EXAMINING THE VALIDITY OF USING A BILINGUAL TEST TO ASSESS FORM TWO STUDENTS' ACHIEVEMENT IN MATHEMATICS

by

S.KANAGESWARI d/o SUPPIAH SHANMUGAM

Thesis submitted in fulfillment of the requirements
for the degree of
Doctor of Philisophy

December 2010

MENYIASAT KESAHAN PENGGUNAAN SATU UJIAN DWIBAHASA UNTUK MENAKSIR PENCAPAIAN PELAJAR TINGKATAN DUA DALAM MATEMATIK

oleh

S.KANAGESWARI a/p SUPPIAH SHANMUGAM

Tesis yang diserahkan untuk

memenuhi keperluan bagi

Ijazah Doktor Falsafah

Disember 2010

ACKNOWLEDGMENTS

I would like to record my heartfelt thanks and gratitude to my supervisor, Dr. Ong Saw Lan for her invaluable intellectual input, her persistent encouragement and continuous guidance. My sincere appreciation also goes to my co-supervisor Associate Professor Dr. Zurida Ismail for the academic assistance and support rendered. I would also like to thank my family especially my mother who has been the pillar of support throughout the duration of completing my thesis and my sister who consistently uplifted my spirit during the most challenging moment of my life. Without them, this thesis is not possible and I would like to dedicate this thesis to my father who is present spiritually for it is his dreams that motivated me to pursue this academic challenge in life.

"Dad, this is for you"

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

APA	American Psychological Association
AERA	American Educational Research Association
LEP	Limited English Proficiency
NCME	National Council on Measurement in Education
MCQ	Multiple Choice Questions
CFA	Confirmatory Factor Analysis
EFA	Exploratory Factor Analysis
IRT	Item Response Theory
CTT	Classical Test Theory
DIF	Differential Item Functioning

ABSTRAK

MENYIASAT KESAHAN PENGGUNAAN SATU UJIAN DWIBAHASA UNTUK MENAKSIR PENCAPAIAN PELAJAR TINGKATAN DUA DALAM MATEMATIK

Tujuan kajian ini adalah untuk membanding kesetaraan konstruk antara ujian Bahasa Inggeris dengan ujian dwibahasa dengan menggunakan kaedah dimensionaliti dan keberbezaan fungsi item (KFI). Di samping itu, kajian ini juga mengkaji perbezaan pencapaian pelajar yang mahir dalam penguasaan Bahasa Inggeris dengan pelajar yang kurang mahir dalam Bahasa Inggeris (PKM) dalam Matematik dari aspek komponen pengiraan dan masalah berayat. Dengan menggunakan pentadbiran berlingkar, semua pelajar dari enam kelas yang telah dipilih mengambil ujian secara serentak. Guru Bahasa Inggeris menggunakan kepakaran mereka untuk mengklasifikasikan pelajar sebagai PKM atau bukan-PKM. Dapatan kajian ini juga membuktikan bahawa kedua-dua ujian adalah setara dari segi item ujian dan statistik ujian itu sendiri, walaupun terdapat item-item keberbezaan fungsi item. Analisis KFI bagi kedua-dua ujian berdasarkan Teori Ujian Klasik, Teori Item Respons dan Model Multidimensional mengesan satu item KFI sepunya yang merupakan item masalah berayat yang memihak kepada pelajar yang menjawab dalam ujian dwibahasa. Tiada item KFI dikesan bagi ujian dwibahasa di kalangan PKM dengan bukan PKM. Dapatan kajian menunjukkan bahawa PKM menunjukkan pencapaian Matematik yang lebih tinggi (min=17.97) dalam ujian dwibahasa apabila dibandingkan dengan ujian Bahasa Inggeris (min=17.59). Namum pencapaian Matematik PKM dalam komponen pengiraan adalah sama untuk kedua-dua ujian (min= 10.15) manakala untuk komponen masalah berayat,

PKM menunjukkan pencapaian Matematik yang lebih tinggi (min=7.82) dalam ujian dwibahasa apabila dibandingkan dengan ujian Bahasa Inggeris (min=7.44). Dapatan kajian ini juga menunjukkan bahawa ujian dwibahasa membantu pelajar-pelajar yang kurang mahir dalam penguasaan Bahasa Inggeris tetapi, ia kurang membantu pelajar-pelajar yang berasal dari Sekolah Rendah Jenis Kebangsaan Cina kerana bahasa pengantar tidak sepadan dengan bahasa pengujian. Walau bagaimanapun, ujian dwibahasa membantu menyemak pemahaman pelajar. Satu lagi dapatan yang menarik ialah PKM masih memilih ujian Bahasa Inggeris kerana bahasa pengantar di kelas padan dengan bahasa pengujian. Dengan itu, dapat disimpulkan bahawa memudahkan struktur Bahasa Inggeris merupakan strategi yang lebih berkesan berbanding dengan ujian dwibahasa. Signifikasi kajian ini ialah ia menerokai ujian dwibahasa sebagai satu ujian akomodasi yang dapat mengukur secara sah pencapaian Matematik pelajar khasnya pelajar PKM bagi menggalakkan kesamarataan dan keadilan semasa menguji pelajar PKM. Implikasi kajian ini ialah ujian Matematik perlu mengurangkan penggunaan dengan meningkatkan penggunaan gambarajah. Kajian ini mencadangkan yang kajian pada masa hadapan perlu meninjau penggunaan ujian Bahasa Melayu-Bahasa Cina dan ujian Bahasa Melayu-Bahasa Tamil bagi membantu pelajar multilingual Malaysia.

ABSTRACT

EXAMINING THE VALIDITY OF USING A BILINGUAL TEST TO ASSESS FORM TWO STUDENTS' ACHIEVEMENT IN MATHEMATICS

The aim of this study is to compare the construct equivalence of the English-only test and the bilingual test by using dimensionality and differential item functioning methods. In addition, this study also examines the differences in the mathematical achievement between the limited English proficiency students' (LEP) and the non-LEP students in both tests for computation and word-problem testlets. By using spiral administration, all the LEP students in the six classes that had been selected sat for the tests simultaneously. The English teachers used their expert judgment to make the LEP and the non-LEP classification for their own students. The findings revealed that the two tests are equivalent at the item level and test level even though there are DIF items. DIF analyses conducted between the two tests based on the Classical Test Theory, the Item Response Theory and the Multidimensional Model detected one common DIF item which is a word problem item that favours the students who answered in the bilingual test. No DIF items were detected between the LEP and non-LEP students for the bilingual test. The findings revealed that that the LEP students' mathematical achievement is higher (mean=17.97) in the bilingual test when compared to the English-only test (mean=17.59). For the computation testlet, they performed equally for both tests (mean=10.15) while for the word problem testlet, LEP students did better in the bilingual test (mean=7.82) when compared to the English-only test (mean=7.44). The results suggest that the bilingual test even though generally helped the LEP students, it

did not render much help to the LEP students who had received their primary education in the Chinese-type schools as the language of instruction did not match the language of assessment. However, the bilingual test served the purpose of a 'comprehension check' for all the students. Another interesting finding is that the LEP students still preferred the English-only test since the language of instruction matched the language of assessment. As such, simplifying the structure of the English language may be more promising than the bilingual test. The significance of this study is that it addressed bilingual testing as a test accommodation that can validly measure students' mathematical achievement particularly to promote equity and fairness in testing the LEP students. The implication of this study is that the Mathematics assessment should reduce the language load embedded in items by using diagrams. This study recommends future studies to explore the usefulness of the Malay-Chinese and Malay-Tamil tests to cater to the multilingual Malaysian students.

CHAPTER 1

INTRODUCTION

1.1 Research Background

The English language is an important mechanism that facilitates the acquisition of knowledge especially in the fields of Science and Technology (Ainan Abdul Samad, 2003; Ain Nadzimah & Chan, 2003). A good command of the English language will enable and assist Malaysian students to use the internet as an access point for acquiring articles which are commonly published in English (Ainan Abdul Samad, 2003; Pandian & Ramiah, 2004; Ministry of Education, 2004; Yoong, 2005). As the future of the world rests in the ever-evolving avalanche of new mathematical, technological and scientific knowledge that is pivotal for new innovative breakthroughs and since these texts are in English, a new policy that reshuffled the language policy of the Malaysian national education system was announced in 2002.

The English-Medium Instruction Policy which was enacted on 19 July 2002, states that as of 2003, the English language would be the language of instruction in the teaching of Mathematics and Science subjects for the Year One, Form One and Lower Six students while the languages of assessment would be in the Malay language and English language (Surat Pekeliling Ikhtisas Bil 11-2002, 2002). The bilingual assessment would be administered to cushion the sudden implementation of English as the language of instruction, (Ainan Abdul Samad, 2003; Ministry of Education, 2004) as there is the perennial issue of questionable levels of language proficiency among

Malaysian students since English is learnt and not acquired (Ain Nadzimah & Chan, 2003).

In a study commissioned by Dewan Bahasa dan Pustaka in 2005 in the wake of this new language policy, students expressed their grave concern in understanding the lessons conducted in English due to their poor grasp of the language and were worried that they would not be able to answer the exam questions in English ("Poor English impedes lessons", 2006a). Similar sentiments were also shared by the teachers who also believed that the students would be able to perform better if the Malay language was retained as the medium of instruction ("Fewer answer Maths and Science in English", 2006c).

Since all assessments measure language skills (American Educational Research Association [AERA], American Psychological Association [APA] & National Council on Measurement in Education [NCME], 1999) language is instrumental in any assessment and for limited English proficiency (LEP) students, when testing is in English, language becomes a measure of that test construct (August & McArthur, 1994). Thus, the test scores may not be an accurate measurement of their ability and as this infidelity compromises with reliability, validity and fairness - the fundamentals of any assessment (AERA et al., 1999; Educational Testing Service [ETS], 2002), assessing them both in English and their native language may be necessary so that their limited English proficiency would not interfere with their performance (AERA et al., 1999). Standard 2.3.11 of International Test Commission [ITC] Test Adaptations Guidelines dictates that the students' language proficiency ought to be determined so that they are fairly tested in the language that they are more fluent in or another

alternative is to administer bilingual assessment (Coyne, 2000). Bilingual test gives provision for the original test items to be translated into the students' dominant language that they are more likely to be proficient in and as such, it removes the unnecessary language barrier (AERA et al., 1999).

Therefore, this research, addressed the validity of using bilingual test to assess Malaysian students' mathematical achievement especially the LEP students whose language of instruction for Mathematics is English.

1.2 Research Rationale

The importance of translated tests cannot be undermined as there will always be a continuous need for tests to be frequently exchanged at the international arena to assess the linguistically diversified population in order to make comparisons across multiple languages and cultures (Sireci, 1999). Gierl (2000) reiterated that the data from the Third International Mathematics and Science Study (from 2003 it was known as the Trends in Mathematics and Science Survey) (TIMSS) 1995 which were administered in 42 different languages were extremely beneficial in making comparisons on the students' performance across 60 countries and it provided crucial information on the effectiveness of the educational policies adopted in those countries.

However, a crucial point worth taking note of is that language dissimilarity is a liability in translated tests because translation does not ensure equivalent test forms as differences can be observed in the test content, item difficulty level, test reliability and test validity (AERA et al., 1999). These discrepancies can result in construct non-equivalence as the construct under measure in one test cannot be generalised to the

construct of a test in another different language and from the psychometric point of view, construct non-equivalence impinges the commonly observed statistical results like group mean levels and correlation (Robie & Ryan, 1996).

In addition, items may function differently for different groups of students with the same ability, resulting in differential item functioning (DIF) as the probability of producing a correct response varies for them for any item under study (Maranon, Garcia & Costas, 1997; Mazor, Clauser & Hambleton, 1994). This is because as Shealy and Stout (1993) argued, abilities other than the target ability that the test intends to measure are likely to influence students' test performance and is reflected in their score. This ability that the test does not intend to measure is termed as a nuisance determinant and violates test validity. This is because the main construct of the test that is intended to be measured which is defined as the primary dimension is compromised. DIF occurs when a secondary dimension is added to the test construct. This secondary dimension can be considered as auxiliary if it is intentionally assessed to be a part of the test construct or a nuisance if it is an unintentional construct. An intentionally assessed construct produces benign DIF in contrast to adverse DIF that is produced by the unintended construct that should not but unfortunately is measured by the test.

DIF items have been detected across different language test forms and thus translating tests items raises issues related to test validity as it does not ensure psychometric equivalence between the different test languages (Gierl & Khaliq, 2001). Language translation is one of the many steps that must be executed when adapting, after which the two tests must show evidence of construct equivalence and the test content remains unchanged. By conducting DIF analyses that can detect DIF items that

function differently for students with the same ability, equivalence between the items of different language versions can be established as items that are flawed by translation or other sources can be identified (Allalouf, Hambleton & Sireci, 1999).

According to AERA et al. (1999), responses that are composed of mixed languages affect test validity. Therefore, when considering fairness and validity in scoring the responses for the same item in different language versions, the equivalence of the words used in the responses need to be determined before scores of students from various linguistic backgrounds can be used interchangeably and before comparisons can be made (Robin, Sireci & Hambleton, 2003). This is because many words have different word frequency rates and difficulty levels in different languages that influence the meanings as even words that carry the closest meaning to another word in another language will still elicit a different meaning which in return contributes to psychometric non-equivalence (AERA et al., 1999).

Using one particular language or a mixture of two languages to answer a question may not elicit the same meaning or create the same effect across languages due to fluidity of meanings, different cultural interpretations or familiarity of words. There is no direct mapping of words in different languages because languages are bound by their own grammatical and syntax rules (Pena, 2007). Harkness and Schoua-Glusberg (1998) affirmed that language is not isomorphic and that meaning will be lost or altered when switching between languages. Therefore, this raises doubts on whether students' responses obtained using predominantly one language is psychometrically equivalent to the response composed of a blend of two different languages for the same item.

In a broader scale, for all the three rounds of TIMSS 1995, 1999 and 2003, Singapore was ranked first. A staggering 93% of the Singaporean students were in the international top half of the TIMSS 1999, even though only 27% of them used the language of assessment at home which is English in contrast to the three-quarter of other participating countries with at least 80% of students who used the language of assessment at home (Dixon, 2005).

Moreover, five Asian countries that seized the top five places were Singapore, the Republic of Korea, Chinese Taipei, Hong Kong SAR and Japan (Mullis, Martin & Diaconu, 2004). A fine-grained analysis revealed that the language of assessment for Mathematics in all these five Asian countries matched the language of instruction. Only Singapore used English while the other four countries used their respective mother tongue (Mullis, Kelly & Haley, 1996).

On the contrary, South Africa which adopted English as the language of instruction but was assessed in its mother tongue, was one of the very poor performing nations for Mathematics in the 1995 and 1999 TIMSS (Mullis et al., 2004). Does this point to the direction that a match between the language of instruction and the language of assessment is vital in producing better results or should the tests be administered in the test takers more proficient language as advocated by AERA et al. (1999)?

Among Malaysian students, which language is a valid measure of their abilities and is the appropriate language of assessment that will not interfere with their mathematical achievement? These questions need serious consideration as Mathematics assessment with linguistic complications is capable of obscuring the mathematical ability of even students whose first language is English (Hargreaves, 1997), what more

Malaysian students whose first language is not English since they come from a Malay speaking environment.

1.3 Problem of Statement

In Malaysia, the majority of students whose first language is not English have limited English proficiency as English is a learnt language and teaching students who are already weak in Mathematics in a language that is foreign to them, only aggravates the difficulty they face in understanding these subjects (Ain Nadzimah & Chan, 2003). This is because in countries where the native language is English, failure to perform in English will contribute to mathematical deficiency as they are functioning in an English-speaking society. But in countries where English is not the native language, performing in English will only intensify students' mathematical deficiency (Callosum, 2005).

In the same vein, there is also the issue of practicality in using English as a medium of instruction to benefit Malaysian students since LEP students form the majority group of the Malaysian student population and they were already finding Mathematics difficult even when Malay was the language of instruction (Yoong, 2005). The National Primary School Examination of 2006 indicated that 70 % of the candidates were weak in English and yet these students in Form One were required to learn Mathematics in English in 2007 (Nor Hashimah Jalaluddin, 2007).

At this juncture, it is important to clearly make a distinction between limited English proficiency students and English Language Learners (ELL). The term LEP should not be confused with ELL. LEP denotes a deficit in English proficiency and LEP

students are non-native speakers of English while ELL carries a positive connotation of students with second language mastery (Abedi & Hejri, 2004) and they are better at conversational English when compared to academic English language that is required in school (Francis, Rivera, Lesaux, Kieffer & Rivera, 2006). In countries where English is the native language, majority are either native English speakers or proficient English speakers who may or may not have mastered the academic English and the minority group of LEP students consisting mainly of immigrants (Pena, 2007). This is quite different with the situation like in Malaysia as among Malaysians, there is a higher density of LEP students than non-LEP students and the non-LEP students may or may not have mastered the academic English. Therefore, using English as the language of assessment has the potential to affect both the LEP and non-LEP Malaysian students even though the former is more linguistically challenged.

Assessing LEP students in English poses great threat as their learning and their performance in any assessment may be inhibited by their weak foundation in English (Abedi, 2006; AERA et al., 1999). This is because the linguistic complexity which is inherent in the test items but is unrelated to the content being assessed, is capable of camouflaging students' true score in any subjects. Students who lack the required proficiency in the language of assessment may perform poorly in that subject (Abedi, 2006; Abedi, Courtney, Leon, Kao & Azzam, 2006). It is therefore, unfair to allow language proficiency to cloud students' performance in any subject and as such, adequate opportunity must be provided to ensure that they are able to fairly demonstrate their performance in Mathematics assessments (Abedi et al., 2006).

Mathematics is particularly very challenging to LEP students as it demands the integration of linguistic knowledge, conceptual knowledge and procedural knowledge. Linguistic knowledge is related to English proficiency while conceptual knowledge is based on the understanding of mathematical concepts that helps to select the correct operations that will direct students to successfully perform the necessary algorithm to reach the final correct answer. Procedural knowledge is the different methods of teaching Mathematics or the different approaches of learning Mathematics that are set by different cultures (Virginia Department of Education, 2004).

In Malaysia, students lacked the level of English proficiency required to fairly demonstrate their mathematical skills especially in word problems as English is their second or third language (Fatimah Hashim & Zarina Ramlan, 2004). This is because students with restricted English language background are subjected to enormous language impediments that threaten their ability to learn and perform effectively especially when English is the language of instruction (Wang, Reynolds & Walberg, 1991). In addition to LEP, non-proficiency in academic English also contributes to students' poor performance in Mathematics especially in word problem items (Brown, 2005).

The release of the 2005 Lower Secondary Assessment or *Penilaian Menengah Rendah* (PMR) results for the first batch of students who had successfully completed a full cycle of studying Mathematics and Science fully in English since Form One, indicated that the passing percentage for Mathematics was 84.1% while English recorded a passing percentage of 73.8%. An issue that raises concern is that more than a quarter (26.2%) failed in English and the passing percentage continued to drop in 2006

to 71.4%. A similar trend was also observed in Mathematics that recorded a drop by 1.3% to 82.8% ("*70% students enjoy learning in English"*, 2006b).

These figures indicate that the performance of students who had received instructions for Mathematics in English seemed to have deteriorated for the English subject over the two year period of 2005 and 2006 instead of improving. Since high academic performance in the English language corresponds to high language proficiency (Nor Azmi Mostafa, 2002), during this duration the continuous drop of the passing percentage for the English subject suggested a decline in the standard of English at a time when teaching and learning of Mathematics was done in English.

Besides that, in this first cohort of the 2005 PMR students only 27% of them answered the questions in English with majority preferring either Malay or a mixture of both languages. However, the candidates who answered in Malay still used the mathematical terms in English as they were more familiar with these terminologies that they had learnt since Form One ("Students still prefer to use Bahasa", 2005). Even though more students of the second batch in 2006 were expected to answer in English, the outcome showed a percentage drop of 2.9% to 24.1% in 2006. Despite being the second batch of students to sit for PMR after studying these two subjects fully in English for three years, 46.2% of the candidates still answered in Malay while 29.7% answered in both languages ("70% students enjoy learning in English", 2006b). Even though this policy was in the second year of implementation, less than a quarter chose to answer fully in English with majority still preferring to answer in the Malay language.

The irony is that despite English being the language of instruction, students still preferred to answer in Malay and despite answering in Malay, they still preferred to use the mathematical terms in English since English was the language of instruction. The issue related to testing is, even though the choice for the language of assessment for Mathematics is between English which is the language of instruction and assessment, and Malay which is the national language and the dominant language of instruction for subjects other than Mathematics, Science, and Technology, can Malaysian students handle English as the languages of instruction and assessment during their Mathematics lessons and respond confidently to the mathematical test items?

This is because it takes between four to seven years to develop the academic English proficiency (Hakuta, Butler & Witt, 2000) and assessing students whose language of instruction was in either Malay, Chinese or Tamil at the primary level (Ong, 2007) suggest that too few years were spent on acquiring the academic English language and therefore, this short duration is not sufficient to fully develop their language proficiency.

Even though the Malay language is retained as one of the languages of assessment, English as the language of instruction may likely interfere with the Malaysian LEP students' mathematical achievement. This is because English has already been adopted as the language of instruction and translating test items from English to other languages may not benefit LEP students (AERA et al., 1999). This is a potential threat to test validity as translating items may confuse students who have learnt these concepts in English and they may not be familiar with the translated terminologies (Abedi, 2006). Therefore, regardless of whether Malay or English

language is enforced as the language of assessment, Malaysian students' limited English proficiency may likely influence their mathematical achievement.

According to AERA et al. (1999), assessing students in their dominant language gives accurate measure of their true ability as they are not impeded by their linguistic disability. Since for Mathematics assessment, language should not interfere with the construct that is intended to be measured which is the mathematical ability, skill or knowledge, this raises an important validity issue which is how accurately do the Mathematics scores reflect students' true mathematical achievement since they are assessed in a language that they have difficulty in understanding? As Bracken and McCallum (1999) pointed out, assessing only the cognitive abilities of students with linguistics impediments have always been a perennial complication in testing. When two constructs like language and mathematical ability are so closely related, how much of the Mathematics test scores is a valid measure of students' mathematical ability and how much of that composite score is due to the language ability?

Furthermore in a multilingual nation like Malaysia, there are three different types of primary schools that emphasise the use of the Malay language, the Chinese language and the Tamil language as the languages of instruction (Ong, 2007) with modest exposure to the academic English language. Therefore, is it valid to test students from varied linguistic background using only one dominant language? Can the bilingual test help to remove the language barrier among Malaysian LEP students and validly measure students' mathematical achievement and thus, help improve LEP students' mathematical achievement? Can the bilingual test be the answer to assess Malaysian

LEP students from a multilingual background by ensuring equity and fairness in testing for them or will it violate test validity and thus, be detrimental to them?

Another potential threat to test validity is the psychometric equivalence between the original test items and the adapted test items (Sireci, 1999). According to Standard D6 by ITC Test Adaptations Guidelines, when a test is translated to different languages, the equivalence of the test versions must be ascertained (Coyne, 2000). Many studies that have been conducted on the equivalence between the original test and its translated version testify that translating a test from one language to another does not guarantee that the adapted version is equivalent to the original test (Geisinger, 1994; Sireci, 1999; Van der Vijver & Hambleton, 1996; Van der Vijver & Poortinga, 1997).

Kester and Pena (2002) elucidated that the psychometric properties of one test is neither shifted to the translated test nor remain unchanged when administrated to a sample other than the intended sample. The psychological processes and constructs of the original tests too may not flow over to the translated test and therefore the two assessment instruments may not be equivalent. In addition, translating or adapting does not ensure psychometric equivalence in terms of content, difficulty level or discrimination level as the meanings of words fluctuate across languages (AERA et al., 1999).

To conclude, putting all these issues into context raises the question of is the bilingual test a valid test accommodation that can measure Malaysian students' mathematical achievement particularly the LEP students? Moreover, will the language of assessment influence students' mathematical achievement especially for word problem items? Are they able to competently answer the word problem items that are

linguistically dense and computation items with minimum linguistic load without any interference from their language skills? Would the different languages of assessment influence students' responses and the psychometric properties of the test items across the different language forms in a way that threatens test validity?

The issues here are, the two different language test booklets should be equivalent so that the scores can be used interchangeably and in assessing Malaysian students' mathematical achievement, their language proficiency should not influence their mathematical skills.

1.4 Research Aim

The main aim of this research is to examine the validity of using bilingual test to assess Malaysian students' mathematical achievement particularly the LEP students.

Therefore the initial part of this study is to establish equivalence for both language versions of the test. Even though the focus is on construct equivalence of the two language tests, psychometric equivalence (also known as statistics or metric equivalence), translation quality, semantic equivalence and cultural relevance will also be examined. The items in the English and Malay languages will be examined for translation quality, semantic equivalence and cultural relevance during the question developmental stage while the construct equivalence and psychometric equivalence will be established during the data analysis phase.

In addition, this research also aims at examining Malaysian learners' mathematical achievement using bilingual test as a test accommodation to determine whether LEP students will particularly benefit from this test accommodation.

1.5 Research Objectives

This research is focused on examining the validity of assessing Malaysian learners' mathematical achievement by using bilingual test as a test accommodation. The objectives of this research are to

1) compare the construct equivalence of the English-only test and the bilingual test by using dimensionality and DIF methods.
2) compare the mathematical achievement between LEP students and non-LEP students in the English-only test.
3) compare the mathematical achievement between LEP students and non-LEP students in the bilingual test.
4) compare the LEP students' mathematical achievement for word problem and computation testlets in the English-only test.
5) compare the LEP students' mathematical achievement for word problem and computation testlets in the bilingual test.
6) compare the LEP students' mathematical achievement in the English-only test and in the bilingual test.
7) compare the non-LEP students' mathematical achievement in the English-only test and in the bilingual test.

1.6 Research Questions

This research will answer several important questions which are

1) Do the English-only test and the bilingual test show comparable construct equivalence?

2) Is there any difference in the mathematical achievement between LEP students and non-LEP students in the English-only test?

3) Is there any difference in the mathematical achievement between LEP students and non-LEP students in the bilingual test?

4) Is there any difference in the mathematical achievement of LEP students for the word problem and computation testlets in the English-only test?

5) Is there any difference in the mathematical achievement of LEP students for the word problem and computation testlets in the bilingual test?

6) Is there any difference in the mathematical achievement of LEP students in the English-only test and in the bilingual test?

7) Is there any difference in the mathematical achievement of non-LEP students in the English-only test and in the bilingual test?

1.7 Research Significance

From a global perspective, international study functions as a 'national report card' (Clarke & Suri, 2003) that transcends the boundaries of physical borders of regions, languages and cultures. According to Hambleton (2002), international studies provide vital information, suggestions and plausible explanations on the quality of education in the participating countries in order to improve their standard of education.

International studies like the TIMSS use tests that have been adapted into different languages so that students' performance in Mathematics and Science can be compared across the participating countries. Beyond using the students' achievement to rank the participating countries, it also provides a wealth of information that addresses

the weaknesses or strengths of the educational policies that have been implemented so that a more effective educational philosophies can be designed in the future (Mullis, Martin, Ruddock et. al., 2005).

In the Malaysian context, the research uncovers the validity issues that surround international testing like the TIMSS administered in Malaysia mainly, the comparability of testing Malaysian students in the English language to the translated items in Malay and the effects of a mismatch between the language of assessment and the language of instruction on test scores. Another issue is the comparability of testing students in the English to the bilingual format and the effects of consequential validity on students.

This research takes heed of Thurlow et al. (2000) calling to focus on test accommodations that involve the largest number of students and of current concern. This research was triggered by the bilingual testing in Malaysia where the bilingual assessment is the widely and most recently adopted test accommodation, targeting all LEP and non-LEP Malaysian students. Therefore, this research is of great significance to the Malaysian education system as functioning in multiethnic, multicultural and multilingual society where there are three types of schools that adopt the Malay language, Chinese language and Tamil language as the mediums of instructions. The research gathered the true picture on the effectiveness of employing the bilingual test as a way to accurately assess students' mathematical ability Being in a society where English is not the native language, the largest student population comprises LEP students and thus, the findings of this study would provide great insights on the

appropriateness and utility of this newly adopted linguistic test accommodation as being either beneficial or even detrimental to the Malaysian LEP students in particular.

The National Association of the Education of Young Children (NAEYC) (2005) recommended that students who show proficiency in both the home language and the English language be assessed in both languages. Therefore in the Malaysian context, this research addressed the extent to which the bilingual test can be an effective test accommodation which is able to validly measure the general students' mathematical achievement particularly to alleviate LEP students' linguistics complication. This is because there is the possibility that the bilingual test may benefit the general students or bring adverse effects on them or benefit one group but be detrimental to another and if so, which group of students. The findings are vital as it would contribute to the issue of test accommodation for the Malaysian education as it would help to decide whether the bilingual test may it be English-Malay, Malay-Chinese or Malay-Tamil, is a cost-effective investment that benefits Malaysian students or cost-ineffective test accommodation that exhausts financial resources and does not benefit the students. As such, it will help to identify other forms of test accommodation that may benefit Malaysian students.

This research also uncovered the utility of using bilingual test on the computation testlet and word problem testlet. Thus, addressed the validity of using a bilingual test within the sub-constructs of Mathematics which are the word problem and computation testlets among Malaysian students in general and LEP students in particular.

As poor test adaptation can change test validity, this research is of significance as it investigated the equivalence of an English-only test and a bilingual test. This research hoped to create awareness among test developers and educators on the need to examine beyond translation equivalence and to consider too the psychometric properties in order to determine the equivalence of the original test with the adapted form.

In addition, whether in the international or national assessments, the test scores provide the grounds to make vital decisions about either the participating countries' or students' achievement. Therefore, this research hoped to highlight the extent to which such decisions are accurate and valid since they are based on tests of different languages.

To conclude, since this is a recently adopted assessment procedure in Malaysia, there isn't much research on this new form of testing in the Malaysian context. Therefore, the findings would be of great significance to the Malaysian education system that may point to whether the bilingual test is a reliable test accommodation that can validly assess students who are non-native speakers of English. The research would bring to light the validity and utility of using bilingual test as a test accommodation to assess Malaysian students' achievement in Mathematics for the computation testlet and word problem testlet.

1.8 Research Limitations and Delimitations

The main limitation of this research is in the selection of schools as the research sample. Only schools where the teachers had completed all the topics assessed in this test which is until Chapter 14 of the From Two syllabus before the first week of October

were selected. This is because the test content covers items from the Form One and Form Two syllabi. The test could not be administered any time later than October as it would coincide with the school examination schedule. In all the schools, only six classes that comprised the high, moderate and low ability students were selected. However, in schools that had less than six Form Two classes, all the classes were selected. Besides, the researcher proposed to randomly select two classes from each ability category. However, due to practical considerations where the schools' administrators approved the use of specific classes, the researcher had to abide by the conditions and administer to the classes that had been selected by the school.

In addition, the computation and word problem mathematical items that were used in this research are multiple choice questions (MCQ). The study only uses MCQ and did not use other types of test items like the constructed-response format. Students' mathematical achievement in short-response items and extended-response items for computation and word problem testlet are not studied. This is because extended-response items pose more challenges to LEP students as they involve a higher degree of reading, thinking and writing skills (Schulte, Elliot & Kratochwill, 2001) due to the intense linguistic density and therefore, can form a new study of its own.

This research utilised teachers for the test translation and test adaptation procedures and not professional translators as recommended by American Translators' Association (2003). Even though Cheung and Cheung (2003) acknowledged bilingual individuals who are proficient in the source and the target languages as exhibiting high competency, American Translators' Association (2003) states that professional translators cannot be substituted for bilinguals as bilinguals are fluent speakers of two

languages who may not have received adequate training in test translation in the written mode. In view of the resources constraints facing this research like manpower and cost, this study used experienced Mathematics teachers who possessed knowledge on assessment literacy.

This research did not gather any information on the students' language background, particularly the language of instruction used in the primary school and the vernacular school-type that they had attended which can be Malay, Chinese or Tamil schools. This information is vital to see whether the Malay-English bilingual test is useful to all LEP students or is only helpful to a particular group of students determined by the language of instruction at the primary levels and if so, which LEP group will benefit.

Even though the researcher took extra steps of holding briefings, and a one-to-one session with teachers to help the English teachers with the LEP designation, there is no objective way of checking whether there was any misclassification.

Besides, the number of items that were covered in this study is only 40. The number of questions that was included is limited due to the practical constraint of time needed to administer the tests. Within an hour, approximately 40 MCQ can be administered. This study does not intend to be over-ambitious as this is a novice attempt in examining Malaysian learners' achievement in Mathematics using bilingual test as a test accommodation since the language policy came to effect. In addition, this study does not want to consume too much of the students' time and disrupt the already planned school programmes by administering long tests. A test with a sufficient number of items that does not compromise validity is the modest intent of this research. In

addition, the sample covered was only students from the states of Penang and Northern Perak due to time constraint. Even though the researcher tried to cover all the zones, this sample is still relatively small and is confined to only a few randomly selected schools.

1.9 Operational Definitions

Limited English Proficiency

A student who has been rated by his English teacher using his expert judgment based on his first-hand experience in dealing with the student for approximately 10 months by grading his student's English oral skills for at least two assessments as required by the School Based English Oral Assessment. In addition, the teacher has scored his written English in at least three school based assessments. A description band (Appendix K) is provided as a guide.

Test Adaptation

Items in the English language are translated into the Malay language and adaptations are made to the names of people and currency to suit the Malaysian culture.

Bilinguals

Individuals who have undergone their formal education using Bahasa Malaysia and English as the languages of instructions.

Malay Speaking Environment

A social setting where Bahasa Malaysia is used as a language of communication due to its revered status as the national language.

Language Proficiency

Focuses on students' demonstration on the usage of language function (communicative skills) in addition to the written language form that is demonstrated by a paper-pencil classroom test.

Construct Equivalence

The degree to which the constructs measured by one test are the same with the constructs measured by another test (AERA et. al, 1999).

Word Problem Item

Word problem item is set in a context and is relatively more loaded with language which may require multiple steps to solve when compared to a computation item.

Computation Item

Computation item involves direct application of mathematical operation that uses either a combination of addition, subtraction, multiplication or division where the item may be presented in the form of language, numbers or operational symbols and is not set in a context.

Semantic Equivalence

The item in the English language and the Malay language carry similar meaning.

Psychometric Equivalence

The bilingual test and the English-only test have similar mean values of item difficulty, item discrimination, point-biserial and the KR 20 index.

Cultural Relevance

The items are appropriate for the Malaysian culture.

1.10 Conclusion

All students despite their personal characteristics should have equal access to demonstrate their ability on the constructs that the test intends to measure. Since all assessments are also measures of language skills, students' limited English proficiency (LEP) may threaten test validity as their mathematical achievement may be masked by their inept English language proficiency. Using bilingual test may be a viable alternative in arresting the issues of unfairness among LEP students who are handicapped linguistically but at the same time, should not provide unnecessary advantage that erroneously benefits the non-LEP students by inappropriately raising their mathematical achievement.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter begins with the definition of test accommodation, test modification, test adaptation and test translation and literature on the researches that have been conducted on the different types of test accommodations for LEP students. The studies that will be of focus are confined to linguistics accommodations related to Mathematics tests, particularly native language assessment and bilingual assessment. The important types of validity related to this study which are content validity, construct validity and consequential validity are also explored. Literature on test equivalence particularly psychometric equivalence which includes item difficulty, item discrimation and point-biserial index, and constructs equivalence are reviewed before moving to test linking and the IRT model. DIF is also discussed before scrutinising the roles of language in assessment and in Mathematics. Literature on word problem items and computation items are examined with a special focus on the problem model proposed by Kintsch and Greeno (Kintsch & Greeno, 1985). The next section highlights the differences in the mathematical achievement when the language of test varies before progressing to the definitions of language proficiency and the TIMSS study administered in Malaysia.

The final section studies the conceptual framework underlying this research which include the Multidimensional IRT model, Problem Model proposed by Kintsch and Greeno (Kintsch & Greeno, 1985) and Cummins' Basic Interpersonal Communicative Skills and Communicative Academic Language Proficiency (Cummins, 1984). The theoretical framework of this study is the Test Theory.