Readability of the STAAR Test is Still Misaligned

Susan Szabo, EdD
Professor
Texas A&M University-Commerce
Commerce, TX

Becky Barton Sinclair, PhD
Associate Professor
Texas A&M University-Commerce
Commerce, TX

Abstract

This study examined the readability of the State of Texas Assessments of Academic Readiness (STAAR) Reading passages for 2018 for grades 3-8. These results were then compared to the authors' first study on the topic, which found that the readability of the STAAR Reading passages were one-to-three years higher than the grade level for which they were written to assess (Szabo & Sinclair, 2012). This study found that some characteristics of the STAAR test had changed since 2012, but many of the reading passages were still misaligned for the targeted grade level.

The term "accountability" has a plethora of meanings. However, in the context of public education the term has come to represent ideas that were first associated with the *No Child Left Behind Act* of 2001 (NCLB, 2002) and later with the College and Career Readiness Standards (Conley, 2010; 2012).These legislative policies have focused on student achievement as a way to produce better student learning. Additionally, the *Every Student Succeeds Act* (U.S. Department of Education, 2015) limited the federal role in defining state accountability (Greene & McShane, 2018).Texas uses the Texas Essential Knowledge Skills (TEKS) to direct student learning and the State of Texas Assessments of Academic Readiness (STAAR) test to determine if students are learning the intended curriculum.

Purpose of the Study

In this age of accountability, assessments have gained a negative reputation. However, criterion-referenced assessments are a valuable tool to help drive instruction and to help students to be successful learners (Sindelar, 2015). Assessments can be powerful in helping teachers plan instruction and in letting the students gauge if their learning is up to the state standards. As the STAAR test is a criterion-referenced test based on the TEKS, it is important to investigate why the passing rates are not higher.

In 2012, it was found that the readability of the STAAR Reading (grades 3-8) passages were written one-to-three grade levels above the grade for which it was intended and that the

questions were written at a higher level as they were either think-and-search questions or on-your-own questions (Szabo & Sinclair, 2012). However, this current study only focused on the quantitative dimensions of text complexity as determined by using various readability formulas. The following question guided our study: How has the readability of the STAAR Reading passages and total overall readability average changed over the past seven years?

Readability and Readability Formulas

Readability is the ease with which a text can be read and understood (Gunning, 1952). Readability determines if any given written text is written clearly and at a comprehensible level. There are both pros and cons toward the use of readability formulas. Both viewpoints have research to support their beliefs (Zmanian & Heydari, 2012).

During the last century, many researchers (e.g. Dale & Chall, 1949; Flesch, 1948; Fry, 1968; McLaughlin, 1969) have addressed the issue of readability and how to calculate it as a way to make classrooms, the work place, and public communications more effective. The purpose of readability formulas is to determine the difficulty of the text so that the reader can determine if the reading material can be read without frustration (Begeny & Greene, 2014). This information helps authors convey complex ideas more clearly and more effectively towards their targeted audience. It also gives the reader advanced knowledge about the text, which may help in determining which book to check-out or to purchase (Zamanian & Heydari, 2012).

However, readability formulas cannot tell if the target audience will understand the text, as they do not measure the context, the reader's prior knowledge, or interest level in the topic or the cohesiveness of the text (Bailin & Grafstein, 2001; Bertram & Newman, 1981; Kirkwood & Wolfe, 1980; Zamanian & Heydari, 2012). Additionally, it was found that tinkering with the text to produce acceptable readability levels may make the text more difficult to understand (Rezaei, 2000).

Nevertheless, today, various readability formulas are commonly used to determine the readability of government documents, educational materials for students, newspapers, magazines and popular literature (Begeny & Greene, 2014). Readability formulas are mathematical in nature and focus on different text features. These features include the number of words in a sentence, the percentage of high frequency words on predetermined grade level word lists, the number of multisyllabic or "hard" words, and/or the number of letters in the text (Bailin & Grafstein, 2001; Begeny & Greene, 2014). For this reason, several formulas should be used and averaged when determining the readability of a selection of text, to account for the differences in formula design (Szabo & Sinclair, 2012).

Methodology

Procedure

In 2012, the researchers investigated the readability level of the STAAR Reading tests for grades 3-8 as well as the types of questions asked (Szabo & Sinclair, 2012). Another readability study on the STAAR Reading passages was done by Lopez and Pilgrim (2016) who found similar results. And in 2015, *HB 743* required the Texas Education Agency (TEA) to modify the STAAR Reading passages reducing, the number of passages that had to be read and the number of questions

that were answered (Huberty, 2015). These changes had third grade students reading four passages and answering 34 questions. Additionally, 2 questions were added to each grade level so that eighth grade students had 44 questions to answer about their reading passages. All students from 4-8 grades read six passages. Because of the previous study and the changes by TEA, this study only focused on the 2018 STAAR Reading passages to investigate what, if any, changes occurred in the readability of the assessment passages.

First, all of the released STAAR Reading passages for grades 3-8 were downloaded from the Texas Education Agency website (2007-2019). The pdf documents were converted into word documents. All photos, graphics, directions and item questions were removed. Line by line editing was done to ensure that line numbers and page numbers did not appear and that there was consistent spacing between all words. Additionally, at each grade level of the STAAR Reading tests, one poem for students to read and interpret was included. However, these poetry passages were not included in this readability study, as the variations in format prevent accurate readability calculations (Fry, 1977).

The reading passages were entered into three different online readability index calculators that linked the readability levels to grade levels. The first calculator used both the Fry and the Raygor readability formulas (ReadabilityFormulas, 2019a). The second calculator used the Flesch-Kincaid, the Coleman-Liau Index, the SMOG Index, the Automated Readability Index (ARI), and the Linsear Write (ReadabilityFormulas, 2019b). The third calculator used the Dale-Chall Formula (ReadabilityFormulas, 2019c). Thus, eight readability formulas were used in the calculation of readability. This approach was necessary, as more than one readability formula should be used to provide a more accurate grade-level indication (Szabo & Sinclair, 2012).

After determining the readability of each of the grade-level reading passage, the results were added and averaged. These scores were then compared to the Reading Consensus Scores (RCS; ReadabilityFormulas, 2019b). It was found that the results were similar. Following are brief descriptions of the readability formulas used in this study.

Flesh-Kincaid. Rudolph Flesch's readability research (1948) made him an early authority in the field and he inspired additional readability formula variations and applications (Kincaid, Fishburne, Rogers, & Chissom, 1975). The formula looks at the number of words and sentence length per 100 words to determine a grade-level reading score. This readability formula was first used by the Department of Defense to determine the difficulty level of technical manuals and today is a standard function on all Microsoft Word products (Zamanian & Heydari, 2012).

Coleman-Liau Index. This formula examines the number of letters in a word and the number of sentences per 300 words in the text. It was created for the U.S. Department of Education to calculate the readability of textbooks for schools (Coleman & Liau, 1975).

SMOG Index. The SMOG, which was created by McLaughlin (1969), was first used to evaluate healthcare material. The formula counts every word with three or more syllables within 30 sentences. This formula is appropriate for fourth grade to college age readers.

The Automated Readability Index (ARI). This formula was created to use computers to calculate the readability of text. The formula uses ratio representing the number of letters per word and the number of words per sentence (Kincaid et al., 1975).

Linsear Write Formula. This formula was developed for the United States Air Force to calculate readability of technical manuals. The formula uses sentence length and the number of words with three of more syllables per 100 words of text (Brewer, 2018).

Fry. The Fry Graph Readability Formula was developed by Edward Fry (1968; 1977). It considers the number of sentences and syllables per 100 words and the mathematical results are then plotted on a graph that was linked to grade levels. He noted that the readability calculator should only be used for prose and not poetry, as there were not enough words in poems.

Raygor. The Raygor Estimate Graph was developed in 1977 by Alton Raygor and examines the number of words that have six or more letters and the number of sentences per 100 words. The results are plotted on a graph that was created to link the mathematic results to a grade level (Baldwin & Kaufman, 1979).

Dale-Chall Formula. The Dale-Chall Formula was unique from other formulas in that it applied a count of "hard words" into its calculations, rather than syllable or letter counts. In 1995, the vocabulary list was expanded and the formula reevaluated (Chall & Dale, 1995; Zamanian & Heydari, 2012).

Readability Consensus. This is an automated feature created by the second free readability calculator (ReadabilityFormulas, 2019b). This formula calculates the number of sentences, words, syllables and characters in text samples and "plugs them into seven popular readability formulas" to provide their own average grade level. This data was not used in this study's average grade calculations, however; the scores are reported in Table 1 for comparison purposes.

Results

The eight different formula calculator results were tabulated and entered into a spreadsheet. Table 1 includes the grade level and passage number (column 1) for each of the 2018 STAAR Reading tests (grades 3-8). Columns 2-9 show the results of each readability formula on each passage for all grade levels. Column 10 shows the average of the readability scores. In column 11 is the Reading Consensus Score (RCS) which was used to check the results. As seen, when comparing the grade level average for each passage from column 10, the RCS average is similar. The final column is the average of all grade level passages to determine the averaged readability of each grade level STAAR Reading test.

Table 1

Readability for Released Grade-Level Passages for 2018 STAAR Reading Tests, Grades 3-8

Grade. Passage	Flesh Kincaid Grade Level	Coleman Liau Grade Level	SMOG Grade Level	ARI Grade Level	Linsear Write Grade Level	Fry Grade Level	Raygor Grade Level	Dale-Chall Grade	Ave for each Passage	RCS	Grade Ave
3.1	2.5	6	3.4	1.6	3.6	3	4	7	3.9	4	
3.2	5.3	7	5.7	4.9	6.1	6	4	5	5.5	6	
3.3									Poem		5.4
3.4	6.2	9	5.5	6.1	6	8	4	9	6.7	7	
4.1	5.9	9	6	5.5	5.9	7	4	8	6.4	7	
4.2	2.7	4	3.7	1	3.9	3	3	6	3.4	3	
4.3	5.5	9	5.9	6.5	6.8	7	6	7	6.7	7	
4.4	6.8	10	7.4	7.2	7.2	9	4	7	7.3	8	5.9
4.5									Poem		
4.6	5.2	8	6.1	4.8	5.6	7	4	6	5.8	6	
5.1	2.9	6	4.2	2	4	3	3	8	4.1	4	
5.2	6.3	7	6	5.9	7.8	7	6	8	6.8	7	
5.3	8.4	9	8	8.7	10.8	9	6	8	8.5	9	
5.4	6.1	8	6.3	5.7	6.7	7	6	7	6.6	7	6.5
5.5									Poem		
5.6	5.8	7	6.3	5.8	7.6	7	6	8	6.7	7	
6.1	5.8	7	6.5	4.9	6.7	7	6	6	6.2	6	
6.2									Poem		
6.3	8.4	11	8	8.4	7.8	10	6	10	8.7	9	
6.4	7.8	11	8.2	7.8	7.5	10	5	9	8.3	9	7.7
6.5	7.1	8	6	6.9	8	7	5	7	6.9	7	
6.6	8.8	10	8.2	9.1	9.5	9	6	8	8.6	9	
7.1	6.1	8	6.6	4.6	5.9	7	6	8	6.5	6	
7.2	9.2	10	8.8	9.4	10.8	11	6	9	9.3	10	
7.3	10	11	9.2	10	11	12	7	10	10.0	10	
7.4	10	10	7.8	10.6	11.4	11	7	11	9.9	10	8.4
7.5									Poem		
7.6	5.9	6	5.9	5.2	7.7	7	6	7	6.3	6	
8.1	6.4	7	5.4	6.5	7.7	7	7	7	6.8	7	
8.2	10.7	12	9.4	10.9	11.4	14	7	10	10.7	11	
8.3	5.3	7	5.6	4.8	6.4	7	6	8	6.3	6	
8.4									Poem		7.7
8.5	10.2	10	9	10.9	12.9	10	7	9	9.9	11	
8.6	4.1	7	5.3	3.2	4.6	4	4	8	5.0	5	

Discussion

This study examined the readability of the 2018 STAAR Reading tests for grades 3-8. First, readability scores on each passage at all grade levels were completed using 8 different readability scales. Next, an average for each reading passage was determined. Third, this average score was then used to determine the average of the overall readability of the STAAR Reading test at each grade level. A variety of readability formulas were used, as each looks at different text features and thus using more than one readability formula is more accurate (Szabo & Sinclair, 2012). Additionally, the passage that was the poem was not used in the calculations (Fry, 1977).

Third Grade

Looking at the individual passages for the 2018 STAAR test, third grade students were asked to read three passages, one poem, and had 34 questions to answer. It was assumed since the STAAR test was given in April, a 3.8-3.9 would be an appropriate readability level for third grade students at the end of the third grade year. When looking at the average readability for the three passages, it was found that two of the reading passages were misaligned with the grade level, as they were written above grade. One reading passage was written at an appropriate reading level. Thus, the average readability for each of the third grade individual reading passages ranged from 3.9-6.7. Finally, when calculating the total average for the three passages, the total average readability for the third grade STAAR Reading passages showed it was written at a 5.4 grade level.

Fourth Grade

Looking at the individual passages for the 2018 STAAR test, fourth grade students were asked to read five passages, one poem, and answer 36 questions. Again, it was assumed since the students were at the end of their fourth grade year, a 4.8-4.9 was an appropriate reading level. However, when looking at the average readability of each passage, all five passages were misaligned, as one passage was written below grade level while four passages were written above grade level. Thus, the average readability level of each of the fourth grade individual reading passages ranged from 3.4 to 7.3. Finally, when calculating the total average for the five passages, the total average readability for the fourth grade STAAR Reading passages was a 5.9 grade level.

Fifth Grade

Looking at the individual passages for the 2018 STAAR test, fifth grade students were asked to read five passages, one poem, and had 38 questions to answer. Again, it was assumed that 5.8-5.9 would be an appropriate readability level at the end of the fifth grade year. When looking at the average readability level of each passage, it was found that all five passages were misaligned, as one passage was written below grade level and four passages were written above grade level. Thus, the average readability for each of the fifth grade individual reading passages ranged from 4.1 to 8.5. Finally, when calculating the total average for the five passages, the total average readability for the fifth grade STAAR Reading passages showed it was written at a 6.5 grade level.

Sixth Grade

Looking at the individual passages for the 2018 STAAR, sixth grade students were asked to read five passages, one poem, and answer 40 questions. Again, it was assumed that a score of 6.8-6.9 would be an appropriate readability level at the end of the sixth grade year. When looking at the average readability level of each passage, it was found that four passages were misaligned, as one passage was written below grade level and three passages were written above grade level. Only one passage was written at the appropriate grade level. Thus, the average readability for each of the sixth grade individual reading passages ranged from 6.2 to 8.7. Finally, when calculating the total average for the five passages, the total average readability for the sixth grade STAAR Reading passages showed it was written at a 7.7 grade level.

Seventh Grade

Looking at the individual passages for the 2018 STAAR test, seventh grade students were asked to read five passages, one poem, and answer 42 questions. Again, it was assumed that 7.8-7.9 would be an appropriate readability level at the end of the seventh grade year. When looking at the average readability of each passage, it was found that all five passages were misaligned, as two passages were written below grade level and three passages were written above grade. Thus, the average readability for each of the seventh grade reading passages ranged from 6.3-10.0. Finally, when calculating the total average for the five passages, the total average readability for the seventh grade STAAR Reading passages showed it was written at an 8.4 grade level.

Eighth Grade

Looking at the individual passages for the 2018 STAAR test, eighth grade students were asked to read five passages, one poem, and had 44 questions to answer. Again, it was assumed that 8.8-8.9 would be an appropriate readability level at the end of the eighth grade year. When looking at the average readability of each passage, it was found that all five passages were misaligned, as three passages were written below grade level and two passages were written above grade. Thus, the average readability for each of the eighth grade individual reading passages ranged from 5.0 to 10.7. Finally, when calculating the total average readability for the five passages, the total readability for the eighth grade STAAR Reading passages showed it was written at a 7.7, almost a year below expected grade level.

Comparison to the 2012 Study

To answer our research question, the results of the two studies had to be compared (see Table 2). It was found that for third grade and eighth grade, the results of the readability calculations were the same. For third grade, the total average readability of the third grade reading passages were still written approximately two grade levels above grade level. Also, the results for eighth grade showed that the eighth grade reading passages are still written one year below grade level.

Table 2

Comparison of Readability Scores from 2012 and 2018 STAAR Reading Tests

	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
2012 study	5th	7th	7th	8th	8th	7th
This study	5th	6th	7th	8th	8th	8th

When comparing the 2012 study to this study, all grade levels had readability scores that were found to be similar. However, the total readability scores for Grade 4 decreased one grade level while the total readability scores for Grade 8 increased one grade level. The total readability scores for Grade 3, Grade 5, and Grade 6 were still two years higher and Grade 7 was one year higher than grade level.

Conclusions

There was some improvement in the readability of the reading passages from 2012 to 2018. However, when looking at the average of the individual reading passages and the total average of the passages, the readability is still too high in most cases. Thus, it is believed that many students may be failing the STAAR test because the passages are written above their grade level.

Third grade students have it the roughest, as the average for the three passages were still written almost two grade levels above the majority of the students' reading abilities, which puts the passages at their frustrational reading level. Although the fourth-seventh grade passage readability scores had gone down a little, the students are still being asked to read above grade level reading passages, even though these passages may be more at their instructional reading level.

Research has shown that text written at both the instructional and frustrational level of readers are too difficult to read independently and inversely impacts learning (Allington, 2006; Rasinski, 1999). When students are forced to read at their instructional and frustrational reading level by themselves, in most cases they will not be successful (Chall, 1983; Treptow, Burns, & McComas, 2007; Walker, 2011). Thus, our students are forced to read and are required to pass a test that has not been written on their grade level. This practice of using above grade-level text can harm a student's self-esteem and confidence in their reading abilities (Hisken, 2011).

Additionally, eighth grade students are given a false sense of accomplishment, as the total average of the five passages is still written below grade level. This could be frustrating for students who pass the 8th grade STAAR, yet struggle with grade-level reading in high school.

Failing high-stakes tests, such as the STAAR, affects students, teachers, and districts in many ways, including the costs of remediation and tutoring programs and materials. The label of "failure" hurts the self-esteem and morale of students and teachers when they are doing their best to cover and learn the material that needs to be taught at each grade level. As, the reading passages for third - seventh grade are written above grade level and for the eighth grade the reading passages are written below grade level, it is implied that the STAAR results may not accurately reflect student mastery of the TEKS. Thus, using the STAAR results as a high-stakes test is not appropriate.

Implications

Accountability and testing are important for schools as they provide important information about learning and how to maximize student success. Therefore, it is essential that high-stakes tests are written at the correct grade level and that teachers use a wide variety of reading passage levels to help students achieve their full reading potential. The results have implications for test development as well as for schools and classrooms.

For Test Development

First, it is suggested that more than one readability formula be used to get a more accurate reading level for passages considered for STAAR test use (Szabo & Sinclair, 2012). Second, the readability of each passage should be identified on each of the reading passages being considered for new STAAR test items to allow the STAAR Teacher Review Committee to be fully informed about the potential test passages. Third, the STAAR is a criterion referenced test. Criterion referenced tests are designed to indicate which skills students have not learned (Haertel, 1985). As the STAAR measures student's learning as determined by their grade level TEKS, the readability levels of the assessment passages should be at the correct grade level.

For the Schools and Classrooms

Students need a plethora of experiences so no matter what type of text passage is on the STAAR, the student has some familiarity with the text content. This can be done "through reading a variety of texts, displaying artwork that illustrates various topics, examining real documents (i.e. Constitution or Bill of Rights), showing educational videos (i.e. History Channel), and taking virtual tours (White House or Smithsonian)" (Szabo & Sinclair, 2012, p. 12).

Students need to be taught how to read hard passages. The passages should be read at least three times, paragraph by paragraph. The first time they read a text is to allow the reader to concentrate on sounding out unknown words. The second time the same text is read, the reader is working on gaining better fluency. Finally, the third time the same text is read, the reader is focusing on comprehension. Students need to learn to use this 3-step reading approach while working on their own whether they are reading a hard textbook or taking STAAR test.

All students should receive instruction to maximize their academic potential, not just the students that cannot pass the test (Lohman & Korb, 2006; Vygotsky, 1978). Thus, it is important that teachers differentiate teaching effectively to provide both remediated and accelerated class experiences. However, accelerated classes does not mean that sixth graders are taught eighth grade material but instead students are given time to delve deeper into the curriculum being taught at their grade level. They are encouraged to use both creative and critical thinking skills to analyze and use information to create projects that show their understanding of the TEKS and how they link various information together (Marzano & Helfebower, 2011).

Finally, guided reading groups need to be used at all grade levels 1-8. This differentiation allows all students to learn yearly and builds the skills students need at their developmental level. Without these skills, students will either fall further behind or not grow at all even if they are average or gifted students (Fountas & Pinnell, 2008; Lohman & Korb, 2006; Szabo & Sinclair, 2012; Tomlinson, Brimijoin, & Narvaez, 2008; Vygotsky, 1978).

References

Allington, R. (2006). *What really matters for struggling readers*? Boston, MA: Pearson/Allyn & Bacon.

Bailin, A., & Grafstein, A. (2001). The linguistic assumptions underlying readability formula: A critique. *Language & Communication*, *21*(1), 285-301.

Baldwin, S., & Kaufman, R. (1979). A concurrent validity study of the Raygor Readability Estimate. *Journal of Reading*, *23*(2), 148-153.

Begeny, J., & Greene, D. (2014). Can readability formulas be used to successfully gauge difficulty of reading materials? *Psychology in the Schools*, *51*(2), 198-215.

Bertram, B., & Newman, S. (1981). *Why readability formulas fail* (Report No. 28). Illinois University, Urbana: Center for the Study of Reading. (Eric Document Service No. ED205915)

Brewer, J. (2018). Measuring text readability using reading level. In M. Khosrow-Pour (Ed.), *Encyclopedia of information science and technology* (4th ed.; pp. 1499-1507). Hershey, PA: IGI Global.

Chall, J., & Dale, E. (1995). *Readability revisited: The new Dale-Chall Readability Formula*. Brookline, MA: Brookline Books.

Chall, J. (1983). *Stages of reading development*. New York, NY: McGraw-Hill Book.

Coleman, M., & Liau, T. (1975). A computer readability formula designed for machine scoring. *Journal of Applied Psychology*, *60*(2), 283-284.

Conley, D. (2012). *A complete definition of college and career readiness*. Eugene, OR: Educational Policy Improvement Center.

Conley, D. T. (2010). *College and career ready: Helping all students succeed beyond high school*. San Francisco, CA: Jossey-Bass.

Dale, E., & Chall, J. (1949). The concept of readability. *Elementary English*, *26*(1), 23.

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, *32*(1), 221-233.

Fountas, I. C., & Pinnell, G. (2008). *When readers struggle: Teaching that works*. Portsmouth, NH: Heinemann.

Fry, E. (1968). A readability formula that saves time. *Journal of Reading*, *11*(1), 513-516.

Fry, E. (1977). Fry's Readability Graph: Clarifications, validity and extension to level 17. *Journal of Reading*, *21*(3), 242-252.

Greene, J., & McShane, M. (2018). *Failure up close: What happens, why it happens, and what can we learn from it*. London, UK: Rowman & Littlefield Publishing Group.

Gunning, R. (1952). *The technique of clear writing*. New York, NY: McGraw-Hill.

Haertel, E. (1985). Construct validity and criterion-referenced testing. *Review of Educational Research*, *55*(1), 23-46.

Hisken, L. (2011). *The correlation between self-esteem and student reading ability, reading level, and academic achievement*. Retrieved from https://jckldigital.ucmo.edu/bitstream/handle/10768/103/Hisken201110_RP_Correlation.pdf?sequence=7&isAllowed=y

Huberty, D. (2015). *House Bill 743*. Retrieved from https://capitol.texas.gov/tlodocs/84R/billtext/html/HB00743I.htm

Kincaid, P., Fishburne, R., Jr., Rogers, R., & Chissom, B. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for navy enlisted personnel*. Orlando, FL: University of Central Florida, Institute for Simulation and Training. (Paper 56c)

Kirkwood, K., & Wolfe, R. (1980). *Matching students and reading materials: A cloze-procedure method for assessing the reading ability of students and the readability of textual materials*. Toronto, Canada: Ontario Department of Education. (ERIC Document Reproduction Service No. ED 195 928)

Lohman, D., & Korb, K. (2006). Gifted today but not tomorrow: Longitudinal changes in ability and achievement during elementary school. *Journal for the Education of the Gifted*, *29* (4), 451-484.

Lopez, M., & Pilgrim, J. (2016). Text complexity: A study of STAAR readability. In E. Martinez & J. Pilgrim (Eds.), *Literacy summit yearbook* (pp. 87-93). Belton, TX: Texas Association for Literacy Education.

Marzano, R., & Heflebower, T. (2011). *Teaching & assessing 21st century skills*. Bloomington, TN: Marzano Research.

McLaughlin, G. (1969). SMOG grading: A new readability formula. *Journal of Reading*, *12*(8), 639-646.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Rasinski, T. (1999). Exploring a method for estimating independent, instructional and frustrational reading rates. *Reading Psychology*, *20*(1), 61-69.

Raygor, A. (1977). The Raygor Readability Estimate: A quick and easy way to determine Difficulty. In P. Pearson (Ed.), *Reading: Theory rap* (pp. 259-263). Clemson, SC: National Reading Conference.

ReadabilityFormulas. (2019a). *Free readability tests using the Fry Graph and the Raygor Estimated Graph*. Retrieved from http://www.readabilityformulas.com/free-fry-graph-test.php

ReadabilityFormulas. (2019b). *Automatic readability checker: Our free text readability consensus calculator*. Retrieved from http://www.readabilityformulas.com/free-readability-formula-tests.php

ReadabilityFormulas. (2019c). *Free Dale-Chall Readability Formula with word list: Original and revised versions*. Retrieved from http://www.readabilityformulas.com/free-dale-chall-test.php

Rezaei, A. (2000). The validity of the Fog-Index of Readability. *Journal of Humanities of Islamic Republic of Iran*, *7*(4), 17-27.

Sindelar, N. (2015). *Teaching and testing: Steps for using test data as a tool for teaching and Learning*. Denton, WA: Apperson. Retrieved from https://www.apperson.com/wp-content/uploads/2015/03/Teaching-Testing-eBook.pdf

Szabo, S., & Sinclair, B. (2012). STAAR reading passages: The readability is too high. *Schooling*, *3*(1), 1-14.

Texas Education Agency. (2007-2019). *STAAR released test questions*. Retrieved from https://tea.texas.gov/student.assessment/staar_released_test_questions/

Tomlinson, C., Brimijoin, K., & Narvaez, L. (2008). *The differentiated school: Making revolution changes in teaching and learning*. Alexandria, VA: Association for Supervision and Curriculum Development.

Treptow, M., Burns, M., & McComas J. (2007). Reading at the frustration, instructional, and independent levels: The effects on students' reading comprehension and time on task. *School Psychology Review*, *36*(1), 159-166.

U.S. Department of Education. (2015). Every Student Succeeds Act (ESSA). Retrieved from https://www.ed.gov/essa

Vygotsky, L. (1978). Interaction between learning and development. In M. Gauvain & M. Cole (Eds.), *Reading on the development of children* (pp. 34-40). New York, NY: Scientific American Books.

Walker, B. (2011). *Diagnostic teaching of reading*. New York, NY: Pearson.

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