Using Qualitative Measurement to Determine Text Difficulty: Relating Experts' Judgment of Reading Text Difficulty and Students' Comprehension Test Results in Grade Four Amharic Reading passages

Mulusew Asrate[1]

Received: 17 September 2022; Accepted; 21 November 2024

Abstract: An objective measurement of text complexity is an important tool in the preparation and selection of texts in teaching reading. This study aims to investigate the extent to which qualitative dimensions of text complexity predict the difficulty level of reading Amharic text. Reading passages included in the fourth-grade Amharic textbook are measured qualitatively by informed experts' judgments, and the result is compared with the comprehension test results of fourth-grade students in two Addis Ababa schools. The computation of the correlation between the two results yields a statistically significant positive relationship in three of the four qualitative dimensions as well as the aggregate result suggesting that qualitative measures can be used as one source of information to determine reading text difficulty in Amharic and other Ethiopian languages which exhibit similar properties.

Keywords: *Text complexity/difficulty, qualitative features of text complexity, comprehension*

[1] Mulusew Asratie Wondem, Department of Amharic Language, Literature and Folklore, College of Humanities, Language studies Journalism and Communications, Addis Ababa University

Introduction

It is well known that reading is one the most important skills in the academic and daily lives of modern individuals. In academic life, it is unquestionable that students' achievement in all subjects is based on their ability to comprehend what they read. In real life, particularly in contemporary society, reading is becoming more and more attached to the daily life of individuals. Today reading is required not only to understand books and newspapers, but also for occasional, yet very important, activities in our daily life. An ordinary resident of Addis Ababa, for example, wakes up in the morning, picks up his/her mobile phone and rushes to one of the social media which he/she follows and reads many short and long texts posted/shared by his/her contacts. Outside the home, the attention of the reader is drawn by roadside banners and advertisements, bus identification numbers, the destination plates of taxis, street names and numbers and the like. In the workplace, he/she reads noticeboards, instructions, letters, manuals, tags, pamphlets, reports, newspapers, books etc. depending on the nature of his/her job. After office hours, he/she may visit some café/restaurant or may go out shopping where he/she has to read the names of the service providers written on top of their gates, menus, tags, manuals, instructions, expiry dates etc.

Reading is now becoming a pattern of modern life and it must be properly acquired as a skill as early as possible. Most governments of various countries nowadays support the creation and development of a reading public and they go as far as outlining policies and strategies in their education system that would enable their citizens to be readers in the languages they deem necessary. Similarly in the Ethiopian education system, reading is one of the major components both in the mother tongue and second and foreign language (English) subjects. Yet, the effectiveness of the education system in developing reading ability is questionable and students' reading performance seems to be declining from time to time. For example, the average results of reading achievement in the Ethiopian National Learning Assessment made in

2000, 2004 2008 and 2016) by NOE/ NEAEA respectively show 64.3, 64.5 and 43.9 and 47 indicating a drastic decline in the third year (NOE 2000, NOE 2004, NEAEA 2012, NEAEA 2016, NOE 2008, RTI 2010).

More recent findings on Early Grade Reading Assessment (EGRA) made by RTI in 2010 (for six major languages: Amharic, Afan Oromo, Tigrigna, Somali, Harari, Sidama), and in 2014 (for two languages: Haddiyyisa and Wolayttatto) also indicate that the problem is deep-rooted. For example, RTI (2010) reported that reading achievement was very low. *When asked to read a simple passage at a Grade 2 level, many regions had more than 30% of Grade 2 and 20% of Grade 3 students unable to do so successfully, with children in the Sidama zone and Oromiya region particularly struggling* (RTI 2010, p 8). RTI also reported that reading comprehension scores were extremely low, with more than 50% of the children in most regions unable to answer a single simple comprehension question.

Following these findings, the MOE in collaboration with partner organizations has implemented various interventions under the umbrella of Reading for Ethiopia's Achievement Developed (READ) project[2]. The intervention is thought to be a holistic approach to advancing student achievement in the early grades. It focuses on professional development for teachers, capacity-building at the Woreda level, curriculum design, inclusive education, higher education capacity for training, and information and communication technology (ICT). Whether the intervention will bring about the long-sought results is yet to be seen by the time of its completion. Yet, there is one important component that should have been part of the intervention through curriculum development; however, it has not been given due attention in the intervention. i.e., text complexity. Text complexity is the level of relative difficulty of a given text in reading and comprehending it. Choosing appropriate texts that suit students at different levels is very important in teaching reading. Reading texts should neither be too complex nor too

[2]https://www.edc.org/ethiopia-read-ii

simple for students. Too complex texts pose unnecessary challenges to students and this may lead to failure and demotivation. Similarly, too simple texts would not present the required challenges and are unable to prepare students for real-life reading.

For a long time, text complexity has been measured quantitatively by counting word length and frequency, as well as sentence length and text cohesion (Pearson 1974/74, Arya 2011, Islami 2014, Siddharthan 2014 among others). In recent times, text complexity has been considered as having qualitative and reader-based dimensions in addition to the quantitative measures (Shanahan 2013, Pearson and Hiebert 2014a, Pearson and Hiebert 2014b, Hiebert 2014, Toyama et. al. 2017, Smith 2022, Mesmer 2021). [3] The qualitative dimension, just like the quantitative, is related to the inherent nature of the texts. It refers to issues such as levels of meaning or purpose; text structure; language conventionality and clarity; and knowledge demands. The reader-based dimension, on the other hand, includes variables specific to particular readers (such as motivation, knowledge, and experiences) and particular tasks (such as purpose and the complexity of the task assigned and the questions posed).

Setting up appropriate text complexity for students in different levels based on the above dimensions is one of the crucial tasks in teaching reading in developed countries. For example, in the US education system, Common Core State Standards for English Language Arts and Literacy are prepared by considering the three dimensions mentioned above. Determining text complexity by considering both qualitative and quantitative dimensions has been given due attention in language teaching research in the last decade. While most of the research

[3] These days the quantitative measurements are being made by the assistance of technology. There are quite a number of technologically supported programs to measure the readability/complexity of texts. As the focus of this paper is on the qualitative measurement, I will not discuss them here. For concise review of such measurements, I refer the reader to National Governors Association Center for Best Practices (NGACBP) and Council of Chief State School Officers (CCSSO) (2010). Common Core Standards for English Language and Arts, Literacy in History/ Social Studies, Science and Technical Subjects. Retrieved from: http://www.corestandards. org/wp-content/uploads/ELA_Standards1.pdf.

literature argues in their favour (Shanahan 2013, Pearson and Hiebert 2014a, Pearson and Hiebert 2014b, Hiebert 2014, Toyama et. al. 2017, Smith 2022, Mesmer 2021), few studies reported none or inverse relationship between text complexity and reading (Amendum et.al. 2017).

In teaching Ethiopian languages, however, the selection of reading texts to a particular grade level has never been made based on objective measurements of text complexity. The only attempts at measuring reading text complexity are that of Getachew (1986) and Megabiw (2009) who tried to measure readability quantitatively. They are attempts to measure the readability levels of texts found in student textbooks by using readability formulas. In addition to their limited scope (that they measure only the quantitative dimension), the readability measures made in these works have theory internal problems. They are based on the mere counting of words and sentences, and they do not consider semantic aspects such as information density. Information density is the amount of information contained within a word or sentence. For example, in Amharic, the reciprocating perfective verb "*tänägaggärnɨbbät*', can be a word, or a sentence consisting of one word. In terms of its information density, it is equivalent to the English sentence "We talked to each other about it". Readability measurements based on word/sentence counting, like the works of Getachew and Megabiw, would count the Amharic word just as a single word or a sentence of one word. On the other hand, they consider the English equivalent as a sentence of six words leading to the conclusion that the Amharic text containing this word is less difficult than the English text containing the sentence with six words, despite both having equal information density.

Measuring the complexity of texts selected for teaching reading in terms of all three dimensions described above is beyond the scope of small-scale research like this. This study is intended to test if the levelling of texts based on the qualitative dimension can serve as a valuable means to determine their difficulty in comprehending. The qualitative dimension of text complexity, as mentioned above, consists of four features: text structure, language conventionality and clarity, knowledge demand and levels of meaning (for literary texts) and purpose (for informational texts). The Appendix A of the US Common Core State Standards for English Language Arts and Literacy (2010, p 5) described each of these features as follows:

> *Text structure* refers to the organization or presentation of the texts. Texts of low complexity tend to have simple, well-marked, and conventional structures, whereas texts of high complexity tend to have complex, implicit, and (in literary texts) unconventional structures. Simple literary texts tend to relate events in chronological order, while complex literary texts make more frequent use of flashbacks, flash-forwards, multiple points of view and other manipulations of time and sequence. Simple informational texts are likely not to deviate from the conventions of common genres and sub-genres, while complex informational texts might be if they conform to the norms and conventions of a specific discipline or if they contain a variety of structures (as an academic textbook or a history book). Graphics tend to be simple and either unnecessary or merely supplementary to the meaning of texts of low complexity, whereas texts of high complexity tend to have similarly complex graphics that provide an independent source of information and are essential to understanding a text. It is essential to note that many books for the youngest students rely heavily on graphics to convey meaning and are an exception to the above generalization.

Language Conventionality and Clarity are related to the vocabulary and sentences of texts. Texts that rely on literal, clear, contemporary, and conversational language tend to be easier to read than texts that rely on figurative, ironic, ambiguous, purposefully misleading, archaic, or

otherwise unfamiliar language (such as general academic and domain-specific vocabulary).

Knowledge Demands refer to the assumption that the reader is expected to have concerning life experiences, culture and literature (literary texts), and subject-matter knowledge (informational texts) to understand the text. In this regard texts that make few assumptions about the extent of readers' life experiences and the depth of their cultural/literary and content/discipline knowledge are generally less complex than texts that make many assumptions in one or more of those areas.

Levels of Meaning (literary texts) or Purpose (informational texts). Literary texts with a single level of meaning tend to be easier to read than literary texts with multiple levels of meaning (such as satires, in which the author's literal message is intentionally at odds with his or her underlying message). Similarly, informational texts with an explicitly stated purpose are generally easier to comprehend than informational texts with an implicit, hidden, or obscure purpose.

This research attempts to investigate whether the complexity level of reading texts in terms of these qualitative features predicts their difficulty to comprehend. In other words, it tries to investigate the relationship between the qualitative measures of text complexity and the comprehension test results. The findings of this and the like research are of great importance to curriculum and textbook developers as well as writers and developers of supplementary reading materials who are often tempted to establish the exact level(s) that a particular text fits into.

Objectives

The general objective of this research is to investigate if the qualitative text complexity measures can be used as a reliable source of information in measuring text difficulty for Amharic and/or Ethiopian languages. Specifically, it tries to investigate the correlation between the qualitative text measures of fourth-grade Amharic reading passages and the results of students' comprehension tests.

Methods and Procedures

Tools

Informed experts' judgment: Informed judgment of experts is used to evaluate the qualitative complexity of texts. This tool is adapted from the Massachusetts Department of Elementary and Secondary Education *Quick Reference Guide: Text Complexity and the Growth of Reading Comprehension.* The tool is used to evaluate the qualitative complexity of texts based on the factors discussed in the section above, namely levels of meaning (literary texts) or purpose (informational texts); text structure including format and text features like illustration, graphics, and page layout; conventionality of vocabulary and sentence structure; as well as knowledge demands. The evaluation of text complexity based on these criteria is given five levels. The levels are assigned for grades as: Level 1(appropriate for grades 3 and 4), level 2 (appropriate for grades 5 and 6) level 3 (appropriate for grades 7 and 8), level 4 (appropriate for grades 9 and 10 and level 5 (appropriate for grades 11 and 12). The questionnaire is prepared in such a way that it describes each of the qualitative criteria in Amharic. The experts are requested to read the criteria first. Then they read the texts one by one; make their judgment in terms of the given criteria' and place them into one of the levels they think is appropriate. Finally, the data is collected by counting the number of experts who judge the texts as appropriate for 4^{th}-grade readers in terms of the respective qualitative criteria.

Various mechanisms were employed to avoid careless and biased judgments of experts and to increase the validity and reliability of the

judgments. Most of the participants did not know each other and they were not informed that the texts were from the fourth grade Amharic textbook. However, by the time the questionnaire was collected, one of the informants revealed that he knew where the texts were taken. Consequently, his questionnaire was excluded from the analysis. The experts were requested to explain their reasons immediately after they made their judgments (see the Appendices). To avoid unnecessary stress, they were given one month time to complete the tasks (approximately equivalent to one text per day). They were paid a compensation of 1000 Birr for the time they spent to fill the questionnaire.

Comprehension test: The comprehension test was also administered throughout the academic year for all grade four Amharic reading passages. The comprehension questions used in the test are directly taken from the textbook unless there is a problem with the question. The subject teachers reviewed each of the questions before administering the test to make sure that the questions necessarily required comprehending the reading text. If they could be answered with common sense, the subject teachers modified them or changed them. Administering the tests and correcting the test papers were all done by the subject teachers after proper explanations were given to them. Spontaneous follow-up was made to make sure that the testing went smoothly.

Sampling

Texts: All reading texts/passages in grade four Amharic (as a mother tongue) were selected for the informed experts' judgment and comprehension test. The reading texts contained in the textbook are of three types. 14 of them are narratives, 6 of them are expository and the other 6 are descriptive essays.

Experts for informed judgment are 26 experts selected from various qualifications and specializations. They include language teachers, literary experts, writers, linguists and journalists. They were considered experts in judging text difficulty because their training and profession are

directly related to texts. They were selected through the snowball method. 25 of them were BA holders while one of them was an MA holder. Their age range was between 25 and 39. 16 of the experts were men while 10 were women. All the informants live in Addis Ababa and are native speakers of Amharic.

Students and teachers: The comprehension test results were collected from sections of fourth-grade students in two schools in Addis Ababa (Biherawi Betemengist and Kiddist Selassie Cathedral primary schools). All the students speak Amharic as their first language. Fourth-grade students are selected because this is the grade level where teaching reading comprehension begins and focusing on the beginning helps to build the blocks from the bottom. In grades 1 through 3, the teaching of the language focuses more on letter and word recognition.

Two Amharic teachers who taught in these sections in the 2021/22 academic year were responsible for the administration and correction of the test.

Data Analysis

The data collected through the instruments mentioned above for each of the texts selected is systematized, organized and computed. The correlation between the qualitative measurements of the texts' complexity and that of the comprehension test results of students was computed by using SPSS version 25.

Findings

Analysis

The informed judgments of experts about the difficulties of the 26 reading passages of the grade four Amharic textbook and the average test results of the students' comprehension test are as follows.

Table 1: Summary of test scores and informed expert judgement

		Results of informed judgment of experts[4]				
Text ID	Average Comprehension test results (10%)	Language conventionality and clarity	Knowledge demands	Levels of Interpretation or purpose	Text structure	Over all judgment
1	8.94	10	8	15	15	40
2	6.47	5	5	5	13	28
3	6.16	10	9	9	14	42
4	7.16	11	5	8	13	37
5	7.39	7	3	3	9	22
6	7.34	1	2	4	6	13
7	6.62	12	10	11	17	50
8	6.21	3	2	3	5	13
9	6.74	2	2	3	4	11
10	5.62	5	2	0	2	11
11	3.40	2	1	2	6	7
12	4.17	2	4	3	3	15
13	5.41	3	1	3	3	10
14	3.61	1	1	1	1	6
15	3.68	1	0	0	1	2
16	5.67	3	1	3	5	8
17	6.15	5	2	3	2	15
18	6.22	0	2	2	6	6
19	6.21	6	4	0	2	16
20	5.63	3	0	0	4	5
21	6.20	3	2	4	6	13
22	5.54	4	4	5	2	19
23	5.58	3	1	0	12	6
24	4.22	9	9	7	2	37
25	2.50	2	2	1	5	7
26	6.09	5	3	5	1	18

[4]Note that the results are calculated by counting the number of experts who judge the texts as appropriate for 4th grade readers in terms of the respective qualitative criteria.

The Correlation between the informed expert judgement results for each of the qualitative[5] measurement criteria and the comprehension test is computed to assess their relationship as shown in Table 2 below. The result yields a statistically significant moderate positive correlation between the comprehension test results and the experts' judgements on three of the qualitative dimensions of text complexity suggesting that qualitative judgements can be used as predictors of text complicity, and a weak positive correlation between the comprehension test results and the experts' judgement on one of the qualitative dimensions.

[5]The use of the term qualitative along with the correlation might seem confusing for readers. Note that the qualitative measurements are quantified based on the experts' judgments. So the correlation is between quantified the experts' judgments and the comprehension test results.

Table 2: Correlations between the experts' judgements and comprehension test results

		test	lan-guage	knowledge	Interpretation	structure	overall
test	Pearson Cor-relation	1	.470*	.338	.548**	.522**	.477*
	Sig. (2-tailed)		.016	.091	.004	.006	.014
	N	26	26	26	26	26	26
language	Pearson Cor-relation	.470*	1	.850**	.768**	.622**	.930**
	Sig. (2-tailed)	.016		.000	.000	.001	.000
	N	26	26	26	26	26	26
knowledge	Pearson Cor-relation	.338	.850**	1	.836**	.583**	.956**
	Sig. (2-tailed)	.091	.000		.000	.002	.000
	N	26	26	26	26	26	26
interpreta-tion	Pearson Cor-relation	.548**	.768**	.836**	1	.663**	.888**
	Sig. (2-tailed)	.004	.000	.000		.000	.000
	N	26	26	26	26	26	26
structure	Pearson Cor-relation	.522**	.622**	.583**	.663**	1	.679**
	Sig. (2-tailed)	.006	.001	.002	.000		.000
	N	26	26	26	26	26	26
overall	Pearson Cor-relation	.477*	.930**	.956**	.888**	.679**	1
	Sig. (2-tailed)	.014	.000	.000	.000	.000	
	N	26	26	26	26	26	26

*. Correlation is significant at the 0.05 level (2-tailed).
**. Correlation is significant at the 0.01 level (2-tailed).

As can be seen from Table 2, the conventionality and clarity of the language of the text and the comprehension test yield a moderate relationship. Pearson's correlation coefficient between these two variables is moderate and positive, r= 0.470, N=26, and the relationship was statistically significant (sig =.016, a = 0.05), suggesting that the experts' judgment on conventionality and clarity of the reading texts can

be a significant predictor for the difficulty (complexity) of reading texts in Amharic.

Similarly, the levels of interpretation (for literary texts) or purpose of the texts (for information texts) and the comprehension test results also show a moderate relationship. The correlation coefficient computed between them shows a moderate positive correlation, *r*=.548, *N*=26, and the relationship is significant (sig =0.004, *a* = 0.01). The result also suggests that the expert judgments regarding the level of interpretation/purpose of the text are a significant predictor of text difficulty/complexity just like the conventionality and clarity of the language of the text.

Likewise, the relationship between the experts' judgment on the difficulty level of the structure of the text and comprehension test results exhibits the same record. It yields a moderate positive correlation coefficient, r=522, N=26, which is statistically significant (sig= 0.006, *a*=0.01). This means that judgment on the structure of the reading text can also serve as a predictor of text difficulty/complexity along with conventionality and clarity of language, and the level/purpose of interpretation.

On the other hand, the experts' judgment about background knowledge demand and the comprehension test show of weak correlation between them. The computation of the correlation coefficient shows a weak positive correlation, *r*= .338, *N*=26. The relationship is not significant (sig=*.091*). This means that the role of background knowledge to serve as a predictor of text complexity is not as strong as the three dimensions we saw above.

In addition to computing the correlation between the experts' judgement and the comprehension test results for each of the qualitative dimensions of the texts, the correlation between cumulative judgements of experts and the comprehension test results was also computed. The result is a moderate positive relationship, r=.477, N=26, which is statistically significant (sig= .014, *a*= 0.05). This means that the

aggregate judgment of experts on the qualitative dimensions of text complexity can serve as a source of information to decide on the level of text difficulty.

To sum up, the computed correlations between the students' comprehension test results and the four criteria of qualitative text difficulty measurements provide a moderate positive correlation for three criteria (language, text structure and levels of interpretation/purpose); and a weak positive correlation for background knowledge demand. The aggregate of all the criteria and the comprehension test generally results in a moderate positive relationship suggesting that expert judgments can be used as one source of information to decide the level of text difficulty.

Interpretation

Generally, the results reported indicate that three of the four qualitative measurement criteria of text difficulty have a moderately positive correlation with the results of the comprehension test. This means that the informed judgement of experts on the qualitative dimensions of texts can be one of the good predictors of the difficulty of reading texts in Amharic. In other words, the result is suggestive of the fact that the qualitative measure of reading texts is a valuable source of information in deciding the level of difficulty/complexity of the texts. In this regard, the findings corroborate with Shanahan (2013). Pearson and Hiebert (2014a), Pearson and Hiebert (2014b), Hiebert (2014b), Toyama et. al. (2017), Smith (2022), and Mesmer (2021) argue that qualitative aspects of text complexity play a role in indicating difficulty level of texts.

However, the finding differs from those studies concerning the role of background knowledge in predicting text difficulty. The reason why background knowledge yields a different result from the other criteria needs further study. From the data collected for this research, one possible idea that could be forwarded to explain this is the impact of technology. The experts' decisions, while considering the knowledge demand to understand reading texts, were mainly based on the

remoteness of the subject matter of the reading texts to the children. Most of the explanations forwarded by experts to explain why they consider a particular text to be less/more difficult in terms of background knowledge contain the idea that “the subject matter is strange/familiar to the daily life of students.”

Contrary to the experts’ assumptions, the recent boom in modern communication technologies such as television, internet, mobile phones etc. may have made information about any issue very accessible to people irrespective of their age. This means that those issues which the experts believe are remote to the children may not have been so remote to the children because they can access them through various media. The weak correlation between the experts’ judgement and the comprehension test could be the result of this mismatch.

On the other hand, the finding of this study contrasts with Amendum et.al. (2017) who reported that reading text difficulty level was inversely related to comprehension. The different results between Amnedum and this study call for further research. Two things that could be a source of difference are the nature of the comprehension questions used to assess the students' understanding of the texts and the age difference between the students may have played a role in this case. For example, Amendum’s subjects were elementary-aged children starting from kindergarten. This may affect comprehension as lower graders are still struggling to understand even the very simple text. Whatever the source of difference between these two studies is, the undeniable truth is that further studies are needed to provide imperial support for the relationship between qualitative text complexity and reading comprehension.

Conclusion

The issue of text difficulty/complexity has been the subject of debate and research in language teaching, particularly in the last decade. Yet, it has never been an issue in the teaching of Amharic and other Ethiopian

languages. On the other hand, as indicated in the introduction, the reading performance of students is decreasing from time to time. The findings in this study can serve as a good initial point and as a source of information concerning the selection of reading texts to improve reading performance in Amharic and other Ethiopian languages. The finding is also a good indication that a systematic/criteria-based selection of reading texts is possible and more helpful to the reading instruction than a random selection of reading texts.

References

መጋቢው ኃይለማሪያም።። (2009)። **በ**2004 **ዓ.ም በተዘጋጀው የ**11**ኛ ክፍል የአማርኛ ቋንቋ መማሪያ መጽሐፍ ውስጥ የቀረቡ ምንባቦች የተነባቢነት ደረጃ። የማስተርስ ቴሲስ፣ አዲስ አበባ ዩኒቨርሲቲ።።**

ጌታቸው አዱኛ::(1986):: “**የ**10**ኛ ክፍል የተመረጡ የአማርኛ ቋንቋ የማስተማሪያ ምንባቦች የብቃት ደረጃ ።።**” **የማስተርስ ቴስስ፣ አዲስ አበባ ዩኒቨርስቲ።።**

Amendum, S., K. Conradi. E. Hiebert (2022). Does Text Complexity Matter in the Elementary Grades? A research synthesis of text difficulty and elementary students’ reading fluency and comprehension. *Educ Psychol Rev*. 30:121–151.

ARYA, D. J. (2011). The effects of syntactic and lexical Complexity on the comprehension of elementary science texts. *International Electronic Journal of Elementary Education, 2011, 4(1), 107-125.*

National Governors Association Center for Best Practices (NGACBP) and Council of Chief State School Officers (CCSSO) (2010). Common Core Standards for English Language and Arts, Literacy in History/ Social Studies, Science and Technical Subjects. Retrieved from:

http://www.corestandards.org/wp-content/uploads/ELA_Standards1.pdf.

National Governors Association Center for Best Practices (NGACBP) and Council of Chief State School Officers (CCSSO) (2010). Appendix A of common core State standards for English language arts & literacy in history/social studies, science, and technical subjects. Massachusetts Department of Elementary and Secondary Education (n.d) Standard

Eslami, H. (2014). The effect of syntactic simplicity and complexity on the readability of the text. *Journal of Language Teaching and Research*, Vol. 5, No. 5, pp. 1185-1191.

Hiebert (2014). What is complex in text complexity? *IRA Essentials*. International Reading Association.

Mesmera, H., E.. Hiebert, J. Cunninghamc & M. Kapaniaa (2021). Does one size fit all? Exploring the contribution of text features, text content, and grade of use on comprehension. *Reading Psychology*, 42: 1, 42–72.

NOE (2000). Ethiopian Baseline National Learning Assessment of Grades 4 & 8 Students. Addis Ababa: NOE/MOE.

NOE (2004). Ethiopian Second National Learning Assessment of Grades 4 & 8 Students. Addis Ababa: NOE/MOE.

NEAEA (2012). Ethiopian 4th National Learning Assessment of Grades 4 and 8 pupils, Analytic Report: Addis Ababa.

NEAEA (2016). Ethiopian Fifth National Learning Assessment of Grades Four and Eight Students. Addis Ababa.

Pearson, D (1975). The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations Reading Research Quarterly, Vol. 10, No. 2, pp. 155-192

Pearson, D and E. Hiebert (2014a), The state of the field: Qualitative analyses of text complexity. *The Elementary School Journal*, 115: 2.

Pearson, D. and E. Hiebert (2014b). Understanding text complexity: Introduction to the special Issue. *The Elementary School Journal*, 115:2.

RTI (2010). Ethiopia Early Grade Reading Assessment: Data Analytic Report- Language and Early Learning. A Report Prepared for USAID/ Ethiopia.

RTI (2014). Ethiopia 2014 Early Grade Reading Assessment Report of Findings. A Report Prepared for USAI/Ethiopia

Shanahan, T (2013). *A Beginner's Guide to Text Complexity.* Sheena Hervey Generation Ready.

Siddhartha, A. (2014) A survey of research on text simplification. *International Journal of Applied Linguistics* 165:2 (2014), 259–298.

Smith C. K & E. Hiebert (2022). What does research say about the texts we use in elementary school? The complexity, type, format, and diversity of texts matter for young readers of all proficiency levels. *Kappan,* 103: 8

Toyamaa, Y., E. Hiebertb, C, & D. Pearsona (2017). An analysis of the text complexity of leveled Passages in four popular classroom reading assessments. Educational Assessment, 22: 3, 139–170.https://www.edc.org/ethiopia-read-ii