mathematics and reading development: Numerical cognition mediates relations between foundational competencies and later academic outcomes. *Journal of Educational Psychology*, *114*(2), 273-288.

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

Srinivasan, R., Lohith, C. P., Srinivasan, R., & Lohith, C. P. (2017). Pilot Study—Assessment of validity and reliability. *Strategic marketing and innovation for Indian MSMEs*, 43-49.

Stansfield, C. W. (2011). Oral translation as a test accommodation for ELLs. *Language Testing*, *28*(3), 401-416. https://doi.org/10.1177%2F0265532211404191

Stone, E. A., & Cook, L. L. (2018). Fair testing and the role of accessibility. In S. N. Elliott, R. J. Kettler, P. A. Beddow, & A. Kruz (Eds.), *Handbook of Accessible Instruction and Testing Practices* (pp. 59-73). Springer.

Suarsana, I., Lestari, I. A. P. D., & Mertasari, N. M. S. (2019). The effect of online problem posing on students' problem-solving ability in mathematics. *International Journal of Instruction*, *12*(1), 809-820.

Sumathi Renganathan. (2016). Educating the Orang Asli children: Exploring indigenous children's practices and experiences in schools. *The Journal of Educational Research*, *109*(3), 275-285.

Sumathi Renganathan, & Kral, I. (2018). Exploring language and education policies for the indigenous minorities in Australia and Malaysia. *International Journal of Multicultural Education*, *20*(1), 138-156. https://doi.org/10.18251/ijme.v20i1.1530

Sweller, J., & Chandler, P. (1991). Evidence for cognitive load theory. *Cognition and instruction*, *8*(4), 351-362. https://doi.org/10.1207/s1532690xci0804_5

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, *4*(4), 295-312.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational psychology review*, *22*(2), 123-138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J. (2016). Working memory, long-term memory, and instructional design. *Journal of Applied Research in Memory and Cognition*, *5*(4), 360-367.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in science education*, *48*(6), 1273-1296. https://doi.org/10.1007/s11165-016-9602-2

Tanius, E., Siregar, S., Kasim, C. M. M., & Abdul Jalil, S. Z. S. (2020). Indigenous (orang asli) primary school mathematics performance in Selangor, Malaysia. *International Journal of Innovative Research in Engineering and Multidisciplinary Physical Sciences, 8*(5), 13-19.

Theens, F. (2019). *Does language matter?: Sources of inequivalence and demand of reading ability of mathematics tasks in different languages* (Doctoral dissertation, Umeå universitet).

Thompson, D. R., & Rubenstein, R. N. (2000). Learning mathematics vocabulary: Potential pitfalls and instructional strategies. *The Mathematics Teacher, 93*(7), 568-574.

Thomson, S. (2018). Achievement at school and socioeconomic background—an educational perspective. *Science of Learning*, *3*(1), 1-2.

Thorndike, R. M., Cunningham, G. K., Thorndike, R. L., & Hagen, E. P. (1991). *(5th ed.). Macmillan Publishing Co, Inc.*

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

Turkan, S., & Oliveri, M. E. (2014). Considerations for providing test translation accommodations to English language learners on common core standards-based assessments. *ETS Research Report Series*, *2014*(1), 1-13.

Unal, Z. E. (2019). *Mathematics vocabulary knowledge of eighth-grade students* (Doctoral dissertation, University of Texas).

Vaismoradi, M., & Snelgrove, S. (2019). Theme in qualitative content analysis and thematic analysis. *Forum: Qualitative Social Research, 20*(3), 1-14. http://dx.doi.org/10.17169/fqs-20.3.3376

Van Ryzin, M. J., & Vincent, C. G. (2017). Use of native language and culture (NLC) in elementary and middle school instruction as a predictor of mathematics achievement. *Journal of American Indian Education*, *56*(2), 3-33. https://doi.org/10.1353/jaie.2017.a798615

Vaughn, S., Roberts, G., Capin, P., Miciak, J., Cho, E., & Fletcher, J. M. (2019). How initial word reading and language skills affect reading comprehension outcomes for students with reading difficulties. *Exceptional Children*, *85*(2), 180-196. https://doi.org/10.1177/0014402918782618

Vidal Rodeiro, C., & Macinska, S. (2022). Equal opportunity or unfair advantage? The impact of test accommodations on performance in high-stakes assessments. *Assessment in Education: Principles, Policy & Practice*, *29*(4), 462-481.

Virgana, V., & Lapasau, M. (2019, October). The Influence of vocabulary mastery and reading comprehension towards performance of students in mathematics. *Journal of Physics: Conference Series*, *1360*, 012001. https://doi.org/10.1088/1742-6596/1360/1/012001

Walkington, C., Clinton, V., Ritter, S. N., & Nathan, M. J. (2015). How readability and topic incidence relate to performance on mathematics story problems in computer-based curricula. *Journal of Educational Psychology*, *107*(4), 1051-1074.

Walkington, C., Clinton, V., & Shivraj, P. (2018). How readability factors are differentially associated with performance for students of different backgrounds when solving mathematics word problems. *American Educational Research Journal*, *55*(2), 362-414. https://doi.org/10.3102%2F0002831217737028

Walkington, C., Clinton, V., & Sparks, A. (2019). The effect of language modification of mathematics story problems on problem-solving in online homework. *Instructional Science*, *47*(5), 499-529. https://doi.org/10.1007/s11251-019-09481-6

Wang, S., Zhang, M., & You, S. (2020). A comparison of IRT observed score kernel equating and several equating methods. *Frontiers in Psychology*, *11*(308). 1-19. https://doi.org/10.3389/fpsyg.2020.00308

White, D. E., Oelke, N. D., & Friesen, S. (2012). Management of a large qualitative data set: Establishing trustworthiness of the data. *International Journal of Qualitative Methods*, *11*(3), 244-258. https://doi.org/10.1177/160940691201100305

White, S. (2012). Mining the text: 34 text features that can ease or obstruct text comprehension and use. *Literacy Research and Instruction, 51*(2), 143–164.

Wickstrom, H., Fesseha, E., & Jang, E. E. (2020). Examining the relation between IEP status, testing accommodations, and elementary students' EQAO mathematics achievement. *Canadian Journal of Science, Mathematics and Technology Education*, *20*(2), 297-311. https://doi.org/10.1007/s42330-020-00088-5

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

Williams, L. J., & Abdi, H. (2010). Post-hoc comparisons. *Encyclopedia of research design*, 1060-1067.

Williams-McBean, C. T. (2019). The value of a qualitative pilot study in a multi-phase mixed methods research. *The Qualitative Report*, *24*(5), 1055-1064.

Willner, L. S., & Mokhtari, K. (2018). Improving meaningful use of accommodations by multilingual learners. *The Reading Teacher*, *71*(4), 431-439.

Wolf, M. K., & Leon, S. (2009). An investigation of the language demands in content assessments for English language learners. *Educational Assessment*, *14*(3-4), 139-159. https://doi.org/10.1080/10627190903425883

Wong, B. W. K., & Abdillah, K.K. (2018). Poverty and primary education of the Orang Asli children. In Joseph C (Eds.), *Policies and politics in Malaysian education: education reforms, nationalism and neoliberalism*, (pp. 54-71). Routledge.

Xu, Y., & Kuti, L. (2024). Accommodating students with exceptional needs by aligning classroom assessment with IEP goals. *International Journal of InclusiveEducation*, *28*(8),1474-1487. https://doi.org/10.1080/13603116.2021.1994662

Yang, X. (2020). Assessment accommodations for emergent bilinguals in mainstream classroom assessments: a targeted literature review. *International Multilingual Research Journal*, *14*(3), 217-232.

Yolcu, A. (2021). Reimagining the citizen and the nation in a globalised world: the case of mathematics education reforms in Turkey. *Research in Mathematics Education*, *23*(3), 278-292. https://doi.org/10.1080/14794802.2021.1993976

Zaleha Ismail, Tan, Y. C., & Nur Amira Muda. (2020). Numeracy competency of year 5 aboriginal students using written and oral tests. *The Mathematics Enthusiast*, *17*(1), 32-62.

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

Zidney, R., Sjöström, J., & Eilks, I. (2020). A multi-perspective reflection on how indigenous knowledge and related ideas can improve science education for sustainability. *Science & Education*, *29*(1), 145-185. https://doi.org/10.1007/s11191-019-00100-x

Appendix A

Written Mathematics Test Blueprint

Year 1 Topics	No. of Items	Item Length			Total
		Short Item Length	Moderate Item Length	Long Item Length	
1.0 WHOLE NUMBERS UP TO 100 (NOMBOR BULAT HINGGA 100)					
1.1 Quantity intuitively (Kuantiti secara intuitif)			1, 2		2
1.2 Number value (Nilai nombor)		4	3		2
1.3 Write numbers (Menulis nombor)			5, 6		2
1.7 Estimate (Menganggar)			26		1
1.8 Round of numbers (Membundarkan nombor)		17			1
1.10 Problem solving (Penyelesaian masalah)			18, 27		2
2.0 BASIC OPERATIONS (OPERASI ASAS)					
2.1 Concepts of addition and subtraction (Konsep tambah dan tolak)		7, 8	19	21	4
2.2 Add within 100 (Tambah dalam lingkungan 100)				28	1
2.3 Subtract within 100 (Tolak dalam lingkungan 100)		9	29	30	3
2.4 Problem solving (Penyelesaian masalah)				31, 32	2
2.5 Repeated addition (Tambah berulang)		10			1
2.6 Repeated subtraction (Tolak berturut-turut)		11			1
3.0 FRACTIONS (PECAHAN)					
3.1 Concept of one over two and one over four in proper fractions (Konsep perdua dan perempat pecahan wajar)		12, 22			2
3.2 Problem solving (Penyelesaian masalah)			33		1
4.0 MONEY (WANG)					
4.1 Notes and coins (Wang kertas dan duit syiling)		13, 14		23	3
4.2 Financial resources and savings (Sumber kewangan dan simpanan)				34	1
4.3 Problem solving (Penyelesaian masalah)			35		1
5.0 TIME (MASA DAN WAKTU)					

5.2 Clock face (Muka jam)		15			1
5.3 Problem solving (Penyelesaian masalah			24	36	2
6.0 MEASUREMENT (UKURAN DAN SUKATAN					
6.1 Relative units to measure length, mass and volume of liquids (Unit relatif untuk mengukur panjang, jisim dan isi padu cecair)		37	38		2
7.0 SPACE (RUANG)					
7.1 Three-dimensional shapes (Bentuk tiga dimensi)		25			1
7.2 Two-dimensional shapes (Bentuk dua dimensi)		16			1
8.0 DATA MANAGEMENT (PENGURUSAN DATA)					
8.2 Pictograph (Piktograf)				20, 39	2
8.3 Problem solving (Penyelesaian masalah)				40	1
Total (Jumlah)	40	15	14	11	40

Appendix B

Educational Planning and Policy Research Division Form

KEMENTERIAN PENDIDIKAN MALAYSIA
BAHAGIAN PERANCANGAN DAN PENYELIDIKAN DASAR PENDIDIKAN
ARAS 1-4, BLOK E8
KOMPLEKS KERAJAAN PARCEL E
PUSAT PENTADBIRAN KERAJAAN PERSEKUTUAN
62604 PUTRAJAYA

TEL : 0388846591
FAKS : 0388846579

Ruj. Kami : KPM.600-3/2/3-eras(14800)
Tarikh : 27 Disember 2022

SUHEYSEN A/L REVINDRAN
NO. KP : 930327105287

9, SOLOK GANGSA 1, GREENLANE
11600 GEORGETOWN
PULAU PINANG

Tuan,

KELULUSAN BERSYARAT UNTUK MENJALANKAN KAJIAN :
EXAMINING THE RECORDED ORAL TESTS AS A TEST ACCOMMODATION IN MEASURING ORANG ASLI PUPILS' MATHEMATICS PERFORMANCE

Perkara di atas adalah dirujuk.

2. Sukacita dimaklumkan bahawa permohonan tuan untuk menjalankan kajian seperti di bawah telah diluluskan dengan syarat :

" KELULUSAN INI BERGANTUNG KEPADA PERTIMBANGAN PENGARAH BAHAGIAN PENGURUSAN SEKOLAH HARIAN, JABATAN PENDIDIKAN NEGERI DAN PENTADBIR SEKOLAH YANG BERKAITAN. PENGUTIPAN DATA TIDAK BOLEH MENGGANGGU AKTIVITI PENGAJARAN DAN PEMBELAJARAN MURID. PENGLIBATAN, PEMERHATIAN SERTA RAKAMAN VIDEO TERHADAP AKTIVITI PENGAJARAN DAN PEMBELAJARAN MURID DI DALAM BILIK DARJAH TIDAK DIBENARKAN. "

3. Kelulusan adalah berdasarkan kepada kertas cadangan penyelidikan dan instrumen kajian yang dikemukakan oleh tuan kepada bahagian ini. Walau bagaimanapun kelulusan ini bergantung kepada kebenaran Jabatan Pendidikan Negeri dan Pengetua / Guru Besar yang berkenaan.

4. Surat kelulusan ini sah digunakan bermula dari **2 Januari 2023** hingga **30 Jun 2023**

5. Tuan dikehendaki menyerahkan senaskhah laporan akhir kajian dalam bentuk *hardcopy* bersama salinan *softcopy* berformat pdf dalam CD kepada Bahagian ini.Tuan juga diingatkan supaya mendapat kebenaran terlebih dahulu daripada Bahagian ini sekiranya sebahagian atau sepenuhnya dapatan kajian tersebut hendak diterbitkan di mana-mana forum, seminar atau diumumkan kepada media massa.

Sekian untuk makluman dan tindakan tuan selanjutnya. Terima kasih.

"BERKHIDMAT UNTUK NEGARA"

Saya yang menjalankan amanah,

Ketua Penolong Pengarah Kanan
Sektor Penyelidikan dan Penilaian Dasar
b.p. Pengarah
Bahagian Perancangan dan Penyelidikan Dasar Pendidikan
Kementerian Pendidikan Malaysia

salinan kepada:-

BAHAGIAN PENGURUSAN SEKOLAH HARIAN
JABATAN PENDIDIKAN PERAK

* SURAT INI DIJANA OLEH KOMPUTER DAN TIADA TANDATANGAN DIPERLUKAN *

Appendix C

Approval Letter from Perak State Education Department

KEMENTERIAN PENDIDIKAN MALAYSIA
Jabatan Pendidikan Negeri Perak
Jalan Tun Abdul Razak
30640 Ipoh, Perak Darul Ridzuan

Tel : 605 501 5000
Faks : 605 527 7273
Laman Web : http://jpnperak.moe.gov.my

"PENDIDIKAN BERKUALITI, SEKOLAH UNGGUL, MURID HOLISTIK"

Ruj. Kami : JPNPk.SPS.USJK.600 1 Jld.8(31)
Tarikh : 26 Januari 2023

Suheysen a/l Revindran
9, Solok Gangsa 1, Greenlane
11600 Georgetown
Pulau Pinang

Tuan,

KELULUSAN UNTUK MENJALANKAN KAJIAN DI SEKOLAH-SEKOLAH DI NEGERI PERAK DI BAWAH JABATAN PENDIDIKAN NEGERI PERAK

Dengan segala hormatnya, perkara di atas adalah dirujuk dan surat tuan yang diterima pada 20 Januari 2023 adalah berkaitan.

2. Sehubungan dengan itu, dimaklumkan bahawa Jabatan Pendidikan Negeri Perak tiada halangan untuk membenarkan pihak tuan menjalankan kajian yang bertajuk ***"EXAMINING THE RECORDED ORAL TESTS ACCOMODATION IN MEASURING ORANG ASLI PUPIL'S MATHEMATICS PERFORMANCE"*** seperti dinyatakan dalam surat tuan dengan syarat-syarat berikut:

2.1 Pihak tuan perlu mendapatkan kebenaran terlebih dahulu daripada Pegawai Pendidikan Daerah dan Pengetua / Guru Besar sekolah berkenaan untuk menggunakan sampel kajian;

2.2 Kajian yang dijalankan hendaklah tidak mengganggu proses pengajaran dan pembelajaran yang telah ditetapkan oleh pihak sekolah;

2.3 Pihak tuan bertanggungjawab menjaga keselamatan dan kebajikan murid dan guru yang terlibat dalam kajian ini;

2.4 Murid, guru dan warga sekolah tidak boleh dipaksa terlibat dalam kajian ini;

2.5 Pihak tuan hendaklah bertanggungjawab menanggung semua kos kajian;

2.6 Pihak tuan dipohon agar menghantar satu (1) salinan laporan kajian dalam tempoh 30 hari ke jabatan ini selepas kajian tersebut dilaksanakan;

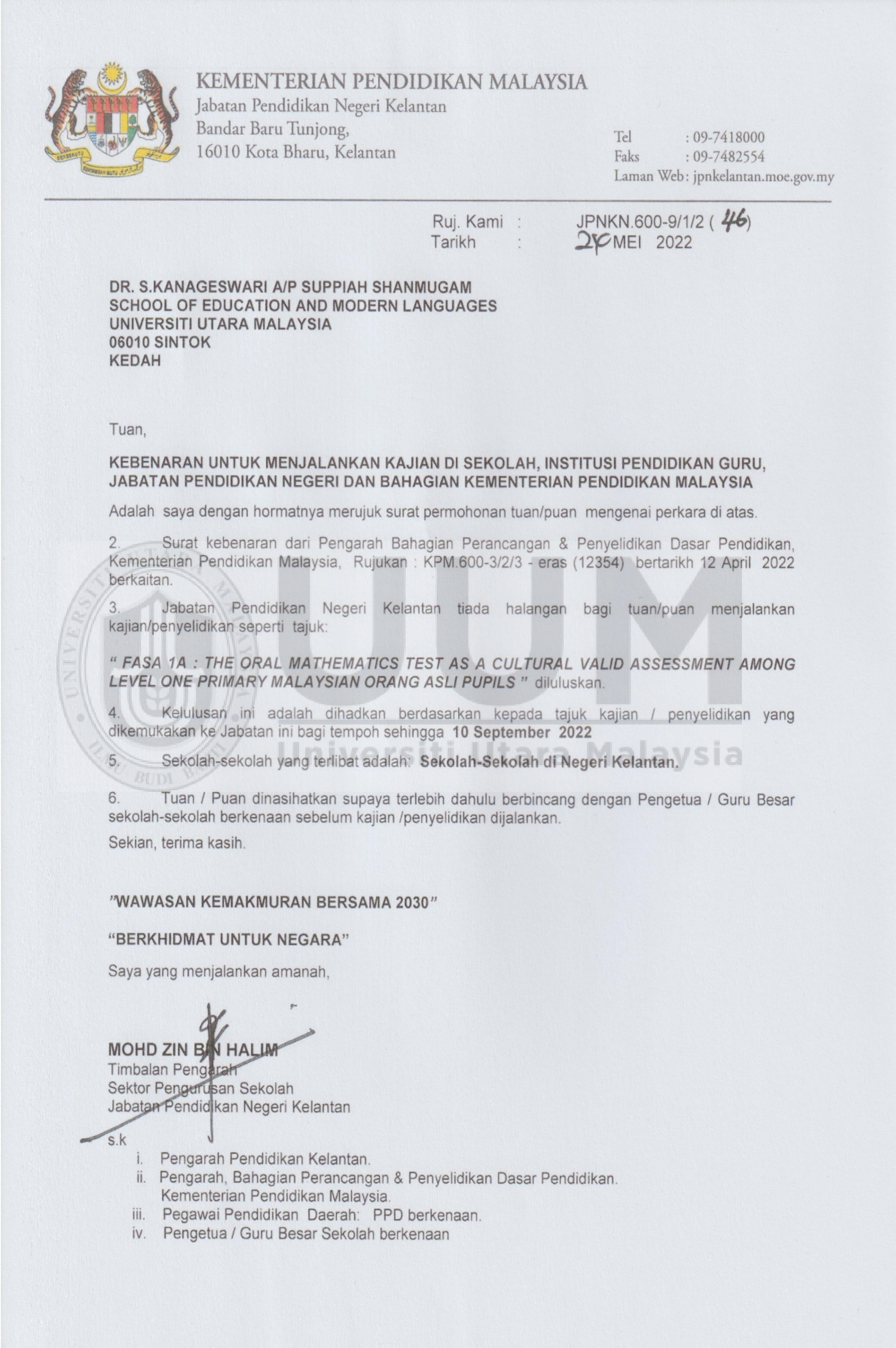
2.7 Tiada sebarang implikasi kewangan kepada Jabatan Pendidikan Negeri Perak, Pejabat Pendidikan Daerah dan pihak sekolah;

.../2

We Deliver

Appendix D

Approval Letter from Kelantan State Education Department



KEMENTERIAN PENDIDIKAN MALAYSIA
Jabatan Pendidikan Negeri Kelantan
Bandar Baru Tunjong,
16010 Kota Bharu, Kelantan

Tel : 09-7418000
Faks : 09-7482554
Laman Web: jpnkelantan.moe.gov.my

Ruj. Kami : JPNKN.600-9/1/2 (46)
Tarikh : 24 MEI 2022

DR. S.KANAGESWARI A/P SUPPIAH SHANMUGAM
SCHOOL OF EDUCATION AND MODERN LANGUAGES
UNIVERSITI UTARA MALAYSIA
06010 SINTOK
KEDAH

Tuan,

KEBENARAN UNTUK MENJALANKAN KAJIAN DI SEKOLAH, INSTITUSI PENDIDIKAN GURU, JABATAN PENDIDIKAN NEGERI DAN BAHAGIAN KEMENTERIAN PENDIDIKAN MALAYSIA

Adalah saya dengan hormatnya merujuk surat permohonan tuan/puan mengenai perkara di atas.

2. Surat kebenaran dari Pengarah Bahagian Perancangan & Penyelidikan Dasar Pendidikan, Kementerian Pendidikan Malaysia, Rujukan : KPM.600-3/2/3 - eras (12354) bertarikh 12 April 2022 berkaitan.

3. Jabatan Pendidikan Negeri Kelantan tiada halangan bagi tuan/puan menjalankan kajian/penyelidikan seperti tajuk:

" FASA 1A : THE ORAL MATHEMATICS TEST AS A CULTURAL VALID ASSESSMENT AMONG LEVEL ONE PRIMARY MALAYSIAN ORANG ASLI PUPILS " diluluskan.

4. Kelulusan ini adalah dihadkan berdasarkan kepada tajuk kajian / penyelidikan yang dikemukakan ke Jabatan ini bagi tempoh sehingga **10 September 2022**

5. Sekolah-sekolah yang terlibat adalah: **Sekolah-Sekolah di Negeri Kelantan.**

6. Tuan / Puan dinasihatkan supaya terlebih dahulu berbincang dengan Pengetua / Guru Besar sekolah-sekolah berkenaan sebelum kajian /penyelidikan dijalankan.

Sekian, terima kasih.

"WAWASAN KEMAKMURAN BERSAMA 2030"

"BERKHIDMAT UNTUK NEGARA"

Saya yang menjalankan amanah,

MOHD ZIN BIN HALIM
Timbalan Pengarah
Sektor Pengurusan Sekolah
Jabatan Pendidikan Negeri Kelantan

s.k
i. Pengarah Pendidikan Kelantan.
ii. Pengarah, Bahagian Perancangan & Penyelidikan Dasar Pendidikan. Kementerian Pendidikan Malaysia.
iii. Pegawai Pendidikan Daerah: PPD berkenaan.
iv. Pengetua / Guru Besar Sekolah berkenaan

Appendix E

WBM Test

Nama: ____

Kelas: ____________________

Jantina:

☐ **Lelaki** ☐ **Perempuan**

MATEMATIK

TAHAP 1

KERTAS 1

1 jam

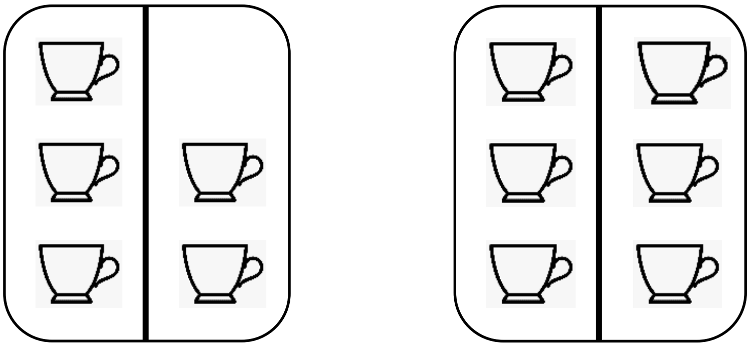
Satu jam

JANGAN BUKA KERTAS SOALAN INI SEHINGGA DIBERITAHU

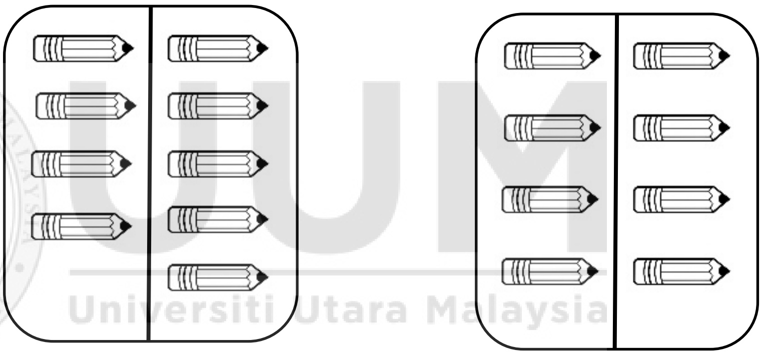
1. Kertas soalan ini dalam Bahasa Melayu bertulis dan secara audio dalam Bahasa Melayu, *Semai/Temiar*.
2. Kertas soalan ini mengandungi 20 item berstruktur.
3. Jawab semua item.
4. Semua jawapan mesti ditulis di ruang yang disediakan.
5. Anda mesti menunjukkan jalan kerja mengira di ruang yang disediakan.

Kertas soalan ini mengandungi 11 halaman bercetak termasuk muka hadapan.

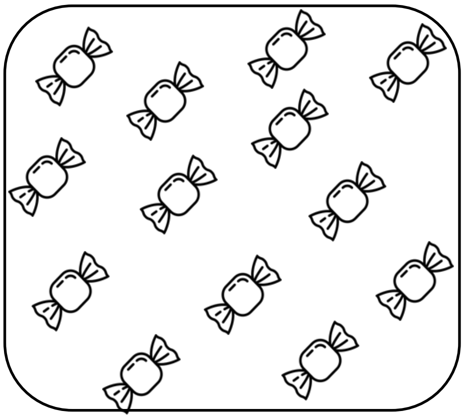
1. **Bulatkan** gambar yang menunjukkan bilangan cawan yang tidak sama banyak.



2. **Bulatkan** gambar yang menunjukkan bilangan pensel yang sama banyak.

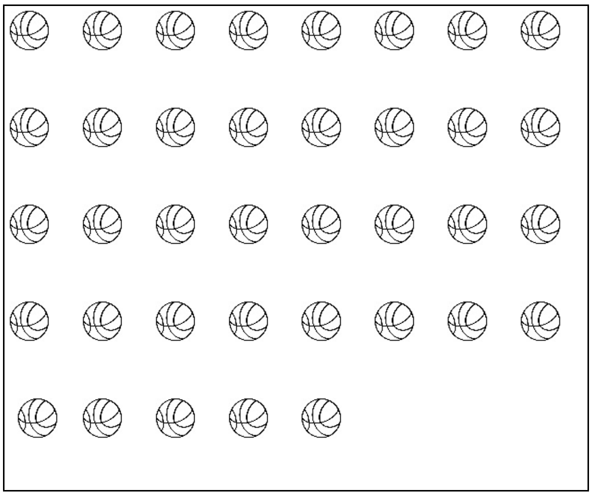


3. **Berapakah** bilangan gula-gula yang terdapat dalam gambar?



=

4. **Bulatkan** jumlah guli yang betul.



36

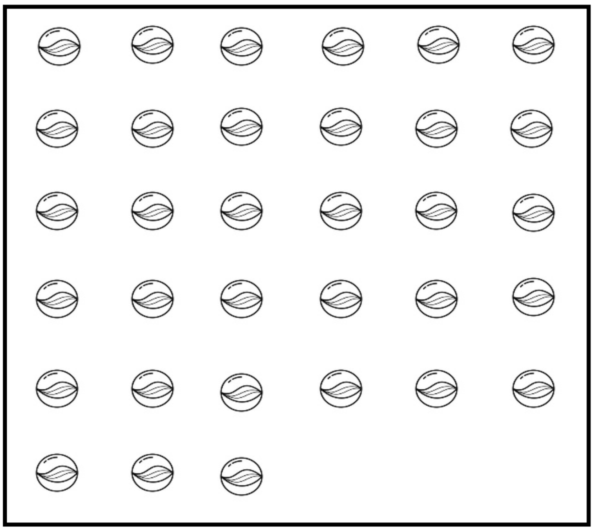
37

38

5. **Tuliskan** nombor yang tertinggal di dalam kotak.

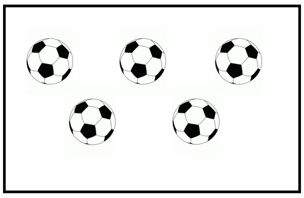
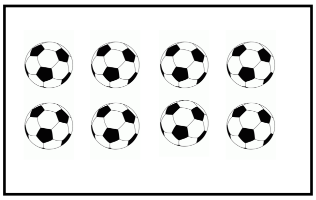
1	2	3	4	5	

6. **Kirakan** jumlah guli yang terdapat dalam gambar.



=

7. **Tuliskan** jumlah bola dalam kotak.

 +  = ☐

8.

5 + 4 = ☐

9.

```
   1 9
—    4
———————

———————
```

10.

6 + 6 + 6 =

11.

6 − 3 − 3 =

12. **Tuliskan** pecahan bagi bahagian hitam.

=

13. **Tuliskan** jumlah duit siling.

=

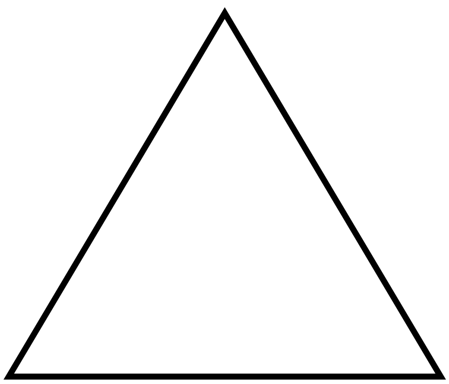
14. **Tuliskan** jumlah wang.

=

15. **Tuliskan** waktu pada jam.

=

16. **Tuliskan** bilangan sisi.



=

17. **Bulatkan** nombor yang paling kecil.

51	15	12	21

18. **Bulatkan** kotak yang betul.

25 dibundarkan kepada puluh yang terdekat.

20	30

19. **Bulatkan** simbol yang betul dalam kotak.

27 [+ | -] 7 = 20

20. **Bulatkan** kumpulan yang mempunyai bilangan yang paling sedikit.

Kumpulan	Bilangan Orang
A	
B	
C	
 **mewakili 1 orang**	

Nama: ____

Kelas: ______________________

Jantina:

☐ **Lelaki** ☐ **Perempuan**

MATEMATIK

TAHAP 1

KERTAS 2

1 jam

Satu jam

JANGAN BUKA KERTAS SOALAN INI SEHINGGA DIBERITAHU

6. Kertas soalan ini dalam Bahasa Melayu bertulis, dan secara audio bagi Bahasa *Semai/Temiar*.

7. Kertas soalan ini mengandungi 20 item berstruktur

8. Jawab semua item.

9. Semua jawapan mesti ditulis di ruang yang disediakan.

10. Anda mesti menunjukkan jalan kerja mengira di ruang yang disediakan.

Kertas soalan ini mengandungi 11 halaman bercetak termasuk muka hadapan

21. **Tuliskan** nombor dan simbol.

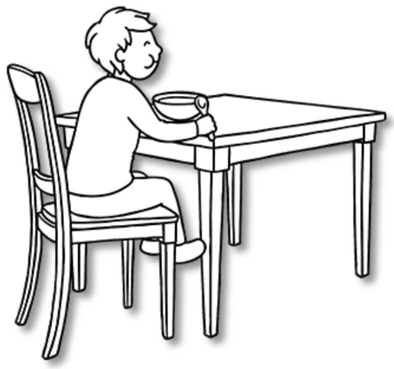
Lima tambah tujuh sama dengan dua belas.

22. **Bulatkan** gambar yang menunjukkan setengah.

23. Ali mempunyai jumlah wang seperti di bawah.

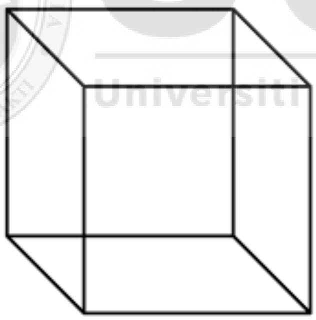
Tuliskan jumlah wang yang ada pada Ali. =

24. **Tuliskan** waktu Ali sedang bersarapan pagi.



=

25. **Namakan** bentuk di bawah.



=

26. **Bulatkan** jawapan yang betul.

49 | kurang | lebih | daripada 43.

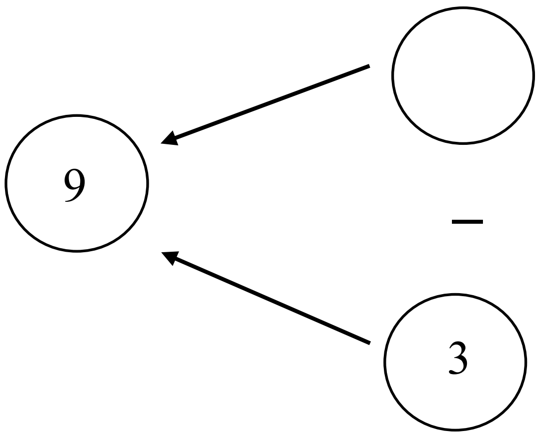
27. **Bulatkan** siapa yang membilang sepuluh-sepuluh?

Farah	:	12,	14,	16
Rozie	:	20,	25,	30
Tajul	:	30,	40,	50

28. Bapa Suzana memberi enam kek dan emak Suzana memberi dua kek.
Berapakah jumlah kek yang ada pada Suzana?

= [ ]

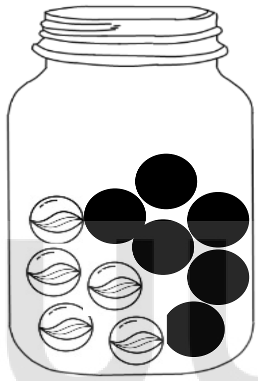
29. **Tuliskan** nombor yang betul dalam bulatan.



30. Ali ada 9 biji kuih. Dia memberi 3 biji kuih kepada adiknya. **Berapakah** baki kuih yang ada pada Ali?

=

31. Di dalam botol ada guli hitam dan guli putih. **Tuliskan** jumlah guli yang terdapat dalam botol.

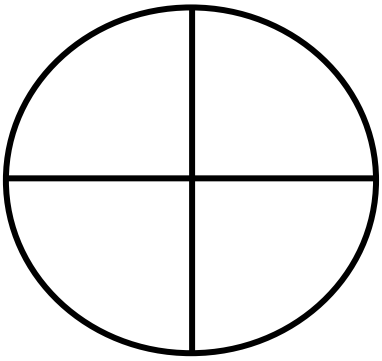


=

32. Bob ada 23 biji telur itik dan Liyana ada 9 biji telur itik. **Berapakah** jumlah telur itik yang ada pada mereka**?**

=

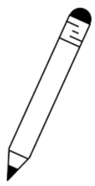
33. **Hitamkan** bulatan di bawah untuk pecahan $\frac{1}{4}$.

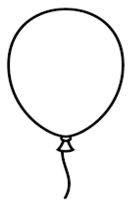


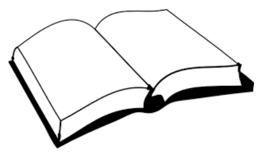
34. Tajul mendapat wang saku sebanyak RM 10. Dia telah membelanjakan RM 8. **Lengkapkan** jadual di bawah.

Wang Saku	**Belanja**

35. Suzana membeli belon dan buku.







90 sen	40 sen	60 sen

Berapakah dia belanjakan? =

36. **Bulatkan** hari Nadia bertugas.

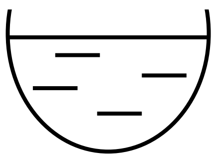
Jadual Tugas Kelas 1 Arif

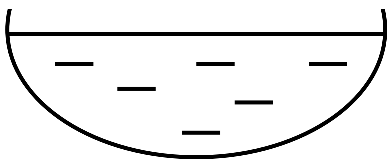
Hari	**Nama**
Isnin	Nadia
Selasa	Bob

37. **Bulatkan** siapa lebih berat?

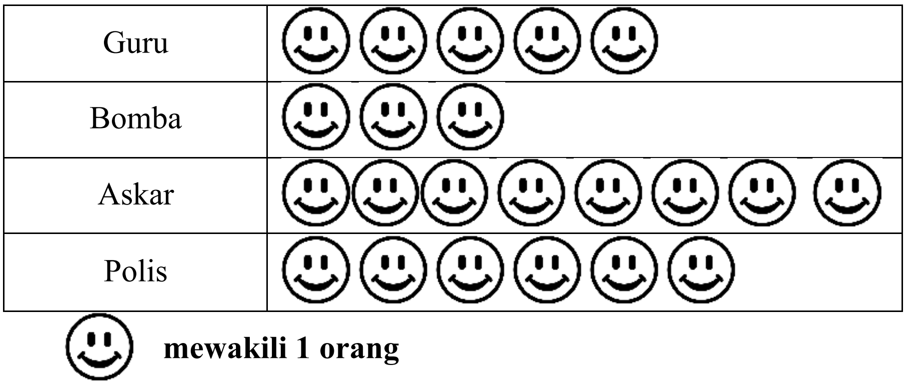
Tajul John

38. **Bulatkan** bekas yang mempunyai air yang lebih banyak.

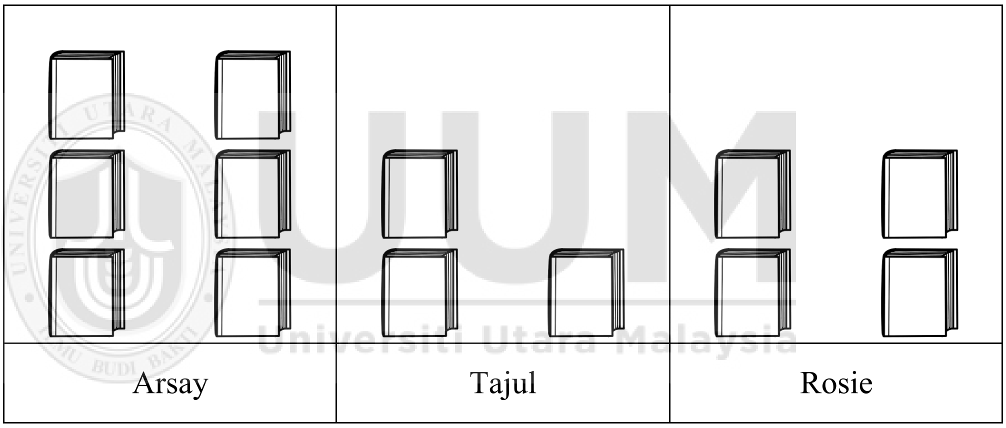




39. **Bulatkan** cita-cita yang paling diminati.



40. **Bulatkan** susunan bagi bilangan buku dalam tertib menaik**.**



A. Arsay → Rosie → Tajul

B. Tajul → Rosie → Arsay

Appendix F

Interview Protocol and Sample Questions

For the session with the mathematics teachers, a brief introduction of the researcher would be provided, and background of this study would be explained. The participants would then be requested to introduce themselves and provide education and professional background.

"Good morning/ afternoon. I would like to thank you for spending the time to attend this interview session. My name is Suheysen, and I am a currently enrolled in the Master of Education programme. Before we start, I would like to provide a background of the research that I am conducting for the *Orang Asli* pupils. This research aims to explore the challenged faced by the mathematics teachers on the current assessment practices for the *Orang Asli* pupils. Before I proceed any further, I would like to start with a brief introduction of yourself…"

Interview Questions

1. How are mathematics assessments implemented for *Orang Asli* pupils?
2. What are the factors that affect the Orang Asli pupils' performance during mathematics assessments?
3. What steps have been taken to address these issues when conducting mathematics assessments or activities in the classroom?
4. What is the teachers' view on the administration of recorded oral tests in *Bahasa Melayu* or *Bahasa Temiar*, or both, during mathematics assessments?

5. What are the teachers' view on the use of the recorded oral test to fairly assess the mathematics performance of the *Orang Asli* pupils?
6. What challenges the *Orang Asli* pupils may face if the recorded oral test was to be implemented as a test accommodation during mathematics assessments?
7. What challenges the teachers and school administrators may face if the recorded oral test was to be implemented as a test accommodation during mathematics assessments?

As we conclude this interview, I extend my heartfelt gratitude for your invaluable insights and candid responses. Your willingness to delve into the discussion on the use of recorded oral tests as a potential accommodation for *Orang Asli* pupils has been particularly illuminating. Your perspectives are integral to the depth of this study. Thank you for your time, commitment, and for sharing your invaluable perspective on this important aspect of testing and education.

Appendix G

Detailed Qualitative Content Analysis for Item Length and Vocabulary Type

Vocabulary Type and Item Difficulty for Short Item Length

The item difficulty for the mathematics test items categorised as short item length were computed and compared across WBM, OBM, and OBT tests. Based on the comparison of item difficulty across the three mathematics tests, the study found that the item difficulty across WBM, OBM, and OBT tests were approximately within the same range for seven test items (Item 4, 7, 8, 10, 11, 17 and 37). On the other hand, five items (Item 9, 11, 13, 16, and 22) were found to be easier when administered the OBM or OBT tests, compared to the WBM test. Lastly, three items (Items 12, 14 and 25) were found to exhibit low item difficulty when administered the OBM or OBT tests compared to the WBM test, whereby the pupils had found the three test items to be much easier in the WBM tests. Table 4.20 presents the item difficulty range for the mathematics test items categorised as short item length for the three mathematics tests.

Table 4.20

Item Difficulty Range based on Vocabulary Type for Short Item Length

Item	**Vocabulary**		**p (WBM)**	**p (OBM)**	**p (OBT)**
	Mathematics Vocabulary	**General Vocabulary**			
4	*jumlah*	bulatkan; *guli; yang; betul*	Easy	Easy	Easy
7	*jumlah*	*tuliskan; bola; dalam; kotak*	Easy	Easy	Easy
8	No vocabulary	No vocabulary	Easy	Easy	Easy

9*	No vocabulary	No vocabulary	Moderate	Easy	Easy
10	No vocabulary	No vocabulary	Easy	Moderate	Easy
11*	No vocabulary	No vocabulary	Moderate	Easy	Easy
12	*pecahan*	*tuliskan; bagi; bahagian; hitam*	Moderate	Hard	Hard
13*	*jumlah; duit syiling*	*tuliskan*	Moderate	Easy	Easy
14	*jumlah; wang*	*tuliskan*	Moderate	Hard	Moderate
15	*waktu*	*tuliskan; pada; jam*	Moderate	Moderate	Moderate
16*	*bilangan; sisi*	*tuliskan*	Moderate	Easy	Easy
17	*nombor*	*bulatkan; yang; paling; kecil*	Easy	Easy	Easy
22*	*setengah*	*bulatkan; gambar; yang; menunjukkan*	Hard	Easy	Easy
25	*bentuk*	*namakan; di; bawah*	Hard	Hard	Hard
37	*lebih; berat*	*bulatkan; siapa; yang; Tajul; John*	Easy	Easy	Easy

*Items showing improved item difficulty in OBM and OBT tests

<u>Vocabulary Type and Item Difficulty for Moderate Item Length</u>

The item difficulty for the mathematics test items categorised as moderate item length were computed and compared across WBM, OBM, and OBT tests. From the 14 mathematics test items categorised as moderate item length, eight items (Item 2, 3, 5, 6, 19, 26, 35, and 38) were found to be within similar item difficulty range, indicating that the pupils' performance was approximately the same across the three tests. A total of four items (Items 1, 27, 29, and 33) showed improved performance when administered the OBM and OBT tests, compared to the WBM test. Table 4.21 presents the item difficulty index values for the mathematics test items categorised as moderate item length for WBM, OBM, and OBT tests.

Table 4.21

Item Difficulty Range for Moderate Items between WBM, OBM, and OBT Tests

Item	**Vocabulary**		**p (WBM)**	**p (OBM)**	**p (OBT)**
	Mathematics Vocabulary	**General Vocabulary**			
1*	bilangan; tidak sama banyak	bulatkan, gambar; yang; menunjukkan; cawan	Moderate	Easy	Easy
2	bilangan; sama banyak	bulatkan; gambar; yang; menunjukkan; pensel	Easy	Moderate	Easy
3	*berapakah; bilangan*	*gula-gula; yang; terdapat; dalam; gambar*	Easy	Easy	Easy
5	*nombor*	*tuliskan; yang; tertinggal; dalam; kotak*	Easy	Easy	Easy
6	*jumlah*	kirakan; *guli; yang; terdapat; dalam; gambar*	Easy	Easy	Easy

Item	Vocabulary		p (WBM)	p (OBM)	p (OBT)
	Mathematics Vocabulary	General Vocabulary			
18	*dibundarkan; puluh yang terdekat*	*bulatkan; kotak; yang; betul; kepada*	Moderate	Easy	Moderate
19	*simbol*	*bulatkan; yang; betul; dalam; kotak*	Moderate	Moderate	Easy
24	*waktu*	*Ali; brsarapan, pagi; tuliskan; sedang*	Moderate	Easy	Moderate
26	kurang; lebih	*bulatkan; jawapan; yang; betul; daripada*	Moderate	Moderate	Moderate
27*	*sepuluh-sepuluh*	*bulatkan; siapa; yang; membilang*	Moderate	Easy	Easy
29*	*nombor; bulatan*	*tuliskan; yang; betul; dalam*	Hard	Moderate	Easy
33*	*hitamkan; bulatan; pecahan*	*di; bawah; untuk*	Moderate	Easy	Easy
35	*berapakah, belajakan*	*belon; buku; membeli; Suzana*	Moderate	Moderate	Moderate
38	*lebih banyak*	*bulatkan; bekas; yang; mempunyai; air*	Easy	Easy	Easy

*Items showing improved item difficulty in OBM and OBT tests

Vocabulary Type for Long Item Length

The item difficulty index for test items categorised as moderate item length were computed and compared across WBM, OBM, and OBT tests. From the 11 mathematics test items that have been categorised under long item length, five items (Item 21, 28, 36, 39, and 40) had obtained similar performance whereby these test items were within the same item difficulty ranges across the three tests. On the other hand, four items (Item 20, 31, 32, and 34) had shown improved performance in the OBM and OBT tests, compared to the WBM tests. The pupils were able to perform better on these items when the recorded oral test in either language was administered. Table 4.22 presents the item difficulty index values for the mathematics test items that are categorised as long item length for WBM, OBM, and OBT tests.

Table 4.22

Item Difficulty Range for Long Items between WBM, OBM, and OBT Tests

Item	**Vocabulary**		**p (WBM)**	**p (OBM)**	**p (OBT)**
	Mathematics Vocabulary	**General Vocabulary**			
20*	*bilangan; paling sedikit*	*bulatkan; kumpulan; yang; mempunyai*	Moderate	Easy	Easy
21	*nombor; simbol; tambah; sama dengan; lima; tujuh; dua belas*	*tuliskan; dan*	Easy	Easy	Easy

Item	Vocabulary		p (WBM)	p (OBM)	p (OBT)
	Mathematics Vocabulary	General Vocabulary			
23	*jumlah; wang*	*Ali; mempunyai; di bawah; tuliskan; yang; ada; pada; seperti*	Moderate	Moderate	Moderate
28	*jumlah; enam; dua*	berapakah; *emak; bapa; kek; memberi; ada; yang; pada; Suzana; dan*	Moderate	Easy	Moderate
30	*baki*	*Ali; ada; biji; kuih; memberi; kepada; adiknya; yang; ada; pada*	Moderate	Easy	Moderate
31*	*jumlah*	*di; dalam; botol; ada; guli; hitam; putih; tuliskan; terdapat; yang*	Moderate	Easy	Easy
32*	*jumlah*	*ada; biji; telur; itik; yang; ada; pada; mereka; Bob; Liyana; berapakah*	Hard	Moderate	Moderate
34*	*wang saku; sebanyak*	*mendapat; lengkapkan; jadual; di; bawah; membelanjakan*	Moderate	Easy	Easy
36	No vocabulary	*bulatkan; hari; Nadia; bertugas;*	Moderate	Easy	Moderate

Item	**Vocabulary**		**p (WBM)**	**p (OBM)**	**p (OBT)**
	Mathematics Vocabulary	**General Vocabulary**			
39	No vocabulary	*jadual; kelas; Isnin; Selasa* *cita-cita; yang; paling; diminati; askar; bomba; polis; guru*	Moderate	Easy	Moderate
40	*susunan; bilangan; tertib menaik*	*bulatkan; buku; dalam; bagi; Tajul; Arsay; Rosie*	Moderate	Moderate	Moderate

*Items showing improved item difficulty in OBM and OBT tests

List of Publication and Conference

Publication

- Suheysen Revindran, S. Kanageswari Suppiah Shanmugam, & Sarimah Shaik-Abdullah. The use of recorded oral tests as test accommodations for assessing mathematics performance among indigenous Orang Asli pupils. *Mathematics Education Research Journal*. (Under review)
- Suheysen Revindran, S. Kanageswari Suppiah Shanmugam, & Sarimah Shaik-Abdullah. The utility of audio mathematics test items as test accommodation for Malaysia's Orang Asli pupils: A mixed-method approach. *Education Assessment*. (Under review)

Conference

- Suheysen Revindran. (2022, November 28-29). *Teacher's perception on STEM education advancement among Malaysia's indigenous pupils: Issues and challenges* [Presentation]. 3rd International Conference of the Southeast Asia International Science-Technology, Engineering and Mathematics (SEA-STEM) 2022, Universiti Terengganu Malaysia, Terengganu, Malaysia.
- Suheysen Revindran, S. Kanageswari Suppiah Shanmugam, & Sarimah Shaik-Abdullah (2023, June 24). *Assessing the viability of audio mathematics test items on mathematics performance among Orang Asli pupils* [Paper Presentation]. 1st Mathematics and Mathematics Education International Conference, Universitas Pendidikan Indonesia, Bandung, Indonesia.

Workshop

- Workshop: Quantitative Data Analysis using SmartPLS 4.0 by Prof. Ramayah Thurasamy. (25– 26 September 2023), Postgraduate Studies Unit, UUM College of Business.