Modern English Education

http://dx.doi.org/10.18095/meeso.2018.19.2.02

Evaluating SLA Textbook Comprehensibility: Comparing Coh-Metrix Results With EFL Teachers' Ratings

Dennis Murphy Odo

Pusan National University

ARTICLE INFO

Received 27 February 2018
Revised 5 May 2018
Accepted 17 May 2018

Examples in: English
Applicable Languages: English
Applicable Levels: Secondary

KEYWORD

readability
second language reading
teacher education
materials evaluation
가독성
제 2 외국어 읽기
교사 교육
자료 평가

ABSTRACT

Target language textbooks are widely used in English teacher preparation programs, but the decision to use them is commonly made with minimal consideration of their difficultly level for the teacher candidates. The purpose of this study was to assess their readability for bilingual pre-service teachers. To accomplish this task, the Coh-Metrix linguistic analysis software was used to analyze the readability and easability of five widely-used SLA textbooks. Additionally, a class of pre-service teachers provided readability ratings of these texts. Coh-Metrix readability indicators showed many of the texts to be at the post-secondary level for L1 speakers. Easability results revealed that texts often lacked many of the features that would make them easier to comprehend. Human readability ratings suggested that many pre-service teachers found them to be moderately challenging revealing somewhat of a mismatch with Coh-Metrix results. EFL teacher educators should consider readability when selecting textbooks for teacher education programs as they may need to provide teacher candidates with access to materials to scaffold their comprehension of these textbooks.

I. INTRODUCTION

Just like the textbook is the backbone of almost any English class, the textbooks used in classes for any program of study play a vital role in determining the content and structure of the courses in the program. If learners struggle to understand their textbooks, it makes their efforts to learn that much more frustrating because an unnecessarily-complicated textbook may cause the learner to ignore, fail to recognize, or misunderstand concepts that form the foundation of their disciplinary knowledge. Indeed, textbooks are only useful if their contents are comprehensible to students (Smith & Taffler, 1992; Snyman, 2004). As well, an understandable textbook helps learners to learn more independently and thus makes more class time available for non-lecture-based teacher-centered learning activities (Razek, Hosch, & Pearl, 1982).

One could argue that the stakes are even higher when it comes to the textbooks used in teacher education programs. The skills and concepts presented in these books largely constitute beginning teachers' professional knowledge base. As well, "In many classrooms the textbook is the chief reading source, the single window into the values and practices of a discipline…thus, in some disciplines, it is textbooks that initially shape [learners'] views" (Johns, 1997, p. 46). Therefore, as teacher educators we must pay careful attention to the suitability of textbooks that we adopt for our classes. In particular, we need to give cautious consideration to the language of these texts and how well it corresponds to the linguistic abilities of our teacher candidates. Failure to do so will likely result in their frustration and sub-optimal learning which will in turn have

Dennis Murphy Odo
The Department of English Language Education Pusan National University 2, Busandaehak-ro 63beon-gil, Geumjeong-gu, Busan, 46241, Korea
Tel: (051) 510-2614 / Email: dmodo@pusan.ac.kr

undesired consequences for their future students. Certainly, inappropriate textbooks are a clear danger in any teacher education program but the issue is only magnified for teachers of other languages when they are users of English as a second language.

Given this danger of inappropriate texts being used with pre-service teachers, the question then becomes: How do we establish whether the texts being used in a particular program are at the right level of comprehensibility for program participants? One method by which to ascertain text difficultly for readers is to use linguistic analysis tools such as Coh-Metrix. Such tools can provide a wealth of information about the text's readability. Information yielded from such an instrument thus can aid teacher educators in deciding whether the texts they are using to prepare new teachers are appropriate.

The subject matter taught in various second or foreign teacher preparation programs varies to some degree. However, in most current programs, some type of course typically exists whose primary aim is to instruct student teachers about the theories, concepts, and research findings relevant to second language acquisition/learning (SLA/SLL). The subject matter in courses of this nature tends to be somewhat more technical and cognitively challenging than many other courses because of the sheer number of new and complex concepts that are presented. Therefore, it is imperative that the materials used to teach these novel concepts are as transparent as possible.

The present research can help to establish whether the textbook materials available to those teaching SLA/SLL courses are largely comprehensible to the target audience of users. This audience includes both native speakers who intend to teach in ESL/EFL or bilingual programs in their own country or other nations across the globe. Of course, native and non-native speakers around the world will be learning from these materials. This aim of evaluating the understandability of these texts can be achieved through comparing results of objective readability analyses (i.e., Coh-Metrix) with target audience members' readability ratings to ascertain how well aligned they are.

As materials designed to support the explication of foundational and complex concepts in language teaching, it is uniquely important that they not be unnecessarily difficult. Unlike the textbooks used in most other disciplines, SLA texts must be written in a way that considers the needs of bilingual non-native speaking learners. This is because, unlike many other kinds of textbooks, a large proportion of their readership is non-native speakers who can only access the information contained in these texts in English.

II. PREVIOUS RESEARCH

1. Textbook Readability and Easability

A number of earlier studies have explored textbook comprehensibility. In these investigations, the distinction has been made between the concepts of text readability and understandability. Readability is considered to be a quality of the text itself and is typically measured by readability formulas (DuBay, 2004). Understandability, in contrast, relates more to the reader's experience or perceptions of the reading text. Thus, it is possible for a text to be judged by readability formulas as readable for learners at a certain level but those readers do not find it to be understandable (Jones, 1997; Plucinski, Olsavsky, & Hall, 2009).

1) Readability

Readability formula results are typically derived from average lengths of sentences and words in a text that provide a rough indicator of text difficulty (DuBay, 2004; Nuttall, 2005). However, the suitability of readability indexes for evaluating L2 texts has been called into question (Brown, 1998; Carrell, 1987). Researchers complain that the formulas fail to take into consideration text organization (Drury, 1985), readers' knowledge and motivation (Armbruster, 1985; Zamanian & Heydari, 2012). These indexes also ignore content difficulty and familiarity, author style, page layout (Armbruster, 1985; Fulcher, 1997) and illustrations (S. Choi, K. H. Kim, Y. B. Lee, J. H. Hong, & E. Cho, 2012).

Nevertheless, a number of other researchers claim that readability formulas help establish the difficulty of L2 reading materials (Y. Choi, 2010; Crossley, Allen, & McNamara, 2011; Greenfield, 2004; Y. Kim & J. H. Ma, 2012). Greenfield (2004) comparison of Japanese university students' reading performance with readability formula text level determinations led him to conclude that the traditional formulas predict the difficulty of texts quite well. Similarly, Crossley, Allen, and McNamara, (2011) acknowledged that "…to some degree, traditional readability formulas do measure levels of text difficulty related to cognitive reading processes such as decoding and syntactic parsing" (p. 97-98).

2) Coh-Metrix Easability

A newer approach to measuring text readability leverages technology to address some of the criticisms of readability formulas. For instance, in addition to readability, the Coh-Metrix web tool[1] has also been designed to measure text "Easability." McNamara and Graesser (2012) explain that "[the] Text Easability components provided by Coh-Metrix go beyond traditional readability measures by providing metrics of text characteristics on multiple levels of language and discourse" (p. 88). Easability evaluates particular text characteristics that more accurately represent current theories of discourse understanding (Kintsch,

[1] More information about Coh-Metrix can be found at http://www.cohmetrix.com/

1998; McNamara, Graesser, Cai, & Kulikowich, 2011) thereby providing insight into the effect of those characteristics on readers' comprehension ability (Jackson, Allen, & McNamara, 2016). Thus, Coh-Metrix "easability" analysis can inform teachers about which specific text features are causing a given reader the most difficulty thereby allowing teachers to select texts that better address students' needs (Jackson et al., 2016).

Coh-Metrix easability is measured through percentiles with a higher percentile demonstrating increased reading ease. Thus, a percentile score of 80% means that 80% of the texts in the comparison corpus are more difficult and only 20% are easier (McNamara et al., 2011).

The five text characteristics measured as part of "easability" are narrativity, syntactic complexity, word concreteness, referential cohesion and deep cohesion (McNamara et al., 2011). Narrativity indicates the text resembles a story with a coherent plot (Schank & Ableson, 1995) and comprehensible vocabulary (Perfetti, 2007). Syntactic simplicity is determined by the number of words, clauses per sentence, number of words before the main verb and similarity of grammar structures throughout the text. Syntactically simple texts are easier to understand. Word concreteness denotes words experienced by the five senses (e.g., spoon) rather than abstract words (e.g., democracy). Texts having many concrete words are easier to read (Jackson et al., 2016). Referential cohesion refers to the explicit connections between words or concepts across sentences in the text (Graesser, McNamara, & Kulikowich, 2011). Absence of referential cohesion causes readers to have difficulty seeing how sentences relate to each other. Deep cohesion refers to how textual information is integrated through logical connecters such as after, because, among others. Lack of deep cohesion indicates limited coherence in textual information (McNamara, Louwerse, McCarthy, & Graesser, 2010).

3) Understandability

Understandability has been described 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). Although they are related, clearly readability and understandability are not entirely synonymous (Blignaut, Wissing, & Van den Berg, 2016; Davidson, 2005). In contrast to readability, understandability or comprehensibility has typically been measured through assessment instruments such as cloze or more subjective means such as readability ratings. In the following section, some of the empirical research related to readability and understandability findings for textbooks used at the secondary and postsecondary level will be reviewed next in order to establish how easy these texts are to read for their target audience.

2. Empirical Research on Textbook Readability and Comprehensibility

1) Secondary-level Text Comprehensibility

Research into the use of content area texts with both L1 and L2 learners has indicated that secondary-level texts were too difficult for learners. Wray and Janan (2013) contend that while the reading demands of postsecondary study, of workforce training and being a good citizen remained stable or risen over the previous 50 years or so, the texts used to teach subjects in American high schools have become less demanding. Their review of the relevant literature led them to conclude that there is a need to develop the reading abilities of all secondary school students. Thus, they call for improving the teaching of reading in secondary schools by incorporating more complex reading materials. Chavkin (1997) analyzed the readability of various secondary science texts designed for L1 readers. He observed that four out of the five high school chemistry textbooks had reading levels beyond high school. Biology textbooks had approximately a 50 percent match with anticipated grade level. This analysis also did not take into account second-language readers who are certain to compose some proportion of many North American high school classes. Another readability study of social studies textbooks used in first-language classrooms at the elementary level indicated that these texts were also above grade-level (Robison, Roden, & Szabo, 2015). In a third study, Reed and Kershaw-Herrera (2016) manipulated levels of readability and cohesion of several informational texts. They then asked high school seniors to read the texts and answer factual and inferential comprehension questions. They found that these secondary learners' reading comprehension is dually influenced by a text's readability and cohesion.

Research in second-language contexts yielded similar results. Regardless of the readability index used, grade seven science texts used in an African middle school were predicted to be excessively demanding for that grade level. Follow up cloze test results based on the texts showed that all texts were at their frustration level (Kasule, 2011). Fry and Smog formula results from analyses of science books used in Jamaican secondary schools where English is a second language for many students revealed that the texts were generally at or below the appropriate language level. Thus, it seemed that they could be used in Jamaican schools with teacher scaffolding. However, results from cloze tests based on text passages were that, for six out of nine textbooks examined, over 60% of the students were at the frustration level when reading the texts. In one instance, 87% of the students were at the frustration level (Daniels, 1996). In contrast to these results, findings reported by Umoke and Nwafor (2015) were that the middle school learners in Nigeria who completed cloze passages based on the textbooks were able to understand the passages. Thus, they concluded that the science textbooks were suitable for these learners.

2) Postsecondary-level Text Comprehensibility

Some doubts have also been raised about L1 post-secondary students' ability to handle the texts they are required to read. Williamson (2008) notes that there have been substantial increases in text demands from high school texts to citizenship materials, workplace materials, community college texts, and university texts. Results from empirical research into the readability of postsecondary materials in L1 contexts indicate that there is considerable variability in these texts and some are too challenging for their readership. Schneider's (2011) investigation of a variety of public speaking textbooks with the SMOG readability index demonstrated that the texts were all above the first year college level for which they were expected to be written. Indeed, one-third of the texts were more than two years beyond the appropriate level. Similarly, the wide range and low readability scores in sources cited in freshman speeches raised concerns for Gray (2012). She expressed misgivings about college freshmen's limited reading abilities effect on their selection of sources when performing research for class assignments. As well, although they generally did not find texts to be overly demanding, (Plucinski et al., 2009) did observe significant differences in the readability of some introductory accounting textbooks. Correspondingly, they noted that some of those books would be too difficult for their target audience. Excessive readability demands have also been linked to student failure. That is, the higher the readability levels of the textbooks used in college business courses, the lower the course averages and the higher the number of D's, F's and student withdrawals were found in those courses (Spinks & Wells, 1993).

Research results for postsecondary L2 students indicate that many are reading at their frustration level. One study investigated accounting textbooks used with marketing students at a university in South Africa. Although the readability formula results and vocabulary test scores showed the texts to be at a suitable level, cloze test results suggested the contradictory finding that the textbook users were reading at their frustration level (Blignaut et al., 2016). Other researchers processed 11 ESP textbooks covering varied humanities and science content with seven different readability formulas to establish their suitability for pre-intermediate Iranian post-secondary ESP students. They determined that the majority of the texts in various ESP textbooks were too difficult for these learners to read (Mohebbi, Nayernia, Nemati, & Mohebbi, 2017).

3) Readability Ratings Research

Readers' ratings have also been compared with text readability indexes in order to establish how well readability formulas reflect readers' judgments of readability (Kanungo & Orr, 2009; Rello, Pielot, Marcos, & Carlini, 2013). Studies of the readability of informational texts have generally shown low correlations between readers' judgements and readability indexes (Pitler & Nenkova, 2008; Yan, Song, & Li, 2006). Pitler and Nenkova (2008) examined which text features best correlate with college students' readability ratings for articles from the Wall Street Journal corpus. They found that surface feature measures such as average number of words per sentence or average characters per word did not predict readability ratings well. However, discourse features such as syntactic and lexical coherence or discourse relations did highly associate with readers' ratings.

Non-specialist users searching for medical information online often have difficulty determining which results are the easiest to read. Investigators asked non-native English speakers to provide their judgments about the word level readability of texts about medical topics that they retrieved from online searches. Low correlations were found between readers' ratings and the results of readability formulas which included the ARI, Flesch, Gunning, except for the Dale-Chall formula (Yan et al., 2006).

Related research has shown discrepancies in the relationship between readability formulas and reader ratings of textbooks. Flesch Reading Ease Index results from economics textbooks were compared with judgments of native and non-native English speakers recruited through Mechanical Turk. The Flesch Index was found not to predict which samples readers will prefer indicating that the readability formula was not an effective predictor of these readers' judgments of text difficulty. Different preferences were also apparent between native and proficient L2 English readers to the text samples suggesting that non-native readers may have unique criteria for what they consider to be a readable text (Tinkler & Woods, 2013).

A similar study was conducted with Iranian EFL teachers and graduate students who completed several readability rating Likert-type scales to rate the difficulty of EFL reading texts according to several text features. The same texts were also processed with Flesch Readability index. Some discrepancies in the results were that readers typically rated the texts as being easy or very easy while the Flesch readability formula indicated that they were fairly easy to very difficult thus seeming to overestimate the difficulty of the texts (Heydari & Riazi, 2012).

3. Research Questions

The preceding review of the extant literature demonstrates that while attention has been given to the readability of textbooks in a variety of contexts and subject areas, limited investigation exists of textbooks used in classes for bilingual pre-service teachers. As well, additional research is needed that compares human judgment and machine analysis of these types of texts. This research is warranted by the fact that these textbooks are being used to teach foundational concepts that form the basis of pre-service teachers' professional knowledgebase. To address this existing gap in the research literature the following research questions have been asked:

1) How comprehensible are second language acquisition textbooks as measured by the presence of text characteristics that indicate reading ease?

2) To what extent are the textbooks judged to be readable by pre-service EFL teachers according to a number of readability rating criteria?

3) How do pre-service EFL teachers' ratings of text-readability compare with readability formula results?

III. METHODS

1. Study 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 10 males and 22 females in the sample whose mean age was 23 years old. 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.

2. Materials

The main materials used in the study consisted of 15 reading passages that were selected from five different textbooks designed to be used as introductory materials to teach second language acquisition/learning concepts to beginning teachers or researchers. Each of the texts was selected based upon several criteria. First, all of the texts included in the Coh-Metrix easability and readability analysis had to be about one of three topics. These three topics were interlanguage, the critical period and learning strategies. The texts used in the study were chosen on the basis of their sharing common subject matter in all of the books that were included in the analysis. Of these topics, interlanguage was selected as the topic used for further investigation of respondents' readability ratings across the textbooks. Additionally, the texts chosen had to stand alone in that the learner could read and comprehend them without requiring access to other parts of the textbook. Text length was also taken into consideration during text selection. They had to be at least 200 words to allow for a valid analysis by Coh-Metrix.

A second instrument used in this research was a questionnaire based on the work of Heydari and Riazi (2012). The survey asks respondents to read the text and rate the its readability according to 10 text characteristics that indicate readability gathered from prior research. Respondents rated the text characteristics on a 10 point Likert scale with 1 indicating the text was easier and 10 that it was more difficult. Due to time constraints and the potential burden of the amount of reading required, one passage from each textbook was selected for a total of five passages all together. Comparable measures of readers' ratings of text readability have been used in other research (e.g., Efklides, Papadaki, Papantoniou, & Kiosseoglou, 1998; Pitler & Nenkova, 2008; Tinkler & Woods, 2013; Yan et al., 2006). The reliability for each of the passage readability ratings from the questionnaire was also calculated with Cronbach's alpha. An alpha coefficient (α) above 0.8 indicates an acceptable level of instrument reliability. The results are reported in Table 1 below.

TABLE 1
Reliability Statistics

Passage	*A*	# of Items
1	0.859	10
2	0.884	10
3	0.881	10
4	0.881	10
5	0.838	10

A final tool used in the research was the Coh-Metrix linguistic analysis software. This software has been used in a growing number of studies to provide a fine-grained analysis of a large number of text characteristics that contribute to the coherence, cohesion and comprehensibility of text. The specific component of the Coh-Metrix tool used in this analysis is known as the "easability" assessor. In essence, the easability assessor consists of five subcomponents which are: narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion. Analysis of these text dimensions provides well-validated indicators of text ease according to a large and rapidly growing number of research studies (Graesser et al., 2011; Jackson et al., 2016; Kintsch, 1998; McNamara et al., 2011; McNamara & Magliano, 2009). Besides the easability tool, the Coh-Metrix also offers the widely used readability scores of Flesch Reading Ease and Flesch-Kincaid Grade level.

3. Procedures

The textbooks to be used in the analysis were selected first. They were chosen based upon their being published within the past ten years, being written to introduce fundamental concepts in SLA to teachers, being published 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.

After the textbooks were designated, the 15 individual texts to be used in the analysis were chosen. The selected texts were then inputted into separate computer text files for analysis using the Coh-Metrix software. Five of these texts were also copied into separate word files to prepare them to be included with questionnaire forms that were

used to establish their readability ratings. Once the text files were prepared and formatted, they were analyzed using the Coh-Metrix web tool to ascertain their readability and easability.

The five texts used for participants' ratings of text readability levels were taken from the same sample of texts as those used for the easability analysis. These five texts were all about the topic of interlanguage to avoid the possibility of undue influence of varied topics on participants' ratings. Interlanguage was also selected as a topic because it is widely-discussed in language teacher education, 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 combined with the readability rating questionnaire. The texts were then piloted with Korean English learners to confirm that survey instructions, items and formatting were all unambiguous and there were no other typographical errors or other issues. Upon satisfactory piloting of the passages, they were then arranged into their booklets and provided to the participants to be rated.

One concern when having participants read the texts to evaluate their readability related to the order in which the texts are presented (i.e., order effects). Other potential sources of bias in respondents' answers are the primacy and recency effects. The issue was that if participants all read the texts in the same order, they might develop background knowledge from reading the earlier texts that could make the later texts in the sequence easier for them to read. These threats were addressed by changing the order of the texts as they were presented through a counterbalanced design. Specifically, a Latin Squares design will be used to avoid order effects.

Another threat to validity was the respondents' random answers to readability preferences because they didn't 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 surveys and the use of verbal and written prompts for respondents to give thoughtful answers that truly reflect their views.

4. Data Analysis

The textual data was analyzed for readability and easability in Coh-Metrix. The software determined readability according to the Flesch Reading Ease formula. Easability was assessed through the Coh-Metrix software's comparison of the texts' narrativity, syntactic simplicity, word concreteness, referential cohesion and deep cohesion with other texts in the Touchstone Applied Science Associates (TASA) corpus. Both the readability and easability indicators provided evidence for the levels of comprehensibility across the various texts used in the analysis. The readability results deliver an overall evaluation of text readability. In contrast, easability results offer some insight into how the texts differ from each other on a more fine-grained level by revealing more subtle differences among them related to the easability characteristics.

Coh-Metrix easability results were evaluated and compared to determine whether there were any discrepancies in level of difficulty across texts according to easability characteristics. Individual readability ratings results were then examined to discover how readers rated the readability of the texts according to various text characteristics. Lastly, participants' overall readability ratings were compared with Flesch reading ease scores in order to understand how well they corresponded with each other.

IV. RESULTS

1. Easability

Findings from the easability analysis revealed that all texts showed low narrativity (see Table 2). Narrative texts typically use everyday oral language to relate stories comprised of familiar characters, events, places, and so forth. Narratives contain more familiar topics, high frequency words, verbs, and adverbs than informational text (Graesser et al., 2011). These characteristics of narrative texts make them easier to comprehend so the relative lack of these features exhibited here indicate that these texts are difficult to comprehend. Most texts also showed low to moderate syntactic simplicity meaning that they were characterized by relatively more words and more complex syntactic structures. This result serves as another indicator of the challenge level in reading these texts. As well, many texts were characterized by low word concreteness which means that they tend to contain abstract words that are

TABLE 2

Coh-Metrix Analysis of Text Complexity Features

		Total number of words	Easability				
			Narrativity	Syntactic Simplicity	Word Concrete	Referential Cohesion	Deep Cohesion
Textbook 1	1	1,198	22.36	50.80	12.30	59.48	40.90
	2	492	25.46	23.58	9.01	51.60	69.50
	3	673	20.90	28.77	29.81	74.54	68.44
Mean		787.66	Low	Low/mod	Low	Mod	Mod
Textbook 2	1	736	21.19	22.96	14.23	51.60	89.25
	2	992	14.23	42.86	19.77	25.48	86.65
	3	865	26.11	21.19	20.33	33.72	99.29
Mean		864.33	Low	Low/mod	Low	Low/mod	High
Textbook 3	1	1,044	19.22	31.21	28.10	22.36	43.25
	2	1,218	48.01	24.51	25.46	59.87	62.55
	3	1,008	64.06	41.29	10.03	65.91	82.89
Mean		1,090	Low/mod	Low/mod	Low	Mod	Mod/high
Textbook 4	1	782	18.41	31.21	48.80	31.56	99.41
	2	631	10.56	36.32	36.32	29.81	54.78
	3	1,572	23.27	67.72	44.04	30.50	72.57
Mean		995	Low	Low/mod	Low/mod	Low/mod	Mod/high
Textbook 5	1	401	27.09	56.36	27.09	17.62	95.15
	2	436	43.25	56.36	14.01	24.51	97.61
	3	408	23.27	73.57	5.48	33.00	97.78
Mean		415	Low/mod	Mod/high	Low	Low/mod	High

Note. Textbook 1 – low easability, high readability; Textbook 2 - …

more difficult to visualize and understand.

Results for cohesion-related text characteristics were a bit more favorable. Low to moderate referential cohesion apparent across that majority of texts analyzed demonstrated that to some extent these texts contain words and ideas that overlap throughout the whole text. Similarly, findings of moderate to high deep cohesion reveal that causal, intentional, and other types of connectives are in place to enable the reader to gain a more coherent understanding of the passages. The overall picture of easability that emerges from the analysis is that the low to moderate results for the majority of these text features point to these texts all being quite challenging to read (see Figure 1).

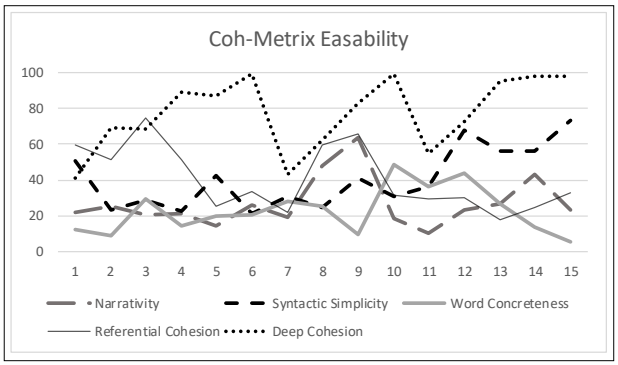

FIGURE 1 Coh-Metrix Easability Percentile Results for Each Text Passage

2. Readability Ratings Criteria

Readability ratings were comprised of the ten individual criteria shown in Table 3. Learner ratings for each of the readability criteria indicate that each of the passages is generally in the moderate readability range. Thus, the findings suggest that they had moderate amounts of difficulty reading the passage according to each of the individual readability criteria. In general, it appears that Coh-Metrix easability results indicate the texts are difficult to read. However, respondents' readability ratings reveal that they are moderately difficult. Thus, there is a discrepancy in the results of the Coh-Metrix computational text analysis and respondents' ratings.

TABLE 3
Readability Ratings of Text Features

Read. Rating Criteria	1		2		3		4		5	
Word fam.[a]	3.813	Mod	4.156	Mod	3.313	Easy	3.625	Easy	4.875	Mod
Word freq.[b]	3.813	Mod	4.531	Mod	3.719	Mod	4.375	Mod	4.438	Mod
Topic/Cont.[c]	4.188	Mod	4.344	Mod	4.125	Mod	3.938	Mod	4.844	Mod
Sent. length[d]	4.781	Mod	4.406	Mod	4.344	Mod	4.188	Mod	6.344	Mod
Pro. density[e]	4.097	Mod	4.333	Mod	4.125	Mod	4.188	Mod	5.094	Mod
Polysemy	3.969	Mod	3.813	Mod	3.719	Mod	4.375	Mod	4.406	Mod
Comp. syn.[f]	4.313	Mod	4.594	Mod	4.469	Mod	4.281	Mod	6.125	Mod
Concreteness	4.594	Mod	4.750	Mod	4.469	Mod	4.656	Mod	5.406	Mod
Imageability	4.750	Mod	4.938	Mod	4.656	Mod	4.531	Mod	6.063	Mod
Con. clar.[g]	5.000	Mod	4.719	Mod	4.531	Mod	4.719	Mod	5.719	Mod

Note. [a]Word familiarity, [b]Word frequency, [c]Topic/Content, [d]Sentence length, [e]Pronoun density, [f]Complex syntax, and [g]Concept clarity

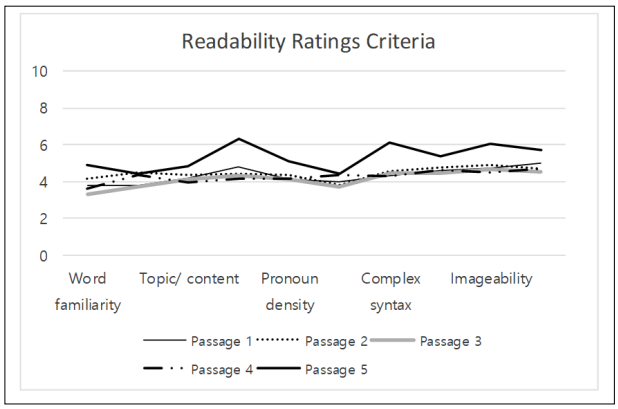

FIGURE 2 Respondents' Readability Ratings Results for Each Text Characteristic

3. Overall Readability Ratings Results

Readability rating surveys were constructed so that higher scores indicated greater difficulty. Results for the overall passage totals are presented in Table 4. When these findings were compared with those from a Flesch readability analysis, they showed that learners found the passages to be moderately difficult. In contrast, the Flesch scores indicated that the same passages were generally difficult to very difficult. Thus, as with the comparisons between readability characteristics ratings and Coh-Metrix easability, it appears that the learners perceived the passages to be easier than the Flesch reading ease scores determined them to be.

TABLE 4
Comparison of Learner Ratings With Flesch Reading Ease Scores

Passage	Readability Rating Score	Difficulty	Flesch Reading Ease Score	Text Difficulty*
1	43.187	Moderate	33.517	Difficult
2	44.312	Moderate	40.908	Difficult
3	41.469	Moderate	54.874	Fairly difficult
4	42.875	Moderate	35.559	Difficult
5	53.313	Moderate	23.578	Very difficult

Note. * Adapted from Finn, S. (1985).

V. DISCUSSION

1. Summary of Results

This study investigated the comprehensibility of second language acquisition/learning textbooks that are commonly used in English teacher preparation programs for bilingual teachers. This research was conducted to consider the level of difficulty of these textbooks for pre-service teachers who are often speakers of English as a second language. Ascertaining the level of challenge of these types of text was accomplished in two ways. First,

Coh-Metrix, an online linguistic analysis tool, was used to evaluate the easability and readability of these texts. The main finding with respect to text easability was that most indicators showed the passages to be rather difficult. The particular sources of difficulty among the texts are likely to be their limited narrativity, complex syntax and abstract vocabulary. Likewise, the Flesch reading ease readability analysis reported that the texts were difficult.

Secondly, students in a pre-service English teacher education program completed a readability rating survey to reveal their perceptions of the readability of the SLA/SLL texts. Overall passage readability ratings indicated that participants found them to be moderate in their readability and additional analysis of ratings for individual text characteristics showed that they were also typically moderately difficult as well.

The readability and easability findings were that these texts were generally difficult. According to the Flesch reading ease scale, they would be at the high school/ college level for native speakers. Likewise, the easability analysis indicated that narrativity, complex syntax and abstract vocabulary were more challenging than a high proportion of texts in the comparison database. Thus, they suggest that these texts might be too difficult for the majority of bilingual pre-service teachers who are expected to learn about the scientific foundation that is supposed to undergird their teaching from them.

In contrast to the Flesch readability and Coh-Metrix easability findings, respondents' readability ratings results reveal that participants found the texts to be moderate in their readability rather than difficult. Thus, they did not appear to perceive the texts to be too difficult to read. A number of possible reasons may explain this apparent mismatch in results. First, these are teacher candidates from a selective university that admits a low percentage of its applicants. Secondly, they are in a department that is very competitive to enter within the College of Arts and Humanities admitting approximately six percent of its applicants. Lastly, they are in the third or fourth year of an English teacher education program. For these reasons, their probable superior English reading skills relative to their peers in other programs may explain their confidence with respect to reading these texts.

2. Connections to Previous Research

The bulk of previous research on textbook readability appears to have relied primarily on readability formulas to verify the readability of texts. This text-centered readability analysis with readability assessed through formulas has demonstrated that both L1 and L2 learners find their secondary-level content area texts to be excessively demanding. Research of L1 second-level readers shows that many chemistry and biology texts had reading levels beyond high school in both their science (Chavkin, 1997) and social studies classes (Robison et al., 2015). One study in an ESL environment that used results of both readability formulas and cloze tests showed that the texts to be appropriate with scaffolding according to the formulas. However, cloze results indicated that the same texts were too demanding for over 60% of the students (Daniels, 1996).

In the postsecondary domain, first-language research revealed that texts used in a variety of different disciplines were also found to be generally too difficult for students. Only texts related to public speaking were found to be suitable for college students (Schneider, 2011). In other research, Gray (2012) examined readability scores in sources cited in freshman speeches and concluded that many students chose short texts with low readability levels calling their ability to comprehend college-level texts into question. Research with introductory accounting textbooks also points to some being more difficult than others (Plucinski et al., 2009). Mohebbi et al., (2017) used a variety of readability formulas to analyze English-for-specific-purposes texts in an EFL context. They observed a serious mismatch between the levels of the students who typically ranged from beginner to pre-intermediate and the texts they used which were mainly suited to L1 college level.

In addition to readability analyses through formulas, reader-centered comprehensibility has also been evaluated through examination of readers' cloze performance on textbook excerpts. Research in secondary school L2 contexts has shown that the majority of science texts investigated were at the frustration level for learners in a Swaziland middle school (Kasule, 2011). In contrast to these findings, Umoke and Nwafor (2015) used cloze to evaluate the readability level of some science textbooks in use in several middle schools in Nigeria. Their results indicated that learners had no difficulty comprehending the texts.

Learners taking accounting classes at a South African university completed cloze tests to determine how well they comprehended their texts. Their cloze performance revealed the textbooks to be at their frustration level for reading (Blignaut et al., 2016). However, these results conflicted with the researchers' other readability formula assessments that showed the texts were at the undergraduate level which would make them appropriate for these learners. Their findings were similar to those of Heydari and Riazi (2012) in that there appeared to be something of a mismatch between their Iranian EFL participants' perceptions of the ease of the passages in their advanced EFL reading textbook and the measures of more objective readability instruments. Heydari and Riazi (2012) observed that their participants generally evaluated the texts as ranging from easy to very easy while the readability formulas suggested that the texts would range from fairly easy to very difficult. Thus, the readability formulas indicated that the texts would be more difficult than the learners perceived them to be.

Readability ratings research by and large presented low correlations for readers' readability perceptions and readability indexes results for various types of texts (Pitler & Nenkova, 2008; Yan et al., 2006). For instance, Pitler and Nenkova (2008) observed that surface text features (e.g.,

words per sentence) correlated minimally with readability ratings while discourse features like coherence had considerably closer associations in non-fiction texts. Comparable findings were noted with health-related informational texts (Yan et al., 2006). Tinkler and Woods (2013) noticed incongruities between readability formula scores and reader ratings of textbooks as well. They also raised the possibility that second-language speaking readers may have different standards for text readability than first language readers. With respect to ratings and readability comparisons in L2 contexts, Heydari and Riazi (2012) also found that EFL readers' ratings differed from readability formula scores with formulas seemingly overestimating text difficulty. Thus, much of the research comparing readability ratings and readability indices exposes frequent inconsistencies between the two. At least one study concluded that formulas overrate text difficulty in comparison to reader ratings (Heydari & Riazi, 2012).

VI. CONCLUSION

Previous readability studies at the secondary and post-secondary levels in L1 and L2 contexts point to textbooks being too difficult for readers. The results of the easability analysis performed in this study are not fully in accordance with those findings. That is, the learners here report that they found the texts to be moderately challenging rather than being too easy or difficult.

The observed mismatch between the teacher candidates' perceptions of the textbooks and readability formula results is more consistent with earlier research findings. While the formulas indicate that the texts are too difficult, the respondents claim that they are not. One reason for this discrepancy could be that the participants in this study are from a selective department at a competitive university. Therefore, they may find these challenging texts to be suitable but they may not represent pre-service EFL teachers more generally. Alternatively, there could be social desirability bias in their survey results. That is, they could be claiming to understand the texts better than they actually do because they feel they are expected to be proficient English users. Only additional research can better establish the cause of the discrepancy between formula and rating survey results detected in the present study.

Some potential limitations to the research should also be acknowledged. First, given the limited sample of texts used in the analysis, we cannot generalize to all second language acquisition textbooks because it is not possible to state with absolute certainty how representative the chosen texts were of the readability of these types of textbooks in general. Second, the research was only conducted in one context with academically well-prepared learners so additional research should be performed at other institutions to compare these findings with English teacher candidates from a wider variety of backgrounds. Lastly, there are both advantages and drawbacks of the self-report methods employed here to measure text comprehension. The benefits are that they offer a cost effective way to gather data that cannot be collected through other means such as observation (Hoskin, 2012). However, we must be aware that these data can be unreliable as respondents under or over-estimate their relevant behaviors (Prince et al., 2008) or may be more concerned with image management (Razavi, 2001). Accordingly, future research should attempt to correct for these potential limitations as well as to enhance our currently insufficient understanding of the respondents' internal judgment criteria through more qualitative investigations such as in-depth interviews.

Some implications from the results of the study are that teacher educators in second and foreign-language contexts should more fully acknowledge the importance of readability among textbook selection criteria when choosing pre-service language teaching course books. This consideration is particularly vital given the sizable bilingual audience of consumers of their textbooks. Teacher educators might even consider incorporating a Coh-Metrix analysis into their textbook selection procedures. As well, they should become more aware of the potential need to provide learners with access to supplementary materials (e.g., helpful websites, YouTube videos or other supporting texts) that they can use to scaffold their learning. These supplementary resources could also be simplified or written in the L1 to better build learners' conceptual understanding before they begin to focus on the challenging language.

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.

Armbruster, B. B. (1985). Readability formulas may be dangerous to your textbooks. *Educational Leadership, 42*(7), 18-20.

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.

Carrell, P. L. (1987). Readability in ESL. *Reading in a Foreign Language, 4*(1), 21-40.

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.

Choi, Seonghee, Kim, Kyong-Hahn, Lee, Yong-Bae, Hong, Ju-Hee, & Cho, Eunkyung. (2012). Developing the Korean EFL Readability Formula (KRF). *English Language & Literature Teaching, 18*(3), 1-24.

Choi, Yunjeong. (2010). Effects of modified and non-mod-

ified text on Korean EFL learners' reading comprehension. *Applied Linguistics, 26*(4), 21-47.

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), 84-101.

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.

Drury, A. (1985). Evaluating readability. *IEEE Transactions on Professional Communication, 28*(4), 11-14.

DuBay, W. H. (2004). *The principles of readability.* Costa Mesa, CA: Impact Information.

Efklides, A., Papadaki, M., Papantoniou, G., & Kiosseoglou, G. (1998). Individual differences in feelings of difficulty: The case of school mathematics. *European Journal of Psychology of Education, 13*(2), 207-226.

Ferris, D. (2006). Does error feedback help student writers? New evidence on the short-and long-term effects of written error correction. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 81-104). New York: Cambridge University Press.

Finn, S. (1985). Unpredictability as correlate of reader enjoyment of news articles. *Journalism Quarterly, 62*(2), 334-345.

Fulcher, G. (1997). Text difficulty and accessibility: Reading formulae and expert judgement. *System, 25*(4), 497-513.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher, 40*(5), 223-234.

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.

Heydari, P., & Riazi, A. M. (2012). Readability of texts: Human evaluation versus computer index. *Mediterranean Journal of Social Sciences, 3*(1), 177-190.

Hoskin, R. (2012, March 3). The dangers of self-report. Retrieved from http://www.sciencebrainwaves.com/the-dangers-of-self-report/

Jackson, G. T., Allen, L. K., & McNamara, D. S. (2016). Common core TERA: Text ease and readability assessor. In S. A. Crossley, & D. S McNamara. (Eds.)., *Adaptive educational technologies for literacy instruction* (pp. 49-68). New York: Taylor & Francis.

Johns, A. M. (1997). *Text, role and context: Developing academic literacies*. Cambridge: Cambridge University Press.

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.

Kanungo, T., & Orr, D. (2009). Predicting the readability of short web summaries. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining* (pp. 202-211). Barcelona: ACM.

Kasule, D. (2011). Textbook readability and ESL learners. *Reading & Writing–Journal of the Reading Association of South Africa, 2*(1), 63-76.

Kim, Youngsu, & Ma, Jee Hyun. (2012). Text readability of the College Scholastic Ability Test and High School English II textbooks. *English Teaching, 67*(4), 195-218.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge: Cambridge University Press.

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., Graesser, A. C., Cai, Z., & Kulikowich, J. M. (2011). Coh-Metrix easability components: Aligning text difficulty with theories of text comprehension. Presented at the *American Educational Research Association Conference,* New Orleans, LA.

McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes, 47*(4), 292-330.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation, 51*, 297-384.

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.

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

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.

Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*(4), 357-383.

Pitler, E., & Nenkova, A. (2008). Revisiting readability: A unified framework for predicting text quality. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 186-195). Honolulu, Association for Computational Linguistics.

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.

Prince, S. A., Adamo, K. B., Hamel, M. E., Hardt, J., Gorber, S. C., & Tremblay, M. (2008). A comparison of direct versus self-report measures for assessing physical activity in adults: A systematic review. *The International Journal of Behavioral Nutrition and Physical Activity, 5*(1), 56.

Razavi, T. (2001). Self-report measures: An overview of concerns and limitations of questionnaire use in occupational stress research. In *Discussion Papers in Accounting and Management Science.* Southampton, UK: University of Southampton.

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.

Rello, L., Pielot, M., Marcos, M.-C., & Carlini, R. (2013). Size matters (spacing not): 18 points for a dyslexic-friendly Wikipedia. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility* (pp.17-24). Rio de Janeiro, BR: ACM.

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.

Schank, R. C. & Ableson, R. P. (1995). Knowledge and memory: The real story. In. R. S. Wyer (Ed.), *Knowledge and memory: The real story* (pp. 1-85). Hillsdale, NJ: Lawrence Erlbaum Associates.

Schneider, D. E. (2011). Assessing the readability of college textbooks in public speaking: Attending to entry level instruction. *Communication Teacher, 25*(4), 246-255.

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: Original article. *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.

Tinkler, S., & Woods, J. (2013). The readability of principles of macroeconomics textbooks. *The Journal of Economic Education, 44*(2), 178-191.

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.

Williamson, G. L. (2008). A text readability continuum for postsecondary readiness. *Journal of Advanced Academics, 19*(4), 602-632.

Wray, D., & Janan, D. (2013). Readability revisited? The implications of text complexity. *The Curriculum Journal, 24*(4), 553-562.

Yan, X., Song, D., & Li, X. (2006). Concept-based document readability in domain specific information retrieval. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management* (pp. 540-549). Arlington, VA: ACM.

Zamanian, M., & Heydari, P. (2012). Readability of texts: State of the art. *Theory and Practice in Language Studies, 2*(1), 43-53.