





tise and research and evaluation experience. Two panelists independently coded each item as either aligned or not aligned. When panelists disagreed, a third panelist independently reviewed the item in question and made a final determination (“adjudicated” items in Table B). When a rating of not aligned

terms of syntactic structure. We report results in terms of grade bands because a text may not “uniquely represent one specific grade” (Nelson, Perfetti, Liben, & Liben, 2012, p. 22). In other words, a text may be appropriate for students in a range of grades, depending on the purpose of the reading task and the student’s reading ability. A passage or item was deemed “readable” if at least two of the three indices used (FK, syntactic simplicity, and narrativity) fell within or below the grade band that encompassed the test’s grade level.

## **Task 2: Item Readability**

Existing research on readability pertains primarily to passages of text. There is little guidance and even less research on evaluating the readability of test items (in terms of complexity of the text used) other than a widespread recognition of the measurement challenges. Because of the lack of research to guide our approach to item-level readability, in a previous study of the 2019 assessments, we compared several methodologies to determine whether we could produce reliable results. In the 2019 study, we examined the effects of including or excluding line breaks between the question and answers, including only the correct answer choice or all answer choices, analyzing items separately or together as a test unit, and more. In implementing the varying approaches to analyzing the text contained in the STAAR assessments, these changes should not alter the ability of students to comprehend the text contained in the items. In other words, the formatting changes are not factors that make a substantive difference in the ease of comprehension of brief texts. In all analyses, we used the same indices to determine readability (FK, syntactic simplicity, and narrativity). If the results were similar no matter the approach to formatting the items, we would have had confidence that our results yielded a reliable estimate of the readability of the items on each test.

However, analyses conducted as part of the 2019 study showed the opposite pattern. When we compared the results from each approach, we found that the values for the three indices shifted substantially. The FK and narrativity indices changed the most from one approach to another; syntactic simplicity was somewhat more stable. Because we did not have confidence in the 2019 item-level results, we concluded that analyzing item readability in a reliable manner was not possible. For the 2020 assessments, we conducted several of the same analyses on a sample of items and found the same pattern emerging. Therefore, we again concluded that analyzing item readability in a reliable manner for this report is not possible.

Unless and until additional research provides clear guidance and evidence of a reliable way to evaluate the complexity of text used in STAAR items, we cannot recommend conducting analyses of the grade-level readability of test items (in terms of the complexity of the text). It is important to note that we were asked to analyze item readability, not item difficulty. An analysis of item difficulty requires a different methodology than an analysis of readability.

## **Task 3: Passage Readability**

Overall, two of the three indices fell within or below the English language arts (ELA) grade band for the test’s grade level for 30 of the 33 passages analyzed. In other words, 91% of passages met the criterion for readability as defined in this study in terms of text complexity (see Table C) when the ELA norms were used. Results for syntactic simplicity fell within or below the specified grade band for 100% of passages, and narrativity results fell within or below the specified grade band for 24% of passages based on the ELA norms. Our initial analysis used the ELA Coh-Metrix norms because passages were from the STAAR Reading and Writing tests. However, many of the passages would be classified as informational texts, a genre more likely aligned with the text samples used to establish the Coh-Metrix social studies norms. When we used the social studies norms to define the upper and lower limits of the grade

band for the test's grade level, two passages did not meet the criterion for text complexity for the grade band. The passages that did not meet the criterion appeared on the grade 3 and grade 6 reading assessments.

**Table C. Percentage of Passages Within or Below the Grade Band**

| Subject                          | FK                      | Syntactic Simplicity     | Narrativity            |                         | 2 of 3 Indices          |                         |
|----------------------------------|-------------------------|--------------------------|------------------------|-------------------------|-------------------------|-------------------------|
|                                  |                         |                          | Based on ELA Norms     | Based on SS Norms       | Based on ELA Norms      | Based on SS Norms       |
| Reading<br>( <i>n</i> = 25)      | 92%<br>( <i>n</i> = 23) | 100%<br>( <i>n</i> = 25) | 32%<br>( <i>n</i> = 8) | 80%<br>( <i>n</i> = 20) | 92%<br>( <i>n</i> = 23) | 92%<br>( <i>n</i> = 23) |
| Writing<br>( <i>n</i> = 8)       | 88%<br>( <i>n</i> = 7)  | 100%<br>( <i>n</i> = 8)  | 0%<br>( <i>n</i> = 0)  | 88%<br>( <i>n</i> = 7)  | 88%<br>( <i>n</i> = 7)  | 100%<br>( <i>n</i> = 8) |
| <b>TOTAL</b><br>( <i>N</i> = 33) | 91%<br>( <i>n</i> = 30) | 100%<br>( <i>n</i> = 33) | 24%<br>( <i>n</i> = 8) | 82%<br>( <i>n</i> = 27) | 91%<br>( <i>n</i> = 30) | 94%<br>( <i>n</i> = 31) |

# Introduction

The Texas Education Agency (TEA) contracted with The Meadows Center for Preventing Educational Risk (MCPER) at The University of Texas at Austin to conduct an independent study of the 2020 State of Texas Assessments of Academic Readiness (STAAR). The study consisted of three tasks:

Task 1: A content alignment study of 17 tests

Task 2: A readability study on questions and answers for 17 tests

Task 3: A readability study on passages for six reading and two writing tests

Task 1 consisted of two subtasks. Task 1A called for an independent study of item alignment to the precoded classification of Texas Essential Knowledge and Skills (TEKS) content standards (i.e., student expectations). Task 1B called for a study of the extent to which the tests as a whole reflect the TEKS for the tested grade or any grade below. For the other two tasks, we applied an evidence-based readability protocol to the items from the 2020 STAAR tests (Task 2) and to the passages from the 2020 Reading tests and Writing tests (Task 3). In the following sections, we describe important background information, our methods, and the results by subject and grade.

## Task 1A

### Background

Subtask 1A called for a study of item alignment to the precoded content standards. For this subtask, we examined the extent to which independent reviewers rated items on the STAAR tests as aligned to the precoded student expectation for the grade and subject being assessed. STAAR test items are “designed to measure the extent to which students have learned and are able to apply the knowledge and skills defined in the state-mandated curriculum standards, the Texas Essential Knowledge and Skills” (TEA, 2018, p. 1). Only specific TEKS are eligible for inclusion on an assessment and can be found in the Eligible Texas Essential Knowledge and Skills<sup>1</sup> documents for each subject and grade. Eligible student expectations are organized by reporting categories that are further delineated into broad knowledge and skills statements and specific student expectations (see Figure 1). Items on the STAAR are written at the student expectation level. An item’s precoded classification indicates both the reporting category and the specific student expectation assessed by that item. As an example, an item with a precoded classification of Reporting Category 1, Student Expectation 3.2A would assess the knowledge and skills in the portion of Figure 1 in bold. Documents indicating the precoded classification for each item are available for released tests.<sup>2</sup>

---

1 Eligible Texas Essential Knowledge and Skills documents can be found on the TEA website:

## Figure 1. Example Reporting Category and Corresponding Student Expectation on the Grade 3 Mathematics Assessment Eligible Texas Essential Knowledge and Skills

---

### Reporting Category 1: Numerical Representations and Relationships

The student will demonstrate an understanding of how to represent and manipulate numbers and expressions.

(3.2) Number and operations. The student applies mathematical process standards to represent and compare whole numbers and understand relationships related to place value. **The student is expected to:**

(A) compose and decompose numbers up to 100,000 as a sum of so many ten thousands, so many thousands, so many hundreds, so many tens, and so many ones using objects, pictorial models, and numbers, including expanded notation as appropriate; Readiness Standard

---

## Mathematics

For the 2020 mathematics assessments, reviewers rated 100% of items as aligned to the precoded student expectations. In grades 3, 4, 7, and 8, both reviewers rated 100% of items as aligned. In grades



**Table 3. Science Assessment Alignment to Precoded Content Standards**

| Grade                    | % Adjudicated       | Final # Not Aligned | Final Rating (% Aligned) |
|--------------------------|---------------------|---------------------|--------------------------|
| Grade 5 ( <i>n</i> = 36) | 0.0 ( <i>n</i> = 0) | 0                   | 100                      |
| Grade 8 ( <i>n</i> = 42) | 0.0 ( <i>n</i> = 0) | 0                   | 100                      |
| TOTAL ( <i>N</i> = 78)   | 0.0 ( <i>n</i> = 0) | 0                   | 100                      |

### Social Studies

As indicated in Table 4, 98% of the 2020 social studies assessment items were aligned to the precoded student expectation. Two items required adjudication by a third reviewer. Following adjudication, one item was rated as not aligned.

**Table 4. Social Studies Item Alignment to Precoded Content Standards**

| Grade                    | % Adjudicated       | Final # Not Aligned | Final Rating (% Aligned) |
|--------------------------|---------------------|---------------------|--------------------------|
| Grade 8 ( <i>n</i> = 44) | 4.5 ( <i>n</i> = 2) | 1                   | 97.73                    |
| TOTAL ( <i>N</i> = 44)   | 4.5 ( <i>n</i> = 2) | 1                   | 97.73                    |

### Writing

Overall, 98% of the 2020 writing assessment items aligned to the precoded content standards. A total of five items—four items in grade 7 and one item in grade 4—required adjudication by a third reviewer. Following adjudication, one item on the grade 7 assessment was rated as not aligned. As indicated in Table 5, the final percentage of items aligned to the precoded content standards was 100% in grade 4 and 97% in grade 7.

**Table 5. Writing Assessment Alignment to Precoded Content Standards**

| Grade                    | % Adjudicated        | Final # Not Aligned | Final Rating (% Aligned) |
|--------------------------|----------------------|---------------------|--------------------------|
| Grade 4 ( <i>n</i> = 25) | 4.0 ( <i>n</i> = 1)  | 0                   | 100                      |
| Grade 7 ( <i>n</i> = 31) | 12.9 ( <i>n</i> = 4) | 1                   | 96.77                    |
| TOTAL ( <i>N</i> = 56)   |                      |                     |                          |



## Science

In the final ratings of item alignment, 100% of items in grades 5 and 8 were rated as aligned to the curriculum, indicating that the 2020 science assessments were aligned with the TEKS for the tested grade levels.

**Table 8. Percentage of 2020 Science Assessment Items Aligned With the TEKS**

| Science                  | % Aligned |
|--------------------------|-----------|
| Grade 5 ( <i>n</i> = 36) | 100       |
| Grade 8 ( <i>n</i> = 42) | 100       |
| TOTAL (                  |           |





Measures of vocabulary load and syntactic structure are available in a number of tools that have been developed to gauge the readability of text. For this study, we processed text through Coh-Metrix (McNamara et al., 2014), a third-generation text analysis tool that provides more than 100 indices of text features, including the FK, syntactic simplicity, and narrativity metrics previously described. Coh-Metrix is used throughout the measurement and evaluation communities for a variety of text analysis purposes (see McNamara et al., 2014, for details on tool development and validation). In a study of seven tools for measuring text complexity, researchers provided evidence to support the validity of using Coh-Metrix to order text according to complexity (Nelson et al., 2012).

We selected Coh-Metrix for several reasons, including the following:

-



reading test is one component of item and test difficulty, but it has not been shown to be central; the link between item difficulty and item readability is even more tenuous. Research on accommodations for students with disabilities has shown that reading test items to students without disabilities (instead of having students read the items on their own) does not affect test performance (Fletcher et al., 2006). These findings suggest that the text complexity of items is not a significant factor in item difficulty for students without disabilities.

Therefore, unless an item's readability is so far beyond a student's reading ability that the item is incomprehensible, measurement experts would expect that a student's mastery of the content standard being tested would be the primary factor in the likelihood of answering an item correctly. When items are written to test knowledge of a particular concept, measurement experts typically focus more on testing knowledge at an appropriate level of difficulty than calibrating the readability of the item. One reason that experts take this approach is because, as previously stated, the concept of readability is not well established for text samples consisting of few words. Additionally, because little research supports item readability as a concept, item developers do not have actionable procedures for writing items to meet a particular grade-level readability. An analysis of item and test difficulty requires a different approach than an analysis of readability. Test and item difficulty must be evaluated using specific methodologies beyond the scope and purpose of this project.

## Task 3: Passage Readability

### Background

We refer the reader to the Text Readability section for background information on passage readability.

### Methods

Text analysis tools are designed to process prose. A primary data source for most text analysis tools, including Coh-Metrix, is a passage's syntactic structure. Poetry, in particular, has an irregular syntactic structure that would result in misleading and invalid estimates of readability. As a result, we excluded 10 reading assessment passages that were either poetry or drama. In grades 3 and 5, we excluded one passage per reading assessment. In grades 4, 6, 7, and 8, we excluded two passages per assessment that did not meet the criteria for inclusion. Appendix B provides an overview of the text preparation process used for this study.

### Results

The following tables present a profile of results for each passage, specifying whether the value of each index (FK, syntactic simplicity, and narrativity) fell within or below the grade band that encompassed the test's grade level. A passage was deemed "readable" if at least two of the three indices met that criterion.

Table 11 summarizes results across the 33 analyzed passages. Results for syntactic simplicity fell within or below the specified grade band for 100% of passages. For narrativity, our initial analysis used the language arts Coh-Metrix norms because passages were from the 2020 STAAR Reading and Writing tests and concluded that 24% of passages fell within or below the specified grade band. However, many of the passages would be classified as informational texts, a genre more likely aligned with the text samples used to establish the Coh-Metrix social studies norms. When we used the social studies norms





| Subject | Grade | Passage | FK  | Syntactic<br>Simplicity | Narrativity           |                      | 2 of 3 Indices        |                      |
|---------|-------|---------|-----|-------------------------|-----------------------|----------------------|-----------------------|----------------------|
|         |       |         |     |                         | Based on<br>ELA Norms | Based on<br>SS Norms | Based on<br>ELA Norms | Based on<br>SS Norms |
| Writing | 7     | p252023 | Yes | Yes                     | No                    | Yes                  | Yes                   | Yes                  |
| Writing | 7     |         |     |                         |                       |                      |                       |                      |
|         |       |         |     |                         |                       |                      |                       |                      |
|         |       |         |     |                         |                       |                      |                       |                      |



from Results Coaching. She holds a bachelor's in education and a master of education in educational leadership and administration.

### **Colleen Reutebuch**

**Senior project manager, researcher, and director, Reading Institute at MCPER**

Reutebuch conducts and manages research and external program evaluation. She has experience directing large-scale, federally funded intervention (Institute of Education Sciences [IES] Goals 2, 3, and 4), external evaluation (Office of Special Education Programs), and professional development and technical assistance projects at the state and national levels (U.S. Department of Education, Texas Education Agency). Currently, she serves as the evaluation project director and co-primary investigator for WestEd's National Center for Systemic Improvement, the National Deaf Center on Postsecondary Outcomes, and Leaders for Literacy and co-investigator on an efficacy and development grant. She executes and directs all aspects of research and program evaluation, including protocol development, data-collection planning, data management, analysis, and reporting. Since 2014, she has worked to identify and capture evidence of program quality and effectiveness. In the field of education for 20 years, she has been an assistant professor of special education, lecturer in special education and reading education, and educational specialist. She has published in peer-reviewed journals on the topics of response to intervention, reading difficulties, and academic enhancements and interventions. She earned a doctorate in special education in 2006 from The University of Texas at Austin. She holds special education, secondary reading, and reading specialist certifications.

### **Paul Steinle**

**Doctoral student, The University of Texas at Austin; research associate, MCPER**

Steinle received his master's in special education from National-Louis University and his bachelor's in anthropology from the University of Notre Dame. He was previously a special education teacher in Chicago Public Schools. His research interests include intensive interventions and response to intervention.

### **Jessica Toste**

**Assistant professor, The University of Texas at Austin; fellow and Board of Directors, MCPER**

Toste received her doctorate in educational psychology from McGill University. She teaches courses on reading instruction, learning disabilities, and special education law. She is a Provost's Teaching Fellow at The University of Texas at Austin and was named one of the 2017 "Texas Ten," nominated by alumni as a professor who inspired them during their time on campus. Her research interests are related to intensive interventions for students with reading disabilities, with a particular focus on data-based decision-making processes and motivation. She was trained in reading intervention research as a post-doctoral fellow at Vanderbilt University (2011–2013) and as a Fulbright scholar/visiting researcher at the Florida Center for Reading Research (2008–2009). She has worked as an elementary school teacher and reading specialist in Montreal, Canada. She serves on the Board of Directors and National Advisory Council of the Gay, Lesbian, & Straight Education Network. She is on the Board of Directors of Disability Rights Texas, the federally designated legal protection and advocacy agency for people with disabilities in Texas, as well the Advisory Board for The University of Texas Charter School System. She volunteers with Court Appointed Special Advocates Travis County as a court-appointed special advocate and guardian ad litem for children who have been abused and neglected.



serves as a co-principal investigator on an IES-funded Research Networks program, a multiyear project focused on the cohesive integration of behavior support within a process of data-based intervention intensification (Project BASIC: 2018–2023). He has also served as principal investigator on an IES-funded Goal-1 Exploration grant (Project CIFOR: 2015–2018) to investigate important associations between

bachelor's in business at San Jose State University in California and her master's in educational technology at Lehigh University in Bethlehem, Pennsylvania. She has more than 25 years of experience working with at-risk populations in the nonprofit and education fields. Her research interests include the prevention of reading difficulties through the systematic implementation of effective instructional strategies.

**Christy Murray**  
Project manager, MCPER







IES-sponsored Tier 2 mathematics interventions and two NSF-sponsored Tier 2 mathematics interventions. He earned his doctorate in special education at The University of Oregon.

**Maria Longhi**

**Project director, MCPER**

Longhi is project director for the Scientific Explorers grant. She has served as associate director of the Texas Literacy Initiative and program director of the Literacy Achievement and Reading to Learn Academies. She has provided high-quality professional development and technical assistance at the state, district, and campus levels in the areas of leadership, assessment, evidence-based literacy practices, and response to intervention. With more than 20 years of experience in the field, she has

## Senior Measurement Advisor

### David J. Francis

Hugh Roy and Lillie Cranz Cullen Distinguished Chair, The University of Houston; director, Texas Institute for Measurement, Evaluation, and Statistics; director, Center for Advanced Computing and Data Systems

Francis is a recipient of the University of Houston Teaching Excellence Award and a former member of the National Institutes of Health Behavioral Medicine study section. His interests include reading acquisition and the identification and prevention of reading disabilities, psychometrics, statistical models for longitudinal data, multilevel models, latent variable models, structural equation modeling, item response theory, and exploratory data analysis.

He is a fellow of Division 5 (Measurement, Evaluation, and Statistics) of the American Psychology Association and current member of the Independent Review Panel for the National Assessment of Title I and the Technical Advisory Group of the What Works Clearinghouse. He collaborates on multiple contracts and grants funded by the National Institute of Child Health and Human Development, IES, the National Institute of Deafness and Communication Disorders, the Texas Education Agency, and the Houston Livestock Show and Rodeo.

# Appendix B:

## Text Preparation Protocol

To prepare text for analysis, a group of researchers did the following:

- Opened assessment documents using Microsoft Edge PDF reader
- Copied and pasted text into plain text files (Coh-Metrix requires each unit to be a separate text file)
- Removed any nontext/nonprose elements—nontext elements included (a) figures, (b) tables, (c) equations, (d) fractions, (e) letter strings used for mathematical notation, (f) footnotes/endnotes, (g) diagrams, (h) instructions for recording answers, (i) ellipses, (j) underscores, (k) pictures, (l) accent marks, and (m) nonstandard characters
- Removed paragraph and sentence numbers
- Removed titles and headings
- Deleted extraneous paragraph breaks that resulted from removing section headings
- Inserted one hard return between paragraphs
- Double-checked punctuation—all text-analysis programs are punctuation sensitive, and removing or placing a period at the beginning of a new paragraph causes text-analysis results to be inaccurate
- Checked text in the Coh-Metrix Corpus Viewer prior to analysis to ensure paragraph and sentence breaks were correct
- Stored files as UTF-8 text
- Included the brief paragraph introducing the writing passages

# References

Benjamin, R. G. (2012). Reconstructing readability: Recent developments and recommendations in the analysis of texts.

# Suggested Citation

The Meadows Center for Preventing Educational Risk. (January 2020). *2020 assessments. Final report: Part 2*. Austin, TX: Author.

