

Reading Skill among Malaysian ESL Lower Secondary Students: Which Girls and Which Boys are Achieving and Underachieving?

Kamal J I Badrasawi

*Kulliyah of Education,
International Islamic University Malaysia,
Kuala Lumpur, Malaysia
kamalbadrasawi@iium.edu.my*

***Noor Lide Abu Kassim**

*Kulliyah of Education,
International Islamic University Malaysia,
Kuala Lumpur, Malaysia
noorlide@iium.edu.my*

Abstract

The literature on educational achievement has shown consistently that boys are underachieving. They are disengaged with learning, and their dropout rates in schools are higher than those for girls. Although the problem of underachievement and disengagement with learning is largely associated with boys, not all boys are underachieving or disengaged with learning, and not all girls are achieving and engaged with learning. There is also strong evidence to suggest that differences within gender are more significant than the difference between gender. Recent research findings have shown that educational performance is highly influenced by socio-demographic factors such as school location, race or ethnicity, socio-economic status, and parents' education. Given that reading is a critical literacy skill for academic achievement and English is an important second language in Malaysia, this study sought to identify groups of Malaysian lower secondary students who are at risk of underachieving in English reading skill. A sample of 944 Malaysian ESL Form 1, 2 and 3 students, randomly selected from national-type schools, participated in the study. A test of English reading skill, consisting of 60 multiple-choice items was used. The Rasch Model analysis as well as selected descriptive statistics were used to answer the research questions. The results showed that students' performance in English reading differed from one group to another, implying that gender did not exclusively influence student performance. Based on the findings, more sound and informed decisions on students' performance in English reading skill and the most effective teaching methods can be made. Qualitative investigation of the factors behind high or low performance among these groups of students is also needed to further understand the influence of these factors on achievement and underachievement .

Keywords: *English reading skill, Rasch Model, demographic factors, underachievement*

INTRODUCTION

The struggle for gender equity has seen a substantial increase in females receiving equal opportunities in education. However, in the last two decades or so, a new disconcerting trend has emerged in many parts of the world. Boys are underachieving in educational attainment.

They are disengaged with learning and their retention rates in schools have dropped considerably. This is seen in Australia where girls are outperforming boys in a number of key subject areas, retention in schools, as well as competence in literacy skills (Collins, Kenway & McLeod, 2000; Hoyt, 2015; Scholes, 2019). A similar pattern is seen in other countries. In the United Kingdom, although there is little difference in performance in Mathematics and Science, girls consistently outperformed boys in literacy skills, particularly writing, at all key stages of schooling (ages, 7, 11, 14 and 16). It was also found that the gap in educational performance is also evident in the secondary schools (Younger et al., 2005). More recently, the gender gap in the United Kingdom is reported to persist (Miele, 2020; Ward & Thurnell-Read, 2019). In Mongolia, a similar trend is seen. Boys are lagging behind considerably in educational performance, particularly in secondary and tertiary education (UNESCO, 2004b).

Voyer and Voyer (2014) conducted a meta-analysis to find gender differences in school marks in elementary, junior/middle, high school and at the university level. They found that female students outperformed male students much more in language courses and less in math courses. Moderator factors were also found in relation to these differences. In Pakistan, Shoaib and Ullah (2019) found that female students outperformed male students in educational results on every level (secondary, higher secondary, graduate and postgraduate levels). They added that this result is similar to trends in other countries in the world, such as in Australia and the United Kingdom (Crawford, Wang, & Andrews, 2016), Denmark (Randolph; 2019), Hong Kong (Kuo, Casillas, & Allen, 2019), New Zealand (Frijters, Brown, & Greenberg, 2019) and Canada (Sammut, Kuruppu, Hegarty, & Bradbury Jones, 2019), Kenya (Okul, Sika, & Olel, 2019), Philippines (Dacuycuy & Dacuycuy, 2019), India (Cheruvalath, 2018; Joshi & Ahir, 2019), Ghana (Oketch, 2019), Cambodia (Rogers & Ruth, 2019), Vietnam (Huong, Tien, Hung, & Loc, 2019) and many others (see also Ullah & Ullah, 2019).

In the Malaysian context, a similar trend is seen. In schools, boys are underperforming in many subject areas, particularly in literacy skills (Ratnawati Mohd-Asraf, Hazlina Abdullah & Ainul Azmin Mat Zamin, 2016), and this continues to the tertiary level, where female students outnumber males (Latifah Ismail, 2015; Nachiappan, Veeran & Andi, 2012; Tienxhi, 2017). The Malaysia Education Blueprint (2013-2025) highlighted that the gender gap is both significant and increasing. Girls consistently outperform boys in almost all levels of education, including tertiary (Ministry of Education Malaysia, 2013). It also emphasized finding ways on how to counter boys' dropouts and make them engaged in education, to contribute to the development of the nation.

Studies on boys' and girls' underachievement and lack of school retention have identified a number of demographic, psychological and systemic factors that exert significant influence on underachievement and lack interest of learning (see Collins et al., 2000; Cortis & Newmarch, 2000; Younger et al., 2005; Hutchison, 2007; Lloyd, 2011; Yu, McLellan and Winter, 2020; Shoaib & Ullah, 2019). Ludicke, Muir and Karen (2019) identified literacy and numeracy barriers, family background factors, lack of engagement, absences, and confidence as key factors of underachievement. It is important to note that though the problem of underachievement and disengagement with learning is largely associated with boys, not all boys are underachieving or disengaged with learning, and not all girls are achieving and engaged with learning (see Collins et al., 2000; Lingard, Martino, Mills & Bahr, 2002). There is evidence

to suggest that differences within groups of boys and girls are more significant than the difference between gender (Collins et al., 2000; Hutchison, 2007; Yu, McLellan & Winter, 2020). In line with Yu, McLellan and Winter (2020) the adoption of “which boys and which girls” approach in addressing the gender gap in educational research is critical. Only through this approach can we determine which boys and which girls are most at risk academically.

To address the problem of underachievement and disengagement with learning, particularly with regard to boys, positive actions have been taken, particularly by developed countries. In Australia and the United Kingdom, a considerable amount of research and official inquiries at the national level have been conducted to empirically ascertain the factors that are associated with underachievement and lack of retention in school (see Collins et al., 2000; Cortis & Newmarch, 2000; Lloyd, 2011; Scholes, 2020). Based on the findings of these studies and inquiries, nation-wide intervention initiatives have been formulated to address this predicament. One important initiative is strengthening boys’ literacy skills. For example, Lloyd (2011) reported in a literature review, strategies that could help boys to be more engaged in education, taking into account the learning context where the influential factors on gender gap might be different. Lloyd quoted examples of these strategies from other studies; such as, whole school approaches (e.g. learning culture, boys ‘classes, behaviour management, tracking, mentoring etc.); in classrooms (e.g. teachers ‘approaches, learning styles, subject specific, alternative curricula etc.); and outside schools (community and family aspiration, father’s influence etc.) (Lloyd, 2011). It is important to add that, contrary to some opinions, all teachers, regardless of their gender, can contribute to the improvement of boys’ engagement and achievement as literacy learners (Watson, 2016).

Given the fact that English language has become the most important language in the world, most countries are improving the English proficiency of their people in general, and school-going children in particular. In Malaysia, English is an important second language and it is used as a medium of instruction in many higher institutions. With the globalization of the English language, it has become more important for Malaysian students at different stages of learning to master the language as mooted in the Malaysia Education Blueprint 2013-2025 (Ministry of Education, 2013). This is especially so for young learners as studies have shown that literacy is highly correlated with academic achievement in later years. Thus, it is important to monitor how much progress Malaysian young learners have or have not made in achieving the required levels of English language skills and identify those who are at risk of underachieving. This study also sought to collect reliable baseline data to form a complete picture related to school children’s performance in reading. Without such data, comparisons across cohorts cannot be effectively achieved, and effective intervention strategies may not be properly formulated.

For the purpose of the study, focus is given to performance in reading. This skill is chosen because it is considered as one of the foremost indicators of being literate (McGee & Richgels, 2000). Any new definition of literacy primarily includes reading skill as students who cannot read and write have difficulties in their studies (Holme, 2004). Specifically, reading skills in English as a second or foreign language is necessary for students’ academic success in their further education (Levine, Ferenz, & Revez, 2000).

The research questions that guide this study are as follows:

1. What are the levels of reading performance of Forms 1, 2 and 3 Malaysian students?
2. Which groups of students are underachieving based on gender, race, school location, SES, and parents' education level?
3. In which reading sub skills are the students achieving and underachieving?

Theoretical Framework

Like other language skills, the assessment of reading should be guided by a clear theory that defines reading skill and an appropriate measurement theory that reflects accurately that definition and its components (Engelhard, 2001). Such relation helps to measure and interpret students' performance in reading skill much more precisely (Alderson, 2000; Engelhard, 2001). With regard to reading skill in a second language, the prime concern is related to the nature of this skill, the identification of its sub skills, and whether these sub skills are attained hierarchically. It is argued that the way of testing or assessing reading is directly influenced by our view towards its nature (Alderson, 2000; Engelhard, 2001; Hedgcock & Ferries, 2009; Hudson, 2007). In short, "the test designers should be aware that their tests reflect their model of the nature of reading, and they should thus seek to ensure that they reflect and build upon what recent research suggests about the process and the product of reading" (Alderson, 2000).

Given the complexities of reading, considerable research has been conducted in L1 and L2 to identify its sub skills and any possible hierarchy of these sub skills, using both quantitative and qualitative methods (Alderson, 2000; Hedgcock & Ferries, 2009; Hudson, 2007; Urquhart & Weir, 1998; Weir & Porter, 1994). The findings of such researches are varied and support two main positions (Engelhard, 2001; Hedgcock & Ferries, 2009). Some researchers suggest that reading is a unitary skill that cannot be divided into identifiable sub skills (Alderson, 1990a; 1990b; Alderson & Lukmani, 1989; Lunzer, Waite & Dolan cited in Urquhart & Weir, 1998; Rost, 1993) while others found that reading is a multi-divisible skill that includes separable and identifiable sub skills (Farhady & Hessamy, 2005; Farhadi & Moeini, 2005; Hughes; 1989; Matthews, 1990; Munby, 1978; Sainsbury, Harison & Watts, 2006; Weir, Hughes & Porter; 1990). In addition, there is no consensus on the number of these sub skills (Alderson & Lukmani, 1989), and research has not consistently supported the notion of "strictly hierarchically ordered reading skills" (Hudson, 2007, p.103). Regardless of this debate, the multi divisibility view of reading is taken for the purpose of the study. Brown (2003) expounded that the skills used in reading appear to be essential consideration in the assessment of reading ability. The idea is mooted by Alderson (2000, p. 122),

For profiling purposes most models of reading make reference to numerous skills or sub processes that occur in reading. At the very least, therefore, students should be tested on a range of relevant skills or strategies with the result possibly being provided in diagnostic, profile-based format.

In this vein, a number of reading taxonomies, frameworks, scales and tests for reporting reading development are found in the literature. For example, Masters and Foster (1997a) developed scales to locate students onto a continuum achievement scale in terms of English literacy in general, and to report the differences of their literacy levels in terms of reading, writing, speaking, listening and viewing, in particular. In this study a similar approach is taken; seven reading sub skills are detailed based on the Malaysian school *English Syllabus* (2003) to identify what sub skills that students can use when answering items in a reading test.

For the measurement purpose, there is an emphasis on using a robust measurement model: namely the Rasch Measurement Model (RMM) theory owing to the fact that RMM has robust advantages over other measurement models, such as the Classical Test Theory (Bachman, 1990; Bond & Fox, 2015; Engelhard, 2001; Hambleton, Swaminathan & Rogers, 1991; Linacre, 2003; William, Patrick, Malcolm & Joseph, 2006). In essence, the Rasch Model is a latent measurement model which transforms raw scores into equal interval linear measures that are invariant of items and persons used in the calibration process. It estimates person ability and item difficulty independently, and calibrates them on the same equal-interval logit scale (Bond & Fox, 2015; 2007; Wright & Stone, 1979). With such logit scale, it is possible to precisely describe how persons differ from each other; that is, it has the ability to locate or display the levels of items difficulty and persons' ability on the same scale (Bond & Fox, 2015; Hambleton et al., 1991; Ingebo, 1997; Granger & Linacre, 2008; Wright & Stone, 1979).

Moreover, the Rasch Model has been used to monitor students' educational growth or progress over time, and to compare groups of students at different years or levels of schooling over time (Masters, 1993; Masters & Foster, 1997a, 1997b; Meiers, 2008; Mossenon cited in McNamara, 1996; Stephanou, Meiers & Foster, 2000; Rowe, 2006). McNamara (1996) demonstrates that visual maps can be produced because of the property of item invariance (independence of item estimates from person characteristics) in Rasch analysis. Such maps can show the progress of student performance on a latent construct, such as reading ability. One good example is the *Item-ability map*, where person ability estimates and items difficulty estimates are calibrated on the same scale. Persons are placed at a relevant position on one side of the scale and item difficulties on the other. This allows comparison of item difficulty and person ability on a given test.

Another example is the *skill-ability map*, where items representing certain skills or levels are calibrated with persons' abilities on the same scale. This map is used in the TORCH tests, '*Test of Reading Comprehension Skills*,' to measure and trace reading development in English Language in Australia (Mossenon cited in McNamara, 1996). Another example is the Australian Language Certificate (ALC) project, which involves tests of reading and listening in seven foreign languages taught in secondary schools. Masters and Forster (1997b) developed a set of achievement scales to report the differences in students' literacy levels in terms of reading, writing, speaking, listening and viewing. Each literacy scale represents a continuum of achievement and is divided into five levels with a sequence of literacy indicators. The most commonly observed behaviours are located at the bottom of each scale and vice versa.

METHODOLOGY

Research Design and Participants

This study used the descriptive research design which commonly involves particular research methodologies and procedures such as tests, surveys, observations, and self-reports (Gay & Airasian, 2000; Wallen & Fraenkel, 2006). Since the major concern of the study is to determine students' performance in English reading skill and identify groups who are at risk of underachieving in this language skill, it utilizes a dual-purpose survey which includes a questionnaire (to gain demographic information) and a test of English reading skill (to collect students' responses on the test items). According to Keeves (2004), in educational settings, tests which require written answers are typically used to measure subject matter achievement.

The population for the study are Forms 1, 2, and 3 students from national-type schools in two states in Malaysia: Wilayah Persekutuan Kuala Lumpur and Selangor. This population of school children was selected for the following reasons. First, the lower secondary level in Malaysia comes at the end of the primary level which lasts six years, and before the upper secondary level which lasts two years. Second, the lower secondary level is considered as the foundation level in which the development of reading and writing skills is necessary for performance in the future at the upper level schooling, tertiary education, and in their future career.

A representative sample was chosen from the population using the multi-stage random sampling procedures. Eleven secondary schools were selected from the aforesaid states. Specifically, five schools were randomly selected from each state, and a boys' school from Wilayah Persekutuan Kuala Lumpur was purposely selected. This is to allow for the examination of differences in performance within groups of boys. For each Form (i.e., grade level), 30 students were randomly selected from each school giving a total of 990 students (30 students x 3 forms per school x 11 schools) (Table 1). This large number of participants is necessary given the demands of the statistical procedures that will be utilized in the data analysis and the fairly large number of variables that will be included in the study. The large sample size is also essential to ensure that relevant subgroups are well-represented. If the sample size is large, it is more likely to represent the population (Wallen & Fraenkel, 2006).

Moreover, it is essential to highlight that the Rasch Model is designed to be sample-distribution independent; hence, the school effect would not be an issue. In addition, the sample included a range of students with different levels of English reading skill, randomly selected from classes with mixed abilities from different schools in urban and rural areas. In the Rasch model analysis, it is also possible to obtain useful results with small samples because Rasch analysis is not dependent on the sample size and it is robust to missing data (Bond & Fox 2015; Linacre, 1994; Granger & Linacre, 2008; Wright & Stone 1979). However, Granger and Linacre (2008) recommend that the most reliable interpretation comes from a sample with at least 50 - 100.

Table 1
Summary of Study Sample in KL and Selangor

Loc.	N%	Gender			Form			Race				
		M	F	Miss	F1	F2	F3	Malay	Chinese	Indian	Other	Miss
KL	497 52.6%	241	254	2	150	140	207	314	132	41	7	3
Kuala Selangor	447 47.4%	207	236	4	149	153	145	277	71	88	6	5
Total	944 100%	448 47.5%	490 51.9%	6 0.6%	299 31.7%	293 31%	352 37.3%	591 62.6%	203 21.5%	129 13.7%	13 1.4%	8 0.8%

Table 1 shows that the total number of students who participated in the study was 944, out of the 990 selected, due to the following reasons. First, in certain schools less than 90 students participated in the study on the test day. Second, 13 cases were deleted because they were considered as invalid. Seven students chose the same option for all items and the others only wrote their names and didn't attempt any item. Table 1 also shows that 497 (52.6%) students were from urban schools and 447 (47.4%) from rural schools; 448 (47.5%) were male students and 490 (51.9%) were females. The number of students for Form 1, 2 and 3 was 299 (31.7%), 239 (31%) and 352 (37.3%) respectively. The highest percentage was for Form 3 students. The number of Malay students was the largest (n= 591; 62.6%) compared to Chinese (n= 203; 21.5); Indians (n= 129; 13.7%) and others (13; 1.4%). It is important to mention that the numbers of students in both gender groups were adequate for comparison.

Research Instrument

The development of the instrument for this study was based on three Reading English language national standardized tests for Form 3 students. These tests usually include specific tasks and reading skills associated with English reading literacy that Form 3 students are expected to possess over time in lower secondary schooling. Based on the Malaysian lower secondary syllabuses of English, the researchers with the help of experts in English language (two university lecturers and two teachers of English) analysed the three standardized tests and came out with item content descriptors. In doing so, they were able to identify the level of item difficulty, sub skills and grade levels that the test items represent. Furthermore, these tests were used by the Ministry of Education and were developed by content experts and teachers from the field. Therefore, content validity would not be an issue.

The choice of the final test items for the reading test was based on two considerations. The first pertains to the results of a pilot study using the common item equating method. The second relates to suggestions given by the language experts. They suggested modifying a few of the items and adding more to the test in order to have enough items for the skills being investigated. Hughes (1989, p. 119) asserts that "successful choice of reading texts depends ultimately on experience, judgment, and a certain amount of common sense". The final test consisted of 60 items representing the following reading sub skills: (1) Ability to infer information from several texts (*Making Inference*); (2) Ability to draw conclusions from several texts (*Drawing Conclusions*); (3) Ability to scan for details in several texts (*Scanning for*

Details); (4) Ability to interpret information from several texts (*Interpreting Information*); (5) Ability to understand figurative language in literary texts (*Understanding Figurative Language*); and (6) Ability to find out meanings of words (*Finding out Word Meanings*). Twenty items selected from the list provided in the syllabus and to be taught within the context of the three areas of language use (*Identifying Grammatical Units*) were also included on the test as they are considered as enabling skills for reading as well as for the other language skills (see Zhang, 2012)

Data Analysis

The one-parameter Rasch Model for dichotomous data was used to examine the psychometric properties of the instrument and answer the research questions. Winsteps version 4.1.0 was used to conduct the Rasch analysis of dichotomous data (Linacre, 2018). The Rasch Measurement Model was utilized since it meets the requirements of fundamental measurement, and it can provide more accurate information on the appropriateness of research instruments and individual items and persons (Bond & Fox, 2015; Wright & Stone, 1979). It also calibrates item difficulty and person ability on one single interval scale for the comparison purposes (Bond & Fox, 2015), where the most able persons and the most difficult items are placed at the upper part of the scale and the least able persons and the easiest items are placed toward the lower part (Bond & Fox, 2015; Wright & Stone, 1979). SPSS version 16 was also used to conduct the descriptive statistics using the interval data produced by the Winsteps software. All analyses of students' performance on the English reading test are depicted in tables and figures in the next section.

RESULTS

Adequacy of the Reading Test

The positive point-measure correlation coefficients provided evidence that items on the test were working together in defining the reading construct. However, one item (Item 50) had a negative correlation (-.04) which might have been the result of lucky guessing, as evidenced by the misfitting/unexpected responses. In addition, the item was the most difficult (2.98 logits) which might have encouraged students to guess. Investigation of individual item fit statistics indicates that the items provided satisfactory fit to the Rasch Model expectations. All the items were within the specified range of Infit mean-square, 0.7 – 1.3 (Bond & Fox, 2015). Of the items with Outfit mean-squares above 1.3, five of them did not depart far from 1.3, whereas 3 items with values less than 0.7 also did not depart far from the desired value 0.7. One of the likely reasons for having relatively large Outfit mean square values was possible guessing by students. No problem was detected with item content or format. This is supported by the mean values and standard deviation for the Infit mean square and Outfit mean square (0.99, 0.12 logit and 1.04, 0.32 logit respectively). It also shows that there is little discrepancy or deviation from the expectation of the Rasch Model (see Green & Frantom, 2002). Together, the fit statistics

and point-measure correlation coefficients provided satisfactory evidence that the items were useful indicators for profiling the reading sub skills and student performance.

The Principal Component Analysis of Standardized Residuals supported the unidimensionality of the reading construct as no secondary factor was extracted. The largest factor extracted from the residuals was 2.6 which had the strength of about 3 items which is insufficient to be considered as violations to unidimensionality. The standardized residual variance explained by the measures for both data and the modeled expectation (29.6 and 29.1 respectively) was only slightly different, supporting the unidimensionality of the reading test. Additionally, disattenuated correlations are one or closer.

As for the construct validity (empirical scaling), all items were well spread along the inquiry scale (i.e., logit scale) defining increasing intensity. This is supported by a high item reliability index (.99). No visible gaps were identified, except for a relatively wide gap at the upper end of the scale because of the extreme value of Item 50. Qualitative investigations showed that this item was difficult because students have to memorize the prepositions used with the verb “aim”. The use of the verb “instilling” in the sentence might have also distracted students’ attention. Being a difficult item, students tried to get the correct answer by guessing. Examining the redundant items, it was found that most of the redundant items were different in the skill measured and text type. For instance, items 57, 12, 18, 32, 36, and 3 look redundant as they are of the same difficulty level, but they actually represent different skills. As the prime concern of the study is to describe what the students can or cannot do, item redundancy in this case is not a serious concern.

The high student reliability coefficient showed that student ability measures will be replicated with a high degree of probability if another comparable set of items is used (Bond & Fox, 2015). Student response analysis showed that a large proportion of the students' responses was within the acceptable Infit MNSQ range. However, Outfit mean square values showed that many students were not responding as the model expected. The major reason for this was lucky guessing rather than carelessness (see Bond & Fox, 2015, Curtis & Boman, 2007). Overall, the patterns of responses of measured examinees were consistent with the expectations of the Rasch model. Finally, although the mean for person ability (0.30 logit) is higher than the mean for item difficulty (0.0 logit), it is safe to say that items and persons are adequately targeted. The relatively high person mean also indicates that the test is relatively easy for the students. In summary, from the above mentioned discussion it can be inferred that the reading test was adequate to give reliable measurement and description related to students’ English reading literacy.

Students’ Performance on the Reading Test

Generally, the reading test was relatively easy for the students since the mean of students’ ability was 0.30 logit, higher than the mean of item difficulty (0.0) (Figure 1). The upper part of the scale indicates the most able students who answered most of the items, while the lower part shows the least able students. Items most often correctly answered (easier items) are positioned towards the lower part and the least correctly answered are positioned towards the upper part of the scale (more difficult items). Additionally, student ability measures spanned

about 6.73 logits (from -2.07 to + 4.66) while item difficulty measures spread was about 5.06 logits (from -2.08 to + 2.98). Most students were distributed between -1 logit and +1 logit.

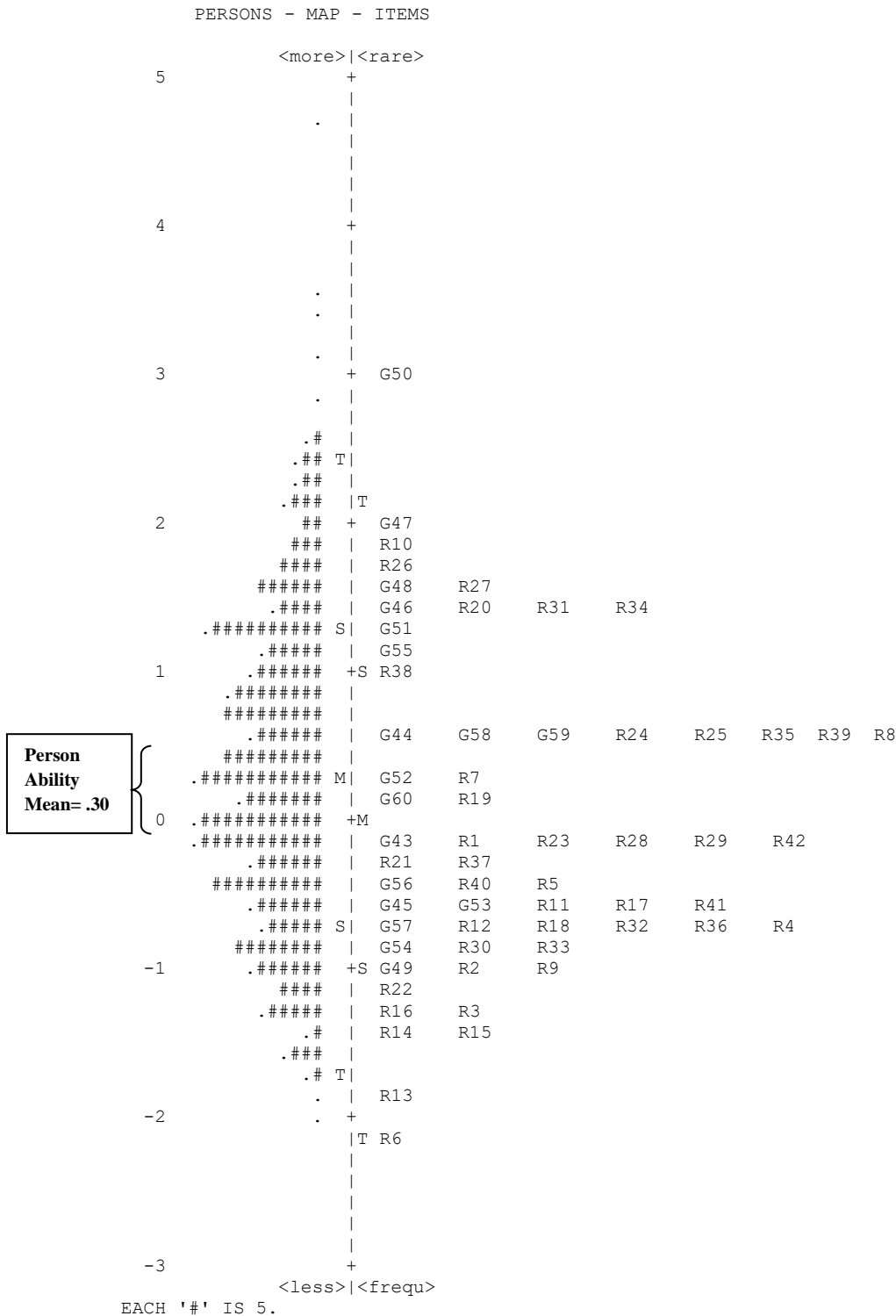


Figure 1: Wright Item-Ability Map: Performance of Lower Secondary Students on Test of English Reading Skill

Figure 2 shows the distribution of all the sub skills associated with the items included in the reading test. All sub skills or categories had different levels of difficulty measures. The most difficult category was related to *interpreting information* (mean = 0.65 logit, SD =1.09 logit), and the easiest one was related to *finding out word meanings* (mean = -0.68 logit, SD = 0.71 logit). Table 2 shows means, standard deviations, medians and mean errors for all sub skill areas. On average, students as a group achieved higher in all sub skill areas except for those related to *interpreting information* and *identifying grammatical units*. However, the items in each sub skill area or category showed different distributions. Some items were positioned at the top while others were placed either in the middle or at the bottom of the scale of inquiry as shown in Figure 2, indicating that students were able as well as unable to apply certain reading sub skills.

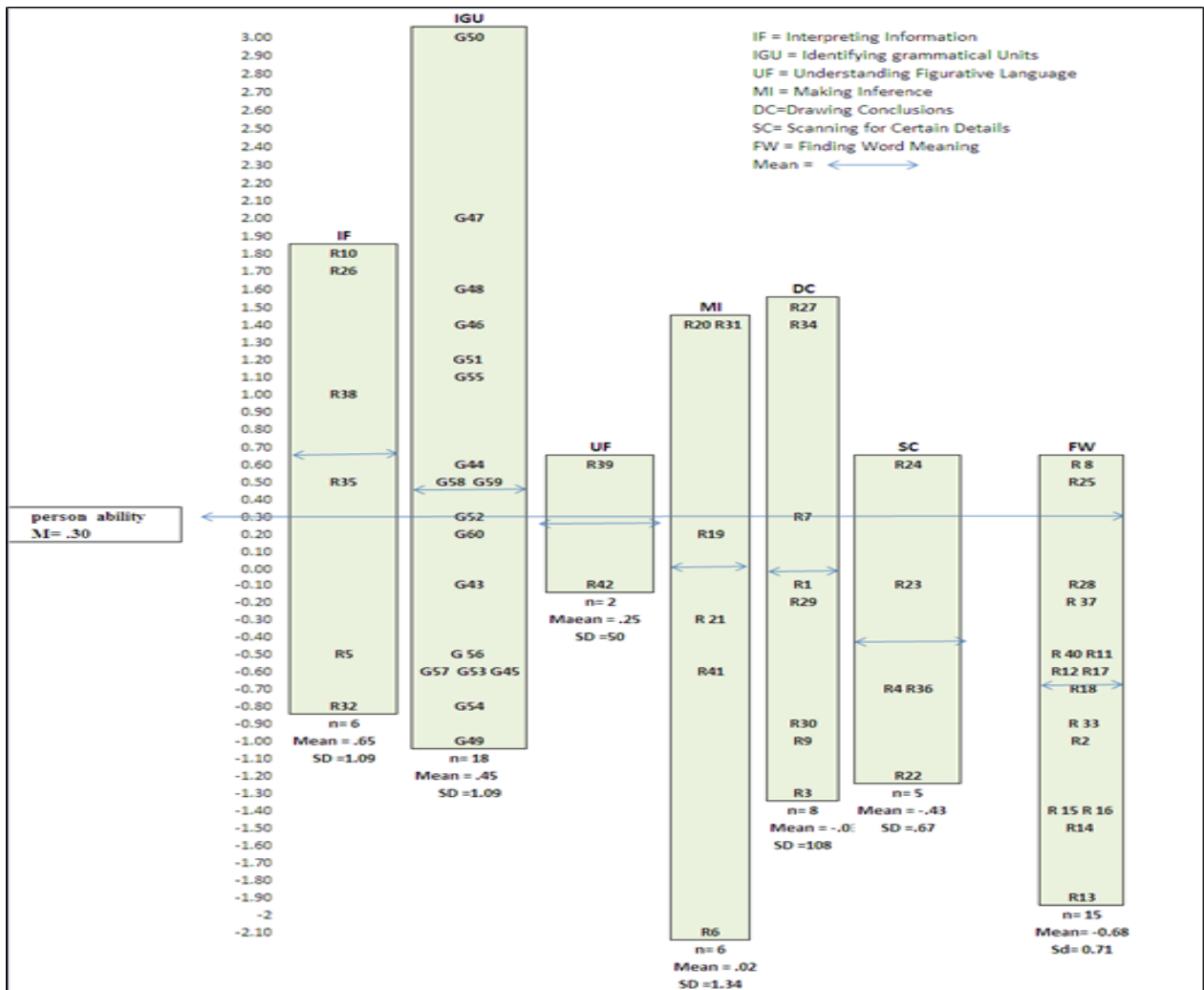


Figure 2: Means of Person Ability and Difficulty Level of Reading Skill Areas

Table 2
Means and Medians for Reading Skill Areas

Broad Skill Areas	Mean (Logit)	Median (Logit)	Std. Deviation	Std. Error of Mean	<i>n</i>
Interpreting Information	0.65	0.77	1.09	0.45	6
Identifying grammatical unit	0.45	0.42	1.10	0.26	18
Understanding figurative language	0.25	0.25	0.51	0.36	2
Making Inference	0.02	-0.05	1.34	0.55	6
Drawing conclusions	-0.03	-0.14	1.08	0.38	8
Scanning for details	-0.43	-0.71	0.68	0.30	5
Finding out word meanings	-0.68	-0.65	0.71	0.18	15

Male and Female Students' Reading Performance

Table 3 shows that, on average, male students as a group performed slightly better than female students. The mean for male students' performance was 0.36 logit with a standard deviation of 1.06 logits, while the mean for female students' performance was 0.25 logit with a standard deviation of 1.00 logit. Figure 3 also shows the difference in distribution of males' and females' reading performance measures. The median estimate for males was about 0.30 logit, and for females, 0.13 logit.

Table 3
Means and Medians of Males' and Females' Reading Performance

Gender	Mean (Logit)	Median (Logit)	Std. Deviation	Std. Error of Mean	<i>n</i>
Male	0.36	0.30	1.06	0.05	448
Female	0.25	0.13	1.00	0.05	490

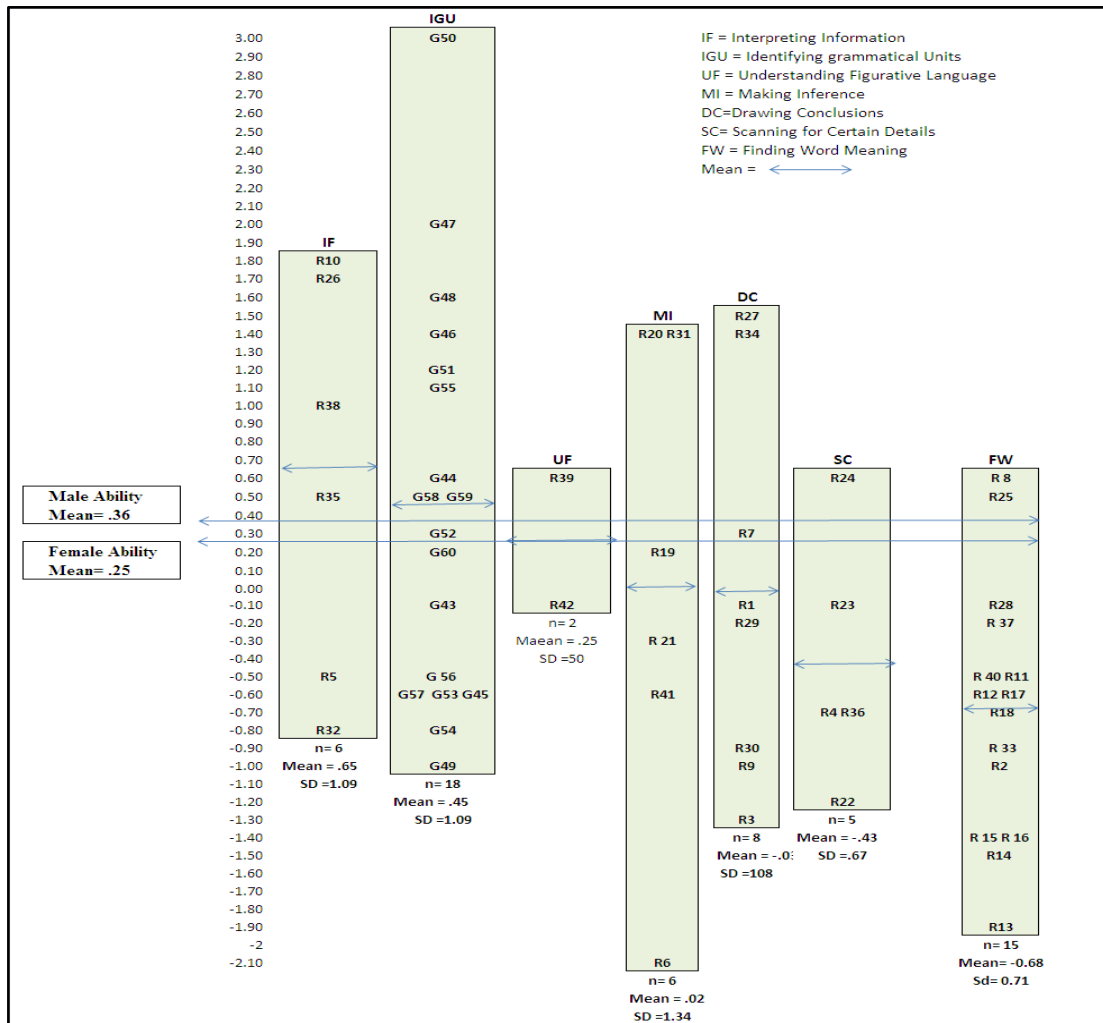


Figure 3 : Males and Females Reading Performance in All Reading Skill Areas

Achievers and Underachievers in English Reading Performance by Demographics

To identify which group of boys and girls are at risk of underachieving, a comparison between all subgroups of male and female students was carried out for each of the demographic variables: Form (i.e., grade level), school location, race, SES, and parents' education levels (Table 4). On average, male students of rural areas performed the lowest. The mean estimate for this subgroup was 0.08 logit. However, there are some rural male students whose performances are comparable to students in the other subgroups. Urban male students generally performed better than the other subgroups with 50% of the students achieving 0.65 logit and above. Form 3 female students achieved the least of all male and female students of other forms. The mean estimate for the subgroup was 0.05 logit. It is interesting that the Form 2 students outperformed the Form 3 students. Indian female students achieved lower than other students. The mean estimate was -0.35 logit. Male Indian boys, however, performed better than Malay girls and slightly better than the Malay boys.

Male and female students of low SES (based on father's job), on average, achieved the lowest performance compared with other students of high and medium SES. The mean estimates were 0.07 logit and 0.13 logit respectively. Interestingly, male students from low SES (mean = 0.07 logit) performed lower than female students from low SES (mean = 0.13 logit).

What is also interesting is that students from the other SES group performed as badly as some of the students of low SES. It is also evident that the higher the SES, the higher the attainment in English reading literacy. A similar pattern is seen for the variable 'mother's job'. The male and female students of low SES achieved less than other students of high and medium SES. The mean estimates for male and female students of low SES were 0.21 logit and 0.23 logit respectively.

For father's education, it can be seen that male and female students whose fathers have either primary or lower secondary qualifications achieved the least. The mean estimates for male and female students whose fathers have primary qualification were -0.14 logit and 0.02 logit respectively; whereas the means for male and female students whose fathers have lower secondary qualification were 0.05 logit and -0.09 logit respectively. The number of students in the *No formal education* category is very small; therefore, the results may not be representative of the actual population. Male students whose mothers have primary qualification achieved less than other groups, followed by female students of mothers who have lower secondary qualifications. The mean estimates were -0.23 logit and -0.10 logit respectively. The same pattern is seen for female students whose mothers have lower secondary qualifications. It is evident that parents' educational qualifications influence their children's performance in English reading skill; if they have higher educational qualifications, their children will do much better on the reading test than other students.

Table 4

Means and Medians of Female and Male Students Estimated Reading Performance by Selected Demographic Variables

Variable / Category		Mean	Median	SD	Mean Error	n
Male	Urban	0.60	0.65	1.12	0.07	243
	Rural	0.08	-0.03	0.93	0.06	211
Female	Urban	0.32	0.22	1.14	0.05	254
	Rural	0.17	0.13	0.84	0.04	236
Male	Form 2	0.44	0.30	1.02	0.30	142
	Form 1	0.37	0.22	1.07	0.21	158
	Form 3	0.27	0.35	1.10	0.35	154
Female	Form 2	0.53	0.41	1.05	0.06	151
	Form 1	0.23	0.13	0.89	0.06	141
	Form 3	0.05	0.05	1.01	0.06	198
Male	Chinese	0.59	0.83	1.22	0.12	102
	Malay	0.27	0.22	0.96	0.06	276
	Indian	0.36	0.19	1.08	0.14	60
Female	Chinese	0.69	0.83	1.17	0.08	101
	Malay	0.22	0.13	0.90	0.04	315
	Indian	-0.35	-0.36	0.87	0.09	69
Male	Upper SES	0.82	0.78	1.19	0.11	98
	Medium SES	0.56	0.64	1.02	0.10	111
	Lower SES	0.07	0.05	0.86	0.08	123
Female	Upper SES	0.75	0.65	1.01	0.12	77
	Medium SES	0.26	0.26	1.01	0.08	155
	Lower SES	0.13	0.05	0.92	0.08	146
Male	Upper SES	0.87	0.65	1.08	0.14	59
	Medium SES	0.58	0.74	1.04	0.14	57
	Lower SES	0.21	0.22	1.02	0.07	244
Female	Upper SES	0.78	0.80	0.98	0.13	54
	Medium SES	0.24	0.47	1.14	0.17	47
	Lower SES	0.23	0.13	0.95	0.06	287

Table 4 Continued

Variable / Category	Mean	Median	SD	Mean Error	n	
Male	Degree	1.06	1.12	0.94	0.11	66
	Diploma	0.72	0.73	1.12	0.16	52
	Upper Sec.	0.31	0.30	1.00	0.08	165
	Lower Sec.	0.05	-0.03	1.05	0.12	79
	Primary	-0.14	-0.28	0.89	0.13	45
	No Formal Edu.	-0.21	0.01	0.89	0.40	5
Male	Degree	1.12	1.22	0.88	0.14	41
	Diploma	0.82	0.73	1.06	0.14	60
	Upper Sec.	0.34	0.31	0.99	0.08	168
	Lower Sec.	0.12	0.01	1.04	0.11	94
	Primary	-0.23	-0.36	0.96	0.14	45
	No Formal Edu.	-0.36	0.68	1.58	0.79	4
Female	Degree	0.89	1.02	1.00	0.20	25
	Diploma	0.67	0.56	1.05	0.16	45
	Upper Sec.	0.43	0.38	0.98	0.07	194
	Primary	0.07	0.01	0.84	0.17	52
	Lower Sec.	-0.10	-0.12	0.97	0.09	108
	No Formal Edu.	-0.39	-0.62	0.82	0.33	6

DISCUSSION

The sample for this study includes slightly more female students 490 (51.9%) than male students, 448 (47.5%). The results indicate that male students' performance on the reading test was higher (mean = 0.36 logit), compared with the performance of female students (mean = 0.25 logit). Regardless of the demographic variables, male students, on average and as a group performed slightly better than female students as a group. The median for males' performance is 0.30 logit and 0.13 logit for females. This unexpected difference might be a result of the demographic variables. It might also have been caused by the inclusion of one boys' school, where the majority of students are high achievers. This result is somewhat inconsistent with the findings of other studies and reports (see Klecker, 2006; Masters & Foster, 1997b; Mullis, & Martin, Kennedy, & Foy, 2007; Tella, Indoshi & Othuon, 2010). In their study, Masters and Foster (1997b) found that year 3 and year 5 female students achieved higher on reading performance than male students. In Kenya, female students generally performed slightly better than male students on the secondary school English curriculum than the males (Tella et al., 2010). In addition, Mullis et al. (2007) highlight that in PIRLS, girls achieved higher than boys in all countries.

However, not all male students were performing better, and not all female students were underachieving. Subgroups of female students performed better than similar subgroups of male students. A comparison of female and male students in terms of their forms showed that Form 3 female students achieved the lowest followed by Form 1 female students and Form 3 male students. The highest performance is for Form 2 female students followed by Form 2 male students and Form 1 male students. This result is possibly influenced by the low results of SMK (E) school, where the majority of the participating students are Form 3 female students. Furthermore, the majority of Form 3 female students might be Indian females or from families with low SES who did not perform well, as discussed later.

As a group, urban students performed better than rural students. Specifically, a comparison of the means for urban and rural subgroups showed that rural male students achieved the least performance followed by rural female students, whereas urban male students achieved the highest performance followed by urban female students. This finding is supported by findings of a research conducted on lower secondary rural students from Johor in which students showed that they rarely use English language (Aziz Nordin, 2005). He further elaborates that this result might be because rural students do not have more opportunities to use the language at home or any place outside the classroom in comparison with urban students. They prefer to use their mother tongue at home and wherever they go. The same reason is mooted by Siti Norliana Ghazali (2008) who argued that rural students have limited exposure and less opportunity to use English outside the classrooms.

It is essential and useful for English learners to interact with each other or with other students outside the classroom to improve their English performance. Another reason might be the negative attitude of students towards learning English language (Nor Azmi Bin Mostafa, 2002). In this respect, Candlin and Mercer (2001) assert that learners' attitude towards learning a target language, its speakers as well as the learning context is one of the major factors that determine their success in learning language. On the other hand, urban students might be more enthusiastic to learn English as they have much exposure to the language as well as more available opportunities to practice it. These findings are in line with findings of other studies, such as Zhang (2006) who argued that differences between rural and urban students could be due to individual characteristics, such as family SES. He adds that rural students suffered from "inferior home and school circumstances" (p. 509) and "had fewer and lower-quality resources than did urban schools in almost all cases" (p. 601). In another study, Cartwright and Allen (2002) found that urban students performed significantly better in reading than students from rural schools in Canada at national and international tests, such as PISA.

With respect to race, Chinese students as a group performed better than Malay students who performed better than Indian students. More specifically, Indian female students showed the least performance followed by Malay female students and Malay male students, while Chinese female students showed the highest performance followed by male Chinese students and Indian male students. These results may be because Chinese students are more exposed to English than other races. They have a strong belief that English language is one of the major keys of their success in their career life. Chinese families may constantly encourage their children to use the language in and outside their homes. These findings are consistent with other studies (Sharifah Md. Nor, 1991; Siti Norliana Ghazali, 2008). In her study, Sharifah Md Nor (1991) found that primary students from the Malay, Chinese, and Indian races in Malaysia are significantly different in their academic achievement, namely English language proficiency; the Chinese students performed better than the Malay and Indian students. The same goes to undergraduate students as found by Jamila Kamal (cited in Nor Azmi Bin Mostafa, 2002). These findings are also congruent with Gillborn & Mirza (2000).

On average, students from high socio-economic backgrounds as a group, performed better than students from medium socio-economic backgrounds, who, in turn, performed better than students from low socio-economic backgrounds. More specifically, male and female students whose fathers' occupations fall in low SES showed the least performance followed by female

students whose fathers' occupations fall in medium SES; while male and female students whose fathers' occupations fall in high SES showed the highest performance followed by male students whose fathers' occupations fall in medium SES. Almost the same pattern is seen for mother's occupation. However, male and female students whose mothers' occupations fall in low SES did better than those whose fathers' occupations fall in low SES.

These findings show a wide gap between students' performance in high and low socio-economic groups. Parents in high SES may have various resources for private tuition; they can send their children to tuition centres, or they can hire teachers to give home tutoring. In addition, they can provide their children with more facilities, such as computers, the internet, electronic dictionaries, etc. These findings are in line with Masters and Foster's (1997b) study in which students from high SES achieved higher than students from medium SES, who in turn achieved higher than students from low SES. The findings are supported by other studies (see Chiu & Chu Hoa, 2006; D'Angiullia, Siegel & Hertzman, 2004; Zhang, 2006) who highlighted that extensive literature has shown that students of low SES significantly achieve less in reading performance than those of higher SES. Furthermore, low-income families have a higher percentage of disadvantaged children in terms of reading achievement than any other socioeconomic groups. High SES families can expend more on educational resources such as books, magazines, and so forth.

For parents' education, students whose fathers have a degree qualification showed highest performance, followed by diploma qualification and upper secondary certificate, while students whose parents have a primary qualification showed the least performance, followed by lower secondary certificate. The number of students in *No formal education* category is very small; therefore, the results may not be representative of the actual population. More specifically, male students whose fathers have a primary certificate achieved the lowest performance among all groups, followed by female students whose fathers have lower secondary certificates. Additionally, female students whose fathers have a primary certificate and male students whose fathers have a lower secondary certificate also under-perform. Almost the same pattern is seen for mother's education.

In general, if parents have higher academic qualifications, their children will do much better on the reading test than other students. These findings might be because parents with high academic qualifications are more concerned with their students' performance, and so they follow them up more regularly. They may have much experience and knowledge about teaching and learning theories that help them to deal with their children in a more effective manner, or they may be capable of giving their children home-tutoring as they are well-educated. Chiu & Chu Hoa (2006) pointed out that literature has shown that less educated parents do not spend much time with their children compared to highly educated parents. They added that more educated parents constantly monitor and supervise their children's progress. These findings go with the findings of other studies (see Mullis et al., 2007; Myrberg & Rosén 2008) who found that parents who have higher levels of education and occupation significantly contribute to their children's good performance.

In summary, students' demographic profile plays a significant role in determining their performance. In his survey, Pandian (2000) highlighted that in Malaysia, people who read often

in English are likely to: live in an urban than in a rural area; belong to a family with a high socio-economic standing; come from a home where there is a greater variety and amount of materials in English, with more influence and reading models at home; attend a school with a greater variety and amount of materials in English, with more teachers who encourage students to read and more friends who read English; be exposed more to English; and have a more positive attitude towards reading in English.

CONCLUSION

Students' performance in English reading varies from one group to another. Unexpectedly, male students, as a group, performed slightly better than females (who were expected to do better on the test items). This result is not consistent with many findings that showed females are outperforming males. Of course, not all male students performed better, and not all female students were underachieving. Investigating male and female students' performance based on the demographic information showed that subgroups of female students did better compared to the same male subgroups. Students of Forms 1 and 2 outperformed Form 3 students, with the lowest performance achieved by Form 3 female students; students from urban areas performed better than students from rural areas, with least performance to rural male students.

With regard to racial identity, students of Chinese race performed better than Malay students, who in turn, performed better than Indian students, with the least performance to Indian female students. SES and parents' educational background also played significant roles in reading achievement. Students of high SES performed better than students of medium SES, who in turn, performed better than low SES, with least performance to male students of lower SES. Students whose parents possess higher educational qualifications did better than students whose parents have lower qualifications. These results indicate that students' performance is also influenced by some other variables, not only gender. Such variables are related to school location, students' racial identity, family SES as well as parents' education. So, it would be imprudent to make any conclusions about male and female performance based on gender only. The factors that cause these differences should be tracked and monitored.

With the application of the Rasch Model, students' performance is more accurately displayed on visual maps. These maps help to determine students' levels and item difficulty on the same interval scale. These maps help identify students' levels of reading performance, that is what skills each group of students have and have not achieved. In principle, the findings of this study support a need of better measurement for students' performances over time. The differences among school performances need more qualitative investigation to identify the factors behind high and low performances of Malaysian lower secondary school students.

REFERENCES

- Alderson, J. C. (2000). *Assessing reading*. United Kingdom: Cambridge University Press.
- Alderson, J. C. (1990a). Testing reading comprehension skills (part 1). *Reading in a Foreign Language*, 6 (1), 425-438.
- Alderson, J. C. (1990b). Testing reading comprehension skills (part 2): getting students to talk about taking a reading test. *Reading in a Foreign Language*, 7(1), 465.
- Alderson, J. C., & Lukmani, Y. (1989). Cognition and reading: Cognitive levels as embodied in test questions. *Reading in a Foreign Language*, 5(2), 253-270.
- Aziz bin Nordin (2005). Students' perception on teaching and learning mathematics in English. *Bulletin Pendidikan Sains dan Matematik Johor*, 14 (1).
- Bachman, F. B. (1990). *Fundamental consideration in language testing*. Oxford: Oxford University press.
- Bond, T., & Fox, C. M. (2015). *Applying the Rasch Model: fundamental measurement in the human sciences*. Routledge.
- Brown, D. (2003). *Language assessment: Principles and classroom practices*. United States of America: Longman
- Candlin, C & Mercer, N. (2001). *English language teaching in its social context*. New York: Routledge.
- Cartwright, F., & Allen, M. (2002). Understanding the rural – urban reading gap. Education, skills and learning Research papers. *Statistics Canada*. Retrieved from www.statcan.gc.ca/pub/81-595-m/81-595.
- Chiu, M. & Chu Hoa, E. (2006). Family effects on student achievement in Hong Kong. *Asia Pacific Journal of Education*, 26 (1), 21–35.
- Collins, C., McLeod, J., Kenway, J. (2000). Factors influencing the educational performance of males and females in school and their initial destinations after leaving school. Canberra: Commonwealth Department of Education, Training and Youth Affairs. Retrieved from <http://bit.ly/AeD7ZL>
- Cortis, N., & Newmarch, E. (2000). Boys in schools: what is happening? Department of Education, Training and Youth Affairs. Retrieved from <http://www.dest.gov.au>
- Curtis, D., & Boman, P. (2007). X-ray your data with Rasch. *International Education Journal*, 8(2), 249-259.
- D'Angiulli, A., Siegel, L., & Hertzman, C. (2004). Schooling, socioeconomic context and literacy development. *Educational Psychology*, 24 (6), 867-883.
- Engelhard, G. (2001). Historical view of influences of measurement and reading theories on the assessment of reading. *Journal of Applied Measurement*, 2(1), 1-26.
- English Syllabus for Forms 1 to 5 (2003). Ministry of Education. Retrieved from www.scribd.com/doc/488142/BInggerisKu

- Mahmood, H. (2014). *Civics and citizenship education in Malaysia: The voice of micro policy enactors*. (Doctoral Thesis, Cardiff University). Retrieved from <http://orca.cf.ac.uk/id/eprint/69576>.
- Farhadi, H., & Moeini, H. R. (2005). Construct validation of reading comprehension skills. *Quarterly Journal of Humanities*, Al-Zahar University.
- Farhady, H., & Hassamy, G. (2005). Construct validity of L2 reading comprehension skills. *Iranian Journal of Applied Linguistics (IJAL)*, 8 (2), 29-53.
- Gay, L. R., & Airsian, P. (2000). *Educational research: Competencies for analysis and application* (6th ed.). Upper Saddle River, NJ: Merrill & Prentice-Hall.
- Gillborn, D & Mirza, S. (2000). Educational inequality: Mapping race, class and gender. *Ofsted*. Retrieved from www.ofsted.gov.uk/Ofsted-home/Publica
- Granger, C. & Linacre, J. M. (2008). What is measurement? Retrieved from www.udsmr.org/Documents/What_Is_Measurement_2008.pdf
- Green, R. & Frantom, C. (2002). Survey development and validation with Rasch Model. A paper presented at the International Conference on Questionnaire Development. A paper presented at the International Conference on Questionnaire Development, Evaluation, and Testing, Charleston, SC. Retrieved http://www.jpsm.umd.edu/qdet/final_pdf_papers/green.pdf
- Hambleton, B. k., Swaminathan, H. & Rogers, H. J. (1991). *Fundamental of item response theory*. London: Sage Publications.
- Hedgcock, J., & Ferris, D (2009). *Teaching readers of English: Students, texts, and contents*. UK: Routledge.
- Holme, R. (2004). *Literacy: an introduction* .UK: Edinburgh University Press Ltd.
- Hoyt, D. (2015). Understanding the minds of boys: Strategies for student engagement. ASCD in Service. Retrieved form <https://inservice.ascd.org/understanding-the-minds-of-boys-strategies-for-student-engagement/>
- Hudson, T. (2007). *Teaching second language reading*. Oxford: Oxford University Press.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Hutchison, D. (2004). A critical evaluation of “raising boys' attainment”. *Educational Psychology in Practice*, 20(1), 3-15. <https://doi.org/10.1080/0266736042000180375>
- Ingebo, G. S. (1997). *Probability in the measurement of achievement*. Chicago, IL: Mesa Press
- Keeves, J. P. (2004). *Educational research, methodology, and measurement: an international handbook* (2nd ed.). England: Pergamon Press.
- Klecker, B. (2006). The gender gap in NAEP fourth-, eighth-, and twelfth- grade reading scores across years. *Free Parents Online*. Received from <http://www.freepatentsonline.com/article/ReadingImprovement/144717435.html>

- Klecker, B. (2006). The gender gap in NAEP fourth-, eighth-, and twelfth- grade reading scores across years. *Free Parents Online*. Received from <http://www.freepatentsonline.com/article/ReadingImprovement/144717435.html>
- Latifah Ismail (2015). Gender gap in higher education: Perspective on factors influencing enrolment in Malaysian universities: A University of Malaya sample. *The Online Journal of Quality in Higher Education*, 2(4), 35-45.
- Levine, A., Ferenz, O., & Revez, T. (2000). EFL academic reading and modern technology: How can we turn our students into independent critical readers? *Teaching English as Second and Foreign Language TESL-EJ*, 4 (4). Retrieved from <http://tesl-ej.org/ej16/a1.html>
- Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7(4), 328.
- Linacre, J. M. (2003). Constructing scientific measurement models. *Rasch Measurement Transactions*, 17 (1), 907.
- Linacre, J. M. (2018). Winsteps & Rasch Measurement (Version 4.1.0) [Computer Software and manual].
- Lingard, B., Martino, W., Mills, M., & Bahr, M. (2002). Addressing the educational needs of boys. Department of Education, Training and Youth Affairs. Retrieved from <http://www.dest.gov.au>
- Lloyd, T. (2011). Boys' underachievement in schools: Literature review. Belfast: CYMS.
- Ludicke, P., Muir, T., & Swabey, K (2019). Identifying and supporting young adolescent academic underachievers in Year 7 and 8 classrooms. *Issues in Educational Research*, 29(2).
- Masters, G. (1993). Rasch measurement and IEA studies. Australian Council for Educational Research. *Rasch Measurement Transactions*, 7(3), 310.
- Masters, G. N., & Foster, M. (1997a). Mapping literacy achievement: Results of the national school English literacy survey. Australian Council for Educational Research. Retrieved from GN Masters, M Forster, Australian Council for - 1997 - research.acer.edu.au.
- Masters, G. N. & Foster, M. (1997b). Literacy standards in Australia. Australian Council for Educational Research. Retrieved from GN Masters, M Forster, Australian Council for - 1997 - research.acer.edu.au.
- Matthews, M. (1990). Skill taxonomies and problems for the testing of reading. *Reading in a Foreign Language*, 7(1), 511-517. Retrieved from nflrc.hawaii.edu/RFL/PastIssues/rfl71matthews.pdf
- McNamara, T. F. (1996). *Measuring second language performance*. New York: Addison Wesley Longman.
- McGee, L. M., & Richgels, D. J. (2000). *Literacy beginnings: Supporting young readers and writers*. Boston: Allyn and Bacon.

- Meiers, H. (2008). A longitudinal investigation of literacy achievement and development in the first three years of school. ACER e-Search. Retrieved from http://research.acer.edu.au/monitoring_learning/13.
- Miele, Holly, "The Lost Boys: A Phenomenological Study of Freshmen Year Failure" (2020). *Doctor of Education (EdD)*. 139. Retrieved from <https://digitalcommons.georgefox.edu/edd/139>
- Ministry of Education Malaysia (2013). Malaysia Education Blueprint (2013-2025). Ministry of Education, Putrajaya, Malaysia.
- Mullis, I., Martin, M., & Foy, P. (2008). TIMSS 2007 International Mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Retrieved from <http://pirls.bc.edu/timss2007/mathreport.html>
- Myrberg, E. & Rosén, M. (2008). A Path model with mediating factors of parents' education on students' reading achievement in seven countries. *Educational Research & Evaluation*, 14 (6), 507- 520.
- Nachiappan, S., Veeran, V. P., & Andi, H. K. (2012). Gender Issues in Malaysian Education: Factors Influencing Male and Female Students Academic Achievement through Cognitive Processes in Public Examinations. *International Journal of Business and Social Research (IJBSR)*, 2(2), 133-150.
- Nor Azmi Bin Mostafa (2002). Bilingualism and the academic achievement of Malay students. International Conference IPBA. Retrieved from kajianberasaskansekolah.files.wordpress.com/pdf
- Pandian, A. (2000). A study on readership behavior among multi-ethnic, multi-lingual Malaysian students. *The International Journal of Learning*. Retrieved April 18, 2011.
- Ratnawati, Mohd-Asraf, Hazlina Abdullah & Ainol Zamin Mat Zamin (2016). Literacy among Malaysian primary schoolers: How do boys perform relative to girls? *International Electronic Journal of Elementary Education*, 9(1), 225-238.
- Rost, D. (1993). Assessing the different components of reading comprehension: Fact or fiction. *Language Testing*, 10 (1), 79-92.
- Rowe, K. J. (2006). Assessment during the early and middle years: Getting the basics right. Australian Council for Educational Research. Retrieved from [KJ Rowe - 2006 - acer.edu.au](http://www.acer.edu.au)
- Sharifah Md. Nor (1991). Faktor-faktor yang berkaitan dengan pencapaian akademik murid darjah satu. *Jurnal Pendidik dan Pendidikan*. Pulau Pinang: Universiti Sains Malaysia. 11, 61-69.
- Shoaib, M., & Ullah, H. (2019). Female and Male Students' Educational Performance in Tertiary Education in Punjab Pakistan. *Pakistan journal of Social Issues*, 10, 83-100.
- Scholes, L. (2019). Differences in attitude towards reading and other school-related activities among boys and girls. *Journal of Research in Reading*, 42(3-4). Retrieved from <https://doi.org/10.1111/1467-9817.12279>

- Scholes, L. (2019). Differences in attitude towards reading and other school-related activities among boys and girls. *Journal of Research in Reading*, 42(3-4). Retrieved from <https://doi.org/10.1111/1467-9817.12279>
- Scholes, L. (2020). Recognition of boys as readers through a social justice lens. *British Journal of Sociology of Education*, 41(7), 975-991.
- Siti Norliana Ghazali (2008). Learner background and their attitudes towards studying literature. *Malaysian Journal of ELT Research*, 4.
- Stephanou, A., Meiers, M., & Forster, M. (2000). Constructing scales for reporting growth in numeracy: The ACER longitudinal literacy and numeracy study. Australian Council for Educational Research. Retrieved from research.acer.edu.au/cgi/viewcontent.cgi?article=1008 pdf.
- Tella, J., Indoshi, F., & Othuon, L. (2010). Relationship between students' perspectives on the secondary school English curriculum and their academic achievement in Kenya. *Educational Research*, 1(9), 390-395.
- Tienxhi, J. Y. (2017). The gender gap in Malaysian public universities: examining the Lost Boys'. *Journal of International and Comparative Education (JICE)*, 1-16.
- Ullah, R., & Ullah, H. (2019). Boys versus Girls' Educational Performance: Empirical Evidences from Global North and Global South. *African Educational Research Journal*, 7(4), 163-167.
- UNESCO (2004b). Gender studies. Retrieved from UNESCO.Org. Retrieved from <http://www.unescobijng.org/view.do?channelId=003002001003#first>
- Urquhart, A., & Weir, C. J. (1998). *Reading in a second language: process, product, and practice*. New York: Longman.
- Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin*, 140(4), 1174–1204. <https://doi.org/10.1037/a003662>
- Wallen, N. E., & Fraenkel, E. W. (2006). *How to design and evaluate research in education* (4th ed.). USA: McGraw- Hill.
- Ward, M.R.M., & Thurnell-Read, T. (2019). Boyhood and belonging. *Boyhood Studies*, 12(2). <https://doi.org/10.3167/bhs.2019.120201>
- Watson, A. (2017). "Male Delivery": a critical investigation of what boys have to say about the influence of male teachers on literacy engagement and achievement. *Educational Review*, 69(1), 102-117. <https://doi.org/10.1080/00131911.2016.1169996>
- Weir, C., & Porter, D. (1994). The multi-divisible or unitary nature of reading: The language tester between Scylla and Charybdis. *Reading in a Foreign Language*, 10 (2), 1-19.
- William, D., Patrick J., Malcolm R., & Joseph M. (2006). Rasch modeling of revised token test performance: validity and sensitivity to change. Retrieved from www.ncbi.nlm.nih.gov/pubmed/16533071
- Wright, B. D., & Stone, M. H. (1979). *Best Design Test: A Handbook for Rasch Measurement*. Chicago: Mesa Press.

- Younger, M., Warrington, M., Gray, J., Rudduck, J., McLellan., Bearne, E., Kershner, R., & Bricheno, P. (2005). Raising boys' achievement. Research Report RR636. Norwich: Department for Education and Skills. Retrieved www.scribd.com/.../20316609-Raising-Boys-Achievement-Research-...
- Yu, J., McLellan, R., & Winter, L. (2020). Which Boys and Which Girls Are Falling Behind? Linking Adolescents' Gender Role Profiles to Motivation, Engagement, and Achievement. *Journal of Youth and Adolescence*, 1-170
- Zhang, Y. (2006). Urban-rural literacy gaps in Sub-Saharan Africa: The roles of socioeconomic status and school quality. *Comparative Education Review*, 50 (5), 581-602
- Zhang, D. (2012). Vocabulary and grammar knowledge in second language reading comprehension: A structural equation modeling study. *The modern language journal*, 96(4), 558-575.