DEVELOPING A THAI READS ENCODER TO GAUGE EFL READING PROFICIENCY OF THAI UNDERGRADUATE STUDENTS

Thapanee Khemanuwong[1]
Abdul Rashid Mohamed[2]
Shaik Abdul Malik Mohamed Ismail[2]

thapaneekhe@hotmail.com

[1]Faculty of Liberal Arts, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand
[2]School of Educational Studies, Universiti Sains Malaysia, Penang, Malaysia

ABSTRACT

Reading is not simply an understanding of graphic symbols, but it is also the interaction between a language and thoughts as it is a message coded in written form that needs to be understood. A very vital skill to master, reading has become an essential component of English as a Foreign Language (EFL) programme that inevitably needs an evaluation part. The component, however, provides no specific information about possessed performance of each student. Presently, there is an existing READS (Reading Evaluation and Decoding System) used as a diagnostic tool in assessing new intake of English reading proficiency at Universiti Sains Malaysia beginning in the academic year of 2014-2015. The existing READS, however, is specifically used on the basis of Malaysian school curricula and apparently has Malaysian culture embedded in its encoder. As a result, the encoder presents a significant amount of contextual biases. This study aims at developing a Thai localised version of READS encoder to serve as an indication of the specific skills each student possesses. A model Thai READS was developed and piloted to more than six hundred Thai undergraduate students in evaluating the students' performance. The study revealed that the Thai variation of READS is effective in evaluating the students' performance and decoding their reading skills.

Keywords: READS, EFL, Encoder, Thailand

INTRODUCTION

This article discusses some significant aspects of the development of a Thai READS encoder that would be utilised in reading ability assessment of Thai undergraduates. Before discussing the development, it is essential to draw a picture of the existing READS, which has currently been employed by Universiti Sains Malaysia (USM). The READS is comprised of three components as shown in Figure 1.
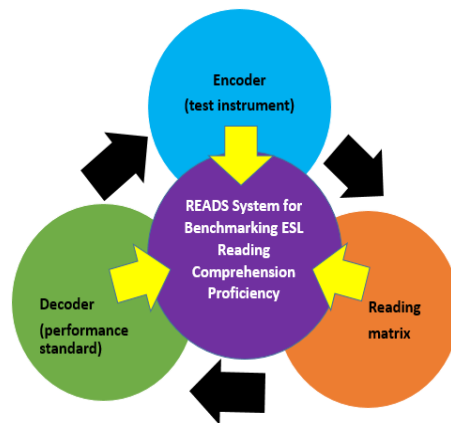
Figure 1. Reading Evaluation and Decoding System (READS)

The READS, as a whole system, is comprised of 60 multiple-choice reading comprehension questions allowing lecturers to administer standardised tests and gather earned marks for further analysis. The questions are proportionally based on distribution of difficulty levels with 25%, 50% and 25% taken as easy, average and difficult, respectively (Mok, 2000, as cited in Abdul Rashid Mohamed, 2010). Each student will be mapped to an appropriate range through a cross-reference from the analysis to a Reading Matrix, whereby primary support intervention can be provided to those students graded as "below average" or "academic warning" (Wasburn-Moses, 2006). Predominantly, Thai EFL lecturers are unable to monitor the students' EFL reading abilities. There have not been any indicators that precisely tell when a particular student is in need of help and how he or she needs it. In other words, it is a summative test without a proper set of criteria. With the lack of adequate information about the standard of reading, it would be impossible to accurately determine their reading abilities. To solve these problems, it would be more practical to have some indications that can identify specific skills each student possesses and provide a score that details the skills to be improved and how to achieve the improvements. This could be of the students' help in paving the way towards autonomous learning of what they have studied in class (Keshavarz. & Ashtarian, 2008).i

Nonetheless, there was a disagreement over reliability and validity of the testing instrument since this test was primarily developed for Malaysian students and thereby involved Malaysian culture, which brought about certain contextual biases. Therefore, the original READS was deemed unfit for deployment amongst Thai target audience. Adapting the existing instrument, rather than developing a new one specifically for the target audience, offers a considerable amount of advantages, as using an adapted version logically permits a greater ability to generalise within an increasingly diverse population (Hambleton, 2005). Still, adapting a testing instrument is a complex task that involves a careful arrangement, especially materials and general validity for the population intended (Cassepp-Borges, Balbinotti, & Teodoro, 2010 as cited in Borsa, Damásio, & Bandeira, 2012). It is also noteworthy that the adaptation of an instrument maintains a cultural fit— being a preparation for practice in different cultural contexts (Hambleton, 2005; Sireci, Yang, Harter, & Ehrlich, 2006).

LITERATURE REVIEW

What is a Contextual Bias?

Human perceptions do not function like a camcorder that constantly captures and stores every piece of data encountered. Instead, the way of perceiving the world and remembering things is shaped by our worldly knowledge and experience. Contextual clues

are also used to help sort, examine and evaluate a slew of information getting inside our senses. The schema theory emphasises that reading comprehension is an interactive process between the reader's previous background knowledge and the text. According to the theory, EFL readers' reading comprehension is not only influenced by how easy or difficult a text is to them but depends more on the level of being able to recall their culturally familiar background knowledge and draw cultural origin clues from the context (Carrell, 1984; Carrell, 1987; Carrell & Floyd, 1989).

The schema theory also points out that test takers' background knowledge could itself have an effect on test performance. The relationship between background knowledge and reading comprehension, in native language, has been investigated extensively, with results revealing that having background knowledge of a text would help make it more understandable (Weber, 1991). Some researchers have also provided evidence for a potential role of background knowledge in reading comprehension in a foreign language. In 1984, Anderson and Pearson conducted a study, which showed that the readers were able to fully comprehend if they already had a bit knowledge of what was in the text. Bernhardt's (2005) findings also suggested that texts containing culturally and contextually familiar content schema were easier to digest. The ability to understand a text is based not only on the reader's linguistic knowledge, but also on worldly, general knowledge and the extent to which it is activated while reading (Yousef, Karimi, & Janfeshan, 2014). Accordingly, EFL students' worldly experience and familiarity can be helpful when it comes to reading comprehension if the reading is culturally or contextually related to them.

Contextual bias involves the complex relationship of informational cues, like worldly knowledge, for instance. It has been known that worldly knowledge influences what is understood from text. Several studies suggested that worldly knowledge is an integral part of the comprehending process. (Bransford & Johnson, 1972; Johnston, 1981). When an ambiguous word is encountered, one with worldly knowledge would be able to come up with the whole meaning. This implies that two individuals with equal reading comprehension abilities but different worldly knowledge would exhibit different degrees of comprehending the same test.

Contextual bias is unquestionably one of the natural influences that potentially dominates a test taker's judgement. Test takers are influenced by this kind of bias in an unwitting and unintentional mode, as familiarity with names, places, things and concepts mentioned in the test questions could help them to comprehend the context more easily, without which familiarity the text could be harder to comprehend. In line with the above, Westin (2006) found that the test takers could better comprehend passages that were contextually related to them. Reading passages that were culturally, nationally and contextually bound to the test takers help them to derive more meanings. They would subsequently be better off with worldly knowledge of context. Studies conducted by Yu (2008) and Orellana and Reynolds (2008) also suggest that contextual settings of passages are instrumental in the test takers' performance since they tend to do better with contextually familiar materials. Remarkably, the test takers, when encountering a familiar item, will approach it with confidence. In contrast, those with no familiarity with context could possibly become confused about the cognitive mechanism and will come up with a wrong answer.

What are the Concerns?

A number of research studies were dedicated to the analysis of how different sources have an impact on the processing of lexically ambiguous words. Based on the research, it is apparent that lexical uncertainty resolution is affected by both contextual information and meaning frequency (Binder & Morris, 1995). Currently, numerous eye movement studies were conducted to determine how lexical ambiguity is processed (Binder & Morris, 1995). In these studies, the readers' eye movements were examined for fixation while they were reading sentences and passages. The results were straightforward that the ambiguous

target words needed higher processing time than did the familiar words. Unfamiliar with those ambiguous words, the test takers could possibly become confused and discouraged. With a limited time slot, wrong answers could be easily picked.

As an integral part of reading comprehension, Carrell (1984) refers to the "interactive parallel process between the reader's background knowledge and the text" and she states: "... much of the meaning grasped from a text is actually not in the text, *per se*, but also in the background knowledge of the reader." In addition, Bernhardt (1991) points out the same concept: "Generic reading tests can be easily biased and, thus, may be incapable of accurately gauging reading abilities". Therefore, an assessment of reading not involving readers' worldly knowledge will fail a validation since it does not achieve the vital factors to comprehension construct.

Undoubtedly, both familiarity with the language and contextual information has an effect on test performance. Carrell (1988) also says: "The most obvious reason why a specific schema may fail to exist for a reader is that the schema is culturally specific and not part of a particular reader's background". A study by Alderson and Urquhart (1985) reveals that based on texts taken from their own subject discipline, students from a particular discipline would achieve better performance on tests than those from other disciplines, that is, the students appear to be in an advantageous position when taking a test with text in their areas of familiarity.

Contextual bias is considered a significant factor that systematically influences the students' test scores. Worldly or background knowledge represents such factor. For students at the limit of linguistic abilities, once a reading passage is outside of their experience, they are adrift on an unknown sea (Aebersold & Field, 1997). When faced with such unfamiliar texts, some students may overcompensate for absent schemata by reading in a slow, text-bound manner; others may overcompensate by wild guessing (Carrell, 1988). Both strategies obviously result in comprehension difficulties. The issue is what to do for the problems. It is advisable to either construct a reading comprehension test with contextual familiarity or less dependent on worldly knowledge or create a valid indicator that measures real comprehension of reading rather than worldly knowledge. Lastly, basic bottom-up practice must not be ignored. EFL students require training of words and structure recognition to activate and improve the schemata they need to comprehend the text.

METHODOLOGY

Adapting the Original READS

Three individuals specialised in English language assessment or particularly knowledgeable about what the instrument assesses were involved in the adaptation of READS. These experts were given a paper-based testing instrument of the original READS to share opinions about biases inherent in the encoder and ways to adjust it. Other essential aspects such as encoder structure and layout and difficulty levels were then assessed. Simultaneously, the assessors made consideration on, for instance, whether the vocabulary words can be applied to Thai context and the target audience and whether they are suited for those the testing instrument was intended for. In a reading passage, for instance, 1 RM was replaced with Thai currency of 10 baht, and Tesco, a name commonly used in Malaysia, was changed to Tesco Lotus, in Thai context. More importantly, the term "Prathom 6" was substituted for the term "USPR" to reflect the Thai environment. It stands to reason that one may do poorly because of lacking worldly knowledge while another one, with adequate worldly knowledge, would outperform.

Word clarity, suitability of typefaces and their sizes, and data preparation for the instrument was also analysed. It is requisite to examine contextual bias existence as some educators feel that it is unsuitable to simply adopt standardised tests of other nations. There has not been any indication that identifes how cultural differences have an influence on

human learning capacity (Boondao, Hurst, & Sheard, 2009). Accordingly, cultural and background differences may lead to assessment inaccuracies (Kim & Zabelina, 2015). Predominantly, the cultural interference in reading comprehension is related to social cognitive functions of the language (Collier & Thomas, 2007). Subject related texts, in particular, may also discriminate against individuals with less background knowledge. The results are compatible with the findings of Lipson (1984) who suggests that partial knowledge can interfere with comprehension if it conflicts with text information. If an adopted test, aimed at measuring reading comprehension, is flawed by a cultural bias, then the test could be gauging students' background knowledge of the test developer's culture rather than their reading comprehension. With cultural interference, the students may answer the test questions wrongly. Lacking schemata previously, they could be experiencing difficulty in comprehending what they read.

In alignment with this, according to the experts' evaluations, the adopted test is not suited for target audiences with cultural and background differences. As an example, according to a passage in the original READS, Malaysian people may have background knowledge pertaining to UPSR (Ujian Pencapaian Sekolah Rendah). UPSR is known as a national assessment taken by all Malaysian primary students at the end of their sixth year. In contrast, the Thai test takers are not accustomed to UPSR since they do not have a schema (background knowledge) of this examination. Associating a part of text with a related matter beyond the passage is not easy, possibly because of their poor worldly or background knowledge (Perez-Amurao, 2011). As a result, without worldly information, they will not be able to link their experience to what they are reading and activate their cognitive mechanism to comprehend it.

Validating a Thai READS

Content validity is for the most part determined by experts (Gay & Airasian, 2003). Three content experts, comprised of one CU-TEP and CU-AAT test writer (CU-TEP and CU-AAT stand for Chulalongkorn University Test of English Proficiency and Chulalongkorn University Academic Aptitude Test, respectively) and two senior TEFL university lecturers, examined the suitability of the questions to test the content validity of the testing instrument. Key domains for the content validity in question include contextual biases, styles of passages, lengths of passages, levels of difficulty, and levels of vocabulary words used.

Each question was examined for Thai READS validation by using IOC (Index of Consistency: IOC). Item Objective Congruence (IOC) indexes were then computed based on the experts' opinions. According to Gay and Airasian (2003), scoring rubrics criteria give satisfactory IOC indexes when over 0.50. The findings revealed that the Thai READS was suitable with rated applicability from the experts of 0.92. Hence, these findings demonstrate that the questions were considered suitable because of high content validity.

Reliability of Thai READS

KR 20 is known as an index of internal consistency of the test. "Internal consistency" refers to consistency of students answering the questions. KR 20 can be understood as measuring the extent to which the test questions consistently provide information about a student's level of content knowledge assessed by the test. Assuming that all the questions on a test share a single content domain, students with a very high level of knowledge domain are expected to answer most questions correctly and those with a very low level of knowledge domain to answer most questions wrongly.

Concerning item consistency of a test, Gay and Airasian (2003) mention that internal consistency reliability provides the information needed. Since the Thai READS consisted of multiple-choice questions, it is unquestionable that the exploitation of Kuder-Richardson formula known as KR20 is most suitable. As the matter of fact, with multiple-choice test

questions, the internal consistency procedure most commonly used was the Kuder-Richardson (Oosterhof, 2001). In contrast, the Coefficient Alpha developed by Cronbach was not used as it is appropriate for calculating internal consistency analysis of Likert-type scale rather than marked as dichotomous choices.

According to a pilot study, a Thai READS test was administered to measure reliability levels. Selected respondents from King Mongkut's Institute of Technology Ladkrabang comprised of 624 respondents with distinct education levels were subjected to a Thai READS test. Analysis of test reliability was performed by using IBM SPSS Statistics version 22 as shown below:

Table 1. Analysis of Test Reliability

Reliability Test	Thai READS Reliability Value	N of Items	N of Respondents
KR20	0.91	60	624

The scores for KR-20 range from 0 to 1.0, where 0 is no reliability and 1.0 is perfect reliability. The closer to 1.0 is considered more reliable.

The Thai READS was 0.91 high in statistical result. Therefore, the Thai READS is more suited to Thai undergraduates as it demonstrates a high KR20 coefficient (e.g. > 0.90), which serves as an indication of a homogeneous test.

Pilot Study

Besides the experts' evaluation of the original and adapted versions of READS, the pilot study was to check and confirm Thai READS applicability. An explanation of methodology used in the pilot study was described as follows:

Respondents

In this pilot study, a group of 59 students with 16 males and 43 females aged between 17-20 attended the pilot session. The respondents were assigned to take both Thai (adapted) and original READS tests one by one on different days. They were expected to complete the tests exactly as it would be done by the target audience. This was also to help decide on time allotments needed for completing the tests.

Instrument

The original and (contextually familiar) Thai versions were employed in the pilot phase. In the Thai one, to remove contextual biases, names were replaced with words that the Thai respondents were more accustomed to.

Procedure

The pilot tests were administered one by one on the same group of respondents on different days. Before taking the tests, the respondents were given a brief explanation on how to do them. They were also assured that their test results would be kept confidential.

Data Analysis and Results

Data analysis was performed by using IBM SPSS Statistics version 22. In the analysis, paired-samples t-test was used to compare the mean scores of the same participant group on matched pairs. This was to find out if there was any significant difference between the respondents' performance on the original and (contextually familiar) Thai versions. A paired-samples t-test determines whether or not a statistically significant difference exists in the mean scores for the Thai and original versions of READS. Simply

put, this was to figure out whether there was a significant difference between the adapted and original READS tests.

Hypotheses are as follows:

H₀: No statistically significant difference exists in the mean scores obtained from the original and adapted READS's tests.

H₁: Statistically significant difference exists in the mean scores obtained from the original and adapted READS's tests.

P-value was used to test the addressed hypotheses. If the P-value is less than (or equal to) 0.05, then null hypothesis is rejected in favour of the alternative hypothesis. Besides, if the P-value is higher than α, then null hypothesis is retained. Level of significance, also denoted as alpha or α, is a probability for rejecting null hypothesis when it is true. A significance level of 0.05, for instance, indicates a 5% risk of concluding that a difference exists when indeed there is no difference. According to a pilot test, the P-value is 0.00, which is less than the significance level, meaning that null hypothesis is rejected. In other words, there is a significant difference between the original READS's test and the adapted one, which can be deduced that the Thai READS test mean scores are higher than those of the original counterpart.

Table 2: Descriptive Statistics

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Thai	43.0678	59	6.11348	0.79591
	Original	38.2542	59	4.93612	0.64263

Table 3: Paired Samples Statistics

Paired Samples Test

		Paired Differences								
					95% Confidence Interval of the Difference					Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df		
Pair 1	Thai - Original	4.81356	5.15108	.67061	3.47118	6.15594	7.178	58		.000

*At significance level of 0.05

In order to answer the hypotheses concerning differences between the respondents' performance on the Thai and original versions of READS, a paired samples t-test was used to analyse the quantitative data. For the Fear of Statistics Test (FOST), a paired-samples t-test was used to determine the impact of intervention on the students' earned marks. There was a statistically significant decline in FOST scores from the Thai READS (M= 43.07, S.D. = 6.11) to the original READS (M= 38.25, S.D. = 4.94), t (58) = 7.18, $p =$. 00 (two-tailed). The mean decline in FOST scores was 4.81 with a 95% confidence interval ranging from 3.47 to 6.16.

The t-test results also demonstrated that the Thai READS test mean scores (M = 43.07, S.D. = 6.11) was significantly higher than those of the original READS (M= 38.25, S.D. = 4.94), t (58) = 7.18, p = .00 (two-tailed), That is, after contextual biases in the original

READS were removed, the students' mean scores in the Thai READS were higher than those of the original READS. Thus, it can be deduced that a statistically significant difference (p< .00) exists in the respondents' performance on the Thai READS test, as opposed to the original version. Therefore, the alternative hypothesis was accepted.

DISCUSSION

As stated previously, the main purpose of this study is to develop an unbiased reading proficiency assessment for Thai undergraduate students. This study adapted the original READS to suit the needs of the Thai environment based on relevant literature that has been reviewed and the opinions received from experts.

In the READS adaptation, the original READS is adopted and only one main factor affecting student test results, which is cultural bias, was adapted. For the most part, cultural bias is considered reading comprehension interference. As mentioned earlier, if an adopted READS test not having student prior schemata is culturally biased, it could be testing student familiarity with Malaysian culture rather than comprehension of reading itself. In this study, contextual bias is found as a major cultural bias in the original READS, so adaptation will be made to make READs familiar in the Thai context.

The present study's findings are supportive of, and in alignment with, those of Sasaki (2000). Sasaki (2000) examined how schemata induced by contextually familiar words could influence students' cloze test-taking process. In this study, with their equivalent English ability, sixty Japanese EFL students were divided into two groups. In the adapted version, unfamiliar words in the original passage were changed to contextually familiar ones. The results revealed that those who read the contextually familiar passages endeavoured to solve more questions and, in general, had a better understanding of the text passages, which has resulted in better performance than those in the unfamiliar group. The t-test results indicated that the familiar group's mean score on the adapted version was significantly higher than the unfamiliar group's (t =3.18, p= .01).

By means of process-oriented methodology and the same cloze tests used by Chihara et al.'s, (1989), Sasaki found that the test takers could improve their scores in a test with contextually familiar materials. Further, he found that not only would the test takers comprehend and hence perform better on contextually familiar cloze tests, but they would become more successful in recalling the content of such texts as opposed to contextually unfamiliar ones. In addition, Horiba (1996) found that linguistic knowledge is supported by the formation of schemata and that the supported linguistic knowledge potentially leads to test takers' better interpretation of texts and hence better performance on tests.

In concert with the above, Sasaki (2000) reported that the respondents who did a contextually familiar cloze test (i.e. familiar group) earned higher scores compared with those who did a contextually unfamiliar one (i.e. unfamiliar group). It was found that the unfamiliar group missed the questions they could undoubtedly have been solved using their background knowledge had the given passages been more familiar. In trying to solve the questions, they might have been overwhelmed by unfamiliar passages. In other words, the test takers in the familiar group attempted considerably more questions than those in the unfamiliar group. Sasaki (2000), therefore, suggested that the test takers' awareness of being able to form a coherent story out of contextually familiar tests might have gotten them highly motivated and hence they got through more questions in the process.

Bachman (1990) mentions that if the purpose intended includes measuring the test takers' linguistic ability, then contextually unfamiliar passages may pose a serious threat to acceptable test score interpretation. In short, passages that were nationally, culturally and contextually related to test takers help them to derive more meaning and thereby they got improved scores compared with those with no familiarity with the passages. (Choy, Lee, & Sedhu, 2014). In such a case, it is suggested that test developers choose texts that are familiar enough to test takers to enable their full utilisation of knowledge. Carrell (1988) and

Horiba (1996) also concluded that substituting contextually unfamiliar words with more familiar ones has certain impact on the respondents' test-taking process.

Although the familiar group cannot statistically be demonstrated that they understood the given passages more correctly than did the unfamiliar one, it can be assumed that they displayed more evidence of "knowledge-based processing". In this study, those who read the contextually familiar version revealed a correct understanding, endeavoured to solve more questions, and in general understood the passages better. This induced better performance than that of those who read the original version. With this, the test developers are advised to consider such a potential obstacle when selecting an appropriate passage.

CONCLUSION

A lot of work has gone into making Thai READS instrument adaption, transparently and systematically. This kind of instrument is not intended to be used as part of in-class grading but is exclusively focused on identifying student performances. The study has shown that the Thai READS instrument can be used by Thai universities to assess the performance of their students at any level. Besides, the instrument can be used at the beginning and the end of the year for progress monitoring. It can then be used in conjunction with a reading matrix and a decoder (reading performance standards) to find out the levels of student performances. Once it is found that a student receives academic warning, for example, corrective measures must be taken as quickly as possible. On the other hand, if a student is at an above average level, enrichment sources are preferably provided.

Limitations

The scope of this study is limited to a test instrument involving Thai culture only.

Implications

From a pedagogical point of view, the results of this study provide implications for (i) test developers; and (ii) test takers. For test developers, this study provides helpful guidance on how to adapt valid and reliable reading comprehension tests. With growing tendency of structured EFL language assessments to fit specific information about performance of students from different cultures, passages questions used in a reading comprehension test are of great importance. Arguably, worldly knowledge of a reading could help the test takers to achieve their highest ability on reading comprehension. On the other hand, a test with too many contextual biases involves substantial knowledge rather than reading proficiency. If the unfamiliar content of a text has an effect on reading comprehension, then it must be considered as a criterion in selecting reading materials and evaluating reading comprehension.

In addition, the findings of this study can be of useful to test takers. The results suggest that using tests with contextually familiar texts may help them to realise their picturesque abilities and can be of helpful to interpreting test scores in a more meaningful way. It is true that they will face fewer comprehension problems and thereby concentrate on how to complete the test given that it is contextually familiar to them. This would, in turn, promote their performance on test taking and, accordingly, unleash their true ability to use the language in a more reliable way. Thus, tests with contextual familiarity are advantageous to the test takers, subsequently helping them to tap their knowledge in the best way possible.

Ideally, the Thai READS can act as a diagnostic, formative and summative assessment, helping lecturers to determine whether or not their students comprehend what they are reading. The lecturers will then be able to provide proper intervention support. Not only serving as an important tool used to figure out reading comprehension aspects in which students are strong or striving, the Thai READS also pinpoints exactly the problems to be addressed. This is because each student is an individual whose pace of learning differs from others'. For this reason, the Thai READS is considered as crucial for helping

educators, lecturers and students themselves to discover their strengths and weaknesses and meet their individual needs.

Suggestions for Further Studies

These advantages are not restricted to standardized reading tests, and similar research on speaking and writing skills should be conducted. Other researchers may apply this study, in a culturally fit manner, to their own geographical regions.

Moreover, in terms of schema knowledge, lecturers can apply relevant worldly knowledge as a prelude to reading, basically as a "schema activation" practice, to help students in unleashing their worldly knowledge in order to comprehend the text.

REFERENCES

Abdul Rashid Mohamed, Lin S. E. & Shaik Abdul Malik Mohamed Ismail (2010). Making more sense of reading scores: developing reading evaluation and decoding system. *English Language Teaching* 3(3), 35-46.

Aebersold, J.A. and Field, M.L. (1997) From Reader to Reading Teacher*.* Cambridge: CUP.

Alderson, J. C., & Urquhart, A. H. (1985). Reading in a Foreign Language. London: Longman.

Alderson, J. C. (2000). Assessing Reading. Cambridge: Cambridge University Press.

Anderson, R. C., & Pearson, P.D (1984). A schema-theoretic view of reading. In P. D. Pearson (Ed.), Handbook of Reading Research (pp. 255-291). New York: Longman.

Bachman, L. F. (1990). Fundamental Considerations in Language Testing. Oxford: Oxford University Press.

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.

Bernhardt, E. (1991). Reading Development in a Second Language: Theoretical, Empirical, and Classroom Perspectives. Norwood, NJ: Ablex Publishing Corp.

Bernhardt, E. (2005). Progress & procrastination in second language reading. In M. McGroarty (Ed.), *Annual Review of Applied Linguistics*, 25, 133-150. New York: Cambridge University Press.

Binder, K. S., & Morris, R. K. (1995). Eye movements and lexical ambiguity resolution: effects of worldly encounter and discourse topic. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. 21: 1186-1196.

Boondao, R., Hurst, A. J., & Sheard, J. I. (2009). Understanding cultural influences: principles for personalized e-learning systems. *International Journal of Human and Social Sciences*, 4(9), 691-695.

Borsa, J. C., Damásio, B. F., & Bandeira, D. R. (2012). Cross-cultural adaptation and validation of psychological instruments: some considerations. Paidéia (Ribeirão Preto), 22(53), 423-432.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 717-726.

Carrell, P. L. (1984). The effects of rhetorical organization on ESL readers. *TESOL Quarterly* 18 (3): 441–69.

Carrell, P.L. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly* 21. (3): 461-481.

Carrell, P.L. (1988). Some causes of text-boundedness and schema interference in ESL reading. Interactive Approaches to Second Language Reading. Cambridge: CUP.

Carrell, P. L. (1989). Metacognitive awareness and second language reading. *The Modern Language Journal*, 73(2), 121-134.

Carrell, P. L., & Eisterhold, J. C. (1983). Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 17(4), 553-573.

Carrell, P. L., & Floyd, P. (1989). Effects on ESL reading of teaching cultural content schemata. *Language Learning*, 37, pp. 88-108.

Cassepp-Borges, V., Balbinotti, M. A. A., & Teodoro, M. L. M. (2010). Tradução e validação de conteúdo: Uma proposta para a adaptação de instrumentos. In L. Pasquali, Instrumentação psicológica: Fundamentos e práticas (pp. 506-520). Porto Alegre: Artmed.

Chihara, T., Sakurai, T. & Oller, J. W. J. (1989). Background and Culture as Factorsin EFL Reading Comprehension. London: Associated University Press, 135-147.

Choy, S. C., Lee, M. Y., & Sedhu, D. S. (2014). Effects of culturally and contextually familiar texts on paraphrasing skills of ESL students. *Malaysian Journal of ELT Research*, 10(1), 1-10.

Collier, V. P., & Thomas, W. P. (2007). Predicting second language academic success in English using the prism model. *International Handbook of English Language Teaching* (pp. 333-348).

Gay, L.R., & Airasian P. (2003). Educational Research: Competencies for Analysis and Applications. New Jersey: Merrill Prentice Hall.

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In R. K. Hambleton, P. F. Merenda, & C. D. Spielberger (Eds.), Adapting Educational and Psychological Tests for Cross-Cultural Assessment (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum.

Harmer, J. (1991). The Practice of English Language Teaching. London: Longman.

Horiba, Y. (1996). Comprehension processes in L2 reading: language competence, textual coherence and inferences. *Studies in Second Language Acquisition*. 18, 433-473.

Keshavarz, M., & Ashtarian, S. (2008). "The relationship between Iranian EFL learners' gender and reading comprehension of three different types of text", in IJAL, 11, 1: 97-113.

Kim, K. H., & Zabelina, D. (2015). Cultural bias in assessment: Can creativity assessment help? *Interdisciplinary Journal of Critical Pedagogy*, 6(2), 129–147.

Lipson, M. Y. (1984). Some unexpected issues in worldly knowledge and comprehension. *The Reading Teacher*, 37, 760-764.

Mok S. S. (2000). Ilmu Pendidikan untuk KPLI (Kursus Perguruan Lepas Ijazah). Subang Jaya: Kumpulan Budiman Sdn. Bhd.

Oosterhof, A. (2001). Classroom Application of Educational Measurement (3rd ed.). Upper Saddle River: Merrill Prentice Hall.

Orellana, M. F., & Reynolds, J. F. (2008). Cultural modeling: Leveraging bilingual skills for school paraphrasing tasks. *Reading Research Quarterly*, 43, 48–65.

Peretz, A., & Shoham, M. (1990). Testing reading comprehension in LSP: Does topic familiarity affect assessed difficulty and actual performance? *Reading in a Foreign Language* 7, 447–455.

Perez-Amurao, A. (2011). The types and nature of question vis-à-vis students' test taking skills as significant indicators of second language examinees' performance on the TOEFL-ITP reading comprehension sub-test. *The International Journal of Research and Review*, 6(2), 39-56.

Sasaki, M. (2000). Effects of cultural schemata on students' test-taking processes for cloze tests: A Multiple Data Source Approach. *Language Testing*. 17(1), 85–114.

Sireci, S. G., Yang, Y., Harter, J., & Ehrlich, E. J. (2006). Evaluating guidelines for test adaptations: A methodological analysis of translation quality. *Journal of Cross-Cultural Psychology*, 37(5),557-567. Doi:10.1177/0022022106290478

Wasburn-Moses, L. (2006). 25 best internet sources for teaching reading. *The Reading Teacher*. 60(1), 70-75.

Weber, R. (1991). Linguistic diversity and reading in American society. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of Reading Research* (pp. 97- 119). New York: Longman.

Westin, E. (2006). Cultural and historical narrative in native and non-native speaker language. In H. L. Andersen, K. Lund, & K. Risager, *Culture in Language Learning* (pp. 45-55). Aarhus: Aarhus University Press.

Yousef, H., Karimi, L., & Janfeshan, K. (2014). The relationship between cultural background and reading comprehension. *Theory and Practice in Language Studies*, 4(4), 707-714.

Yu, G. (2008). Reading to summarise in English and Chinese: A tale of two languages? *Language Testing*, 25(4), 521-551.