

Making Sense of Reading Scores with Reading Evaluation and Decoding System (READS)

Abdul Rashid Mohamed, Lin Siew Eng & Shaik Abdul Malik Mohamed Ismail

School of Educational Studies

Universiti Sains Malaysia, Penang, Malaysia

E-mail: rich@usm.my

Abstract

It is imperative that teachers need to assess their learners' reading proficiency. Incidentally, most assessments developed and conducted by teachers merely discriminate who among their learners are performing better than the other. These assessments deplorably provide nothing more than norm reference data. Nonetheless, this is the only information that teachers have pertaining to the reading proficiency of their learners. A test score preferably should supply teachers with analytical information of what learners can or cannot do. The test scores should allow teachers to determine where the learners' position in their reading development. It is long overdue for teachers and those in the education enterprise to take a closer look behind test scores and their learners' precise abilities. As such, it was felt that a system that not only provides test scores but matches learners' performance against a benchmark and divulge their precise reading abilities should be developed. This article traces the development of a Reading Evaluation and Decoding System (READS) comprising an Encoder, Analyser and Decoder components. A prototype system was first devised based on the Malaysian school curriculum. Next, a model encoder was developed and piloted on more than three thousand students. Their scores were then used to develop the Analyser. Finally, the Decoder was developed based on data gathered from the respondents. The three components of READS were then calibrated and refined through more tests for accuracy. The study found that READS is a reliable system to evaluate learners' performance and decode their reading abilities.

Keywords: Encoder, Analyser, Reading Matrix, Decoder, Calibrated

1. Current Assessment Practices

Despite the equal importance attached to both categories of receptive (reading and listening) and productive skills (speaking and writing), reading is traditionally considered to be a vital requisite for the effective acquisition of knowledge and a major source of input for both writing and speaking. It appears that reading instruction and assessment has always been a prominent component in the Malaysian school curriculum. However, there is a major drawback in contemporary assessment systems as the grades or test scores obtained constitute the only source of information that the teachers have concerning the reading abilities of their learners. Contiguous to this shortcoming is the fact that the scores often provide ambiguous and superficial descriptions of reading capabilities. Currently, national standardised tools of assessments in Malaysia within the formal school system assign candidates a single grade irrespective of the fact that the assessment framework appraises a candidate's language proficiency in each of the 4 language skills. Candidates are assigned a generic grade ranging from A to F that encompasses all 4 skills. As these grades are transferred from raw scores, they imply that some learners perform better than others. However, the opaque nature of the grades and the ambiguity of the prevailing descriptors constrain the ability of pedagogues and learners alike on how to accurately interpret the actual reading ability of learners. This dichotomy in performance and ambiguity in outcomes is alluded to by Kubiszyn and Borich (2000:33) who noted that; *"Too often teachers, parents, and others report that they know little or nothing about learners after testing than before"* which to a certain extent nullifies, Temple, Ogle, Crawford & Freeppon's, (2008) contention that assessments are able to provide teachers and parents with feedback on learners' progress and identify the areas where the learners require further support.

Assessments are not merely tools for gathering data for scoring purposes but are vital conduits that inform and guide the teaching and learning process (Farr, 2003). In reiterating this feedback function, Weeden, Winter and Broadfoot (2002:18), opine that "the purpose of assessment is to improve standards, not merely to measure them". Johnson and Costello (2005), note that assessments must incorporate feedback mechanisms that can facilitate future improvements in performance since assessment has implications for what is learned. In a similar vein, Baker (ed.) (2004) and Baguley (2001), stress that performance measurement are empirical indicators of current performance and predictors of future outcomes. It can be surmised that assessment output provides a snapshot of not only prevailing ability but also latent potential which can plausibly be manifested in future outcomes through the application of facilitative learning initiatives and palliative pedagogical measures (Lanski,

Ehlers-Zavala, Daniel and Sun-Irminger, 2006; Cobb, 2005; Tomlinson, 2009). Implicit in this comprehensive approach is the principle that, effective practices derived by informative assessments can improve instruction and learning as noted by Wilcox (2005).

It is obvious from the foregoing that within the reading domain, ESL teachers in Malaysia, as elsewhere, should be equipped with effective evaluative tools that can elicit accurate information with regard to their learners' actual reading abilities. This is especially pertinent to pedagogues as such information serves as a vital source of feedback through which teachers can continuously monitor progress within the reading classroom in order to ensure their consonance with their learners' abilities and potential (Allen, 2000). Perceptually, such an approach would also be in sync with the procedures of assessment as adopted by Routman (2003), wherein data is collated, evaluated and modifications are made to existing methodologies, materials and activities through such evaluation.

This paper purports to explicate in detail the conceptualisation and implementation of a Reading Evaluation and Decoding System (READS) designed to provide a comprehensive and accurate picture of a learner's actual capabilities within the reading classroom.

2. About READS

The purpose of this study is to establish a Reading Evaluation and Decoding System (READS) which functions to evaluate the learners' reading abilities, analyse the results and utilise them to obtain an in-depth and holistic understanding of an individual's reading capabilities. Zarrillo (2007) states that such processes of data collation and interpretation constitute assessment. Theoretically, the approach adopted in this study was also consanguineous with that of Routman's (2003), in which data is collected, and evaluated before the findings are utilised to make the necessary adjustments in teaching instruction, materials and activities so as to meet the needs and demands of learners.

READS comprises of three components as shown in Figure 1. Theoretically, READS allows a teacher to administer the standardised test, compile the scores and subject them to analysis. The analysis obtained can then be cross-referenced with the Reading Matrix to determine the band in which the reader belongs. Subsequently, teachers, learners and other relevant stakeholders in the learning enterprise can refer to the Descriptors of Reading Abilities guideline to decode the learners' ESL reading abilities. Through this procedure, students identified as 'below standard' or at 'academic warning' can then be provided early intervention assistance through the application of the appropriate reading methodology and materials (Wasburn-Moses, 2006).

3. Snipped of READS Developmental Processes

The development of READS basically encompassed three stages. The first stage involved the development of the Encoder or Test Instrument while the second stage portrayed the development of the Analyser or Reading Matrix. Finally, in the third stage, the Decoder or the Descriptors of Reading Abilities were established.

3.1 Getting Started: How to Develop the Encoder

The first stage involved the establishment of the Reading Performance Indicators before the development of a Prototype Test. The Prototype Test was then subjected to refinement and further modification subsequent to its validity and reliability being determined before it was used as the Test Instrument. Figure 2 illustrates how the Encoder or Test Instrument was developed.

3.2 Establishing the Reading Performance Indicators (RPI)

Essentially, the Reading Performance Indicators (RPI) refers to the detailed descriptions of the specific ESL reading skills that the learners need to master at different levels. The rationale for the development of the RPI was predicated on the premise that there was a need to have a standardised reference guide as a framework from which a reliable and valid Test Instrument could be conceptualised. The input for the RPI was mainly sourced from the relevant primary and secondary English Language Syllabuses, Barrett's Taxonomy of Reading Comprehension, textbooks and past year examination papers.

The Malaysian English Language Syllabuses (primary and secondary) were selected as one of the main sources of input in line with Rabbini's (2002) postulation that a syllabus acts "... as a guide for both teacher and learner by providing some goals to be attained". However, for the purpose of this study, only the reading objectives were focused upon. In the Malaysian English Language Syllabuses, the components on reading outline the reading skills learners need to acquire in order to achieve comprehension skills. This includes skills such as skimming, scanning, summarising, predicting, inferring and other relevant sub skills of reading (Ministry of Education Malaysia, 2003).

The skills and sub-skills suggested by the syllabuses were then matched with three of the major classifications of comprehension abilities as found in The Taxonomy of Reading Comprehension proposed by Barrett (cited in Alderson and Urquhart, 1984). The reference to this taxonomy was based on the fact that its conceptualisation of the cognitive and affective aspects of reading comprehension was deemed to be comprehensive, constructive, functional and eminently suitable for use in the development of a reading test. The three important comprehension reading sub-skills selected to develop the Test Instrument included: literal; reorganisation and; inferential comprehension.

The sub-skills of reading in the textbooks and past year examination papers were also carefully examined in the development of the RPI. During the developmental process, constant cross-referencing with the syllabus revealed a lot of similarities, since, on most occasions the syllabus is presented in the recommended textbooks (Harmer, 1991). Next, information extracted from the Malaysian English Language Syllabus and Barrett's Taxonomy of Reading Comprehension was re-categorised to develop the RPI. The results of the reclassification are illustrated in Table 1.

3.3 Developing the Prototype Test

The Reading Performance Indicators formed the foundation on which a Prototype Generic Standardised Reading Comprehension Test, also known as the Prototype Test was developed.

The development of the Prototype Test was predicated on the principle that reading standards are dynamic as reading proficiency invariably improves with proper instruction. Hence, learners at any educational level are postulated to be at different stages on the reading development continuum, i.e., learners come with diverse abilities, interests and attitudes (Ediger, 2009) with differing rates of progress. Each question consisted the assessment of three different types of reading skills namely; literal, reorganisation and inferential.

3.4 Establishing Content of Prototype Test

The Compendium Volume III (Ministry of Education Malaysia, 1991) stipulates that if a multiple-choice format is used in a test, there should be a minimum number of multiple-choice items with the recommended number being fifty. For this Prototype Test, we decided to use 60 multiple-choice questions to increase the reliability of the test.

The 60 multiple-choice reading comprehension questions comprised of 15 Year 6 level (Primary School Assessment Examination) questions which constituted 25% of the test, 30 Year 9 level (Lower Secondary Assessment) questions which constituted 50% of the test questions and 15 Year 11 level (Malaysian Certificate of Education) questions encompassing 25% of the test questions. The proportion of the questions was based on the distribution of the difficulty level with 25% being designated as easy, 50% as average and 25% as difficult (Mok, 2000).

3.5 Validating the Prototype Test

Basically, content validity is determined by expert judgement (Gay and Airasian, 2003). Five content experts comprising of three experienced examiners of Year 6, Year 9 and Year 11 English Language papers and two senior university lecturers (Teaching of English as a Second Language) analysed the suitability of the questions to test the content validity of the test instrument. The parameters of content validity examined included: type of texts, length, difficulty level and questions, vocabulary, the rubrics and distractors. The findings indicate that the content validity of the test was high and the questions were deemed to be appropriate. Another pilot study comprising 100 selected respondents from Year 7 to Year 11 of a specific school was conducted to determine the construct validity of the Prototype Test.

3.6 Ensuring the Reliability of Prototype Test

Gay and Airasian (2003), state that internal consistency reliability provides valuable information regarding item consistency in a single test. Since the Prototype Test is a multiple-choice test, it was decided that the utilisation of the Kuder-Richardson formulae particularly KR20 was more appropriate. This is because the Kuder-Richardson methods are more sensitive to sources of internal consistency (Oosterhof, 2001); a fact corroborated by Popham (2002) who noted that when a test consists of multiple-choice items, the most commonly used internal consistency procedure was the Kuder-Richardson method. On the other hand, the Coefficient Alpha, as developed by Cronbach was not used because it is more suitable for computing internal consistency of a set off test items which comprised short-answers or essay in which each of the items were scored using a range of points rather than marked as correct or incorrect. The KR20 of this test was found to be within the range of 0.78 to 0.85 for all educational levels which is consistent with the findings of Diederich, as

cited in Oosterhof (2001:74), who proposes that “if a teacher’s test requires a full class period to complete, its Kuder-Richardson reliability should be between 0.60 and 0.80”.

3.7 Scoring Procedure for Prototype Test

A pilot study was conducted in two schools on 120 respondents (60 respondents from each school) in order to select the best scoring procedure. The test was conducted on learners with similar reading abilities using different scoring or marking schemes. This was designed to identify the most appropriate scoring procedure to be used for the Test Instrument. The test conducted in School A was marked using the ‘traditional’ marking scheme where one mark is awarded for every correct answer and zero for an incorrect answer while the test conducted in School B was marked using the ‘partial-credit’ marking scheme (OECD, 2006), where students who provide ‘almost correct’ answers receive partial credit. In this scheme, 3 marks were awarded for the most accurate answer, 2 marks for the almost accurate answer, 1 mark for the next most accurate answer and 0 for an incorrect answer. It was found that when the test was marked using the ‘partial-credit’ marking scheme, the KR20 was only 0.465 indicating a very low question reliability. This implied that the ‘partial-credit’ marking scheme was inappropriate for use in this test. Furthermore, the distractors were basically not constructed to be marked using the ‘partial-credit’ marking scheme. Ultimately, the ‘traditional’ marking scheme was adopted for use after preliminary grading yielded a KR20 score of 0.703.

3.8 Time Allocation for Prototype Test Administration

A pilot study was conducted to determine the time taken to complete the Prototype Test. 90 selected respondents from Year 10 from a specific school comprising of 30 high performers, 30 average performers and 30 low performers sat for the Prototype Test. The time taken by the respondents to complete the Prototype Test is as recorded in Table 2. Based on the data, the amount of time allotted to complete the Test Instrument was premised on the average of the total time taken by the three groups surveyed i.e., 70 minutes.

3.9 The Encoder

After the necessary modifications were made to the Prototype Test, the Test Instrument or Encoder was used to measure the ESL reading performance of secondary school learners from Year 7 to Year 11. In order to benchmark the learners’ reading abilities, the Test Instrument was administered and the learners’ scores then matched to the Reading Matrix to identify the Performance Standards of the learners. The next section will explain how the Reading Matrix was developed.

3.10 How to Develop the Analyser

A Reading Matrix or Analyser refers to a chart which acts as a reading indicator to indicate the reading abilities of learners at a particular educational level, i.e. in this study, from Year 7 to Year 11 in Malaysian secondary schools. Basically, the Reading Matrix was developed through a combination of two perspectives on Progression through the levels as propounded by Horton (1990) and Green (2002). The fundamental criteria of Progression namely; levels of proficiency and age, which are vital components in determining progression, were co-opted into our model.

The purpose of this Reading Matrix is to gauge learners’ reading abilities through a number of bands. These bands are indicators which are explicated upon by detailed descriptors of what the learners could or could not do. In order to obtain accurate and optimal information pertaining to their learners’ performance, it is recommended that the Reading Matrix be used in conjunction with the Test Instrument and the Performance Standard (together with the Descriptors of Reading Abilities of Band 1 to Band 6). This will generate reliable data that can be further utilised as a basis to revise prevailing methodologies, refine materials as well as modify learning activities so as to meet actual learning needs and demands.

3.10.1 Components of the Reading Matrix

The Reading Matrix acts as a reading indicator to indicate whether the learners are below or above the relevant reading benchmarks (Refer to Figure 3).

3.10.2 The Performance Standards

The four levels of reading performance of Year 10 respondents were developed based on the Performance Bands and Reading Performance Indicators. The rationale being for this choice is based on Tomlinson’s (2009) observation that, a cohort of children will not show the same reading levels with some reading at grade level while others a little above or below grade level, and still others, at levels far exceeding or well below the relevant benchmark.

In this study, the learners in Year 10 who are in Band 5 would be classified as 'meet standard'. If the Year 10 learner is in Band 6, he would be classified as 'above standard'. Learners in Year 10 who are in Band 4 would be classified as 'below standard' and learners who are in Band 1, Band 2 and Band 3 would be classified as 'academic warning'.

The following are the four levels of reading performance specifically developed to suit Malaysian secondary school learners as adapted from the Prairie State Achievement Examination (PSAE) standards (Illinois State Board of Education, 2004). The Reference Standards of the PSAE are more holistic in orientation in that it includes learners' knowledge and skills in comprehending a variety of literary and informational texts whereas the Malaysian Performance Standards focus only on informational texts. Table 3 illustrates the descriptors of the developed Performance Standards for Year 10 respondents.

3.10.3 Determining the Cut Scores for Performance Bands

After the Prototype Test was administered and the scores obtained, they were grouped into clusters in order for Performance Bands to be determined. As these Performance Bands should as closely as possible reflect the actual performance of those taking the Prototype Test, great care was taken to ensure that the cut score for each band was accurate and able to identify the actual reading capabilities of the learners.

According to the American National Association of State Board of Education (1999), it was implausible that a single 'best' method could exist given the fact that determining acceptable performance levels and rendering them as cut scores inevitably involved arbitrary decisions arising from subjective assessments. Consequently, no absolute, unequivocal cut scores can be plausibly yielded which in turn implies that there exists no single correct or true score (Wylie and Tannenbaum, 2006). To establish the cut scores of the Performance Bands, a comparison of the reading performance of the high, average and low performers of each educational level (Year 7 to Year 11) was conducted. It was determined that 6 bands be developed to represent the learners' reading abilities and the number of bands approximate the 6 years a learner would need to complete the normal public school examination cycle; Primary School Assessment Examination, Lower Secondary Assessment and Malaysian Certificate of Education. Figure 4 reveals that learners in Year 6 should be in Band 1, learners in Year 7 should be in Band 2, learners in Year 8 should be in Band 3, learners in Year 9 should be in Band 4, learners in Year 10 should be in Band 5 and learners in Year 11 should be in Band 6.

This section explains how the cut scores for the Performance Bands (Band 1 to Band 6) were established. 1430 respondents comprising high, average and low performers from Year 7 to Year 11 sat for the Prototype Test as shown in Table 4.

Essentially, the cut scores for Band 1 to Band 6 were developed based on the z-score. The combined mean score of Year 7 to Year 11 respondents was 29.4 and the standard deviation (sd) was 11.9. The mean and the relevant raw scores were then rounded to the nearest whole number. The scores for the various performance bands were then calculated based on z-score as shown in Table 5. Finally, the respondents were categorised into the 6 bands thus generated based on their reading performance in the Prototype Test.

As an illustration, a raw score of 41, as shown in Figure 5, would be assigned a z-score of +1.00 sd because it is one standard deviation above the mean. In contrast, a raw score of 6 would be given a z-score of - 2.00 sd because it is two standard deviations below the mean. Figure 5 illustrates how the cut scores were developed based on the z-score. After determining the cut scores of the performance bands, we could then identify the specific reading abilities of the learners by referring to the Reading Matrix.

3.11 How to Develop the Decoder

The Descriptors of Reading Abilities for Band 1 to Band 6 function as the Decoder. These indicators were developed based on the respondents' reading performance in the Test. The development of the Descriptors of Reading Abilities involved the utilisation of North's 'Reading Scale for the Council of Europe Framework' as cited in Alderson (2000:132-134).

The Descriptors of Reading Abilities was primarily devised to describe a learner's reading performance and provide pedagogues and other relevant stakeholders with a holistic picture of what learners could or could not do in a certain band. Thus, it serves as a useful diagnostic tool for determining the learners' reading abilities. The performances of the learners in the written reading comprehension test were described in terms of their ability to answer literal comprehension, reorganisation and inferential comprehension questions and their scores were then represented within a Band 1 to Band 6 scoring range. Finally, the descriptor also provides the relevant stakeholders within the feedback loop to evaluate their own contributions or lack thereof in a more objective and

empirical fashion. In short, information on the reading abilities obtained can provide teachers with powerful insights on instruction (Bishop, Reyes & Pflaum, 2006).

A quantitative analysis of respondents' test results was conducted to identify the respondents' reading abilities. Learners from Year 11 of 5 selected schools were identified and consequently the descriptors were developed. The reading abilities were interpreted based on the quantitative data and qualitative data. The qualitative data were gathered through a series of interviews. Two respondents from each band were selected to be queried as to what they were capable of. The researcher went over all the questions with the respondents over a number of days. Their responses were tabulated, analysed and interpreted. An extract of the main findings is appended in Appendix 1. This information was then combined with information obtained from the quantitative data. The combination of the two forms of data resulted in a more comprehensive Descriptors of Reading Abilities. Appendix 2 shows an extract of the newly formulated descriptors.

4. How to Use READS

The procedures for the application of READS are outlined as follows:

Step 1: Conduct the Test. Learners are given 70 minutes to complete the Test.

Step 2: Use the test scores to identify the learners' reading abilities. The total score of each test is 60 marks. From the test scores, the learners are categorised into the various bands (Band 1 to Band 6) (Refer to Table 6).

Step 3: Identify the learners' reading abilities by using the Reading Matrix. Match the learners' reading performance against the Reading Matrix and then correlate them to the Performance Standards and Descriptors of Reading Abilities, ranging from Band 1 to Band 6.

Step 4: ESL teachers can refer to the Performance Standards to find out what learners from different performance levels of reading achievement could or could not do. Next, refer to the Descriptors of Reading Abilities of Band 1 to Band 6 to identify the learners' specific reading abilities.

4.1 How to Use the Reading Matrix

Table 7 and Table 8 explain how to use the Reading Matrix to identify a learner's reading ability.

For example, the reading ability of Amy, who is a Year 10 learner should correspond to Band 5 to "meet standard" but in this case, Amy is "above standard" by one band because she is in Band 6. (Refer to Performance Standards and Descriptors of Reading Abilities of Band 6). By referring to the Performance Standards, it is noted that learners who are "above standard" at their educational levels demonstrate advanced knowledge and skills in reading. Thus, the English language teacher should expose her to reading texts one level higher than her reading ability. This should be done to augment her future progress.

5. Conclusion

With this Reading Evaluation and Decoding System, ESL teachers can now identify where their learners are i.e. their current reading abilities. The test scores provide invaluable information that can inform teachers, school, parents and learners about the learners' specific reading abilities. At the same time, it can also help diagnose reading achievement problems and inform decision making in the classroom, the school, the district and the state.

References

- Alderson, J.C. (2000). *Assessing reading*. Cambridge: Cambridge University Press.
- Alderson, J. C. & Urquhart, A. H. (1984). *Reading in a foreign language*. Essex: Longman.
- Allen, J. (2000). *Yellow brick roads: Shared and guided paths to independent reading*. Portland, ME: Stenhouse.
- Baguley, P. (2001). *Performance management in a week*. London: Hodder & Stoughton.
- Baker, M. W. (2004). Assessment, diagnosis, and evaluation. In Baker, M. W. (Ed.), *Preparing reading professionals: A collection from the International Reading Association* (pp.193). Newark: International Reading Association.
- Bishop, P.A., Reyes, C. & Pflaum, S.W. (2006). Read smarter, not harder: Global reading comprehension strategies. *The Reading Teacher*, 60(1), 66.
- Cobb, C. (2005). Effective instruction begins with purposeful assessments. In Barrentine, S. J., & Stokes, S. M. (Eds.), *Reading assessment: Principles and practices for elementary teachers* (2nd ed.) (pp.21). Newark: International Reading Association.
- Ediger, M. (2009). For an effective reading program. *Reading Improvement*, 46(3), 119.

- Farr, R. (2003). Building useful instructional reading assessments. *New England Reading Association Journal*, 39 (1), 1.
- Gay, L.R. & Airasian P. (2003). *Educational Research: Competencies for analysis and applications*. New Jersey: Merrill Prentice Hall.
- Green, S. (2002). Criterion referenced assessment as a guide to learning - The importance of progression and reliability. *In a paper presented at the Association for the study of evaluation in education in Southern Africa International Conference, Johannesburg*. 10 July – 12 July 2002. Online. Accessed on 12 July 2004. <http://www.ucles-red.cam.ac.uk/Conferencepapers/ASEESA2002.PDF>
- Harmer, J. (1991). *The practice of English Language teaching*. London: Longman.
- Horton, T. (ed). (1990). *Assessment debates*. London: Hodder & Stoughton.
- Illinois State Board of Education (2004). *Prairie State Achievement Examination (PSAE): Teachers' handbook 2003-2004*. Illinois: Illinois State Board of Education. Online. Accessed on 4 July 2004. <http://www.isbe.net/assessment/PDF/2004/PSAETchHndbk.pdf>
- Johnson, P. & Costello, P. (2005). Principles for literacy assessment. *Reading Research Quarterly*, 40(2), 256.
- Kubiszyn, T. & Borich, G. (2000). *Educational testing and measurement: Classroom application and practice*. New York: John Wiley & Sons Inc.
- Lenski, S. D., Ehlers-Zavala, F., Daniel, M.C. & Sun-Irminger, X. (2006). Assessing English-language learners in mainstream classrooms. *The Reading Teacher*, 60(1), 33.
- Ministry of Education Malaysia. (1991). *Compendium: A handbook for ELT Journal teachers, Volume Three*. Kuala Lumpur: Ministry of Education.
- Ministry of Education Malaysia (2003). *Huraian Sukatan Pelajaran: Bahasa Inggeris Tingkatan 4 (Curriculum Specifications English Language Form 4)*.
- Mok, S.S. (2000). *Ilmu pendidikan untuk KPLI (Kursus Perguruan Lepas Ijazah)*. Subang Jaya: Kumpulan Budiman Sdn. Bhd.
- National Association of State Board of Education (NASBE) (1999). Issues in brief: Setting cutscores on large-scale assessments. *The State Board Connection*. Vol. 19. No.2 July.
- OECD (2006). *Assessing scientific, reading and mathematical literacy: A framework for PISA 2006*. Online. Accessed on 25 November 2008. www.oecd.org/dataoecd/63/35/37464175.pdf.
- Oosterhof, A. (2001). *Classroom application of educational measurement* (3rd ed). Upper Saddle River: Merrill Prentice Hall.
- Popham, W. J. (2002). *Assessment: What teachers need to know*. Boston: Allyn and Bacon.
- Rabbini, R. (2002). An introduction to syllabus design and evaluation. *The Internet TESL Journal*, 8(5). Online. Accessed on 25 November 2008. <http://iteslj.org/Articles/Rabbini-Syllabus.html>
- Routman, R. (2003). *Reading essentials: The specifics you need to teach reading well*. Portsmouth: Heinemann.
- Temple, C., Ogle, D., Crawford, A. & Freppon, P. (2008). *All children read: Teaching for literacy in today's diverse classrooms*. Boston: Pearson Education, Inc.
- Tomlinson, C.A. (2009). Intersections between differentiation and literacy instruction: Shared principles worth sharing. *New England Reading Association Journal*, 45(1), 28.
- Wasburn-Moses, L. (2006). 25 Internet sources for teaching reading. *The Reading Teacher*, 60(1), 70.
- Weeden, P., Winter, J. & Broadfoot, P. (2002). *Assessment: What's in it for schools?* London: Routledge Falmer.
- Wilcox, B. L. (2005). Changing attitudes on assessment. In Barrentine, S. J. & Stokes, S. M. (Eds.), *Reading assessment: Principles and practices for elementary teachers* (2nd ed.) (pp.102). Newark: International Reading Association.
- Wylie, E. C. & Tannenbaum, R. J. (2006). *TOEFL® Academic speaking test: Setting a cut score for international teaching assistants*. ETS Research Reports: Princeton. Online. Accessed on 19 July 2008. www.ets.org/Media/Tests/TOEFL/pdf/ngt_itastandards.pdf
- Zarrillo, J. (2007). *Are you prepared to teach reading: A practical tool for self-assessment*. Upper Saddle River: Pearson Education, Inc.

Table 1. Sub-skills of Reading

| Skill categories in Barrett’s Taxonomy | Skill specifications in the Malaysian English Language Syllabus/textbook | Barrett’s Taxonomy of Reading Comprehension |
|---|--|--|
| Literal Comprehension (concern with information stated explicitly in the text) | i. scan for certain details ii. identify main ideas in a simple text iii. identify supporting details in a simple text | i. Recognition of details ii. Recognition of main ideas iii. Recognition of sequence iv. Recognition of comparisons v. Recognition of cause and effect relationships |
| Reorganisation (analysing, synthesising, and organising information that has been stated explicitly) | i. extracting supporting details ii. acquiring the meaning of words by understanding contextual clues iii. identifying simple cause and effect | i. Classifying ii. Outlining iii. Summarising (paraphrasing or condensing a selection) iv. Synthesising (Consolidating information from more than a single source) |
| Inferential Comprehension (using information explicitly stated along with one’s own personal experience as a basis for conjecture and hypothesis) | i. making inferences ii. drawing conclusions in simple texts | i. Inferring supporting detail ii. Inferring main ideas iii. Inferring sequence iv. Inferring comparisons v. Inferring cause and effect relationships |

Table 2. Pilot Test to Find Out Time Taken to Complete Test for Year 10 Learners

| Educational Level | Average Time taken by: | | |
|---|------------------------|--------------------|----------------|
| | High Performers | Average Performers | Low Performers |
| Year 10 | 60 minutes | 75 minutes | 85 minutes |
| Average time taken to complete test by all groups | 70 minutes | | |

Table 3. Performance Standards for Year 10 Respondents

| | |
|------------------|--|
| Above Standard | <ul style="list-style-type: none"> Learners able to achieve the learning outcomes related to the sub-skills of reading to be achieved by learners in Year 11. Learners able to fulfil the requirements specified in the Malaysian English Language Syllabus and Barrett’s Taxonomy of Reading Comprehension. |
| Meet Standard | <ul style="list-style-type: none"> Learners able to achieve the leaning outcomes related to the sub-skills of reading to be achieved by learners in Year 10. Learners able to meet the requirements specified in the Malaysian English Language Syllabus and Barrett’s Taxonomy of Reading Comprehension. |
| Below Standard | <ul style="list-style-type: none"> Learners not able to achieve the learning outcomes related to the sub-skills of reading to be achieved by learners in Year 10. Learners have gaps in reading at their educational level (Year 10), partially meeting the requirements specified in the Malaysian English Language Syllabus and Barrett’s Taxonomy of Reading Comprehension. |
| Academic Warning | <ul style="list-style-type: none"> Learners not able to achieve the leaning outcomes related to the sub-skills of reading in Year 7, Year 8 and Year 9. Learners have major gaps in reading at their educational level (Year 10), not meeting the requirements specified in the Malaysian English Language Syllabus and Barrett’s Taxonomy of Reading Comprehension |

Table 4. Number of Respondents for Each Educational Level

| Educational Level | Year 7 | Year 8 | Year 9 | Year 10 | Year 11 | Total |
|-----------------------|--------|--------|--------|---------|---------|-------|
| Number of Respondents | 307 | 298 | 332 | 248 | 245 | 1430 |

Table 5. Establishing Scores for Bands

| Bands | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 |
|--------|--------------------|------------------------------|--------------------------------|--------------------------------|--------------------------------|-----------------------|
| Scores | Below 6 (0 – 6) | Between 6 and 18 (7 – 18) | Between 18 and 29 (19 – 29) | Between 29 and 41 (30 – 41) | Between 41 and 53 (42 – 53) | Above 53 (54 – 60) |

Table 6. Performance Bands and the Scores

| Bands | Band 6 | Band 5 | Band 4 | Band 3 | Band 2 | Band 1 |
|--------|---------|---------|---------|---------|--------|--------|
| Scores | 54 - 60 | 42 - 53 | 30 - 41 | 19 - 29 | 7 - 18 | 0 - 6 |

Table 7. Amy’s Reading Performance

| Learner | Year | Score | Band |
|---------|------|-------|------|
| Amy | 10 | 55/60 | 6 |

Table 8. Charting Reading Performance

| LEVELS | | | | | | |
|---------|---|---|---|---|---|-----|
| Year 11 | | | | | | X |
| Year 10 | | | | | X | Amy |
| Year 9 | | | | X | | |
| Year 8 | | | X | | | |
| Year 7 | | X | | | | |
| BANDS | 1 | 2 | 3 | 4 | 5 | 6 |

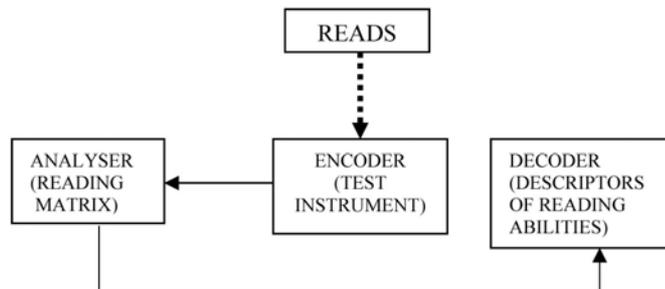
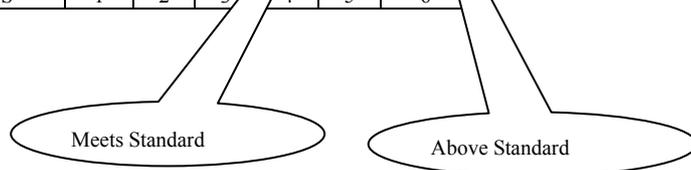


Figure 1. Reading Evaluation and Decoding System

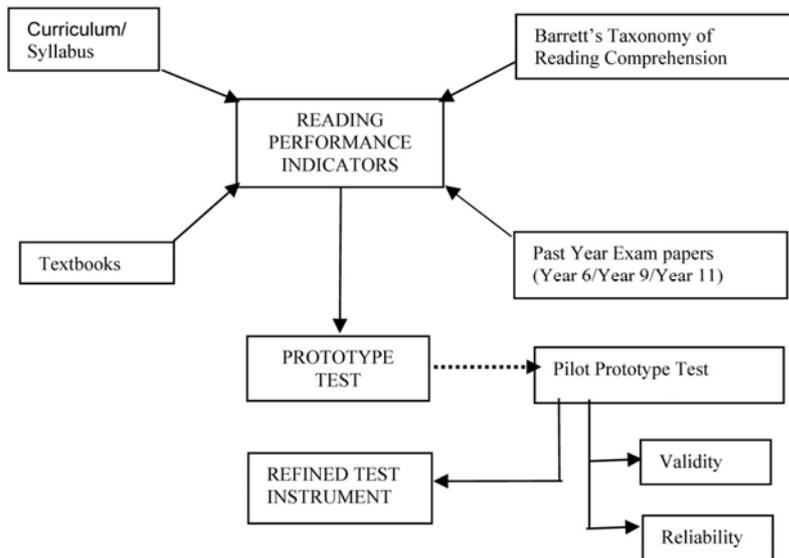


Figure 2. Developing Test Instrument

| LEVELS | | | | | | |
|---------|---|----------|----------|----------|----------|----------|
| Year 11 | | | | | | X |
| Year 10 | | | | | X | |
| Year 9 | | | | X | | |
| Year 8 | | | X | | | |
| Year 7 | | X | | | | |
| BANDS | 1 | 2 | 3 | 4 | 5 | 6 |

Figure 3. Reading Matrix

X indicates meet standard

- i. Levels refer to the educational levels of the learners, i.e. Year 7 to Year 11 in a Malaysian secondary school
- ii. Bands refer to the performance indicators which indicate the learners’ reading abilities.

| | | | | | | |
|--------------------|--------|--------|--------|--------|---------|---------|
| Educational levels | Year 6 | Year 7 | Year 8 | Year 9 | Year 10 | Year 11 |
| Performance bands | Band 1 | Band 2 | Band 3 | Band 4 | Band 5 | Band 6 |

Figure 4. Performance Bands and Educational Levels

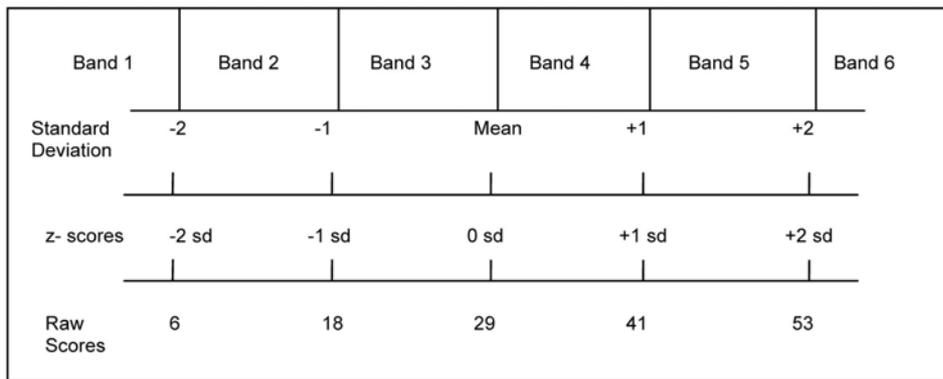


Figure 5. Cut Scores based on z-scores

Appendix I. Data Notes from Interview

| Difficulty Levels | Items | Responses (Simplified Text) | | | | | |
|----------------------------|---|--|--|--|---|--|--|
| | | BAND 6 (54 – 60) raw score | BAND 5 (42 – 53) raw score | BAND 4 (30 – 41) raw score | BAND 3 (19 – 29) raw score | BAND 2 (7 – 18) raw score | BAND 1 (0 – 6) raw score |
| Literal (Year 6 Level) | How did you choose your answers for questions number 6, 8, 11, 12 and 55? | Can find directly stated information easily. | Scan for directly stated information in the text. | Find answer directly stated in the text. | Find answer directly stated in the text but not sure of the answer. | Can understand only a few phrases. Guess answer. | Can understand only a few words. Guess answer. |
| | What did you do to get the answer for question number 1? | Skim text for look for the main idea. | Skim text to identify the gist. | Rereads to look for the main idea. | Skim text to look for the answer. | Can understand only a few phrases. Guess answer. | Can understand only a few words. Guess answer. |
| Literal (Year 9 Level) | What about questions number 16, 17, 21, 27,29,31, 33, 36, 39 and 42, how did you get the answers? | Can locate answer directly stated in text | Read text to find directly stated information in the text. | Identify key words to locate key information. | Identify the key words in the question and text. | Do not understand the question. Guess answer. | Do not understand the text and question. Answer through pure guessing. |
| | For questions number 20, 26, 41 and 46, how did you arrive at the answers? | Skim text to locate for information | Skim the text to identify the answer. | Skim text to locate key information. | Reread text to locate answer. Not sure of the answer. | Do not understand the question. Guess answer. | Do not understand the text and question. Answer through pure guessing. |
| Literal (Year 11 Level) | How did you get the answers for questions number 47, 48, 51, 52, 53 55, 56 and 59? | Skim text to locate information. | Read text to scan for the answer. | Can locate answer from the text but not aware of the specific meaning of some details. | Do not know the difference of some specific details. | Cannot find answer in the text. Guess answer. | Do not understand the question. Simply guess answer. |
| | What about question number 46, how did you get the answer? | Skim text to look for main ideas. | Skim text to locate answer. | Skim text to locate key information. | Reread text to locate main idea. | Cannot find answer in the text. Guess answer. | Do not understand the question. Guess answer. |

Appendix 2. Descriptors of Reading Abilities: Literal Sub-skill

| Difficulty Levels | Sub-skills of reading | BAND 6 (54 – 60) raw score | BAND 5 (42 – 53) raw score | BAND 4 (30 – 41) raw score | BAND 3 (19 – 29) raw score | BAND 2 (7 – 18) raw score | BAND 1 (0 – 6) raw score |
|-------------------------|---|---|--|---|---|---|---|
| Literal (Year 6 Level) | Identifying supporting details in texts | Can locate the supporting details very well . Can find directly stated information easily. | Can locate the supporting details very well . Can scan directly stated information in the text. | Can locate the supporting details very well . Can find answer directly stated in the text. | Can locate the supporting details very well . Can find answer directly stated in the text but not sure of the answer. | Can locate the supporting details satisfactorily . Can understand only a few phrases. Guess answer. | Can hardly locate the supporting details. Can understand only a few words. Guess answer. |
| | Identifying main ideas in texts | Can locate the main ideas very well . Skim text to look for answer. | Can locate the main ideas very well . Skim text to identify the answer. | Can locate the main idea very well . Reread text to skim for the answer. | Can locate the main ideas very well . Reread text to look for the answer. | Can locate the main ideas satisfactorily . Can understand only a few phrases. Guess answer. | Can hardly locate the main ideas. Can understand only a few words. Guess answer. |
| Literal (Year 9 Level) | Identifying supporting details in texts | Can locate the supporting details very well . Can locate answer directly stated in the text. | Can locate the supporting details very well . Can scan directly stated information in the text. | Can locate the supporting details very well . Can identify key words to locate the answer. | Can locate the supporting details moderately well . Can identify key words in the question and answer. | Can hardly locate the supporting details. Do not understand the question. Guess answer. | Can hardly locate the supporting details. Do not understand the text and question. Answer through pure guessing. |
| | Identifying main ideas in texts | Can locate the main ideas very well . Skim text to look for answer. | Can locate the main ideas very well . Skim text to identify the answer. | Can locate the main ideas very well . Skim text to locate the answer. | Can locate the main ideas moderately well . Reread text to locate the answer. Not sure of the answer. | Can hardly locate the main ideas. Do not understand the question. Guess answer. | Can hardly locate the main ideas. Do not understand the text and question. Answer through pure guessing. |
| Literal (Year 11 Level) | Identifying supporting details in texts | Can locate the supporting details very well . Skim text to look for answer. | Can locate the supporting details very well . Read text to scan the answer. | Can locate the supporting details satisfactorily . Can locate answer from the text but not aware of the specific meaning of some details. | Has difficulty locating the supporting details. Do not know the difference of some specific details. | Can hardly locate the supporting details. Cannot find answer in the text. Guess answer. | Can hardly locate the supporting details. Do not understand the question. Answer through pure guessing. |
| | Identifying main ideas in texts | Can locate the main ideas very well . Skim text to look for the answer. | Can locate the main ideas very well . Skim text to locate the answer. | Can locate the main ideas satisfactorily . Skim text to locate information. | Has difficulty locating the main ideas. Reread text to locate the answer. | Can hardly locate the main ideas. Cannot find answer in the text. Guess answer. | Can hardly locate the main ideas. Do not understand the question. Answer through pure guessing. |