



A Comparison of Readability and Understandability in Second Language Acquisition Textbooks for Pre-service EFL Teachers

Dennis Murphy Odo

Pusan National University, South Korea

Teachers must be able to comprehend their textbooks in order to fully benefit from their preparation programs. However, it is yet unclear whether textbooks used to teach challenging content to pre-service non-native speaking English teachers might be beyond their ability. To investigate this question, a textbook readability analysis was performed with the Coh-Metrix text analysis tool. Textbook understandability was also assessed by means of maze passages given to teacher candidates in a pre-service English teacher education program. Readability analyses results indicated that the majority of the texts are excessively challenging for L2 readers at the level of the participants. Understandability findings revealed that many texts were able to be read independently but that all texts could be understood with scaffolding. Contrary to prior understandability research findings that showed that these textbooks were too difficult, this study found them to be suitable for their audience. However, a correlation between text understandability and course grades indicates that ability to comprehend the text could relate to course performance. Implications are that EFL teacher educators must recognize the importance of readability for textbook selection criteria and provide learners with access to other supplementary materials to scaffold their learning.

Keywords: readability, second language reading, teacher education, materials evaluation

Introduction

The present research investigates the readability and understandability of second language acquisition (SLA) textbook materials for non-native speaker teachers in an EFL context. Textbooks have long been recognized as the main vehicle for conveying information in most educational institutions (Cline, 1972; Johns, 1997). Of course, they must be comprehensible to be useful to students (M. Smith & Taffler, 1992; Snyman, 2004), so understandability is an important consideration when adopting a new textbook (K. J. Smith & DeRidder, 1997). However, limited research currently exists regarding the understandability of textbooks used in a wide range of educational contexts.

Research into textbook understandability with pre-service second language (L2) teachers is an issue worthy of concern for a number of reasons. Firstly, investigations in a number of other domains have revealed that many first language (L1) and L2 readers struggle to comprehend their textbooks (Blignaut, Wissing, & Van den Berg, 2016; Daniels, 1996; Kasule, 2011). Failure to understand textbooks has serious consequences because these learners tend to perform less well academically (Spinks & Wells, 1993).

Accordingly, adequate consideration of the readability of college textbooks is necessary to meet the needs of students who have limited reading skills (Schneider, 2011). In their research with postsecondary L2 readers, Tinkler and Woods (2013) observed that these readers prefer text features that differ from

those of native speakers. Thus, they concluded that the linguistic background of the intended audience is an important determinant of readability. Brown (1998) reached similar conclusions, arguing that readability formulas that are specifically designed for L2 readers better account for the unique variables they bring to texts such as L1 background and knowledge of rhetorical organization among others.

Among L2 textbook readers, one group that has received limited research attention is non-native L2 teachers (Moussu & Llorca, 2008). These teachers often use L2 textbooks in the course of their teacher preparation. Among the various textbooks that they use, some are certain to be more difficult to understand than others. For example, in teacher education programs that offer conversation classes, these L2 textbooks are likely to be less relatively cognitively demanding. However, most teacher preparation programs offer classes in topics that are considerably more complex and technical, such as SLA. In these instances, the dense and complicated contents of these texts may pose great difficulty for L2 readers.

As an illustration, target language texts about SLA are written to support the exposition of complex yet foundational concepts in language teaching and learning. Authors of these texts need to consider the needs of non-native speaking learners because such a substantial proportion of their audience are not native speakers. Authors' failure to produce texts that meet their readers' needs can cause learners to misunderstand foundational concepts in the discipline (Berti, 1991). These misapprehensions may then be passed on to their students in the form of the teachers' ineffective classroom practice. As well, comprehensible textbooks enable independent study (Razek, Hosch, & Pearl, 1982), and thus offer potentially greater depth of learning.

To ensure that preservice teachers are gaining adequate access to the texts that they need, teacher educators require a means by which to firmly establish a given textbook's understandability. Fortunately, innovative tools have recently been developed that can better assist textbook users in determining the readability of materials. An example of such a tool, known as Coh-Metrix, was selected for use in the current analysis. This newly-developed online linguistic analysis system provides detailed and valuable information about a text's readability that had previously been unavailable. Access to these tools can facilitate teacher educators' selection of appropriate course texts.

Literature Review

Textbook Readability and Understandability Research

A number of researchers note the distinction between the concepts of readability and understandability. Both are relevant to the present study. These attributes are related, but they are not the same thing. "Readability refers to the qualities of writing which are related to reader comprehension" (Plucinski, Olsavsky, & Hall, 2009, p. 119), that is, the features of the text that make it more or less difficult to understand. In general, the concept is centered on the text rather than the reader and is stable for the same text. Various methods can be used to evaluate readability, but readability formulas tend to be quite commonly used. Additionally, "readability is a prerequisite of understandability; but [it] does not guarantee individual understandability" (Jones, 1997, pp. 105–106).

In contrast, understandability measures text-reader interaction and is a broader concept. It is also more centered on the reader than on the text. It has been defined as "the ability of readers to comprehend ... textbooks and to complete the act of communication initiated by the writers of those textbooks" (Adelberg & Razek, 1984, p. 109). Jones (1997) explains that understandability depends on both text characteristics, such as grammar, and on reader attributes, such as background knowledge and reading proficiency. Consequently, different readers will likely understand the same text to varied degrees due to their diverse personal characteristics. This clear division between readability and understandability therefore compels us to take both concepts into consideration when evaluating how suitable particular texts might be for their intended readers (Blignaut et al., 2016; Davidson, 2005).

Defining and measuring readability

DuBay (2004) states that “readability is what makes some texts easier to read than others” (p. 3). Readability formulas are often used to help determine the appropriateness of a textbook for its anticipated readers (Blignaut et al., 2016). Each technique that been used to measure readability has typically focused on text difficulty indicators (e.g., word or sentence length) that are then arranged according a continuum that indicates where a particular text falls relative to other texts with respect to those pre-defined indicators (Tanaka-Ishii, Tezuka, & Terada, 2010). Common text characteristics that have been selected as indicators of readability typically include vocabulary and structural complexity (Chen, 2016). Readability formulas mainly differ with respect to the weight assigned to the text features included (Wait, 1987).

Research support for readability formulas

In their review of readability research, Zamanian and Heydari (2012) point out that some controversy surrounds the use of readability formulas. Both supporters and detractors of readability formulas can garner evidence to defend their position. They also maintain that, contrary to widespread criticisms, plenty of evidence exists in favor of the validity of readability formulas. Another confirmation of the validity of readability formulas includes the fact that their results correlate with comprehension assessed by traditional multiple-choice questions, comprehension assessed by cloze passages, and oral reading errors (Fry, 1989). DuBay (2004) offers additional support for their validity by pointing out that their results are widely accepted in many types of research within a variety of diverse academic fields and in court proceedings. He adds that “over 80 years of research and testing have contributed to the worldwide use in many languages of the readability formulas” (p. 1).

As with the use of readability formulas for L1 readers, some controversy also exists for their use with L2 reading materials. Nevertheless, here as well, formulas are claimed to accurately measure text readability for L2 readers (Zamanian & Heydari, 2012). Greenfield (2004) maintains that while formulas have been developed that are better suited to L2 texts, “...we have strong evidence that the new formulas have only a narrow, if any, advantage over the time-tested traditional formulas” (p. 18).

Although readability formulas have their supporters, they have also received considerable criticism over the years. For instance, most formulas consider only sentence length and word difficulty, which only provides dubious results in terms of accuracy or usefulness (Bailin & Grafstein, 2001). That is, they rely on sentence-level difficulty (i.e., vocabulary and syntax) to estimate structure-level difficulty (e.g., inference load or story structure) when the two often do not correlate well (Oakland & Lane, 2004). They also ignore factors such as text structure, cohesion, complexity of ideas, and schemata, while also failing to take into account readers’ motivation, interest, perseverance, and purpose for reading (Armbruster, Osborn, & Davison, 1985; Zamanian & Heydari, 2012). They likewise overlook content difficulty and familiarity, the organization of ideas, the authors’ style, and page layouts (Armbruster et al., 1985). Text decodability, the number of words, the image quality, font size, and so forth can also influence the beginning reader’s initial reading experiences and skill development (Benjamin, 2012). Indeed, Fulcher (1997) contends that “factors which are not considered in traditional readability formulae are more important determinants of text accessibility” (p. 497). Bailin and Grafstein (2001) also point out that the real danger is that the specious numerical precision and seeming objectivity of formula scores misleads educators to believe that these scores should supersede their own professional judgement.

As well, many researchers have commented on the wide range of readability scores or grade levels that are produced by different formulae for the same text (Armbruster et al., 1985; Begeny & Greene 2014; Maslin, 2007). This phenomenon has also been observed with L2 readers (Chen, 2012). Addressing this concern, Chen (2012) points out that while readability results from various formulae differ slightly, the rank of the various texts remain the same. Thus, formulas can still be useful tools for determining relative difficulty amongst various texts. Dubay (2004) concurs adding that agreement among formulas is much less important than their consistency in predicting readability across a range of texts.

All of these limitations for traditional assessments of readability should cause researchers and practitioners to consider newer alternatives that are available, to supplement traditional formula results at the very least. For example, qualitative methods of assessing readability, such as through readability scales (e.g., see Chall, Bissex, Conard, & Harris-Sharples, 1996), may be more appropriate for those who are learning to read. Similarly, relatively simple formulas were useful in the past, but researchers have since developed more sophisticated tools that are consistent with the psycholinguistic reading theories that can account for linguistic variables related to readers' cognitive processing (Crossley, Greenfield, & McNamara, 2008). These up-to-date tools allow us to accurately measure the readability for a wider variety of more complex texts such as informational texts (Benjamin, 2012).

Cloze and maze as measures of text understandability

Cloze has been defined as "encompassing any procedure that omits portions of a text or discourse and asks readers or listeners to resupply the missing elements" (Oller & Jonz, 1994, p. 3). Scholars have praised cloze tests as a relatively easily developed and administered assessment to measure readability and global reading skills for ESL learners (Cohen, 1994; Hurley & Tinajero, 2001). A number of researchers who have used cloze tests have concluded that they are a valid and useful measure of text-level reading comprehension for L2 readers (Saeedi, 2016; Tabatabaei & Mirzaei, 2014; Williams, Ari, & Santamaria, 2011; Yamashita, 2003).

Chavez-Oller, Chihara, Weaver, and Oller (1985) contend that skeptics of cloze, such as Markham (1987), McKamey (2006), and Saito (2003), base their conclusions on faulty research designs that use texts that are not suitable for demonstrating the ability of cloze to capture text-level reading knowledge. Alternatively, others like Shanahan, Kamil, and Tobin (1982), incorrectly assume that they can prove a null hypothesis that there is no difference in readers' scores on contextualized and de-contextualized cloze passages. However, other researchers have acknowledged that the types of words deleted in cloze tests can affect inter-sentential processing (Alderson, Clapham, & Wall, 1995; Fotos, 1991; Hale et al., 1988).

Several adaptations have been developed for cloze to address criticisms of the method. One adaptation most relevant to this discussion is the multiple-choice cloze or maze. This version of the cloze was initially conceived in response to the criticism that answers could be ambiguous and thus call into question the test's validity. Maze restricts the number of answers available because the learner has to select from options instead of filling blanks. Maze also enables rapid scoring by raters with minimal expertise (Brown, 2005). Several studies support the use of the maze procedure as a valid assessment tool for L2 learners (Hale et al., 1988; Kumazawa, 2016; Sadeghi, 2010; Sattarpour & Ajideh, 2014).

Empirical research on readability and understandability

Research into the use of content area texts with L2 learners has indicated that secondary-level texts are too difficult. Chavkin (1997) analyzed the readability of various secondary L1 science texts and found that many high school chemistry and biology textbooks had reading levels beyond high school. Social studies textbooks used in L1 classrooms also had readability results that were above grade level (Robison, Roden, & Szabo, 2015).

Findings from L2 settings were comparable. Kasule (2011) noted that, regardless of the readability index used, science texts were challenging for their grade seven level intended readership. Subsequent cloze test results with target audience students showed that all texts were at their frustration level. Readability results from analyses of science books used in another study of Jamaican secondary learners demonstrated that the texts were appropriate with some scaffolding. However, accompanying cloze tests results revealed over 60% of the students were at the frustration level when reading the texts (Daniels, 1996). Conflicting results from another study of middle school learners in Nigeria indicated that those who completed cloze passages based on the textbooks were able to understand the passages. Thus, the authors concluded that the science textbooks were suitable for these learners (Umoke & Nwafor, 2015).

At the high school level, Reed and Kershaw-Herrera (2016) had Spanish L1 speaking seniors read several informational texts and answer comprehension questions about them. They found that readability and textual cohesion influenced these secondary English L2 learners' reading comprehension.

Much L1 research indicates that texts used at the postsecondary level are too difficult for learners. An investigation of texts used by L1 students in a variety of non-technical subject areas at a community college reported that over 50% of the students had reading levels below the textbooks readability scores (Cline, 1972). Results were similar for more recent research in specific subject areas. For instance, significant differences existed in the readability of some introductory accounting textbooks (Plucinski et al., 2009). All of the freshmen-level public speaking textbooks analyzed in another study had readability levels above the first year of college (Schneider, 2011). Likewise, the wide range of readability scores in sources cited in freshman speeches raised questions related to college freshmen's reading abilities and how that affects their selection of sources when they conduct research for class assignments (Gray, 2012). Spinks and Wells (1993) also emphasize the importance of textbook readability because it affects student performance. They found that the higher the readability levels (i.e., difficulty) of the textbooks used in college business courses are, the lower the course averages are and the higher the number of D's, F's, and student withdrawals from those courses are.

As with L1 learners, many postsecondary L2 students are reading at their frustration level. Cloze test results from one study showed that over 90% of South African post-secondary users of accountancy textbooks were reading texts beyond their proficiency level (Blignaut et al., 2016). Correspondingly, Iranian undergraduate and graduate English for specific purposes (ESP) students at beginner and pre-intermediate English proficiency levels were also using ESP texts from various subject areas that were beyond their current ability (Mohebbi, Nayernia, Nemati, & Mohebbi, 2017).

Although some research has investigated the readability of textbooks used in a variety of contexts, numerous questions remain. At present, it is becoming apparent that textbooks are excessively demanding for many L1 and L2 learners at both secondary and post-secondary levels. However, limited research is currently available regarding the suitability of textbooks used in the professional preparation programs like teacher education. This is especially the case in second and foreign language contexts. Access to this knowledge is important because of the deleterious impact that inappropriate textbooks have been shown to have on learners (e.g., Spinks & Wells, 1993). Therefore, the present study will investigate the readability and understandability of textbooks commonly used in English teacher preparation programs for bilingual teachers to ascertain whether these texts are at their appropriate level. The research questions are as follows:

- (1) What are the readability levels of excerpts from introductory SLA textbooks?
- (2) Were there significant differences among pre-service teachers' performance on understandability assessments for passages taken from each of the various textbooks?
- (3) To what degree is course performance related to reading performance with SLA textbooks?

Method

Participants

The participants were undergraduate pre-service English teachers from a research-intensive university located in a major metropolitan area in South Korea. There were 9 males and 27 females in the sample, whose mean age was approximately 23 years old. They had all been studying English since elementary school. Their English proficiency generally ranged from intermediate to upper intermediate. They were all third- and fourth-year students in a four-year undergraduate pre-service program that prepared them to become secondary-level English teachers.

Materials

The textbooks to be used in the analysis were selected first. Information for the textbooks used in the analysis can be found in Table 1 below. They were chosen based upon their being published within the past ten years, being written to introduce fundamental concepts in SLA to teachers, being produced by a well-known publisher, and their relatively widespread use as evidenced by their Amazon rankings. Three passages were taken from each textbook for the total of fifteen all together to be included in the Coh-Metrix analysis.

TABLE 1
Second Language Acquisition Textbooks Used for the Analysis

Textbook	Year	Publisher	Pages
A	2011	Cambridge	214
B	2014	Wiley-Blackwell	288
C	2008	Routledge	316
D	2015	Oxford	365
E	2008	Routledge	320

The two readability formulas used in this analysis were taken from Coh-Metrix version 3.0 (<http://www.cohmetrix.com/>). This is an online web tool that leverages the computational ability of technology to provides a more comprehensive readability analysis than traditional readability formulas by analyzing higher-level features of language and discourse (Graesser, McNamara, & Kulikowich, 2011; Graesser, McNamara, Louwerse, & Cai, 2004; McNamara, Graesser, McCarthy, & Cai, 2014). The Flesch-Kincaid Grade Level readability formula translates the Flesch Reading Ease Score into a grade level so that high numbers indicate increased difficulty (Coh-Metrix, n.d.). The Flesch-Kincaid Grade level has been supported as a valid measure of text readability (Greenfield, 2004), although growing debate exists regarding the veracity of this claim (Crossley, Skalicky, Dascalu, McNamara, & Kyle, 2017; De Clercq et al., 2014).

The Coh-Metrix L2 readability formula (RDL2) produces a readability score based on indices of variables that represent psycholinguistic and cognitive reading processes (Crossley, Allen, & McNamara, 2011). RDL2 includes information about word and sentence difficulty in the text (i.e., the same text features as traditional formulas), as well as word overlap and sentence syntax similarity to measure overall text cohesion (Crossley et al., 2008). Inclusion of these additional important text features give it a clearer relationship with psycholinguistic constructs related to reading comprehension (Crossley et al., 2011).

Instruments

In addition to the regular passages, 5 of the original 15 passages analyzed in the readability analysis were transformed into maze passages (see Table 2 below for maze understandability criteria). Maze is a modified cloze technique that presents learners with options rather than having them supply the answers themselves. It is assumed to be a measure of reading comprehension ability (Williams et al., 2011). The five texts selected were all about the topic of interlanguage to avoid varied topic selection being a source of spuriousness. The texts were adapted according to the recommendations of Guthrie et al. (1974) and Wright, (2013). The first sentence of the text was left in its original form. For the remainder of the text, every seventh word was replaced with three randomly ordered options that the reader had to choose from. One of these words was from the original text and the other two were distractors that were either the same part of speech or not. The distractors were also recycled words from the original passage so that there was overlap with the original specialized vocabulary from the text. This procedure was accomplished with the Maze Passage Generator web tool found at <http://www.interventioncentral.org/teacher-resources/test-of-reading-comprehension>. Following the guidance of Parker, Hasbrouck, and Tindal, (1992), passages

generally contained between 300-400 words and between 30-33 items each.

TABLE 2
Maze Comprehension Levels

Maze score	Level
85% or more	Independent level – learners can read the passage without assistance
50-84%	Instructional level – learners can read the passage with some assistance
0-49%	Frustration level – language is too difficult for readers

Note: These criteria were developed by Guthrie, Seifert, Burnham, and Caplan. (1974)

Research findings have supported the validity and reliability of maze tests. Maze has significantly predicted performance on standardized state reading assessments for English learners (Ives Wiley & Deno, 2005; Kim, Vanderwood, & Lee, 2016). Maze also predicts overall reading achievement as measured by standardized reading comprehension assessments with older L2 learners (Saeedi, 2016; Sattarpour & Ajideh, 2014). Bensoussan and Ramraz (1984) found that maze was as effective as multiple-choice questions for reading comprehension tests for postsecondary L2 learners in Israel. Hale, Rock, and Jirele (1989) studied over 400 TOEFL test takers and found a “relatively high correlation between total MC [multiple choice] cloze score and total TOEFL score [which] attests to the concurrent validity of the method... Also, the [multiple-choice] cloze procedure appears to provide assessment that is as reliable as the TOEFL” (p. 65).

Maze has also been found to be a useful measure of expository text comprehension (Gillingham & Garner, 1992; Tichá, Espin, & Wayman, 2009). Maze passages created from content area textbooks identify students who struggle with academic content reading and correlate with other reading and science assessments (Johnson, Semmelroth, Allison, & Fritsch, 2013; Tolar et al., 2012).

Procedure

The readability analysis was performed through the following steps. The textbooks were selected, then 15 individual texts used in the readability analysis were chosen. The texts were selected for inclusion based upon their being about topics that were common to all of the textbooks included in the analysis. The topics of the texts were: the critical period hypothesis, interlanguage, and learning strategies. The selected texts were then inputted into separate computer text files for analysis using the Coh-Metrix software. After the text files were prepared, they were analyzed using the Coh-Metrix web tool to establish their readability. Five of these texts were also copied into separate word files to prepare them to be adapted into the maze passages used to establish their understandability.

The five texts used for the maze understandability assessment were taken from the same sample of texts as those used for the readability analysis. These five texts were all about the topic of interlanguage to avoid the possibility of undue influence of varied topics on participants' performance. Interlanguage was also selected as a topic because it was present in all of the texts, and the lengths of the relevant passages were generally more similar than other topics. After selecting the five texts, they were each adapted into a maze passage. The texts were then piloted with English learners to check the instructions, items, and formatting. They were then arranged into their booklets and administered to the participants.

One concern when having participants complete the maze passages to evaluate text understandability related to the order in which the texts are presented (i.e., order effects). If participants all read the texts in the same order, they might develop background knowledge from reading the earlier texts that could affect their performance on the subsequent texts in the sequence. To address these threats, the texts were presented through a counterbalanced Latin Squares design. Another threat to validity was respondents' failing to take the experiment seriously. Experiments that involve unsupervised subjects are particularly susceptible to this problem (Oppenheimer, Meyvis, & Davidenko, 2009). This threat was mitigated through the researcher's presence during the completion of the maze assessments and the use of verbal and written prompts for respondents to give what they believe to be the most correct answers.

Data Analysis

The textual data were analyzed for readability via the Coh-Metrix web tool. The software determined readability according to the Flesch-Kincaid and RDL2 formulas. Both of the readability indexes deliver an overall evaluation of text readability and provide an indication of how the texts compared with each other. In the case of the Flesch-Kincaid index, readability is indicated on a grade-equivalent scale. So a readability of 10 reveals that the text is at a Grade 10 level according to the North American grading system. In contrast, the RDL2 is not leveled according to grade. Lower scores point to the text being more challenging to comprehend. Therefore, a text that scores at 10 is more difficult to read than one that scores at 15.

Maze passage results were used to evaluate the understandability of the texts. First, item scores for each passage were totaled and then compared with maze criteria to ascertain whether participants were at the independent, instructional, or frustration reading level for each passage. Next, a series of Friedman's tests were performed to determine whether there were significant differences in numbers of readers at both the independent and instructional level across the various textbooks. Finally, a Spearman's correlation was performed to ascertain whether there was a relationship between course grades and performance on the various maze tests assessments.

Results

Readability Results

Table 3 below displays word and sentence count summaries and results of readability analyses for both the Flesch-Kincaid and Coh-Metrix readability formulas. The Flesch-Kincaid readability analysis revealed that texts are within the Grade 12 to Grade 16 readability level, which puts them within the range for college-level texts that would be used by native speakers (see Figure 1).

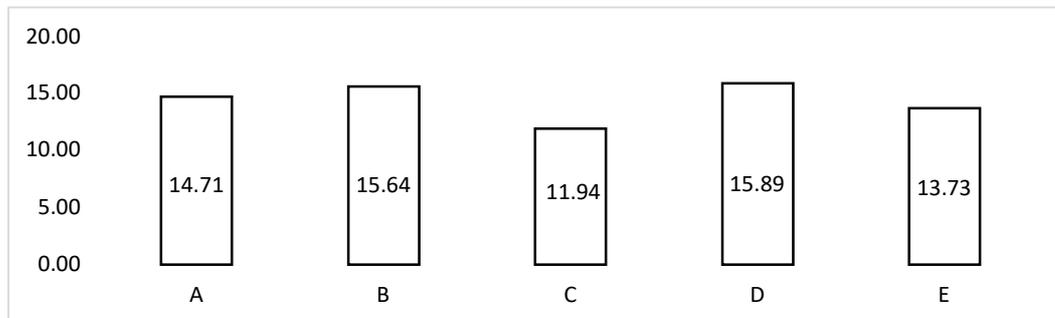


Figure 1. Mean Flesch-Kincaid readability per textbook.

Results from the RDL2 readability analysis echo those of the Flesch-Kincaid. RDL2 results differ from that of Flesch-Kincaid in that RDL2 does not report grade-level results, and higher readability scores in RDL2 indicate increased text readability (Goldin & Carlson, 2013). That is, “the metric is reversed scaled wherein higher numbers are related to more readable texts” (Crossley, 2017, personal communication). As exhibited in Figure 2 below, RDL2 readability scores range from 7.75 to 10.73 for four of the five textbooks thus demonstrating that these texts are rather difficult for L2 readers. As with the Flesch-Kincaid results, textbook C is somewhat easier for L2 learners to read than the others.

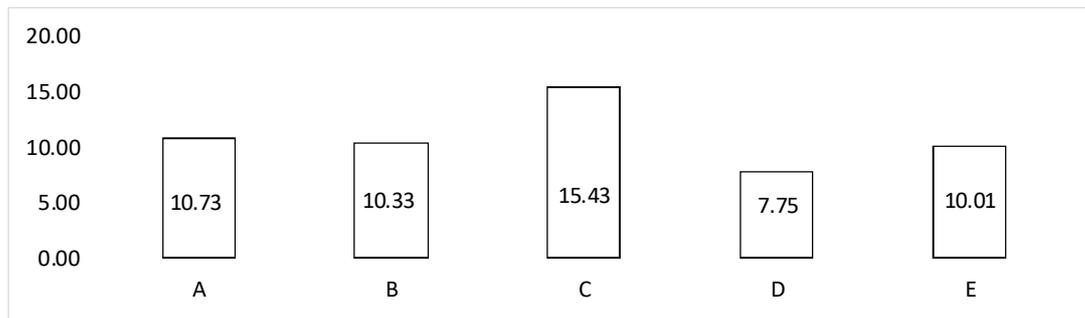


Figure 2. Mean RDL2 readability per textbook.

Findings from both of the readability analyses highlight the relatively high reading proficiency demands that all of the texts reviewed in this analysis present for L2 learners. The readability results concur that these passages are largely written for readers with a high level of reading proficiency and that they would be quite difficult for L2 readers to comprehend.

TABLE 3
Analysis of Text Readability Results

		Text features			Readability	
		Total number of words	Number of words/sentence M (SD)	Number of sentence/paragraph M (SD)	Flesch-Kincaid	Coh-Metrix
Textbook A	1	1198	23.03 (10.20)	2.6 (1.50)	13.39	13.02
	2	492	25.89 (14.09)	3.8 (2.04)	15.01	8.5
	3	673	25.88 (15.24)	3.71 (1.6)	15.72	10.68
Mean		787.66	24.93	3.37	14.7	10.73
Textbook B	1	736	30.66 (11.53)	3.42 (1.51)	16.65	12.58
	2	992	24.80 (14.24)	2.85 (1.46)	14.46	9.06
	3	865	26.21 (14.59)	1.94 (1.78)	15.80	9.36
Mean		864.33	27.22	2.73	15.63	10.33
Textbook C	1	1044	23.72 (12.92)	4.88 (3.01)	13.37	13.52
	2	1218	24.85 (14.37)	6.12 (2.9)	11.78	17.28
	3	1008	19.01 (9.67)	2.94 (2.38)	10.67	15.48
Mean		1090	22.52	4.64	11.94	15.42
Textbook D	1	782	26.96 (10.76)	5.8 (3.42)	15.31	9.37
	2	631	27.43 (12.31)	5.75 (1.89)	16.78	4.14
	3	1572	27.10 (15.95)	3.22 (2.43)	15.58	9.75
Mean		995	27.16	4.92	15.89	7.75
Textbook E	1	401	20.05 (9.56)	5.0 (2.16)	13.30	11.07
	2	436	22.94 (13.87)	3.16 (2.40)	13.25	10.3
	3	408	18.54 (9.47)	2.75 (2.43)	14.64	8.66
Mean		415	20.51	3.63	13.73	10.01

Maze Performance Results

As shown in Table 4 below, all of the passages are at the instructional level for learners, and many of them are at the independent level. That is, the majority of learners appear to be able to read many of the texts independently. In the case of texts C and E, almost all of the learners were at the independent level in their responses. Texts A, B, and D appear to be a bit more challenging in that there was a higher proportion of learners who scored at the instructional level. None of the texts were at the frustration level for any of the learners. Thus, it appears that learners could read all of the sample texts independently or with some scaffolding support.

TABLE 4
Maze Descriptive Statistics and Maze Readability by Passage

	A	B	C	D	E
N	36	36	36	36	36
No. of items	30	36	30	31	33
Mean	24.8	32.1	28.3	27.0	30.2
Median	25.0	33.5	29.0	27.0	30.0
Std. Deviation	2.62	3.88	2.21	2.26	1.93
Minimum	20.0	23.0	18.0	22.0	26.0
Maximum	29.0	36.0	30.0	31.0	33.0
85% or more Independent	47.2% (17)	69.4% (25)	91.6% (33)	69.4% (25)	83.3% (30)
50-84% Instructional	52.7% (19)	30.5% (11)	8.3% (3)	30.5% (11)	16.6% (6)
0-49% Frustration	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)

The assumption of data normality was violated, so a non-parametric one-way repeated-measures Friedman's test was employed in order to ascertain whether there were statistically significant differences between respondents' maze score performance across the five text samples. Differences in maze scores indicate discrepancies in passage comprehensibility. Prior to performing the Friedman's test, the maze scores were scaled as proportions in order to account for each of the maze passages having a slight difference in the number of items that it contained.

The Friedman's test identified statistically significant differences in mean scaled maze scores among the five maze passages, $X^2(4) = 51.85, p < .001$. Wilcoxon signed-rank tests post hoc comparisons with a Bonferroni correction determined that there were significant differences in maze passages scores for the A ($Mdn=0.83, IQR=0.133$) and B ($Mdn=0.93, IQR=0.138$) passages, $Z = -3.473, p < .001$. Differences were also observed between A ($Mdn=0.83, IQR=0.133$) and C ($Mdn=0.96, IQR=0.033$) passages, $Z = -4.461, p < .001$, as well as A ($Mdn=0.83, IQR=0.133$) and E ($Mdn=0.90, IQR=0.090$), $Z = -4.871, p < .001$. Additional differences were found between C ($Mdn=0.96, IQR=0.033$) and D ($Mdn=0.87, IQR=0.112$), $Z = -3.930, p < .001$, along with D ($Mdn=0.87, IQR=0.112$) and E ($Mdn=0.90, IQR=0.090$), $Z = -3.473, p < .001$.

In contrast, no statistically significant differences in maze score performance were found between passages A ($Mdn=0.83, IQR=0.133$) and D ($Mdn=0.87, IQR=0.112$), $Z = -2.860, p = .004$, passages B ($Mdn=0.93, IQR=0.138$) and C ($Mdn=0.96, IQR=0.033$), $Z = -2.615, p = .009$, or B ($Mdn=0.93, IQR=0.138$) and D ($Mdn=0.87, IQR=0.112$), $Z = -1.392, p = .164$. The same result was also noted for passages B ($Mdn=0.93, IQR=0.138$) and E ($Mdn=0.90, IQR=0.090$), $Z = -.652, p = .514$, as well as passages C ($Mdn=0.96, IQR=0.033$) and E ($Mdn=0.90, IQR=0.090$), $Z = -2.658, p = .008$. Therefore, it can be concluded that meaningful differences could be observed in learners' passage comprehensibility between five pairs of passages, but no meaningful difference could be found for five other pairs of passages.

The percentage of respondents who read each passage at the independent, instructional, and frustration levels was also calculated to determine whether there were large discrepancies in the proportion of test takers who read at the independent level across the five maze passages. Results from this analysis could help to establish whether some texts were easier for respondents to read independently than others. Findings were that the proportion of respondents who comprehended each of the passages at the independent level varied considerably among the passages. Passage A had the lowest percentage of respondents who read it at the independent level (47.2%), while passage C had the highest percentage (91.6%). The classification of reading levels on the passages is displayed in Table 5.

TABLE 5
Correlations of Course Grade and Maze Passage Score Variables

	Passage A	Passage B	Passage C	Passage D	Passage E
Course Grade	.531	.479	.188	.397	.457
p value	<.001	.003	.271	.016	.005

An additional correlational analysis was conducted comparing final course grades with learner maze assessment scores for each passage. The purpose of this analysis was to investigate whether there is a connection between participants' comprehension of these various types of related content area texts and their performance in a course focused on that specific content area. The existence of such a correlation would suggest the possibility of a relationship between course text comprehension and course performance. Assumption tests for the Pearson correlation revealed that the course grade variable and variables for passages A, D, and E had a non-normal distribution. Therefore, Spearman's rho was used for correlations calculated with these variables. As can be seen in Table 4 above, in four of the five maze passages, there is a statistically significant relationship between course grades and maze passage scores. Results of the Spearman correlation were that the variables of grades and passage A were significantly correlated with one another, $r_s = .531$, $n = 36$, $p = <.001$. The same was true for correlations between course grades and passage B, $r_s = .479$, $n = 36$, $p = .003$, grades and passage D, $r_s = .397$, $n = 36$, $p = .016$, as well as course grades and passage E, $r_s = .457$, $n = 36$, $p = .005$. However, the course grade and passage C scores were not significantly correlated, $r_s = .188$, $n = 36$, $p = .271$. In general, these results indicate that those who performed well on maze passages tended to have higher course grades, and those who performed less well on the maze tended to have lower course grades.

Discussion

Flesch-Kincaid and RDL2 readability results revealed that these texts are demanding for L2 readers. Understandability findings demonstrated that all learners read at either the instructional or independent level for all texts. Thus, the texts were comprehensible with scaffolding support. Some significant differences were observed in learners' maze performance across passages, and substantial discrepancies were found in the proportions of learners who read at the independent level among the texts. This finding suggests that there were some differences among the textbook passages in terms of their difficulty as measured by maze assessments. Lastly, correlations between course grades and maze performance demonstrated significant correlations across four of the five texts. This finding suggests that the texts were more challenging to read for those with lower grades in the course. Similar results were observed in a business course by Spinks and Wells (1993), who noted that higher college textbook readability levels corresponded with lower course averages and higher proportions of D's, F's, and student withdrawals.

Results also show somewhat of a discrepancy between Coh-Metrix judgments of text readability and learners' performance. On the one hand, the Flesch-Kincaid readability formula determined that these texts were generally written at the college level for L1 readers. Likewise, the L2 readability results assigned rather low scores to all of the texts, indicating that they would be quite demanding for L2 readers. There was some correspondence between maze and readability results in that the maze results showed that textbooks C and E were the most comprehensible, and the Flesch-Kincaid readability results concurred. The L2 readability analysis also agreed that C and E were the most readable, but in the case of E, not by a large margin. However, the disagreement in findings is that, according to the readability results, the textbooks are excessively difficult for readers to comprehend, while the maze results show that they are not.

These findings conflict to some extent with those of previous research. According to various readability indices, science texts were anticipated to be challenging for African middle school students. Follow up cloze test results confirmed the textbooks texts were indeed at the frustration level (Kasule, 2011). In

another study, readability formula results of science books used in Jamaican secondary schools with a high proportion of ESL students were that the texts were generally suitable. Cloze tests adapted from the same texts showed that they were at the frustration level (Daniels, 1996). However, Umoke and Nwafor (2015) presented one conflicting finding. Three computer studies textbooks used in a Nigerian state's junior secondary schools were examined to evaluate their understandability for local learners. Students had high cloze scores, so the authors concluded that the texts had high readability.

At the post-secondary level, L2 South African university-level business students completed cloze tests to determine the readability of their accountancy textbooks. Cloze results established that the text was at the frustration level of over 90% of the participants (Blignaut et al., 2016). In another study, chapters of ESP textbooks published by the Iranian government were randomly selected representing majors including humanities, science, medicine, and engineering. They were then subjected to an analysis with seven different readability formulas. Findings were that most texts ranged from difficult to very difficult to read and thus they were not suitable for Iranian post-secondary students studying English for specific purposes (Mohebbi et al., 2017).

Findings from previous readability analyses demonstrate that most studies found textbooks to be too difficult to read. That is, in many secondary and postsecondary contexts that include both L1 and L2 learners, their textbooks are beyond their ability level. The results for readability analyses conducted for the present study correspond with these findings. In contrast, results from previous understandability (i.e., cloze) analyses were somewhat mixed in that some show that texts are too difficult for learners (Blignaut et al., 2016; Daniels, 1996; Kasule, 2011), while at least one study reported that texts were comprehensible (Umoke & Nwafor, 2015).

The findings from the current study correspond to those of Umoke and Nwafor (2015). That is, according to maze test results, the texts commonly being used to teach the core skills and concepts within the field of SLA largely match the reading abilities of the pre-service teacher candidates assessed, provided that some candidates are granted some additional scaffolding. On the other hand, these results also correspond to other studies in that there is often an incongruence between the findings of the readability formulas and other tools used to assess learner's comprehension. As noted above, several other studies report that results from readability formulas and cloze tests do not match (Blignaut et al., 2016; Daniels, 1996).

Conclusion

The present study explored the readability and understandability of SLA textbooks used to teach preservice English teachers. The main findings for readability results were that the texts reviewed in this analysis were at the post-secondary level, and therefore were excessively demanding for L2 readers to understand. In contrast, text understandability as indicated by maze passages revealed that they were at the participants' independent or instructional level, although there were moderate correlations between text understandability as measured by maze passages and course grades. Thus, the texts could be read independently or with some support and were therefore comprehensible. These inconsistent results were similarly reported in other research.

Limitations in the research also merit some discussion. First, the results of this study cannot be generalized to all textbooks used in pre-service English teacher education programs because it is unclear how representative the texts were of textbooks used in other programs. Second, other potential influences on course grades and maze test performance (i.e., topic/content familiarity, general academic performance, motivation to be in the course, other individual differences, etc.) were not accounted for in the present analysis, so future research needs to address these oversights. Lastly, although this is only conjecture, the maze passages used in the study may have lacked sufficient difficulty, so additional research with somewhat stricter task criteria may be warranted. Learners performed well on maze tests produced according to the criteria suggested by Wright (2013). However, given these results, further scrutiny of the

test items led to the speculation that some of the questions could have been solved by deducing which distractors did not grammatically make sense rather than through actually having to comprehend the content of the passage. Thus, these criteria may have been too lenient. Additional follow up research could be conducted using the somewhat more stringent criteria, such as Parker et al. (1992), to establish whether learners produce comparable results.

Bearing in mind these limitations and the exploratory nature of the study, some implications for teacher educator practice are still worth noting. First, pre-service teacher educators in EFL teacher preparation programs should conduct readability analyses of the textbooks they plan to adopt for their courses taught in the target language. This type of analysis is vital given the sizable bilingual audience of consumers of their textbooks and the demands of a target language mediated course. Second, teacher educators should provide learners with access to other supplementary materials that they can use to scaffold their learning, such as helpful websites, videos on websites such as YouTube, or other supporting texts. These resources could also be simplified or written in the L1 to better build users' conceptual understanding before they begin to focus on the language. Alternatively, if learners have access to electronic versions of the texts, they might use online dictionaries or translation software such as Google Translate to better scaffold their understanding.

The Author

Dennis Murphy Odo is an Associate Professor in the Department of English Education at Pusan National University in South Korea. His research interests include second language reading, teacher development, technology, and use of the L1 in the L2 classroom. Some of his recent work has been published in *Modern English Education Journal*, *The Journal of Literacy Research*, and *English Language Teaching*.

Department of English Language Education
College of Education
Pusan National University
Pusan, 46241, South Korea
Tel: +82 515102614
Fax: +82 515823869
Email: dmodo@pusan.ac.kr

References

- Adelberg, A. H., & Razek, J. R. (1984). The Cloze procedure: A methodology for determining the understandability of accounting textbooks. *Accounting Review*, 59(1), 109–122.
- Alderson, J. C., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge, UK: Cambridge University Press.
- Armbruster, B. B., Osborn, J. H., & Davison, A. L. (1985). Readability formulas may be dangerous to your textbooks. *Educational Leadership*, 42(7), 18–20.
- Bailin, A., & Grafstein, A. (2001). The linguistic assumptions underlying readability formulae: A critique. *Language & Communication*, 21(3), 285–301.
- Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of text difficulty. *Educational Psychology Review*, 24(1), 63–88.
- Bensoussan, M., & Ramraz, R. (1984). Testing EFL reading comprehension using a multiple-choice rational cloze. *The Modern Language Journal*, 68(3), 230–239.
- Berti, A. E. (1991). Capitalism and socialism: How 7th graders understand and misunderstand the information presented in their geography textbooks. *European Journal of Psychology of*

- Education*, 6(4), 411–421.
- Blignaut, A. S., Wissing, G.-J., & Van den Berg, K. (2016). Using readability, comprehensibility and lexical coverage to evaluate the suitability of an introductory accountancy textbook to its readership. *Stellenbosch Papers in Linguistics*, 46(1), 155–180.
- Brown, J. D. (1998). An EFL readability index. *JALT Journal*, 15(2), 7–36.
- Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment* (2nd ed.). New York, NY: McGraw-Hill.
- Chall, J. S., Bissex, G. L., Conard, S. S., & Harris-Sharples, S. H. (1996). *Qualitative assessment of text difficulty: A practical guide for teachers and writers*. Cambridge, MA: Brookline Books.
- Chavez-Oller, M. A., Chihara, T., Weaver, K. A., & Oller, J. W. (1985). When are cloze items sensitive to constraints across sentences? *Language Learning*, 35(2), 181–206.
- Chavkin, L. (1997). Readability and reading ease revisited: State-adopted science textbooks. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 70(3), 151–154.
- Chen, A. C.-H. (2016). A critical evaluation of text difficulty development in ELT textbook series: A corpus-based approach using variability neighbor clustering. *System*, 58, 64–81.
- Cline, T. A. (1972). Readability of community college textbooks and the reading ability of the students who use them. *Journal of Reading Behavior*, 5(2), 110–118.
- Cohen, A. D. (1994). *Assessing language ability in the classroom* (2nd ed.). Boston, MA: Heinle and Heinle.
- Coh-Metrix. (n.d.). Retrieved December 22, 2017, from <http://www.cohmetrix.com/>
- Crossley, S. A., Allen, D. B., & McNamara, D. S. (2011). Text readability and intuitive simplification: A comparison of readability formulas. *Reading in a Foreign Language*, 23(1), 86–101.
- Crossley, S. A., Greenfield, J., & McNamara, D. S. (2008). Assessing text readability using cognitively based indices. *TESOL Quarterly*, 42(3), 475–493.
- Crossley, S. A., Skalicky, S., Dascalu, M., McNamara, D. S., & Kyle, K. (2017). Predicting text comprehension, processing, and familiarity in adult readers: New approaches to readability formulas. *Discourse Processes*, 54(5–6), 340–359. <https://doi.org/10.1080/0163853X.2017.1296264>
- Daniels, D. (1996). A study of science textbook readability. *Australian Science Teachers Journal*, 42(3), 61–65.
- Davidson, R. A. (2005). Analysis of the complexity of writing used in accounting textbooks over the past 100 years. *Accounting Education*, 14(1), 53–74.
- De Clercq, O., Hoste, V., Desmet, B., Van Oosten, P., De Cock, M., & Macken, L. (2014). Using the crowd for readability prediction. *Natural Language Engineering*, 20(3), 293–325.
- DuBay, W. H. (2004). *The principles of readability*. Costa Mesa, CA: Impact Information.
- Fotos, S. S. (1991). The cloze test as an integrative measure of EFL proficiency: A substitute for essays on college entrance examinations? *Language Learning*, 41(3), 313–336.
- Fry, E. B. (1989). Reading formulas: Maligned but valid. *Journal of Reading*, 32(4), 292–297.
- Fulcher, G. (1997). Text difficulty and accessibility: Reading formulae and expert judgement. *System*, 25(4), 497–513.
- Gillingham, M. G., & Garner, R. (1992). Readers' comprehension of mazes embedded in expository texts. *The Journal of Educational Research*, 85(4), 234–241.
- Goldin, I. M., & Carlson, R. (2013). Learner differences and hint content. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (eds.), *International conference on artificial intelligence in education* (Vol. 7926, pp. 522–531). Heidelberg, Germany: Springer.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods*, 36(2), 193–202.
- Gray, C. J. (2012). Readability: A factor in student research? *The Reference Librarian*, 53(2), 194–205.
- Greenfield, J. (2004). Readability formulas for EFL. *JALT Journal*, 26(1), 5–24.

- Guthrie, J. T., Seifert, M., Burnham, N. A., & Caplan, R. I. (1974). The maze technique to assess, monitor reading comprehension. *The Reading Teacher*, 28(2), 161–168.
- Hale, G. A., Rock, D. A., & Jirele, T. (1989). *Confirmatory factor analysis of the Test of English as a Foreign Language* (TOEFL Research Report No. 32). Princeton, NJ: Educational Testing Service.
- Hale, G. A., Stansfield, C. W., Rock, D. A., Hicks, M. M., Butler, F. A., & Oller JR, J. W. (1988). *The relation of multiple-choice cloze items to the Test of English as a Foreign Language* (TOEFL Research Report No. 26). Princeton, NJ: Educational Testing Service.
- Hurley, S. R., & Tinajero, J. V. (2001). *Literacy assessment of second language learners*. Boston, MA: Allyn & Bacon.
- Ives Wiley, H., & Deno, S. L. (2005). Oral reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education*, 26(4), 207–214.
- Johns, A. M. (1997). *Text, role and context: Developing academic literacies*. Cambridge, UK: Cambridge University Press.
- Johnson, E. S., Semmelroth, C., Allison, J., & Fritsch, T. (2013). The technical properties of science content maze passages for middle school students. *Assessment for Effective Intervention*, 38(4), 214–223.
- Jones, M. J. (1997). Methodological themes: Critical appraisal of the Cloze procedure's use in the accounting domain. *Accounting, Auditing & Accountability Journal*, 10(1), 105–128.
- Kasule, D. (2011). Textbook readability and ESL learners. *Reading & Writing-Journal of the Reading Association of South Africa*, 2(1), 63–76.
- Kim, J. S., Vanderwood, M. L., & Lee, C. Y. (2016). Predictive validity of curriculum-based measures for English learners at varying English proficiency levels. *Educational Assessment*, 21(1), 1–18.
- Kumazawa, T. (2016). Factors affecting multiple-choice cloze test score variance: A perspective from generalizability theory. *International Journal of Language Studies*, 10(1), 15–30.
- Markham, P. L. (1987). Rational deletion cloze processing strategies: ESL and native English. *System*, 15(3), 303–311.
- McKamey, T. (2006). Getting closure on cloze: A validation study of the "rational deletion" method. *Second Language Studies Paper*, 24(2), 114–164.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press.
- Mohebbi, H., Nayernia, A., Nemati, M., & Mohebbi, B. (2017). Readability of ESP textbooks in Iran: A neglected issue or a taken-for-granted one? *Journal of Teaching English for Specific and Academic Purposes*, 4(3), 641–654.
- Moussu, L., & Llurda, E. (2008). Non-native English-speaking English language teachers: History and research. *Language Teaching*, 41(3), 315–348.
- Oakland, T., & Lane, H. B. (2004). Language, reading, and readability formulas: Implications for developing and adapting tests. *International Journal of Testing*, 4(3), 239–252.
- Oller, J. W., & Jonz, J. (1994). *Cloze and coherence*. Cranbury, NJ: Bucknell University Press.
- Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology*, 45(4), 867–872.
- Parker, R., Hasbrouck, J. E., & Tindal, G. (1992). The maze as a classroom-based reading measure: Construction methods, reliability, and validity. *The Journal of Special Education*, 26(2), 195–218.
- Plucinski, K. J., Olsavsky, J., & Hall, L. (2009). Readability of introductory financial and managerial accounting textbooks. *Academy of Educational Leadership Journal*, 13(4), 119–127.
- Razek, J. R., Hosch, G. A., & Pearl, D. (1982). Readability of accounting textbooks. *The Journal of Business Education*, 58(1), 23–26.
- Reed, D. K., & Kershaw-Herrera, S. (2016). An examination of text complexity as characterized by readability and cohesion. *The Journal of Experimental Education*, 84(1), 75–97.
- Robison, T., Roden, T., & Szabo, S. (2015). Readability levels show that social studies textbooks are

- written above grade level reading. *Journal of Teacher Action Research*, 1(2), 100–111.
- Sadeghi, K. (2010). Cloze validation against IELTS reading paper: Doubts on correlational validation. *Journal of English Language Teaching and Learning*, 217(2), 131–154.
- Saeedi, M. (2016). Construct validity of multiple-choice cloze test and cloze-elide test in testing reading comprehension among Iranian EFL learners. *International Research Journal of Applied and Basic Sciences*, 10(6), 632–637.
- Saito, Y. (2003). Investigating the construct validity of the cloze section in the examination for the certificate of proficiency in English. *Spain Fellow Working Papers in Second or Foreign Language Assessment*, 1, 39–82.
- Sattarpour, S., & Ajideh, P. (2014). Investigating the relatedness of cloze-elide test, multiple-choice cloze test, and C-test as measures of reading comprehension. *Journal of English Language Teaching and Learning*, 6(13), 89–112.
- Schneider, D. E. (2011). Assessing the readability of college textbooks in public speaking: Attending to entry level instruction. *Communication Teacher*, 25(4), 246–255.
- Shanahan, T., Kamil, M. L., & Tobin, A. W. (1982). Cloze as a measure of intersentential comprehension. *Reading Research Quarterly*, 17, 229–255.
- Smith, K. J., & DeRidder, J. J. (1997). The selection process for accounting textbooks: General criteria and publisher incentives—a survey. *Issues in Accounting Education*, 12(2), 367–384.
- Smith, M., & Taffler, R. (1992). Readability and understandability: Different measures of the textual complexity of accounting narrative. *Accounting, Auditing & Accountability Journal*, 5(4), 84–98.
- Snyman, M. (2004). Using the printed medium to disseminate information about psychiatric disorders. *African Journal of Psychiatry*, 7(4), 15–20.
- Spinks, N., & Wells, B. (1993). Readability: A textbook selection criterion. *Journal of Education for Business*, 69(2), 83–87.
- Tabatabaei, O., & Mirzaei, E. (2014). Correlational validation of cloze test and c-test against IELT. *Journal of Educational and Social Research*, 4(1), 345–356.
- Tanaka-Ishii, K., Tezuka, S., & Terada, H. (2010). Sorting texts by readability. *Computational Linguistics*, 36(2), 203–227.
- Tichá, R., Espin, C. A., & Wayman, M. M. (2009). Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading-aloud and maze-selection measures. *Learning Disabilities Research & Practice*, 24(3), 132–142.
- Tinkler, S., & Woods, J. (2013). The readability of principles of macroeconomics textbooks. *The Journal of Economic Education*, 44(2), 178–191.
- Tolar, T. D., Barth, A. E., Francis, D. J., Fletcher, J. M., Stuebing, K. K., & Vaughn, S. (2012). Psychometric properties of maze tasks in middle school students. *Assessment for Effective Intervention*, 37(3), 131–146.
- Umoke, C. C., & Nwafor, C. E. (2015). Examination of the readability level of some approved science textbooks in use in junior secondary schools in Ebinyi state of Nigeria. *British Journal of Education*, 3(11), 26–31.
- Wait, S. S. (1987). *Textbook readability and the predictive value of the Dale-Chall, comprehensive assessment program, and cloze* (Unpublished doctoral dissertation). Florida State University, Tallahassee, FL. Retrieved from Dissertation Abstracts International (357A)
- Williams, R. S., Ari, O., & Santamaria, C. N. (2011). Measuring college students' reading comprehension ability using cloze tests. *Journal of Research in Reading*, 34(2), 215–231.
- Wright, J. (2013). Test of reading comprehension maze passage generator. Retrieved December 17, 2017, from <http://www.interventioncentral.org/teacher-resources/test-of-reading-comprehension>
- Yamashita, J. (2003). Processes of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, 20(3), 267–293.
- Zamanian, M., & Heydari, P. (2012). Readability of texts: State of the art. *Theory and Practice in Language Studies*, 2(1), 43–53.