Instrument Development Procedures for Silent Reading Measures

Kimy Liu

Krystal Sundstrom-Hebert

Leanne R. Ketterlin-Geller

Gerald Tindal

University of Oregon

Behavioral Research and Teaching

Published by

Behavioral Research and Teaching
University of Oregon • 175 Education
5262 University of Oregon • OR, 97403-5262
Phone 541-346-3535 • Fax 541-346-5689
http://brt.uoregon.edu

Abstract

The purpose of this study was to develop and gather validity evidence for silent reading fluency

passages. A number of passages were written following a traditional story grammar structure

(character, setting, events) and placed on a computer for students to read silently. We describe in

detail, the manner in which content-related evidence was established and then present a number

of statistical analyses conducted to evaluate the technical adequacy of these measures. The

outcomes support the test development process and reflect a series of measures that have

potential for use in measuring elementary and middle school students' silent reading fluency.

Introduction

Although unintended, students' silent reading ability is frequently tested on many standardized reading comprehension tests. On these tests, students quietly read passages and respond to comprehension questions. Even though the behavior that is being evaluated is students' responses to the questions, students' ability to decode text is implicitly measured. As such, developing silent reading skills is an important component to successfully demonstrating reading comprehension skills in many settings.

Because of the importance of developing silent reading skills, we developed and pilot tested a series of Silent Reading Fluency (SRF) tests. These tests measured students' rate of reading text to themselves. The SRF measures were administered individually to students on computers. Students read four short paragraphs (SRF-Paragraphs) and six sentences (SRF-Sentences), one at a time, to themselves. When they finished reading each paragraph or sentence, they clicked the "next" button, and the following paragraph or sentence appeared. The speed with which they read the paragraphs and sentences was calculated by timing how long it took to finish each passage. Students received a silent reading fluency score when they finished each passage, which was stated as Words per Minute (W.P.M.). For example, if it took a student six seconds to read a 20-word grade-level sentence, then the speed of this student's silent reading fluency was 200 words per minutes, or 200 WPM. This strategy is unlike most other oral reading fluency measurement strategies where time is fixed and behavior free to vary.

The SRF measures were developed for and administered to students in grades 3 through 8. Students were given six grade-level sentences and four grade-level paragraphs on the computer. The four SRF-Paragraphs in each grade level followed one story line, while the six SRF-Sentences in each grade level followed another story line. Each story line had a main

character and followed a modified narrative plot structure. The SRF-Paragraphs ranged in length from 75-78 words and the SRF-Sentences ranged in length from 19-25 words. Each paragraph and sentence was written in the mid-range for its respective grade level according to the Flesch-Kincaid Readability Index. For example, a third-grade paragraph would have a Flesch-Kincaid Grade Level between 3.3 and 3.6, while an eighth-grade paragraph would have Flesch-Kincaid Grade Level between 8.3 and 8.6.

In the following sections, we describe the process and results of internal and external reviews of SRF Measures, as well as a pilot study for examining the technical adequacy of the measures. The purpose of the internal and external reviews was to gather content related evidence for validity. The purpose of pilot study was to determine whether the test items were functions appropriately and reliably differentiating high-performing students from low-performing students.

Methods

Setting and Participants

Silent Reading Fluency measures were first developed and then administered to students in grades 3-8 attending public schools in two mid-size towns of the Pacific Northwest. In all, 125 Grade three students, 98 Grade four students, 145 Grade five students, 97 Grade six students, 109 Grade seven students and 103 Grade eight students took the Silent Reading Measures.

Testing took place either in the schools' computer labs or in the teachers' classrooms if they had a mobile laptop lab. Two trained research assistants administered the test. All students took the test on a computer. Testing took place during a four-week window from the last week of February 2007 through the third week of March 2007.

Instrument Development

Qualifications of the internal reviewer. The internal reviewer was a third year Ph.D. student in Special Education at the University of Oregon. She had an elementary teacher certificate for grades one to six and finished her licensure training for special education teachers. The internal reviewer worked as an ESL teacher for two years, during which time she helped developed a reading curriculum for ESL students. She also worked as an elementary school teacher for two years and a freelance curriculum developer for six months. Her in-depth knowledge in reading, instructional design and her work experience with diverse student populations allowed her to provide constructive feedback on our instrument development, particularly in the issues that are related to content validity, clarity of direction, and bias against students with limited English proficiency and students with disabilities.

Internal review procedures and results. After the test items were written and edited, the internal reviewer evaluated the measures before distributing for an external review. For each measure, the internal reviewer assessed for readability, grade-level appropriate language, grade-level appropriate concepts, length and flow of sentences, and possible biases. To verify the item specifications, the internal reviewer reported the range of grade-level readability using the Flesch-Kincaid readability formula as well as the sentence length. She also inspected whether the wording and topics of the sentences and paragraphs were appropriate for the indicated grade level. Next, the internal reviewer commented on any gender, cultural, or linguistic biases. Finally, the internal reviewer shared suggestions for changes to be made to the paragraphs and sentences.

As an example of the recommendations, consider the following third grade sentence: *"Al's dad said he would buy him a new toy that he wanted when he turned ten years old."* This

sentence may be confusing to young readers. The pronoun *he* could refer to *Al* or *Al's dad*. The internal reviewer suggested changing Al's dad to Al's mom in the sentence, so the confusion over the pronouns could be avoided. Another example was an eighth grade paragraph that centered on a basketball game and contained several sport-specific terms or phrases such as "dribble down the court" or "score a basket." The internal reviewer noted that basketball is primarily an American sport and is not played worldwide. In order to make the paragraph more accessible to students with diverse cultural backgrounds, the item writer changed the basketball game to a soccer game, a sport familiar to a wider audience. The sport-specific terms also were changed to terms that are more general. Although the content was altered, the paragraph retained the appropriate grade-level readability rating, yet provided more access in terms of content to students from diverse cultural backgrounds.

Qualifications of external reviewers. Six teachers working in local schools reviewed the passages for the grade level in which they were currently teaching with the following experiences and credentials.

1. An eighth-grade teacher in a local middle school, who previously taught sixth grade. He had taught for four years and holds a Master's of Education degree.

2. A special-education teacher in a local elementary school who had taught for 20 years. She held a Master of Education degree and previously taught special education in middle school.

3. A fifth-grade teacher in a local elementary school who held a Master of Education degree and had taught for two years.

4. A teacher who had taught for 28 years and was teaching seventh-grade at a local middle school at the time of the review. She held a Master of Arts degree and previously taught sixth-grade and third-grade.

5. A reading specialist at a local elementary school who had taught for 23 years. She previously taught fourth- and fifth-grades and was pursuing a Master's Degree in Educational Leadership and Administrative Licensure at the time of the review.

6. A local sixth and seventh grade middle school teacher who held a Master of Education degree and was in his first year of teaching.

External review procedures and results. In addition to the internal review, several local public-school teachers reviewed the measures. They reviewed the items for grade-level appropriateness in terms of content and readability. External reviewers examined the language and vocabulary of the passages for grade-level appropriateness, any concepts described in the passages for grade-level appropriateness, the clarity of writing, and finally, any potential bias. Teachers rated the sentences on a Likert scale of 1-4 for each criterion. A rating of 1 indicated that the criterion was not at all met/appropriate, a rating of 2 indicated that the criterion was somewhat met/appropriate, and rating of 3 indicated that the criterion was met/appropriate, and a rating of 4 indicated that the criterion was exceptionally met/appropriate. In addition to rating the measures using the Likert scale, the teachers also provided feedback and suggestions to improve the quality of the measures. In all, the reviewing teachers found the vocabulary and content appropriate for the grade-level and reported no evidence of bias.

A total of three paragraphs or sentences received a score of less than 3 (appropriate). In these instances, the item writer made revisions accordingly. For example, one teacher commented on the inappropriateness of certain vocabulary words for ELL students, so the item writer substituted those words for more commonly used synonyms. This was completed without changing the grade-level readability of the passage significantly. When a teacher indicated that the vocabulary wasn't challenging enough for the grade-level, then the item writer chose more

challenging synonyms. In one case, the item writer wasn't able to substitute more difficult vocabulary of a measure without lowering the grade level rating significantly, resulting in a measure that didn't meet test specifications. In that case, the item writer kept the original vocabulary as the measure had passed the internal review process successfully. Another example included a lower rating in terms of clarity, and the reviewer indicated phrases that were vague or confusing. The item writer then revised those phrases to make them more clear to readers. Items that received a rating of 3 (appropriate) were evaluated on a case by case basis, and the item writer made revisions where necessary to incorporate teachers' suggestions or to make the passages more grade-level appropriate.

Pilot Testing of Measures

With all passages written and reviewed both internally and externally, the following procedures were used to ascertain the technical adequacy of the measures. Students took the silent reading passages in a computer lab; each student logged into the website and was directed to click on a passage title when they were ready to begin. Once they read the passage, they clicked on the "done" button.

Scoring

A simple algorithm calculated the number of words in each passage before calculating the students' words per minute (WPM) score. The algorithm was a simple code that compacts any extra sequential spaces, then splits up the passage on every space encountered, and counts the number of parts.

Word count algorithm. The word-counting algorithm has been implemented as a simple function, which receives a string of passage text as its only parameter, and returns an integer representing the total words counted. Upon execution, the code first trimmed the string argument

of any leading or trailing white space characters (such as an ordinary space character, a tab, new line, carriage return, etc). Next, a regular expression was used to match any consecutive white space characters within the string, and compact them into a single space character. That single space character was then used as a delimiter to split the entire string into an array of substrings, each formed by using the delimiter as a boundary. The function then returned the count of the total number of elements in the array.

The student's scores on Silent Reading Fluency measures were the estimated number of words the students read in one minute. The computer captured the raw number of seconds elapsed (to the third decimal place) once the students clicked "done." The scores were determined by two factors: (a) the number of words in each sentence or paragraph and (b) the amount of time it took a student to finish reading a given sentence. The WPM score was determined by first calculating the total number of words in the passage, then dividing by the number of seconds elapsed before students click "done," and finally, multiplying that number by 60. For example, it took a student six seconds to finish reading a 20-word sentence; the estimated number of words per minute for this student when given a comparable readability passage was 200 words per minute.

Scoring accuracy. The accuracy of the word-count algorithm was analyzed by comparing hand-counted items versus algorithm-counted items. The accuracy of the algorithm was analyzed to determine if it could be trusted for these measures, and to determine if it could be used in the future for other measures. This would enable the same scoring procedures to be used in the future, as the computer algorithm contributes to scoring efficiency.

After running a comparison analysis, the accuracy between the hand-counts and algorithm-counts was 99.8%. Of the 60 different items (which included 2,573 words total), the

counts didn't match exactly only 4 items; each of those was only off by one word. The algorithm didn't match exactly for two reasons:

　　1. For three of the four inaccurate items, the measures in the database accidentally had extra spaces after the end of the passage. So, the algorithm included that extra space when breaking apart the words, and viewed the passage as one word longer than correct count. This issue was addressed by adding just one line of code that "trims off" any leading or trailing white space to the existing algorithm.

　　2. The other miscount was caused by one sentence: "She sorted them into two piles--toys that she wanted to keep and toys that she wanted to give away." The "piles--toys" was counted as a single word instead of two, because of no breaking space. This example was determined to represent infrequent punctuation that would occur very rarely in items, so it was not modified.

Data Analyses

　　Each student had 10 SRF scores: Four WPM for each SRF-Paragraph and six for each SRF-Sentence. Prior to analyzing the data, researchers removed spurious data. Scores were omitted that exceeded 400 WPM for students in grades 3-5 and 600 WPM for students in grades 6-8. Scores above these rates were considered to be invalid based on the following evidence:

1. Adults can read silently with good comprehension at or above 500 words per minutes (Blayne, 1945).

2. People with speed reading training can easily read up to 700-1100 words per minutes.

3. National norms for oral reading fluency are considerably lower (Hasbrouck & Tindal, 2005).

After imposing the top thresholds, only a small number of extreme scores were eliminated and the overall remaining observed scores were normally distribution. The mean scores of words per minutes were adjusted without the undue influence of these extreme scores.

Results

The authors obtained the mean, minimum and maximum values of the observed words per minute (WPMs) for each grades. The data indicated the overall increase of the mean words per minute from Grade 3 to Grade 8. See Appendix D Tables D1-6 for descriptive statistics for each SRF-Paragraph and SRF-Sentence by grade.

Item characteristics were calculated using the Rasch Item Response model. The data indicated the narrow range of measures across the passages within the same grade, which means the paragraphs and sentences were comparable in the item difficulty even though the number of words differ in the sentence-passages and the paragraph-passages (See Appendix E Tables E1-6 for item difficulty estimates for each SRF-Paragraph and SRF-Sentence by grade.).

The estimated item difficulty obtained from Rasch Item Response Model reported the outfit mean squares of each test item. The items were considered "productive items" if the values of outfit mean squares values fell within .5 and 1.5. The data indicated that all sentences and paragraphs were deemed productive items.

Discussion

The SRF measures reflected the speed of students' silent reading. The internal and external review indicate that the SRF-Paragraphs and SRF-Sentences contain grade-level appropriate content, language, and are free of bias. Pilot study results showed consistency in item difficulty of all paragraphs and sentences within the same grade level, which indicated that the measures contained comparable forms. Because there were no anchoring persons or items in the

test design, we have not displayed comparison of students' SRF observed scores across the grade levels. Pilot study data also showed that the length of the passages didn't influence the difficulty or performance of students, regardless of whether they were reading sentences or paragraphs.

One limitation of the design was that silent reading measures could not determine if students indeed read every word. The existence of extreme scores indicated that some students did not follow the test administration direction of reading every word of the given passage. It was impossible to know how many students skipped words. However, the IRT fit-statistics indicated that more skillful students took less time to finish reading the passages; less skillful students took more time. Few unexpected patterns were observed in which less skillful students finished the sentences in less time. This discovery suggested that few students click the "done" button without reading every word in the test items. However, we were not able to confirm or disconfirm this assertion.

Reading for understanding has been the primary purpose of academic reading. The word *fluency* implied the dual standards: *accuracy* and *speed*. The current design did not include a measure of comprehension. While research on oral reading fluency indicates a link to reading comprehension, no such link has been established for silent reading fluency measures.

References

Blayne, T. C. (1945). Building comprehension in silent reading. *The Modern Language Journal, 29*(4), 270-276.

Embretson, S. E., & Reise, S.P. (2000). The new rules of measurement. *Item Response Theory for Psychologists.* Mahwah, NJ. : Lawrence Erlbaum Associates.

Hasbrouck, J., & Tindal, G. (2005). Oral Reading Fluency: 90 Years of Measurement (Tech. Rep. No. 33). Eugene, Oregon: University of Oregon, College of Education, Behavioral Research and Teaching.

Appendix A: Internal Review Rating Form

Passage File Name	Readability?	Language?	Concepts?	Test Specifications?	Bias?	Suggestions
Paragraphs						
Sentences						

Appendix B: External Review Rating Form

Passage File Name	Language? Vocabulary?	Concepts?	Clarity?	Bias?	Suggestions
Paragraphs					
Sentences					

Appendix B: Directions for Administration

Project INFORM: Directions for Administration

My name was _____. Today we are going to work with you in math and reading. You will work on the computer to complete the tasks.

Please your very best job on the reading problems and math problems. Sometimes you will be asked to read some sentences or some paragraphs. Please make sure you read every word on the sentences or paragraphs. I know it might be tempting to NOT read the sentences or paragraphs, but we really want to make sure you read all of the words. Sometimes you will be asked questions about the reading that you do, so please do your best reading.

To solve the math problems, you will have scratch paper and a pencil to use.

Each of you will have a different set of tasks so it was very important to FOLLOW DIRECTIONS. Some of you might have some more stories and fewer math problems or more math problems and fewer stories. So it doesn't matter who finishes first.

When you finish, please raise your hand and we will excuse you.

Is everyone ready? Does anyone have any questions? [wait for questions]

When we excuse you to the computers, please find your teacher's name on the list. Then find your name on the list. DO NOT GO ON UNTIL WE HAVE CHECKED TO MAKE SURE YOU HAVE ALL OF THE CORRECT INFORMATION.

Are you ready? [excuse students one at a time; helper will help them get set up on the computer]

What do you do first? *Find your teacher's name, then your name.*
What do you do after you have selected your name? *Wait for the teacher.*
What do you do when you're finished? *Raise your hand.*

Appendix C: Comparison of Word Counts: Hand-Counted versus Algorithm-Counted

Table C1.

Items	Item 1 Hand Count	Item 1 Code Count	Item 2 Hand Count	Item 2 Code Count	Item 3 Hand	Item 3 Code
G3 Para. 1	74	74	78	78		
G3 Para. 2	78	78	78	78		
G3 Sent. 1	19	19	19	19	19	19
G3 Sent. 2	19	19	19	19	19	19
G4 Para. 1	79	79	78	78		
G4 Para. 2	75	75	76	76		
G4 Sent. 1	19	19	19	19	19	19
G4 Sent. 2	19	19	20	20	19	19
G5 Para. 1	75	75	76	76		
G5 Para. 2	77	77	74	74		
G5 Sent. 1	21	21	20	20	21	21
G5 Sent. 2	22	22	23	23	19	19
G6 Para. 1	75	76	82	83		
G6 Para. 2	76	75	76	76		
G6 Sent. 1	21	21	20	20	21	21
G6 Sent. 2	20	20	24	24	20	20
G7 Para. 1	77	78	78	78		
G7 Para. 2	76	76	75	75		
G7 Sent. 1	19	19	20	20	23	23
G7 Sent. 2	19	19	18	18	22	22
G8 Para. 1	76	76	75	75		
G8 Para. 2	77	77	75	75		
G8 Sent. 1	21	21	23	23	22	22
G8 Sent. 2	24	24	21	21	24	24
Totals	1158		1167		248	

Appendix D: Mean, Minimum and Maximum Words per Minute for Each Test Item[1]

Table D1.
Grade 3 Silent Reading.

	Paragraphs				Sentences					
Items	1-1	1-2	2-1	2-2	1-1	1-2	1-3	2-1	2-2	2-3
Number of Valid Entries	123	123	124	124	125	123	123	124	124	125
Number of Missing Data	116	116	115	115	114	116	116	115	115	114
Mean	125.68	127.21	120.28	122.21	115.30	125.68	127.21	120.28	122.21	115.30
SD	56.93	58.15	53.85	69.25	60.74	56.93	58.15	53.85	69.25	60.74
Minimum	12.03	16.82	10.99	11.08	9.91	12.03	16.82	10.99	11.08	9.91
Maximum	321.74	293.16	283.62	387.96	392.29	321.74	293.16	283.62	387.96	392.29

Table D2.
Grade 4 Silent Reading.

Items	1-1	1-2	2-1	2-2	1-1	1-2	1-3	2-1	2-2	2-3
Number of Valid Entries	96	97	98	96	97	97	97	98	97	97
Number of Missing Data	88	87	86	88	87	87	87	86	87	87
Mean	167.98	158.59	143.74	165.33	134.27	134.09	153.35	131.03	145.36	160.09
SD	76.09	65.44	61.53	64.23	60.81	51.88	61.44	60.71	64.79	66.85
Minimum	34.54	25.56	26.23	34.74	31.8	19.56	25.58	17.27	15.26	55.36
Maximum	348.17	292.04	294.41	293.93	278.46	266.54	280.24	283.3	368.21	365.62

---

[1] Items 1-4 reflect two successive paragraphs (1-2) and items 5-10 reflect successive sentences within paragraphs (1-3)

Table D3.
Grade 5 Silent Reading.

Items	1-1	1-2	2-1	2-2	1-1	1-2	1-3	2-1	2-2	2-3
Number of Valid Entries	142	142	143	143	142	141	141	145	144	144
Number of Missing Data	131	131	130	130	131	132	132	128	129	129
Mean	176.04	186.65	168.37	191.64	162.69	168.85	188.02	160.48	174.28	158.34
SD	81.29	67.51	82.35	94.07	81.92	75.08	72.47	83.36	62.56	68.93
Minimum	25.94	55.73	46.17	67.05	23.8	48.91	37.91	22.17	52.13	23.2
Maximum	436.6	458.2	523.57	548.35	503.4	563.91	539.38	567.25	441.18	493.72

Table D4.
Grade 6 Silent Reading.

Items	1-1	1-2	2-1	2-2	1-1	1-2	1-3	2-1	2-2	2-3
Number of Valid Entries	97	96	97	97	97	97	97	97	97	97
Number of Missing Data	116	117	116	116	116	116	116	116	116	116
Mean	196.61	190.30	176.77	178.53	154.92	166.94	158.72	152.11	174.08	175.47
SD	94.87	87.19	76.99	73.66	88.64	77.94	86.74	67.30	65.39	58.65
Minimum	15.32	52.64	39.06	63.43	29.29	34.55	24.04	39.52	37.78	51.46
Maximum	453.15	455.38	361.82	422.34	436.14	439.88	541.24	454.89	373.15	339.46

Table D5.
Grade 7 Silent Reading.

Items	1-1	1-2	2-1	2-2	1-1	1-2	1-3	2-1	2-2	2-3
Number of Valid Entries	108	109	109	109	109	109	109	109	109	108
Number of Missing Data	98	97	97	97	97	97	97	97	97	98
Mean	207.84	214.54	205.45	217.06	168.08	214.52	224.34	173.74	195.35	203.12
SD	34.04	11.2	38.63	46.29	33.58	45.99	67.63	25.12	39.72	34.15
Minimum	486.54	473.4	445.57	485.75	453.64	588.24	578.86	391.21	472.03	483.87
Maximum	108	109	109	109	109	109	109	109	109	108

Table D6.
Grade 8 Silent Reading.

Items*	1-1	1-2	2-1	2-2	1-1	1-2	1-3	2-1	2-2	2-3
Number of Valid Entries	102	102	101	102	103	101	102	102	102	102
Number of Missing Data	110	110	111	110	109	111	110	110	110	110
Mean	212.12	201.69	205.97	211.22	182.27	205.04	199.99	173.05	188.82	216.25
SD	19.8	28.93	35.65	37.22	19.48	24.62	32.33	21.24	37.11	30.11
Minimum	499.18	458.44	499.14	488.28	455.37	440.19	435.36	463.92	431.51	449.58
Maximum	108	109	109	109	109	109	109	109	109	108

Appendix E: Estimated Item Difficulty Obtained from Rasch Item Response Model

Table E1.
Grade 3.

ENTRY	MEASURE	COUNT	SCORE	OUT. MSQ	OUT. ZSTD	OBS MATCH	EXP. MATCH
Paragraph1-1	-0.01	116	13507	0.88	-0.92	0.9	1.3
Paragraph1-2	-0.01	118	14189	0.77	-1.82	0.8	1.3
Paragraph 2-1	-0.01	120	13782	1.15	1.14	0	1.3
Paragraph 2-2	0	116	12611	1.26	1.78	0	1.4
Sentence 1-1	0	122	13425	1.44	2.95	1.6	1.3
Sentence 1-2	-0.02	121	15252	1.14	1.1	1.7	1.2
Sentence 1-3	0	124	13240	0.94	-0.43	0	1.4
Sentence 2-1	0.02	122	11146	1.31	2.11	2.5	1.4
Sentence 2-2	0.01	123	12529	1.25	1.79	1.6	1.4
Sentence 2-3	0.01	124	12771	0.95	-0.3	2.4	1.4

Table E2.
Grade 4.

ENTRY	MEASURE	COUNT	SCORE	OUT. MSQ	OUT. ZSTD	OBS MATCH	EXP. MATCH
Paragraph1-1	-0.01	116	13507	0.88	-0.92	0.71	0.9
Paragraph1-2	-0.01	118	14189	0.77	-1.82	0.77	0.8
Paragraph 2-1	-0.01	120	13782	1.15	1.14	0.7	0
Paragraph 2-2	0	116	12611	1.26	1.78	0.71	0
Sentence 1-1	0	122	13425	1.44	2.95	0.65	1.6
Sentence 1-2	-0.02	121	15252	1.14	1.1	0.71	1.7
Sentence 1-3	0	124	13240	0.94	-0.43	0.72	0
Sentence 2-1	0.02	122	11146	1.31	2.11	0.61	2.5
Sentence 2-2	0.01	123	12529	1.25	1.79	0.62	1.6
Sentence 2-3	0.01	124	12771	0.95	-0.3	0.69	2.4

Table E3.
Grade 5.

ENTRY	MEASURE	COUNT	SCORE	OUT. MSQ	OUT. ZSTD	OBS MATCH	EXP. MATCH
Paragraph1-1	0	114	16501	0.95	-0.33	0.76	2.6
Paragraph1-2	-0.01	116	18998	0.82	-1.43	0.78	2.6
Paragraph 2-1	0	129	19137	1.01	0.08	0.72	0
Paragraph 2-2	-0.01	116	18006	0.77	-1.87	0.76	3.4
Sentence 1-1	0.01	126	18018	1.44	3.13	0.7	3.2
Sentence 1-2	0	123	18315	1.07	0.6	0.73	1.6
Sentence 1-3	-0.01	119	19842	0.97	-0.22	0.76	1.7
Sentence 2-1	0.01	123	16596	1.64	4.33	0.6	0
Sentence 2-2	-0.01	130	20962	0.97	-0.21	0.75	0.8
Sentence 2-3	0.01	132	19076	0.81	-1.66	0.79	2.3

Table E4.
Grade 6.

ENTRY	MEASURE	COUNT	SCORE	OUT. MSQ	OUT. ZSTD	OBS MATCH	EXP. MATCH
Paragraph1-1	-0.01	75	11580	1.3	1.76	0.64	1.3
Paragraph1-2	-0.01	76	11649	0.85	-0.97	0.74	1.3
Paragraph 2-1	0	78	11325	0.75	-1.67	0.74	2.6
Paragraph 2-2	-0.01	83	12868	0.68	-2.29	0.79	1.2
Sentence 1-1	0.02	85	10883	1.34	2.02	0.66	1.2
Sentence 1-2	0	84	11951	1.14	0.9	0.64	1.2
Sentence 1-3	0.01	87	11857	0.94	-0.38	0.75	0
Sentence 2-1	0.01	92	12997	1.39	2.41	0.6	0
Sentence 2-2	-0.01	86	13442	0.89	-0.73	0.68	0
Sentence 2-3	-0.01	83	13033	0.76	-1.66	0.69	3.6

Table E5.
Grade 7.

ENTRY	MEASURE	COUNT	SCORE	OUT. MSQ	OUT. ZSTD	OBS MATCH	EXP. MATCH
Paragraph1-1	0	82	13418	0.76	-1.65	0.77	2.4
Paragraph1-2	0	80	13260	1.26	1.55	0.67	1.3
Paragraph 2-1	0	80	13034	0.56	-3.32	0.83	2.5
Paragraph 2-2	0	77	12534	0.82	-1.15	0.76	2.6
Sentence 1-1	0.02	96	13920	1.36	2.34	0.69	1
Sentence 1-2	0	76	12297	1.46	2.55	0.67	1.3
Sentence 1-3	-0.01	74	12764	0.71	-1.88	0.8	1.4
Sentence 2-1	0.01	89	12796	1.03	0.25	0.76	2.2
Sentence 2-2	0	90	15013	0.91	-0.58	0.77	2.2
Sentence 2-3	0	84	13797	0.98	-0.11	0.71	0

Table E6.
Grade 8.

ENTRY	MEASURE	COUNT	SCORE	OUT. MSQ	OUT. ZSTD	OBS MATCH	EXP. MATCH
Paragraph1-1	0	75	12344	0.98	-0.04	0.64	1.3
Paragraph1-2	-0.01	85	14592	0.89	-0.67	0.67	0
Paragraph 2-1	0	79	13213	1.07	0.47	0.61	1.3
Paragraph 2-2	-0.01	76	12909	0.82	-1.15	0.72	0
Sentence 1-1	0.01	83	12078	1.33	2.04	0.61	1.2
Sentence 1-2	0	78	12674	0.98	-0.07	0.68	3.8
Sentence 1-3	0	79	12600	0.99	-0.01	0.6	0
Sentence 2-1	0.01	86	12091	1.57	3.4	0.54	0
Sentence 2-2	0.01	76	11030	1	0.04	0.63	0
Sentence 2-3	-0.01	72	12388	0.88	-0.73	0.72	0