

2019 Assessments

Final Report: Part 1

OSP# 201902572-001

December 2, 2019



The Meadows Center
FOR PREVENTING EDUCATIONAL RISK

Executive Summary

[Redacted text]

[Redacted text]

Table B. Item Alignment to Precoded Content Standards by Subject

Subject	Total Items	% Adjudicated	# Not Aligned as Coded	% Aligned as Coded	% Aligned to Grade-Level TEKS
Mathematics	<i>N</i> = 222	1.4 (<i>n</i> = 3)	1	99.55	100
Reading	<i>N</i> = 234	3.4 (<i>n</i> = 8)	0	100	100
Science	<i>N</i> = 78	2.6 (<i>n</i> = 2)	1	98.72	100
Social Studies	<i>N</i> = 44	6.8 (<i>n</i> = 3)	3	93.18	100
Writing	<i>N</i> = 56	10.7 (<i>n</i> = 6)	3	94.64	100

Task 1B. Test Alignment to the TEKS

To evaluate the extent to which the tests reflected the TEKS, we used the item ratings from Task 1A and calculated the percentage of items aligned with the TEKS. For this subtask, we classified an item as aligned if it addressed a student expectation from the tested grade or any grade below. Therefore, if an item was rated as not aligned to the precoded standard for Task 1A, but the alternative student expectation provided by the reviewer was from the tested grade or any grade below, we considered that item aligned for Task 1B. All the alternative expectations provided by reviewers for the eight items rated as not aligned to the precoded student expectation in Task 1A were from the tested grade’s TEKS. As a result, the data indicate that across grade levels and subjects, all tests included in this study were aligned with the TEKS content standards for the grade level tested.

Tasks 2 and 3: Readability Study

For Tasks 2 and 3, we applied a readability rubric to the text based on the most recent research in this area. For this study, we processed text through Coh-Metrix (McNamara, Graesser, McCarthy, & Cai, 2014), a sophisticated text-analysis tool that provides many indices of text features. To evaluate readability of the STAAR, we used three Coh-Metrix indices: a measure of word and sentence length and difficulty (the Flesch-Kincaid [FK] grade-level estimate of readability), a measure of syntax (syntactic simplicity), and a measure of vocabulary load (narrativity). For each index, we determined whether the results fell within or below a grade band, defined as the tested grade and the two adjacent grades (i.e., +/- one grade). The syntactic simplicity and narrativity results for a passage are linked to the readability levels of passages that have been determined to be suitable for students at different grade levels. For example, a test passage with a syntactic simplicity score in the grade 4 to grade 5 band is comparable to passages written for and previously rated as readable for students in grades 4 and 5 in terms of its 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, other than a widespread recognition of the measurement challenges. Because of the lack of research to guide our approach to item-level readability, we compared several methodologies to determine whether we could produce reliable results. For example, 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, our results showed the opposite pattern. When we compared the readability 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 do not have confidence in these results, we were forced to conclude 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 item readability, we cannot recommend conducting analyses of the grade-level readability of test items. 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 35 passages analyzed. In other words, 86% of passages met the criterion for readability as defined in this study (see Table C) when the ELA norms were used. Results for syntactic simplicity fell within or below the specified grade band for 97% of passages, and narrativity results fell within or below the specified grade band for 31% 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, only one passage did not meet the criterion for readability. The passage that did not meet the criterion appeared on the grade 7 reading assessment.

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
Writing (N = 8)	88% (n = 7)	100% (n = 8)	25% (n = 2)	88% (n = 7)	88% (n = 7)	100% (n = 8)

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

Methods

Item Rating Protocol

To determine item alignment with precoded content standards, MCPER leveraged the content area expertise of its staff and a related faculty members. MCPER is a collaboration of researchers from multiple disciplines who have conducted research, professional development, and program evaluation on a national level and across districts in Texas. MCPER has conducted research funded by the Institute of Education Sciences, National Institutes of Health, and the National Science Foundation in a range of domains, including reading and language arts, social studies, science, and mathematics. In addition, MCPER's partner center, the Vaughn Gross Center for Reading and Language Arts, is a leader in state literacy initiatives and research. A panel of qualified staff members (see Appendix A) with content knowledge and research and evaluation experience rated items. Before the panel began this work, they completed a self-paced training to review the legislation mandating the study,³ the STAAR program, the TEKS organization, and the rating rubric. Because MCPER staff members are trained in handling confidential data, able to do their work on secure university-owned IT assets, and not involved in administering STAAR tests in schools, test security is maintained more easily than if external raters were involved in rating items.

For this task, we defined items as the question, answer choices, and any accompanying passages, maps, graphs, charts, or figures. In each subject area (reading, mathematics, social studies, science, and writing), 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 1). We selected reviewers with leadership roles on research studies or professional development projects to serve as third reviewers. Third reviewers were able to render an unbiased, expert judgment because they had not previously rated the items in question. When a rating of not aligned was assigned, the reviewer indicated the reason(s) for the rating and provided an alternative student expectation that more closely aligned with the knowledge and skills addressed in the item, if one existed. TEA-provided resources used for this task included the 2019 STAAR tests, the Eligible Texas Essential Knowledge and Skills documents for each grade and subject reviewed, and the precoded classifications listed on the STAAR Student Expectations Tested documents.

3 86th Legislature, HB3, Sec. 39A.907: Assessment Instrument Study

Alignment Definition

For the purposes of this task, *alignment* was defined as agreement between the knowledge and skills assessed by the item and those encompassed in the precoded content standard. When items are aligned with the content standard, students who have mastered the knowledge and skills in the corresponding student expectation would be expected to answer the item correctly. Aligned items may address only a portion of the precoded standard. For example, an item aligned with standard 4.11B (“Students are expected to distinguish fact from opinion in a text and explain how to verify what is a fact”) may address only the first skill listed (distinguishing fact from opinion) and still be aligned. In addition, we used the TEA guidelines (TEA, 2015) to explain to reviewers that examples following the terms “such as” and “including” do not represent the only examples that may provide the basis for an item. Items considered not aligned assess knowledge and skills that are not associated with the precoded student expectation.

Rater Reliability

To assess rater reliability following training, we used sets of STAAR items from the 2018 assessments. Items were taken from multiple grade levels within each subject area and assembled into a sample set of 12–15 items for each subject area. Each rater’s reliability was tested using items from the subject area to which the rater was assigned. To establish the practice item sets as a gold standard for assessing reliability, multiple members of the project team reviewed these items and determined that they were aligned to the associated content standard. The threshold for acceptable rater reliability was set at 90%; raters had to identify at least 90% of items in the practice set as aligned to the precoded student expectation. Of the 15 total raters, 14 achieved this level of reliability; one rater achieved 80% agreement. The project team provided additional clarification to this rater regarding the definition of alignment to the standards and the guidelines for determining whether an item was aligned before this rater rated the 2019 items.

Results

In the following sections, we report the results of reviewers’ independent ratings of item alignment for each subject area and grade level. Tables in each subject area section indicate (a) the percentage of items that had discrepant ratings and were subsequently adjudicated by a third reviewer, (b) the total number of items rated as not aligned, and (c) the percentage of items with a final rating of aligned after adjudication.

Mathematics

For the 2019 mathematics assessments, reviewers rated 99.55% of items as aligned to the precoded student expectations. In grades 3, 5, and 8, both reviewers rated 100% of items as aligned. In grades 4, 6, and 7, one item per assessment required adjudication by a third reviewer. Following adjudication, one item on the grade 7 assessment was rated as not aligned. As indicated in Table 1, the final percentage of mathematics items rated as aligned to the precoded content standards following adjudication ranged from 98% to 100%.

4 See Report Addendum for information on nonaligned items, including the rating rationale and alternative student expectation(s).

Table 1. Mathematics Item Alignment to Precoded Content Standards

Grade	% Adjudicated	Final # Not Aligned	Final Rating (% Aligned)
Grade 3 (n = 32)	0.0	0	100
Grade 4 (n = 34)	2.9 (n = 1)	0	100
Grade 5 (n = 36)	0.0	0	100
Grade 6 (n = 38)	2.6 (n = 1)	0	100
Grade 7 (n = 40)	2.5 (n = 1)	1	97.50
Grade 8 (n = 42)	0.0	0	100
TOTAL (N = 222)	1.4 (n = 3)	1	99.55

Reading

For the 2019 reading assessments, 100% of items across grades 3–8 assessments were aligned to the precoded content standards. In grade 8, both reviewers rated 100% of items as aligned. In grades 3–7, a total of eight items required adjudication by a third reviewer (grades 4, 6, and 7 = one item each; grade 3 = two items; grade 5 = three items). Following adjudication, 100% of items across grades 3–8 were rated as aligned (see Table 2).

Table 2. Reading Item Alignment to Precoded Content Standards

Grade	% Adjudicated	Final # Not Aligned	Final Rating (% Aligned)
Grade 3 (n = 34)	5.9 (n = 2)	0	100
Grade 4 (n = 36)	2.8 (n = 1)	0	100
Grade 5 (n = 38)	7.9 (n = 3)	0	100
Grade 6 (n = 40)	2.5 (n = 1)	0	100
Grade 7 (n = 42)	2.4 (n = 1)	0	100
Grade 8 (n = 44)	0.0	0	100
TOTAL (N = 234)	3.4 (n = 8)	0	100

Science

Overall, 99% of items on the science assessments were aligned to the precoded content standard. In grade 5, both reviewers rated 100% of items as aligned. In grade 8, two items required adjudication by a third reviewer. Following adjudication, one item on the grade 8 assessment was rated as not aligned. As indicated in Table 3, final science item ratings after adjudication were 100% for grade 5 and 98% for grade 8.

Table 3. Science Assessment Alignment to Precoded Content Standards

Grade	% Adjudicated	Final # Not Aligned	Final Rating (% Aligned)
Grade 5 (n = 36)	0.0	0	100
Grade 8 (n = 42)	4.8 (n = 2)	1	97.62
TOTAL (N = 78)	2.6 (n = 2)	1	98.72

Social Studies

As indicated in Table 4, 93% of the 2019 social studies assessment items were aligned to the precoded student expectation. Three items required adjudication by a third reviewer. Following adjudication, the three items were 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)	6.8 (<i>n</i> = 3)	3	93.18
TOTAL (<i>N</i> = 44)	6.8 (<i>n</i> = 3)	3	93.18

Writing

Overall, 95% of the 2019 writing assessment items aligned to the precoded content standards. A total of six items—four items in grade 4 and two items in grade 7—required adjudication by a third reviewer. Following adjudication, two items on the grade 4 assessment and one item on the grade 7 assessment were rated as not aligned. As indicated in Table 5, the final percentage of items aligned to the precoded content standards was 92% 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)	16.0 (<i>n</i> = 4)	2	92.0
Grade 7 (<i>n</i> = 31)	6.5 (<i>n</i> = 2)	1	96.77
TOTAL (<i>N</i> = 56)	10.7 (<i>n</i> = 6)	3	94.64

Task 1B

Background

Subtask 1B called for a study of the extent to which tests as a whole reflect the TEKS for the tested grade or any grade level below. When rating item alignment to the precoded student expectations for Task 1A, raters considered the item and any accompanying passages, figures, graphs, etc. Because the ratings considered information about the test as a whole, we were able to leverage data from Task 1A to answer the question of test alignment to grade-level TEKS.

Methods

To determine the extent to which the tests reflect the TEKS, we used the item ratings from Task 1A and calculated the percentage of items aligned with the TEKS. However, for this subtask, we classified an item as aligned if it addressed student expectations from the tested grade or any grade below. In other words, if an item was rated as not aligned to the precoded standard for Task 1A but the alternative student expectation provided by the reviewer was from the tested grade or any grade below, we considered that item aligned for Task 1B.

Results

For each content area and grade level, we report the percentage of items aligned to the TEKS for the tested grade or any grade below. When the third reviewer rated an item as not aligned, we used that reviewer's explanation and alternative student expectation(s) in our analyses.

Mathematics

In the final ratings of 2019 mathematics item alignment, one grade 7 item was rated as not aligned to the precoded student expectation. However, the alternative student expectation provided by the reviewer was also within the grade 7 standards. Therefore, results indicate that the 2019 mathematics assessments were aligned with the TEKS from the tested grade levels.

Table 6. Percentage of 2019 Mathematics Assessment Items Aligned With the TEKS

Mathematics	% Aligned
Grade 3 ($n = 32$)	100
Grade 4 ($n = 34$)	100
Grade 5 ($n = 36$)	100
Grade 6 ($n = 38$)	100
Grade 7 ($n = 40$)	100
Grade 8 ($n = 42$)	100
TOTAL ($N = 222$)	100

Reading

In the final ratings of item alignment, 100% of reading items were aligned to the precoded student expectation, indicating that all 2019 reading tests across grades 3–8 were aligned to the TEKS from the tested grade levels.

Table 7. Percentage of 2019 Reading Assessment Items Aligned With the TEKS

Reading	% Aligned
Grade 3 ($n = 34$)	100
Grade 4 ($n = 36$)	100
Grade 5 ($n = 38$)	100
Grade 6 ($n = 40$)	100
Grade 7 ($n = 42$)	100
Grade 8 ($n = 44$)	100
TOTAL ($N = 234$)	100

Science

In the final ratings of item alignment, 100% of grade 5 science items were rated as aligned to the precoded student expectation. One grade 8 item was rated as not aligned to the precoded student expectation. However, the reviewer indicated that the item was better aligned with an alternative student

5 See Report Addendum for information on nonaligned items, including the rating rationale and alternative student expectation(s).

Second-generation readability formulas such as Lexile (MetaMetrics), the New Dale-Chall (Chall & Dale, 1995), and Advantage/TASA Open Standard (ATOS; Renaissance Learning) built on the traditional for

sure vocabulary load because it describes the extent to which a text is “likely to contain more familiar oral language that is easier to understand” (McNamara et al., 2014, p. 85), which closely aligns with the notion of vocabulary load. The index is labeled “narrativity” because narrative (storylike) passages are characterized by frequent use of words acquired earlier in the development of language comprehension. Researchers have found that although the average narrativity score is higher for language arts texts than it is for social studies and science texts within each grade band, narrativity scores decrease (i.e., text becomes less narrative) as a function of grade level, regardless of the subject area (Graesser, McNamara, & Kulikowich, 2011). Therefore, the narrativity index is an appropriate measure of vocabulary load for different text types and provides a robust estimate of the network of attributes that contribute to a text’s vocabulary load.

An additional limitation of the first-generation readability formulas that is not addressed in the second-generation formulas is the complexity of the syntax, or structure, of a text. The Coh-Metrix index of syntactic simplicity is used as a third component of our approach to evaluating readability because it represents the “degree to which the sentences in a text contain fewer words and use simpler, familiar syntactic structures that are less challenging to process” (McNamara et al., 2014, p. 85). Syntax influences text comprehension, and research indicates that a measure of syntax can be used to rank texts in order of complexity (Graesser et al., 2011).

The FK, syntactic simplicity, and narrativity indices provide a balanced and complete perspective on the readability of passages and items on the STAAR tests. Together, they represent text characteristics that indicate the relative difficulty of the text. Research indicates that a measure of syntax can be used to rank texts in order of complexity (Graesser et al., 2011).

framework for assessing the readability for these types of stimuli exists. Appendix B provides an overview of the text preparation process used for both items and passages.

Because of the lack of research to guide our approach to item-level readability, we compared several methodologies to determine whether we could produce reliable results. The methodologies we implemented included the following:

- Analyzing each item separately as a single paragraph of text, with some items containing one sentence and others containing multiple sentences. We removed line breaks in the item text and implemented this approach in two ways, first including only the correct answer choice and then, for a subsample of items, including all answer choices.
- Analyzing a sample of items separately, retaining any line breaks, resulting in some items having multiple one- or two-sentence paragraphs. We implemented this approach including all answer choices.
- Analyzing all of the test items in each STAAR test as a unit to determine the test's overall readability, with each item formatted as a single paragraph. On the reading and writing assessments, we implemented this approach both including and excluding the passages in the assessment.
- Analyzing a sample of items in the order in which they appeared on the test and again ordered by item type (e.g., stem and leaf items).

In implementing these varying approaches to analyzing the text contained in the STAAR assessments, the changes we made should not alter the ability of students to comprehend the text contained in the items. The presence or absence of line breaks, the inclusion of correct and incorrect answer choices,

Healthcare

Homecare services	80%
Medical equipment	75%
Pharmacy services	90%
Hospital care	85%
Insurance coverage	95%
Mental health services	70%
Nursing homes	82%
Physical therapy	78%
Public health programs	88%
Rehabilitation services	76%
Specialty care	84%
Telemedicine	92%
Urgent care centers	72%
Wound care services	80%
Workplace health programs	77%
Women's health services	86%
Yield	90%

Overall performance metrics

Summary statistics

In writing, seven of eight passages (88%) met the criteria of having two or three indices fall within or below the grade band for the test's grade level using the ELA norms for narrativity. One passage in grade 4 did not meet the criteria. When the social studies norms for narrativity were applied, all eight passages met the criteria for readability.

In reading, 23 of 27 passages (85%) met the criteria of having two or three indices fall within or below the grade band for the test's grade level using the ELA norms for narrativity. Three grade 4 passages and one grade 7 passage did not meet the criteria. Using the social studies norms for narrativity, 26 of 27 passages (96%) met the criteria for readability. One passage from the grade 7 reading test did not meet the criteria.

Appendix A: Review Panelists and Advisors

Review Panelists

Reading

Christy Austin

Doctoral candidate, The University of Texas at Austin; research associate, The Meadows Center for Preventing Educational Risk (MCPER)

Austin worked for 2 years as a first- and second-grade teacher at Rawson Saunders, a private school for students with dyslexia. Prior to teaching at Rawson Saunders, she worked as a special education coordinator and assistant principal at Knowledge is Power Program Camino Academy in San Antonio, Texas. She was responsible for developing and monitoring the implementation of individualized education programs for students receiving special education services, developing and monitoring the services provided to students on 504 plans, managing student discipline, coaching and supervising teachers, and presenting professional development related to special education, school culture, and discipline. Austin also spent 2 years as a life-skills teacher at Chase's Place, a school for students with moderate to severe developmental disabilities. She received a bachelor's in humanities from Trinity University. She received a master of education in special education from The University of Texas at Austin, specializing in learning disabilities and behavioral disorders. She is particularly interested in research in the area of reading interventions. She currently coordinates the initiative Behavior and Academic Supports: Integration and Cohesion.

Michelle Lambert-Yuhasz

Senior field trainer/analyst, MCPER

Lambert-Yuhasz has been an educator for 21 years, 14 of which she has spent supporting literacy in schools. Her support spans content areas and has involved reading and writing connections, small-group instruction, and interventions. She has assisted several districts with the implementation of a coaching model, including six Texas juvenile facilities, and she served as a state trainer-of-trainers for the 2016–2017 Literacy Achievement Academies for first and third grades. She is a certified teacher in grades 1–8, prekindergarten to grade 12 special education, English as a second language, and prekindergarten to early childhood. She also is a certified principal. She is currently obtaining a certification in adult training and development. In addition, she has level 3, or advanced level, training in coaching from Results Coaching. She holds a bachelor's in education and a master of education in educational leadership and administration.

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.

Mathematics

Rene Grimes

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

Grimes received her master's from The University of Texas at Arlington in mind, brain, and education with a focus on the cognitive and psychological aspects of learning. She received her bachelor's from the University of North Texas with a focus on early education and English as a second language. She is also certified in special education. She is interested in the cognitive and neurological aspects of mathematical learning difficulties. In particular, she is interested in identifying classroom prevention and intervention methods for early childhood through blended learning. Grimes previously worked in public and private schools in both general education and co-taught classrooms for preschool children with disabilities, and for prekindergarten, first-, and second-grade students. She has worked with adults and children on the autism spectrum, as well as their families, in private education settings and in their homes. She is a member of the Fort Worth Museum of Science and History Autism Advisory Board, which supports the museum in implementing programs for children with autism and their families.

Nancy Lewis

Researcher and project manager, MCPER

Lewis works on data-related research projects funded by the Institute of Education Sciences (IES) and National Institutes of Health. She has served as a key researcher and methodologist for numerous applied education research projects involving research design and data analysis, meta-analysis, program evaluation, survey construction, and survey data analysis. Her expertise includes advanced statistical techniques such as hierarchical linear modeling, structural equation modeling, and regression-discontinuity analysis. She completed the IES-sponsored methods training program in cost-effectiveness and benefit-cost analysis conducted by the Center for Benefit-Cost Studies of Education in May 2016. She has a doctorate in educational psychology and master's in program evaluation from The University of Texas at Austin, a master's in clinical psychology from Wheaton College, and a bachelor's in psychology from Northwestern University.

Greg Roberts

Associate director, MCPER; executive director, Vaughn Gross Center for Reading and Language Arts

Roberts directs all data-related activities for the centers. He is or has been a principal investigator, co-principal investigator, or lead methodologist on more than 20 research, development, and technical assistance grants and contracts funded by IES, National Institutes of Health, National Science Foundation (NSF), and Office of Special Education Programs, among others. Trained as an educational research psychologist, with expertise in quantitative methods, he has more than 90 peer-reviewed publications in multidisciplinary Tier 1 journals using structural equation models, meta-analysis, multi-level models, and explanatory item response theory. He holds a master's and doctorate in educational psychology from The University of Texas at Austin and a bachelor's in special education from North Texas State University. He taught sixth-grade math for 6 years.

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 worked closely with directors, administrators, literacy coaches, and teachers to build capacity and im

Kim Rodriguez

Senior field trainer/analyst, MCPER

Rodriguez earned her master's in special education from The University of Texas at Austin in 2000. She currently supports data collection and reporting tasks for this evaluation and for the National Center on Systemic Improvement. Previously, she worked at the Vaughn Gross Center for Reading and Language Arts on both research and evaluation projects. She holds Texas teacher certifications in elementary and special education.

Writing

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 (IES Goals 2, 3, and 4), external evaluation (Office of Special Education Programs), and professional development and technical assistance projects at the

algebra readiness. She was awarded the Presidential Early Career Award for Scientists and Engineers in 2019. Her research interests include developing and testing interventions for students with mathematics disabilities, with a special emphasis on peer tutoring, word-problem solving, mathematics writing, and the symbols and vocabulary within mathematics. She has a master's and doctorate from Vanderbilt University.

Reading

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 (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 disabilities, 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.

Sharon Vaughn

Professor, The University of Texas at Austin; executive director, MCPER

Vaughn is the Manuel J. Justiz Endowed Chair in Education. She was the editor-in-chief of the *Journal of Learning Disabilities* and the co-editor of *Learning Disabilities Research and Practice*. She is the recipient of the American Educational Research Association Special Interest Group Distinguished Researcher Award and The University of Texas Distinguished Faculty Award. She is the author of numerous books and research articles that address the reading and social outcomes of students with learning disabilities. She earned her doctorate in education and child development at The University of Arizona.

research also includes investigating teachers' use and uptake of evidence-based teaching practices. As a principal investigator or co-principal investigator, he has been awarded more than \$26.5 million in funding from the U.S. Department of Education and National Science Foundation. He currently serves as a principal investigator on two DRK-12 Design and Development projects funded by NSF to design and test innovative mathematics (Precision Mathematics: 2015–2019) and science (Scientific Explorers: 2017–2021) interventions for struggling learners in first and second grades. He also serves as a co-principal investigator on two IES-funded Goal-3 Efficacy Trials (Fusion: 2016–2020; NumberShire Level-1: 2016–2020) to test the impact of Tier 2 mathematics interventions on student mathematics outcomes. Additionally, he 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 malleable factors of instruction and student academic outcomes within an archival, multi-intervention observation dataset collected during the course of four IES-funded efficacy trials. He has published 40 peer-reviewed publications and led the design and development of four 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 worked closely with directors, administrators, literacy coaches, and teachers to build capacity and implement sustainable literacy practices. She holds an M.Ed. in elementary reading and a B.B.A. in management. Prior to her work at MCPER, she served for 15 years as a bilingual teacher and district literacy coach. Her interests include implementation science, teacher effectiveness, and second-language acquisition.

Social Studies

Kim Rodriguez
Senior field trainer/analyst, MCPER

Rodriguez earned her master's in special education from The University of Texas at Austin in 2000. She currently supports data collection and reporting tasks for this evaluation and for the National Center on Systemic Improvement. Previously, she worked at the Vaughn Gross Center for Reading and Language Arts on both research and evaluation projects. She holds Texas teacher certifications in elementary and special education.

Writing

Michelle Lambert-Yuhasz
Senior field trainer/analyst, MCPER

Lambert-Yuhasz has been an educator for 21 years, 14 of which she has spent supporting literacy in schools. Her support spans content areas and has involved reading and writing connections, small-group instruction, and interventions. She has assisted several districts with the implementation of a

coaching model, including six Texas juvenile facilities, and she served as a state trainer-of-trainers for the 2016–2017 Literacy Achievement Academies for first and third grades. She is a certified teacher in grades 1–8, prekindergarten to grade 12 special education, English as a second language, and prekindergarten to early childhood. She also is a certified principal. She is currently obtaining a certification in adult training and development. In addition, she has level 3, or advanced level, training in coaching from Results Coaching. She holds a bachelor's in education and a master of education in educational leadership and administration.

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,

References

Suggested Citation

The Meadows Center for Preventing Educational Risk. (December 2019). *2019 assessments. Final report: Part 1*. Austin, TX: Author.

Report Addendum

2019 Assessment Items Rated as Not Aligned to Precoded Content Standards

Mathematics

Grade	Unique Item #	Reason for Nonalignment
7	I2549	Use of the word <i>fluently</i> . Typically, <i>fluently</i> means an immediate response given under timed conditions. This item aligns better with 7.3(B).

Science

Grade	Unique Item #	Reason for Nonalignment
8	I4653	Question asks the reader to make a prediction “about the immediate future” which aligns more closely to 8.11(B) and its emphasis on short and long term changes.