Readability of Reading Passages in English Textbooks and the Thai National Education English Test: A Comparative Study

Thanaporn Srisunakrua
Language Institute
Thammasat University, Thailand

Tipamas Chumworatayee
Language Institute
Thammasat University, Thailand

Abstract

Readability has long been regarded as a significant aspect in English language teaching as it provides the overall picture of a text's difficulty level, especially in the context of teaching and testing. Readability is a practical consideration when making decisions on materials to match a text with target readers' proficiency. However, few studies have compared the readability levels of teaching and testing materials in terms of the difficulty of passages. The present study, therefore, aims to explore the readability levels and the linguistic characteristics of reading passages in English textbooks and the Thai National Education English Test based on three readability formulas and eight aspects of linguistic characteristics as provided by the Coh-Metrix computational tool. Two sets of corpora were generated and analyzed by using Coh-Metrix as the main instrument. The obtained data from the reading passages compiled in the English textbooks and the Thai national education English test were compared to explore the significant differences. The results revealed a mismatch in the readability levels and linguistic characteristics. Passages from the English textbooks are easier than those used in the English test. It is recommended that all stakeholders in both teaching and testing administration be aware of the different levels of readability between reading passages. More considerations should be made when preparing the teaching and testing materials because a suitable difficulty level will ensure that students receive the most benefit from the materials. Moreover, an incongruity could affect students' learning and testing performance.

Keywords: Coh-Metrix, English tests, English textbooks, readability, reading passages

Introduction

In readability research, the ultimate goal is generally focused on making sure that a text or reading passages matches the target readers' proficiency. It is believed that matching the difficulty level and readers' proficiency will support language learning and development (Adams, 2009; Mesmer, 2005; Moje, 2006). Most studies in this area, therefore, focus on estimating the difficulty level of reading materials to make a suitable match with the target readers. However, reading passages are not only found in teaching materials. In a testing context, reading passages are mainly used in reading comprehension tests, and their linguistic characteristics are regarded as a neglected area (Solnyshkina, Harkova, & Kiselnikov, 2014). Relatively few studies have been published in the field of language testing, especially those focusing on a comparison of the readability levels of teaching and testing materials. The present study, therefore, aims to fill this gap by comparing the readability levels of reading passages comprised in teaching and testing materials. Using the Coh-Metrix computational tool, three types of readability formulas were employed: Flesch Reading Ease, Flesch-Kincaid Grade Level, and Coh-Metrix L2 Readability. On top of that, eight linguistic characteristics proposed by McNamara, Grasser, McCarthy, and Cai (2014), namely, narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion, verb cohesion, connectivity, and temporality, were included in the analysis. It is believed that the findings would be beneficial to all parties involved in teaching and testing contexts.

Literature review

Readability can be defined as a measure to predict text difficulty using different kinds of readability formulas (Davies, 1995). Klare (1963) and Pikulski (2002) define readability as an indicator or a measure of the ease or difficulty of text comprehension. Alderson and Urquhart (1984) expand the definition by indicating that the level of ease or difficulty of texts is determined through the analysis of the features or various aspects of a text. Nuttall (2005) views that these features originate from both structural and lexical features.

Readability research involves studies related to the prediction of text difficulty level through the analysis of the text's features that might facilitate or obstruct the comprehension of the text. Many scholars have tried to develop and try out readability formulas in order to find the most suitable way to predict text readability level.

Readability is generally measured by using readability formulas. Flesch Reading Ease and Flesch-Kincaid Grade Level are the two most common and practical methods used in estimating difficulty level (Solnyshkina et al.: 2014). These two formulas assess the difficulty level based on the word and sentence length in the target reading text (Flesch, 1948). The assumption is that texts consisting of longer words and lengthy sentences tend to require more time to process, making them more challenging to understand (Graesser et al., 2001). These two formulas are very popular among educators due to their practicality and the evidence that they employ objective criteria in assessing the difficulty level (Zamanian & Heydari, 2012). However, there are some critics. Kirkwood and Wolfe (1980) claim that the variables behind the formulas are based on the surface level of the text, which can possibly be invalid. A text with jumbled sentences can be easy to read because it consists of familiar words, and it is short. Schriver (1989) and Dreyer (1984) also argue that the formulas disregard the whole text aspects and ignore the flow of ideas throughout a reading text. Apart from using readability formulas to estimate the difficulty level, a tool needs to be

developed that could assess the challenges of a reading text at the word, sentence, and deeper levels of language. Toward the end, Crossley, Salsbury, McCarthy, and McNamara (2008) developed a unidimensional readability formula called the Coh-Metrix L2 Readability formula, which incorporates a deeper analysis of the cohesion between sentences into the formula. This formula is claimed to produce the more valid and objective results, which not only describe the superficial characteristics of a text but also deeper levels of discourse in the algorithm.

Meanwhile, the Coh-Metrix Easability Components provide a better picture of text difficulty based on the linguistic characteristics of the reading text, which are narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion, verb cohesion, connectivity, and temporality (McNamara et al., 2014).

The narrative features are the characteristics of a reading text that focus on telling a story via characters, events, and places. They are closely related to word familiarity, world knowledge and everyday oral language. Narrative text is easier to comprehend than informational text (Graesser, Olde, & Klettke, 2002)

Syntactic simplicity reflects the degree of words per sentence, and the familiarity and simplicity of the syntactic structures of the sentence. Sentences that contain more words and complex structures are more challenging to process (McNamara, Graesser, McCarthy, & Cai, 2004).

Word concreteness analyzes the characteristic of words included in a text. A text that contains a higher number of concrete and meaningful words will enhance comprehension. Concrete words are better at evoking mental image than abstract words, making texts with a greater number of abstract words more difficult to comprehend (McNamara et al., 2014)

Referential cohesion reflects the overlapping of words and ideas across sentences and the entire text, forming explicit connections throughout a text and making it more cohesive. A highly cohesive text is typically less challenging to read because of the explicit connections between ideas (McNamara & Graesser, 2012).

Deep cohesion refers to the degree to which a text has causal and intentional connectives. These types of connectives can make causal and logical relationships more explicit, enabling readers to better understand the meaning of a text. A text that contains more explicit connectives is easier to process since it can reduce the need for inference while reading (McNamara et al., 2014).

Verb cohesion refers to the analysis of overlapping verbs in a reading text. These repeated verbs, usually found in narrative texts and texts for young readers, make a text more coherent, which facilitates situation model understanding (McNamara, Graesser, & Louwerse, 2012).

Connectivity reflects the explicit use of adversative, additive, and comparative connectives in a reading text.   The use of connective words can make the logical connections in a text more explicit, facilitating reading comprehension (McNamara et al., 2014).

Temporality shows the consistency of tense and aspect used in a reading text. The more consistent the text is, the easier it is for readers to process and understand (McNamara et al., 2014).

Incorporating these linguistic characteristics with traditional readability formulas using Coh-Metrix allows for better prediction of text difficulty and a more accurate picture of readability level since they cover both the surface and deep levels of text analysis.

Research questions
1. What are the readability levels of reading passages in English textbooks (CPET) and the Thai National Education English Test (CONET)?
2. What are the linguistic characteristics of reading passages in English textbooks (CPET) and the Thai National Education English Test (CONET)?

Research methodology
Two sets of corpora were built for data analysis. The first corpus was a collection of 155 reading passages in English textbooks prescribed and certified by the Office of Basic Education Commission. These textbooks are used as the main teaching resources and materials for the English subject in M.6 (Grade 12). The second corpus was a collection of 20 reading passages in seven Thai National Education English Tests.

Each passage included in the corpora was computationally analyzed using Coh-Metrix (http://tool.cohmetrix.com/). The results were then calculated to find the average percentages of the two sets of corpora. To answer the first research question, three types of readability formulas were selected: Flesch Reading Ease, Flesh-Kincaid Grade Level, and Coh-Metrix L2 Readability. For the second research question, eight linguistic characteristics (narrativity, syntactic simplicity, word concreteness, referential cohesion, deep cohesion, verb cohesion, connectivity, and temporality) were analyzed. The results were averaged to obtain the mean values and then further analyzed using t-test to find out the significantly different aspects of the linguistic characteristics.

Results
Research Question 1
To answer the first research question, "What are the readability levels of reading passages used in the English textbooks (CPET) and the Thai National Education English Test (CONET)?", three indices of the Coh-Metrix concerned with readability formulas were used as tools: Flesch Reading Ease, Flesch-Kincaid Grade Level, and Coh-Metrix L2 Readability. In interpreting the values for Flesch Reading Ease and Coh-Metrix L2 Readability, a higher value represented a less difficult reading passage whereas a lower value indicated a more difficult reading passage. For the Flesch-Kincaid Grade Level formula, a higher value represented more difficult reading passages, which are probably used in a higher grade level; meanwhile, a lower value indicated less difficult reading passages, which are used in a lower grade level. The following table illustrates a comparison of the average readability levels obtained from the three readability formulas.

Table 1. *Comparison of the average readability levels of passages from CPET and CONET: Means (M) and Standard Deviation (SD)*

Index	CPET		CONET	
	Mean	**SD**	**Mean**	**SD**
Flesch Reading Ease	<u>**70.684**</u>	12.070	<u>*60.189*</u>	13.935
Flesch-Kincaid Grade Level	<u>**6.889**</u>	2.255	<u>*8.696*</u>	2.416
Coh-Metrix L2 Readability	<u>**16.476**</u>	6.076	<u>*10.932*</u>	3.339

Table 1 displays the average readability levels obtained from the two corpora (CPET and CONET). As can be seen, all three readability formulas yielded congruent results. The results showed that the readability level of the reading passages in CPET was easier for the readers than those of CONET.

The average readability values from the Flesch Reading Ease formula (M=70.684 SD=12.070 > M=60.189 SD=13.936) showed that the mean value of CPET was higher than that of CONET, indicating that the reading passages from CPET are easier than CONET. The values from Coh-Metrix L2 Readability formula yielded the same results, indicating that CPET is easier than CONET (M=16.476 SD=6.076 > M=10.932 SD=3.339). The results were in line with the ones obtained from Flesch-Kincaid Grade Level formula, indicating that the reading passages in CPET are easier, at the approximate level of grade 6, whereas the passages in CONET are at grade 8 (M=6.889 SD=2.255 < M=8.696 SD=2.416).

Research Question 2

Coh-Metrix was also used as the main instrument to answer the second research question "What are the linguistic characteristics of reading passages used in English textbooks (CPET) and the Thai National Education English Test (CONET)?" The results were analyzed to find the significantly different values using a t-test. The following table displays the results.

Table 2. *The Linguistic Characteristics of CPET and CONET: Means (M) and Standard Deviation (SD)*

Linguistic characteristics	CPET		CONET			RESULTS	
	Mean	**SD**	**Mean**	**SD**	**t**	**CPET easier**	**CONET easier**
Narrativity	**57.585**	29.427	33.753	21.859	-4.390*	✓	
Syntactic simplicity	45.929	23.676	47.716	45.930	0.706	-	-
Word concreteness	67.781	27.180	**72.446**	23.725	0.732*		✓

Referential cohesion	**30.554**	26.621	19.294	26.087	-1.784*	✓	
Deep cohesion	56.079	32.383	**64.668**	24.000	1.440*		✓
Verb cohesion	45.988	29.735	43.713	28.198	-0.324	-	-
Connectivity	10.115	19.127	**25.629**	28.694	2.351*		✓
Temporality	**58.798**	32.383	45.670	29.131	-1.724*	✓	-

Note: * = significantly different value

As illustrated in Table 2, the six aspects of linguistic characteristics that were found to be significantly different between CPET and CONET were narrativity, word concreteness, referential cohesion, deep cohesion, connectivity, and temporality. However, the mean values obtained from the other two aspects, namely, syntactic simplicity and verb cohesion, were not significantly different.

In interpreting the results shown in Table 2, the higher values represented a lower difficulty level. As can be seen in the table, three linguistic characteristics showed that the passages in CPET have a lower readability level than those of CONET: narrativity (M = 57.585 SD = 29.427 > M =33.753 SD = 21.859), referential cohesion (M=30.554 SD = 26.621 >M = 19.294 SD = 26.087), and temporality (M = 58.798 SD = 32.383 > M = 45.670 SD = 29.131). Meanwhile, CONET contained three linguistic characteristics that support ease in comprehension: word concreteness (M = 72.466 SD = 23.725 > M = 67.781 SD = 27.180), deep cohesion (M = 64.668 SD = 24.000 > M =56.079 SD = 32.383), and connectivity (M = 25.629 SD = 28.694 > M = 10.115 SD = 19.127).

Discussion
Research question 1
The results indicate incongruent readability levels between the teaching and testing materials. It seems that the teaching materials that Thai students are exposed to are easier than the reading passages on the tests, which might have an effect on the test results. In the view of Shohamy (1993) and Hughes (2013), test results can have crucial consequences for various stakeholders, such as students, teachers, and the schools. The most important stakeholders are the students as they are directly affected by test results (Pan & Newfields, 2012). It seems more than a little bit unfair for students to take achievement tests that are more difficult than the material they have been taught at their grade level. This disparity between readability levels may be a cause of anxiety, which Aydin (2009) claims has a significant effect on students' learning and testing performance. A study conducted by Lunrasri (2014) on the perceptions of students toward the national English test (O-NET) showed that most students fear getting a low score. Additionally, the students stated that the content on the test was more difficult than what they were taught in class, especially the vocabulary part. Some students reported that there were words they had never seen or had never been taught before. This may encourage students to resort to tutoring schools in order to prepare for the national test, raising questions about fairness since tutoring is less affordable for those with lower incomes. This concern is given credence by Goodman (2017), who found that urban students in Thailand were likely to score higher on tests than rural students

Arab World English Journal   262
www.awej.org
ISSN: 2229-9327

because they may have more access to educational technology and private tutoring. Messick (1996) also discussed the social consequences of testing, contending that 'consequential validity' can be part of the broader concept of test validity.

Teachers can also be affected due to pressure from schools, students' parents, or even students themselves. In the view of Alderson and Wall (1993, p.117), tests may force teachers to do what "they would not otherwise necessarily do", potentially impacting their teaching practice. For example, teachers may be tempted to forego the prescribed teaching materials and use previous tests or mock tests as teaching materials (Cheng, 2003). Lunrasri (2014) found that some Thai teachers reported negative washback from high-stakes tests on their teaching practices, claiming they had to spend more time preparing students for the test rather than teaching the content prescribed by the curriculum. This may rob students of the chance to improve their language skills and overall proficiency.

The schools or institutions might also be affected by the discrepancy in the difficulty levels of the teaching and testing materials. According to the standards set by the Office for National Education Standards and Quality Assessment (ONESQA), the O-NET results are used as one of the twelve criteria for the evaluation of a school's quality. The results of this high-stakes test may thus have a great impact on the teaching and learning practices in schools, as well as the school administration. According to Goodman (2017), many of the stakeholders, specifically the principals, view O-NET scores as the most important criterion. Low O-NET scores can cause a school to fail the quality evaluation, damaging its reputation. School principals therefore put great effort toward improving students' scores. Negative effects from high-stakes testing are not limited to Thailand. A study conducted by Sundayana, Meekaeo, Purnawarman, and Sukyadi (2018) found that schools in Thailand and Indonesia set up special test preparation programs to enhance students' test-taking skills before the actual test. Further evidence was obtained in the study of Lunrasri (2014), who determined that school principals in one province of Thailand directed teachers to conduct special classes to go over old test exams, as well as tutor and train students to cope with the actual test.

Research question 2

Apart from the incongruent readability levels in the CPET and CONET, some differences were also discovered in the linguistic characteristics of the reading passages. Specifically, the narrativity scores indicate that the CPET is easier in terms of the genre of reading passages. This suggests that the CPET contains more passages that have the characteristics of narrative text than those found in CONET. According to McNamara et al. (2014), a text with high narrativity tends to have more familiar oral language that is easier to understand. Moreover, Ismail and Yusof (2016) concluded that passages containing more characteristics of narrativity are likely to be easier to process than reading materials that are more informational, especially for younger readers.

In addition, there is more evidence of referential cohesion in CPET than in CONET, with the passages in the former containing more overlapping words and ideas across the sentences. As a result, the concepts and content are more explicit, which can support the comprehension process and make the passages in CPET less challenging. This explicit coreference enables readers to

make connections and understand the relationships between propositions, clauses, and sentences across the whole passages (Halliday & Hasan, 1976; McNamara & Kintsch, 1996). On the contrary, texts with low referential cohesion can be problematic and cause frustration for readers who read independently (Isamail & Yusof, 2016).

Further evidence that CPET contains features that aid in the comprehension of reading passages is provided by the high scores on 'temporality,' which reveals the cues of the time used in the passages. Cues of time are related to the genre of narrativity because stories are likely to be narrated through time sequences. Moreover, the high temporality scores also suggest consistency with respect to tense and aspect within the reading passages. This facilitates the comprehension process since the readers do not have to worry very much about tense while processing the reading passages.

Based on the six indices of the Text Easability Component scores, CONET has three features that are found more in less difficult reading passages. The first one is 'word concreteness'. Passages in CONET have a higher number of concrete words, which means the words are more meaningful than those found in CPET. This may facilitate readers as concrete words evoke mental images and enable them to create a situation model in the reading process. Passages containing words that are more concrete also give readers more time to use their working memory to process and comprehend what they read (Perfetti, 2007). Moreover, Silfhout (2014) contends that texts containing many concrete words are more interesting, and this can better facilitate comprehension than abstract words. The higher word concreteness in CONET seems sensible because students have to take the test by themselves without any support from others within a limited period of time. In contrast, reading passages in classroom teaching materials may contain a higher number of abstract words because students can ask for clarification from their teachers.

There is also greater evidence of deep cohesion in CONET, which means the reading passages contain more causal and intentional connectives. This makes the logical relationships in the passages clearer for the readers, possibly enabling them to infer the meaning of a text. Reading passages with explicit relationships and global cohesion are easier for readers to comprehend because the logical relationships between ideas in the text are made explicitly via connective words. (McNamara et al, 2014). Thus, students do not have to make many inferences.

CONET also features greater 'connectivity,' which refers to the use of adversative, additive, and comparative connections. The use of explicit connective words can reduce the need to make inferences in the comprehension process, potentially supporting the ease of reading**.** Graesser, McNamara, Louwerse, and Cai (2004) state that texts which are high in connectives can aid in the process of making connections between the concepts presented in a text and readers' existing knowledge, resulting in a clearer and more coherent mental representation. Silfhout, Evers-Vermeul, and Sanders (2015) also determined that reading a text with explicit connective words could lead to a higher performance in comprehension tasks than reading implicit or non-connective texts. The evidence suggests greater use of connective words in the reading passages in CONET. Graesser, McNamara, and Louwerse (2003) found that some cohesive devices have a distinctive role in presenting the rhetorical structure of the passages. The uses of adversative, additive, and comparative connectives show that the passages in CONET are mainly focused on

giving information, indicating that the reading passages are more expository than narrative, in contrast to CPET.

No significant differences were found with respect to the other two linguistic characteristics, syntactic simplicity and verb cohesion. Syntactic simplicity refers to the number of words per sentence and the syntactic structure of the sentences in the passages. The insignificant differences found between the passages used in teaching and testing may have resulted from the limitation of passage length. It is common to see similarity in terms of sentence complexity and length in the reading passages contained in teaching and testing materials. The results showed that CPET and CONET are not different in terms of syntactic simplicity. For verb cohesion, which reflects the degree of overlapping verbs in the reading passages, the results also showed insignificant differences between CPET and CONET. According to McNamara et al. (2014), this feature is less related to the ease of the text.

Besides providing evidence on ease in reading comprehension, the linguistic characteristics found in CPET and CONET also bring to light the outstanding characteristics of the reading passages. In CPET, the higher incidences of narrativity and temporality are relevant, as this suggests that the reading passages in English textbooks are usually narrative. The higher values in 'referential cohesion' also support the narrative features of the reading passages. The use of overlapping words, especially those related to content words and noun and pronoun references, can support the narrative features since they show how the story and ideas relate to each other throughout the reading text. It can be briefly summarized that the linguistic characteristics found in CPET provide strong evidence that the reading passages in English textbooks are narrative.

Meanwhile, the analysis of the linguistic characteristics in CONET revealed higher values in terms of deep cohesion and connectivity, two aspects that are related to the evidence of the use of explicit connectives in the reading passages. The types of connective words can illustrate the genre of reading passages. For example, CONET has connective words, which show causal, logical, additive, and comparative relationships of ideas throughout the reading passages. It can be inferred that the passages in CONET are more informational whereas CPET has more narrative texts.

In sum, both CPET and CONET contain linguistic characteristics that support the ease in comprehension process. Moreover, the linguistic characteristics found in both corpora illustrate the distinctive features of the reading passages used for teaching and testing materials. However, more narrative texts are employed in a teaching context, whereas texts that are more informational are used as test materials.

Implications

The results of the study provided strong evidence of incongruent readability levels between the reading passages used in teaching materials (CPET) and testing materials (CONET). As the latter, in this context, is designed to be an achievement test, it should accurately assess students' mastery of course content and their depth of learning. Flateby (2014) concludes that if a test reflects the content and the level of cognitive demand, valid and reliable results on students' achievement should be obtained; on the other hand, if a test is far more challenging than the material students

have been taught in class, the test results may be less valid. Therefore, more consideration needs to be given to the process of materials selection.

In order to construct a valid test, the designers need to conduct an in-depth review of the aims of the core curriculum and the content provided in the teaching materials. The revision might employ a tool to check both the readability levels of the materials and the outstanding linguistic characteristics of the main teaching resources. The results obtained from this investigation will provide insight into the teaching materials used in courses, which can serve as the basis for designing a national achievement test that matches the content to be assessed. The test results would provide an accurate assessment of the actual performance of the students, enhancing the validity of the test.

Limitation
The present study focused on analyzing the readability levels of the reading passages used in teaching (CPET) and testing (CONET) contexts. However, when estimating the difficulty level of the reading texts, other fundamental considerations, such as the readers and the reading tasks or activities, should be taken into account. The interaction of the readers and the target reading texts in terms of reading proficiency levels, motivation, and reading purposes has an effect on the comprehension process. Moreover, the requirements of the reading tasks also influence how readers tackle a text. These factors were beyond the scope of the present study. Therefore, interpretation or generalization of the results of the present study should be undertaken with careful consideration as it aimed to estimate the difficulty level of the reading texts by considering only one fundamental factor: the text.

Conclusion
The study aimed to analyze the readability of the reading passages using three readability formulas and eight linguistic characteristics as the main instruments. The results revealed incongruent readability levels between the two corpora. Moreover, the passages used in each corpus also had distinctive linguistic characteristics that facilitated reading comprehension. Based on the results of the study, it is recommended that test developers take the readability level and the unique linguistic characteristics of reading passages into consideration when designing a high-stakes test.

About the Authors
Thanaporn Srisunakrua is a full-time lecturer at the Department of Language Studies, School of Liberal Arts, King Mongkut's University of Technology Thonburi. She is currently a Ph. D. Candidate at Thammasat University majoring in English Language Teaching. Her research interests are in the areas of Reading and English language teaching.
ORCID : https://orcid.org/0000-0003-1487-9493

Dr. Tipamas Chumworatayee (Associate Professor) teaches both post- and undergraduate courses at the Language Institute, Thammasat University. She obtained her Ph.D. from Department of Reading, College of Education, Texas Woman's University, Texas, USA. Her main interests include ELT methodology, ELT teacher training, EFL reading- strategy instruction, and EFL reading-strategy awareness-raising.

References

Adams, M.J. (2009). The challenge of advanced texts: The interdependence of reading and learning. In E.H. Hiebert (Ed.). *Reading more, reading better* (pp. 163-189). New York, NY: Guilford Press.

Alderson, J. C. & Urquhart, A.H. (1984). *Reading in a Foreign Language*. London: Longman.

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, *14*, 115-129.

Aydin, S. (2009). Test anxiety among foreign language learners: A review of literature, *Journal of Language and Linguistic Studies*, *5*, 127-137.

Cheng, L. (2003). Looking at the impact of a public examination change on secondary classroom teaching : A Hong Kong case study. *Journal of Classroom Interaction*, *38*(1), 1-10.

Crossley, A.S., Salsbury, T., McCarthy, P.M., & McNamara, D.S. (2008). LSA as a measure of coherence in second language natural discourse. In V. Sloutsky, B. Love, and K. McRae (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 1906-1911). Washington, DC: Cognitive Science Society. doi: 10.17509/ijal.v8il.11478

Davies, F. (1995). Introducing Reading. *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, *38*, 173-177.

Dreyer, L.G. (1984). Readability and responsibility. *Journal of Reading*, *32*, 334-338.

Flateby, T. L. (2014). *A guide for writing and improving achievement tests*. Statesboro: Georgia Southern University.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(3), 221-233

Goodman, J. (2017) Unpacking the narrative of educational failure: Thailand in the standardized testing era. *13th International conference on Thai studies: Globalized Thailand? Connectivity, Conflict, and conundrums of Thai studies*. 15-18 July 2017, Chiang Mai, Thailand. Retrieved from http://www.icts13.chiangmai.cmu.ac.th/pdf_abstract.php?abs_id=222 September 26, 2018.

Graesser, A. C., Karnavat, A.B., Daniel, F. K., Cooper, E., Whitten, S.N., & Louwerse, M. (2001). A computer tool to improve questionnaire design. In *Statistical Policy Working Paper 33, Federal Committee on Statistical Methodology* (pp. 36-48). Washington, DC: Bureau of Labor Statistics.

Graesser, A. C., McNamara, D. S., & Louwerse, M. M. (2003). What do readers need to learn in order to process coherence relations in narrative and expository text? In A. P. Sweet & C. E. Snow (Eds.), *Rethinking reading comprehension* (pp. 82–98). New York: Guilford.

Graesser, A.C., McNamara, D.S., Louwerse, M.M. & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments and Computers*, *36*, 193-202

Graesser, A.C., Okde, B., & Klettke, B. (2002). How does the mind construct and represent stories? In M.C. Green, J.J. Strange, and T.C. Brock (Eds.), *Narrative Impact: Social and cognitive foundations* (pp. 231-263), Mahwah, NJ: Erlbaum.

Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

Hughes, A. (2003). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.

Ismail, A., & Yusof, N. (2016). Readability of ESL picture books in Malaysia. *Journal of Nusantara Studies, 1*(1), 60-70. Retrieved from http://dx.doi.org/10.24200/jonus.vol1iss1pp60-70

Kirkwood, K.J., & Wolfe, R.G. (1980). *Matching students and reading materials: A cloze-procedure method for assessing the reading ability of students and the readability of textual materials*. Toronto: Ontario Department of Education.

Klare, G.R. (1963) *The measurement of readability*. University of Iowa Press, Ames, IA.

Lunrasri, Y. (2014). *Washback effects of the Ordinary National Educational Test (O-NET) on English language teaching and learning in ninth grade*. Chulalongkorn University, Bangkok, Thailand.

McNamara, D.S. & Graesser, A.C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P.M. McCarthy, & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation, and resolution* (pp. 188-205). Hershey, PA: IGI Global.

McNamara, D.S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247-287.

McNamara, D.S., Graesser, A.C., & Louwerse, M.M. (2012). Sources and text difficulty: Across genres and grades. In J.P. Sabatini, E. Albro, and T. O'Relley (Eds.), *Measuring up. Advances in how we assess reading ability* (pp. 89-116). Plymouth, UK: Rowman & Littlefield Education.

McNamara, D.S., Graesser, A.C., McCarthy, P.M. & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments and Computers*, Vol. 36. 193-202

McNamara, D.S., Graesser, A.C., McCarthy, P.M. & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix.* New York: Cambridge University Press.

Mesmer, H. A. (2005). Text accessibility and the struggling reader. *Reading and Writing Quarterly*, *47*, 235-258. doi:10.1080/10573560590523603

Messick, S. (1996). Validity and washback in language testing. *Language Testing, 13*(3), 241-256. Retrieved from http://dx.doi.org/10.1177/026553229601300302

Moje, E.B. (2006). Motivating texts, motivating contexts, motivating adolescents: An examination of the role of motivation in adolescent literacy practices and development. *Perspectives*, *32*, 10-14.

Nuttall, C. (2005). *Teaching reading skills in a foreign language*. Oxford: Macmillan Education.

Pan, Y. C. & Newfields, T. (2012) Do tests promote changes in listening and reading skills?: Evidence from a Taiwanese EFL context. *The International Journal of foreign language teaching*. Retrieved from http://ijflt.com/images/ijflt/articles-june-2012/pan-newfield.pdf

Perfetti, C. (2007). Reading Ability: Lexical Quality to Comprehension, *SCIENTIFIC STUDIES OF READING*, *11* (4), 357-383, doi:10.1080/10888430701530730

Pikulski (2002). *Readability*. Retrieved from https://www.eduplace.com/state/author/pikulski.pdf

Schriver, K.A. (1989). Evaluating text quality: The continuum from text-focused to reader-focused method. *IEEE Transactions on Professional Communication*, *32* (4), 238-255.

Shohamy, E. (1993). The power of tests: the impact of language tests on teaching and learning. *NFLC Occasional Papers*, Washington DC. *56*. Retrieved from https://files.eric.ed.gov/fulltext/ED362040.pdf

Silfhout, G. (2014). *Fun to read or easy to understand? Establishing effective text features for educational texts on the basis of processing and comprehension research.* Retrieved from https://dspace.library.uu.nl/handle/1874/300805

Silfhout, G. V., Evers-Vermeul, J. & Sanders, T. (2015). Connectives as processing signals: how students benefit in processing narrative and expository Texts, *Discourse Processes*, *52* (1), 47-76, doi:10.1080/0163853X.2014.905237

Solnyshkina, M. I., Harkova, E. V., & Kiselnikov, A. S., (2014). Comparative Coh-Metrix analysis of reading comprehension texts: Unified (Russian) state exam in English vs Cambridge first certificate in English. *English Language Teaching*, *7* (12), 65-76.

Sundayana, W., Meekaeo, P., Purnawarman, P., and Sukyadi, D. (2018). Washback of English national exams at ninth-grade level in Thailand and Indonesia. *Indonesian Journal of Applied Linguistics*. *8* (1). 167-176 Retrieved from http://ejournal.upi.edu/index.php/IJAL/article/view/11478

Zamanian, M. and Heydari, P. (2012). Readability of text: State of the Art. *Theory and Practice in Language studies*. *2* (1).43-53.