

# Independent Evaluation of Colorado READ Act Materials



Approved Assessment List and Advisory List for  
Instructional Programming and Professional Development

Linda Friedrich, Bryan Hemberg, Masha R. Jones, and Rachel Tripathy,  
WestEd and Robin Wisniewski, RTI International

June 22, 2021

Contract Number :148306



## Colorado READ Act EVALUATION



© 2021 WestEd. All rights reserved.

Suggested citation: Friedrich, L., Hemberg, B., Jones, M. Tripathy, R., & Wisniewski, R. (2021). *Independent Evaluation of the Colorado READ Act Materials*. WestEd.

WestEd is a nonpartisan, nonprofit agency that conducts and applies research, develops evidence-based solutions, and provides services and resources in the realms of education, human development, and related fields, with the end goal of improving outcomes and ensuring equity for individuals from infancy through adulthood. For more information, visit [WestEd.org](https://www.wested.org). For regular updates on research, free resources, solutions, and job postings from WestEd, subscribe to the E-Bulletin, our semimonthly e-newsletter, at [WestEd.org/subscribe](https://www.wested.org/subscribe).



# Table of Contents

<b>EXECUTIVE SUMMARY</b> .....	<b>1</b>
ASSESSMENTS .....	2
INSTRUCTIONAL PROGRAMS .....	3
PROFESSIONAL DEVELOPMENT PROGRAMS.....	5
<b>INTRODUCTION</b> .....	<b>6</b>
THE COLORADO READ ACT .....	7
EVALUATION OF THE READ ACT.....	8
PURPOSE AND ORGANIZATION OF THIS REPORT .....	9
SCIENTIFIC FOUNDATIONS OF READING PROFICIENCY IN EARLY ELEMENTARY GRADES.....	12
<b>APPROVED ASSESSMENTS</b> .....	<b>20</b>
CRITERIA USED FOR REVIEW .....	21
DESCRIPTION OF REVIEW CATEGORIES FOR READ ACT COMPLIANCE .....	21
DESCRIPTION OF CRITERIA DRAWN FROM ADDITIONAL PROFESSIONAL STANDARDS.....	27
DATA COLLECTION AND METHODS .....	32
RESULTS AND DISCUSSION .....	34
DISCUSSION OF CROSS CUTTING THEMES AND ISSUES THAT EMERGED FROM THE ASSESSMENT REVIEWS .....	36
CDE’S PROCESSES FOR IDENTIFYING ITEMS FOR APPROVED AND ADVISORY LIST .....	41
<b>INSTRUCTIONAL PROGRAMS ON THE ADVISORY LIST</b> .....	<b>44</b>
TYPES OF INSTRUCTIONAL PROGRAMMING MATERIALS .....	45
DESCRIPTION OF REVIEW CATEGORIES .....	45
DATA COLLECTION AND METHODS .....	50
RESULTS AND DISCUSSION .....	52
CDE PROCESSES FOR IDENTIFYING ITEMS FOR APPROVED AND ADVISORY LISTS.....	60
<b>PROFESSIONAL DEVELOPMENT PROGRAMS ON THE ADVISORY LIST</b> .....	<b>64</b>
DESCRIPTION OF REVIEW CATEGORIES .....	65
DATA COLLECTION AND METHODS .....	67
RESULTS AND DISCUSSION .....	68
CDE PROCESSES FOR IDENTIFYING ITEMS FOR APPROVED AND ADVISORY LISTS.....	71
<b>SUPPORTING ENGLISH LEARNERS</b> .....	<b>74</b>
PROVIDING LANGUAGE SUPPORTS .....	75
DIFFERENTIATION FOR ENGLISH LEARNERS IN PROFESSIONAL DEVELOPMENT .....	79
<b>CONCLUSIONS AND RECOMMENDATIONS</b> .....	<b>82</b>
SUMMARY FINDINGS .....	85
SPECIFIC RECOMMENDATIONS .....	87
<b>REFERENCES</b> .....	<b>94</b>
<b>LIST OF APPENDICES</b> .....	<b>99</b>



## Executive Summary

The Colorado State Legislature passed the Reading to Ensure Academic Development (READ) Act in 2012 and updated the Act in 2019. The revised Act requires an independent evaluation to identify and assess strategies that the state and local districts and schools have taken to support Colorado students in achieving proficiency in reading.

This report focuses on the findings related to Approved Assessments, advisory lists for Instructional Programming and Professional Development, and the Colorado Department of Education's (CDE) processes for selecting materials for these lists.

## Key findings

- **The materials the Colorado Department of Education approved for use with READ Act funds meet the minimum requirements in SB 19-199.**
- **All approved assessments met the minimum summary threshold for compliance with the SB 19-199 required elements: 6 “fully” meet, 7 “largely” meet, and 2 “partially” meet the requirements of the READ Act.**
- **For our overall summary rating 46 instructional programs received “fully met,” 18 received “largely met,” 3 programs received “partially met,” and 1 was not rated.**



## Executive summary

The importance of achieving early-grade reading proficiency for later student academic success is well documented. The Colorado State Legislature responded to this challenge by passing the Colorado READ Act (Reading to Ensure Academic Development) in 2012. Most recently, the Legislature updated READ Act requirements with its 2019 revision. These revisions included requirements for an independent evaluation. This report focuses on the findings related to Approved Assessments, advisory lists for Instructional Programming and Professional Development, and the Colorado Department of Education’s (CDE) processes for selecting materials for these lists.

The overall conclusion from the review of assessments, instructional materials, and professional development programs is **that the materials CDE approved for use with READ Act funds meet the minimum requirements in SB 19-199** (see Exhibit ES.1).

Exhibit ES.1. Summary of Ratings				
	Fully Met	Largely Met	Partially Met	Did Not Meet / Not Rated
Assessments	6	7	2	0
Instructional Programs	46	18	3	0 / 1
Professional Development Programs	6	0	0	0

In the remainder of this executive summary, we describe findings for each type of material, and present some broad recommendations. The concluding chapter to this report contains more detailed findings and recommendations.

### Assessments

All approved assessments met the minimum summary threshold for compliance with the SB 19-199 required elements: 6 “fully” meet, 7 “largely” meet, and 2 “partially” meet the requirements of the READ Act. Some assessment vendors provided more robust, organized, and comprehensive evidence than others, resulting in some assessments receiving higher ratings



## Executive summary

---

than others within criteria. The most challenging criteria for vendors were “is evidence-based” and “is aligned with the preschool through elementary and secondary education standards for reading adopted by the state board pursuant to section 22-7-1005.” Those assessments that received a “partially” meets summary rating did so because of a lack of alignment evidence.

All approved assessments also either “fully” meet (4), “largely” meet (9), or “partially” meet (2) additional professional standards examined in this evaluation. Some assessment vendors provided more robust, organized, and comprehensive evidence than others, resulting in some assessments receiving higher ratings than others within criteria. The most challenging criteria for vendors were “The assessment development and review processes are designed and implemented to remove bias against all students” and “The assessment offers appropriate accommodations so all students can be fairly and accurately assessed.” Those assessments that received a “partially” meets summary rating did so because of a lack of bias and fairness evidence.

Four overarching recommendations derive from these findings:

1. Establish a baseline for an appropriate evidence base
2. Require externally conducted standards-alignment studies
3. Consider a higher threshold for classification accuracy
4. Ensure that approved assessments provide evidence that they are
  - 1) fair and free of bias and 2) include appropriate accommodations

## Instructional Programs

CDE’s instructional program review process is rooted in empirical evidence and reflects both historical and current understandings of how the science of reading can be applied effectively in classroom practice. Of 123 instructional programs that were reviewed by CDE, 69 were ultimately approved or partially approved. Our evaluation of these instructional programs concluded that, by and large, these programs met the core requirements outlined in the READ Act.



## Executive summary

---

By and large, the instructional programs met the core requirements outlined in the READ Act. For our overall summary rating 46 instructional programs received “fully met,” 18 received “largely met,” 3 programs received “partially met,” and 1 was not rated. All instructional programs met the minimum threshold for evidence – a clear logic model rooted in the science of reading - that they have the potential to make a positive impact on students’ reading outcomes. All but one program demonstrated the presence of skill development in the areas of phonemic awareness, phonics, reading fluency, and reading comprehension (as applicable), with 49 fully meeting criteria for explicit and systematic skill development. All core programs met the minimum requirements for including texts on core academic content to assist students in maintaining or meeting grade-appropriate proficiency in academic subjects in addition to reading. While all instructional programs included some form of embedded assessment, only four programs submitted formal evidence related to their assessments’ reliability and validity. Our recommendations include:

1. Provide additional guidance to vendors for supporting comprehension and inference beyond the text (e.g., text-to-text connections)
2. Consider guidance about diverse authorship and representation that meets or exceeds research thresholds
3. Consider a process for designating programs that offer full differentiation for English Learners
4. Consider lifting requirements for formal validity and reliability requirements for assessments embedded in instructional programs and instead offering guidance for how state-approved assessments can / should be used in conjunction with these more informal tools



### Professional Development Programs

All six professional development programs on the advisory list were in compliance with all SB 19-199 required elements. All professional development programs met the minimum threshold for evidence – a clear logic model; one vendor submitted formal research documenting its impact on student outcomes. Each program is rooted in the science of reading, having the potential to make a positive impact on students’ reading outcomes. Each of the programs have rigorous evaluation throughout the course, testing teacher knowledge of reading instruction and pedagogy. Our recommendations for professional development programs are:

1. Consider incorporating revised criteria for professional development program’s logic models and theories of action based on more recent research syntheses
2. Consider requiring vendors to include a higher proportion of items that assess pedagogical knowledge and pedagogical content knowledge on their assessments and possibly performance-based assessments
3. Consider expanding focus on support for English learners





# 1

## Introduction

The 2019 revision of the READ Act (SB 19-199) includes a provision mandating that an independent, external evaluation of the READ Act program be conducted over a five-year period.

The multi-year evaluation is now under way and is being conducted by an independent research team led by WestEd that includes APA Consulting and RTI International.

The key legislative goals for this evaluation are as follows:

- **Help state policymakers and district leaders understand the impacts of READ Act funding and support on students, families, schools, and districts**
- **Determine the extent to which CDE's processes resulted in approved assessments and advisory lists for instructional programming and professional development that are consistent with READ Act requirements**
- **Provide feedback on how CDE's processes for selecting assessments, instructional programming, and professional learning might be improved**



### The Colorado READ Act

The importance of achieving early grade reading proficiency for later student academic success is well documented. In fact, researchers and education leaders consider achievement of reading proficiency by the end of the third grade to be crucial to a child's future academic success and financial independence.<sup>1</sup> To help schools and districts support all children in achieving this goal, the Colorado State Legislature passed the Colorado Reading to Ensure Academic Development) Act (READ Act) in 2012; this replaced the Colorado Basic Literacy Act.<sup>2</sup> The READ Act provides local education providers (LEPs), including school districts, with funding and support to aid literacy development for kindergarteners through third-grade students, especially those identified with “significant reading deficiencies” (SRDs) who are at risk of not reading at grade level by the end of third grade.

Under provisions of the READ Act, schools test students using reading assessments approved by the Colorado State Board of Education.<sup>3</sup> Schools are then required to develop individual READ Act plans that identify a pathway for reaching grade-level proficiency for those designated as having a significant reading deficiency. The READ Act specifies certain components required in all student READ Act plans; however, each plan must be tailored to meet individual student needs.

In addition to specifying that the Colorado State Board of Education approve a set of reading assessments, the READ Act also charges the Colorado

---

<sup>1</sup> Hernandez, D. J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation*. The Annie E. Casey Foundation; Fiester, L. (2013). *Early warning confirmed: A research update on third-grade reading*. The Annie E. Casey Foundation. <https://www.aecf.org/resources/double-jeopardy/>

<sup>2</sup> The READ Act includes many of the same elements as the CBLA, including a focus on K-3 literacy, assessment, and individual plans for students reading below grade level with the addition of: (1) funding to support these efforts, (2) requirements for parent communication, and (3) an explicit focus on students identified as having a significant reading deficiency.

<sup>3</sup> <https://www.cde.state.co.us/coloradoliteracy/readact/resourcebank>



Department of Education (CDE) with creating advisory lists of instructional programming<sup>4</sup> and professional development programs<sup>5</sup> that are scientifically based and evidence based. LEPs may use READ Act funds to purchase instructional programming from the advisory list; LEPs may purchase instructional programs that are not on CDE's advisory list if they do not use READ Act funds. With the 2019 revision of the READ Act, the legislature requires all K-3 teachers to complete evidence-based training in teaching reading by January 31, 2022. The professional development programs on CDE's advisory list allow teachers who successfully complete the professional development to meet this requirement.

### Evaluation of the READ Act

The 2019 revision of the READ Act (SB 19-199) includes a provision mandating that an independent, external evaluation of the READ Act program be conducted over a five-year period (see *2020 Annual Report on the Colorado READ Act* for an overview of updates in SB 19-199).<sup>6</sup> The multi-year evaluation is now under way and is being conducted by an independent research team led by WestEd that includes APA Consulting and RTI International.

The key legislative goals for this evaluation are as follows:

1. Help state policymakers and district leaders understand the impacts of READ Act funding and support on students, families, schools, and districts
2. Determine the extent to which CDE's processes resulted in approved assessments and advisory lists for instructional programming and professional development that are consistent with READ Act requirements
3. Provide feedback on how CDE's processes for selecting assessments, instructional programming, and professional learning might be improved

---

<sup>4</sup> <https://www.cde.state.co.us/coloradoliteracy/advisorylistofinstructionalprogramming2020>

<sup>5</sup> <https://www.cde.state.co.us/coloradoliteracy/readactprofessionaldevelopmentevidenceteachertraining>

<sup>6</sup> <https://www.cde.state.co.us/cdedepcom/readactreport>



This report summarizes findings and data gathered during the first year of the legislatively mandated evaluation for goals 2 and 3. The report relies on multiple sources of information, including

1. Materials collected by the independent evaluation from vendors
2. Materials submitted by vendors to CDE as part of the review process, as available (instructional programming and professional development only)
3. Independent program and research reviews from *EdReports* and *What Works Clearinghouse*
4. Publicly available documentation of CDE's review processes and timelines
5. Interviews with CDE staff who lead the review processes

Future reports will examine the implementation and impact of instructional programming and professional learning on student outcomes. Because this report focuses on newly approved programs, it is not feasible to examine impact in this report.

### **Purpose and Organization of this Report**

In this report, the evaluation team describes the evaluation of 15 approved assessments, 69 instructional programs on the advisory list,<sup>7</sup> and six professional development programs on the advisory list. Key data and information presented in this summary report for assessments, instructional programs, and professional development programs include: (a) the evidence-base for the assessments or programs; (b) scientifically based reading skills; and (c) assessment and program-specific requirements. The summary report describes the processes used, results with lessons learned, and recommendations.

---

<sup>7</sup> CDE approved 50 unique instructional programs, but some were approved for multiple categories and therefore counted more than once in our total (e.g., *Spalding: The Writing Road to Reading* was approved as a core program, a supplemental program, and as an intervention program, so it is counted three times).



This report focuses on providing initial answers to research questions for each type of material. It starts with a general literature review, then describes the processes used in the evaluation, findings with discussion of lessons learned, and ends with conclusions and recommendations. Of special interest was the inclusion of English learner support in the instructional programs and professional development programs. The background, results, and related discussion of lessons learned related to English learner support are synthesized in a separate chapter (Chapter 5).

It is also important to note limitations in this report. First, the question of whether programs resulted in a growth to standard is not addressed in this report; as implementation and outcome data become available this will be a major topic of analysis in future reports. CDE's review of instructional programs and professional development programs was completed in Spring 2020, with lists beginning to be made public on a rolling basis beginning in May 2020. The complete list of Spanish-language programs was not made public until September 2020. This means that districts would not have started to use these programs until at least Fall 2020. Therefore, we are not reporting on district's use of these programs at this time, nor are we able to report on programs' impact on growth to standard. Second, limited documentation exists for CDE's previous review and selection processes, and members of CDE's staff had not participated in prior selection processes. Thus, the evaluation team was unable to systematically explore how selection processes for items on the current list differ from CDE's past processes.

### **Research Questions**

This first-year report addresses the following questions pertaining to assessments, instructional programs, and professional development programs.

### **Approved Assessments**



## Introduction

---

1. Do all items on the approved assessment list meet the requirements of the READ Act?
2. Do all items on the approved assessment list meet additional professional standards of quality?

### **Instructional Programming Advisory List**

1. Do all items on the advisory list for instructional programming meet the requirements of the READ Act?
2. Do all items on the advisory list for instructional programming meet additional professional standards of quality?

### **Professional Development Advisory List**

1. Do all items on the advisory list for professional development meet the requirements of the READ Act?
2. Do all items on the advisory list for professional development meet additional professional standards of quality?

### **CDE's Processes for Identifying Items for Approved and Advisory Lists**

1. To what extent were CDE's processes and selection criteria aligned with criteria outlined in the READ Act?
2. What other criteria, if any, did CDE take into consideration when identifying items on the lists?
3. In addition to CDE staff, what were the qualifications of those involved in the selection processes, how were they identified, and what training did they receive (if relevant)?
4. What other professional standards or criteria might CDE consider in the future?

### **Analytic Frameworks Used for Review**

The criteria used in this evaluation for reviewing assessments, instructional programs, and professional development programs derive from READ Act statutory language, updated regulatory and nonregulatory guidance, the Colorado Academic Standards for Reading, Writing, and Communicating, the



*What Works Clearinghouse*, *EdReports*, and other related policies and guidance. Additional criteria used by the evaluation derive from information provided by an Expert Advisory Panel convened for this project as well as professional standards for evaluating assessment, instructional program, and professional development quality (see Appendix A.1-A.3 for detailed rubrics). By anchoring the evaluation protocols in related statute, guidance, and regulations, these evaluation protocols provide a transparent and consistent framework to determine READ Act compliance. The rubrics (see Appendix A.1-A.3) list the required elements, criteria for the elements, ratings for the evidence, what evidence is needed, and the workflow for the reviewers.

### **Scientific Foundations of Reading Proficiency in Early Elementary Grades**

Decades of research have demonstrated the importance of reading proficiency in the early elementary grades. Around third grade, students transition from developing foundational reading skills (“learning to read”) to using reading as a tool for acquiring information (“reading to learn”; Adams, 1990). These early years are a critical time for intervening to support struggling readers since students who do not have the ability to read independently by third grade are at risk of falling behind academically in subsequent grades. Longitudinal studies have shown that students with low reading test scores in third grade are less likely to complete high school (Lloyd, 1978), failing to graduate on time at a rate four times higher than their proficient peers (Hernandez, 2012).

Recognizing the importance of reading in the early grades, the United States Congress asked the National Institute of Child Health and Human Development to establish a National Reading Panel (NRP) that would perform a comprehensive and informed synthesis of the research around effective methods for teaching children to read. In 2000, the 14-member Panel released its report, identifying five instructional components that are essential for early-grade reading development: phonemic awareness, phonics, reading fluency, vocabulary, and



## Introduction

---

reading comprehension (Langenberg et al., 2000). In a minority view included with the report, Panel member Joanne Yatvin cautioned Congress about interpreting the NRP findings as definitive, claiming that the scope of topics that NRP examined was biased and narrow, and that the Panel had neither the time nor resources to conduct analyses with the rigor required to answer their research questions with certainty. Still, the NRP findings have had substantial influence on both policy and practice, as the five essential components of reading have become widely accepted as best practices in reading instruction.

Following the passage of the No Child Left Behind Act in 2001, and its emphasis on increased instructional time for reading, numerous funding and policy initiatives emerged aimed at raising early-grade reading proficiency rates. At the federal level, Reading First provided roughly one billion dollars in grants annually from 2002 through 2008 to support the instructional practices recommended by the NRP (US Department of Education, 2015). At the state level, at least 26 states have passed reading laws since 2000 that are aimed at providing financial support, accountability measures, procedural requirements, and interventions that will improve third-grade reading proficiency rates (Center on Enhancing Early Learning Outcomes, 2019). Most of these laws reference or require “scientifically based” reading instruction, interventions, and curricula, although by the time many of these laws were passed the five essential components of reading had already been adopted by major publishers and teacher training programs in response to the NRP Report (Herlihy et al., 2009).

With the proliferation of curricula, interventions, teacher professional development programs, and assessments centered around these five essential components has come a large body of empirical research aimed at determining the efficacy of targeting these components. In fact, there have been so many studies on early reading instruction and intervention that researchers have been able to conduct meta-analyses whereby the authors attempt to identify all high-quality studies on a given topic and use statistical modeling to produce a more





accurate impact estimate than any one study alone could provide. What follows is a short summary of recent meta-analytic findings on each of the five essential components of reading for Pre-K through third-grade students; all five components are included in the READ Act.

1. *Phonemic awareness* is the ability to notice, distinguish, and manipulate the individual sounds in spoken words (Liberman et al., 1974) (e.g., the word “juice” has three phonemes, “j-,” “ooo”, and “sss”) – is a strong predictor of students’ later reading abilities (e.g., Share et al., 1984; Snider, 1997). Research indicates that explicit instruction is highly effective in promoting the development of phonemic awareness skills, and leads to moderate improvements in reading overall (Bus & Van Ijzendoorn, 1999; Ehri, Nunes, Willows, et al., 2001). Longitudinal studies have shown that interventions focused specifically on supporting phonemic awareness were found to have lasting impacts on student reading proficiency, showing a greater effect one year after the end of the interventions than interventions focused more on phonics (Suggate, 2016).
2. *Phonics* is an instructional approach where students learn to sound out and blend letters in order to decode a word (which is a different skill than understanding what that word means). Explicit and systematic teaching of phonics has been shown to improve student decoding, spelling, and comprehension to a statistically greater degree than instruction without a focus on phonics (Ehri, Nunes, Stahl, et al., 2001; Jeynes, 2008). Research on phonics instruction specifically for low-performing readers similarly finds systematic phonics instruction to improve reading outcomes (Mcarthur et al., 2018). Explicit phonics instruction was found to have a smaller effect over time than instruction focusing on phonemic awareness and comprehension (Suggate, 2016).



3. *Fluency* refers to the relative degree of ease and automaticity with which letters are understood as words, words are understood for their meaning, and comprehension of a subject is derived from that meaning (Wolf & Katzir-Cohen, 2009). At higher levels of reading fluency, mental attention can be devoted to comprehension rather than the mechanics of reading, and fluency is therefore considered a critical link between word analysis and text comprehension. The developmental definition of fluency makes it difficult to study empirically, and evidence around the effectiveness of interventions and approaches to support fluency is mixed. There is some evidence that repeated reading and the modeling of reading (either in person or via audiobook) can improve fluency and comprehension (Chard et al., 2002; Stevens et al., 2017), but more rigorous empirical research is needed to understand how to best improve reading fluency in the early grades.
4. *Vocabulary* instruction represents an important component of reading comprehension because understanding text requires the construction of meaning from known words (Kamil, 2004). There is strong consensus that size of a student's vocabulary is predictive of how well they will understand what they read (e.g., Scarborough, 2001). Recent research indicates that interventions supporting vocabulary development are effective in improving expressive and receptive vocabulary (Marulis & Neuman, 2010). There is evidence that such interventions are also effective in improving comprehension of texts aligned with the intervention, but there are fewer studies finding that these interventions improve generalized reading comprehension (Elleman et al., 2009; Wright & Cervetti, 2017). Multidimensional approaches to learning words (e.g., providing contextual information around a set of words) tend to have a stronger impact on student reading comprehension than instruction focused on definitions (Stahl & Fairbanks, 1986; Wright & Cervetti, 2017).



5. *Reading comprehension* is the overall goal of reading instruction and occurs when students can process the text they read, derive meaning from it, and integrate that meaning with what they already know. Gough & Tunmer's (1986) influential model describes successful reading comprehension as dependent upon two foundational components: decoding and linguistic comprehension. Others have argued that fluency is a third critical component for supporting text comprehension (Joshi & Aaron, 2000; Solari et al., 2018). While some meta-analytic reviews show that decoding (García & Cain, 2014) and linguistic comprehension are each important predictors of reading comprehension, others found the effects to be small or inconclusive (Mcarthur et al., 2018). Part of the challenge in studying the effect of foundational components on reading comprehension is that the most important components for reading change with students' age. In elementary school, for example, reading ability is largely based on print knowledge and phonological awareness, whereas in middle school reading accuracy and linguistic comprehension play a larger role in overall comprehension (Storch & Whitehurst, 2002). It is not surprising then that studies show interventions focused on phonemic awareness to be most appropriate for students entering elementary school; interventions focused on phonics and fluency to have greatest effects in first and second grade; and interventions targeting comprehension overall to be most effective for third grade and beyond (Suggate, 2016).

Effective reading comprehension is dependent upon a complex and not entirely understood network of foundational skills that shift in their importance with a student's age and individual learning needs. In other words, when it comes to reading instruction one size does not fit all – and certain groups that have historically struggled with reading in the early grades require support and intervention beyond the typical reading curriculum. Effective reading instruction for English Learners and students with disabilities, for example, shares many



## Introduction

elements of reading instruction for proficient readers, but also includes additional practices and supports for these groups. Research shows that English Learners benefit from frequent and intentional instruction focused on oral language development – in other words, including modifications and support to ensure that students understand the words and concepts they read (Goldenberg, 2020). Additionally, multiple systematic reviews of research have found that models focused on simultaneously strengthening students’ home language and their English skills have been more effective than models that focus on English alone (Greene, 1998; Slavin & Cheung, 2005). Consequently, we would expect effective reading instruction for young English Learners to include modifications that help them understand a language that is new to them, likely by utilizing native language supports or bilingual resources (see Chapter 5 for more on supporting English Learners).

While students with disabilities comprise a heterogenous group with different challenges and needs, research has shown specific instructional strategies benefit reading outcomes for many students in this group, including sustained multi-year interventions, one-on-one or small group instruction, systematic instruction on foundational reading components, and abundant opportunities for practice and feedback (Berkeley et al., 2010; Vaughn & Wanzek, 2014). We would expect effective reading instruction for young students with disabilities to incorporate personalized, targeted reading interventions that allow for supported practice of foundational skills.

Despite efforts to tailor instruction and improve reading outcomes for at-risk groups like English Learners and students with disabilities, national reading outcomes for these groups have not improved in the last decade; on average, English Learners and students with disabilities in fourth grade score far below even the “Basic” reading benchmark as measured by the National Assessment of Educational Progress (NAEP) (National Center for Education Statistics, 2019). Unfortunately, race and socioeconomic background are also predictors of student



## Introduction

---

reading ability. While White and Asian students' fourth grade reading scores have hovered at or around the NAEP "Proficient" benchmark, Black and Hispanic students' scores fall around or below the NAEP "Basic" benchmark. Students who are not classified as economically disadvantaged tend to score near the NAEP "Proficient" benchmark, while students from economically disadvantaged backgrounds score, on average, around the NAEP "Basic" benchmark. These disparities in early elementary reading scores are alarming and the achievement gaps are not narrowing, underscoring the need for effective instruction and resources that work specifically to support at-risk groups.

Even with an ever-expanding body of research on reading mechanics and instructional best practices, most large-scale early literacy interventions have not produced the desired positive impacts on student reading achievement. Only a handful of rigorous impact evaluations have been conducted for large federal and state level reading initiatives, and they present mostly similar findings: some impact on instructional practices, but no impact on student reading performance. Following the Reading First funding initiative, for example, the Department of Education commissioned a study to examine the impact of Reading First on student reading proficiency. While the study found that teachers in Reading First schools received more professional development for reading instruction and spent more instructional minutes on the five essential components of reading, no impact on student reading performance was detected (Gamse et al., 2008). More recently, North Carolina State University evaluated the impact of the state's Read to Achieve program, aimed at grade-level reading mastery for all third-grade students. The study found no significant impacts on student reading achievement for students altogether, or for demographic subgroups (e.g., low income students or students with a disability) (Weiss et al., 2018).

One exception to these interventions which seemingly failed to impact student reading performance is Oregon's Reading First program, implemented from 2003 through 2009, which was shown by a rigorous multi-year evaluation to



## Introduction

---

have improved student reading scores for students in kindergarten through third grade (Baker et al., 2007). This comprehensive evaluation analyzed data from three different cohorts of students over three years. A staggered implementation rollout (i.e., the first cohort began their Reading First activities in Year 1, the second cohort began in Year 2, etc.) allowed researchers to examine not only year-to-year impact, but also to analyze the magnitude of impact as schools became more experienced with the intervention. The Oregon Reading First evaluation found that schools receiving Reading First funding were more effective in improving student reading outcomes each year they implemented the intervention – in other words, they got better with experience. This finding is consistent with literature on effective educational interventions that has found consistent, sustained interventions to produce impacts of greater magnitude than short interventions (Borman & D’Agostino, 1996). These findings suggest evaluations of state reading policies and programs may need to be focused on longer-term outcomes in order to identify impacts on student reading performance.



# 2

## Approved Assessments

The evaluation team used the criteria established by the READ Act to build its evaluation and review rubric.

The team added additional criteria to the evaluation and review rubric to provide supplemental information on the assessments. These additional criteria draw from established research on principles and characteristics for identifying high-quality assessments.

### Key findings:

- Overall, the approved assessments either fully or largely met the requirements outlined in the READ Act.
- Six of the 15 assessments reviewed fully met all READ Act requirements.
- However, for each required element, vendors provided evidence of varying levels of quality.



## Criteria Used for Review

This section defines the criteria used for consistently reviewing all approved assessments. The assessment rubric (see Appendix A.1) lists the required elements, criteria for the elements, ratings for the evidence, what evidence is needed, and the workflow for the reviewers.

## Description of Review Categories for READ Act Compliance

As the foundation for the categories for the review of assessments, the evaluation team used the criteria established by the READ Act to build our evaluation and review rubric. Exhibit 2.1 summarizes the READ Act requirements and sample evidence, while the text following the exhibit provides definitions for each criterion as well as a more extensive list of evidence considered.

Exhibit 2.1 READ Act Requirements for Approved Assessments and Example Evidence			
Criterion Number	Short Name	SB 19 – 199 Requirement	Example Evidence Reviewed
1	Evidence-based	Is evidence-based or scientifically based (22-7-1209 (2)(b)(I)(A))	<ul style="list-style-type: none"> <li>Simulation evaluation results</li> <li>Theory of action about how the assessment is intended to work</li> </ul>
2	Standards-aligned	Is aligned with the preschool through elementary and secondary education standards for reading adopted by the state board (22-7-1209 (2)(a)(II)(A))	<ul style="list-style-type: none"> <li>Independent (nonvendor) alignment study results</li> <li>Vendor-completed alignment study results</li> </ul>
3	Validity	Each of the recommended reading assessments is valid . . . proven to effectively . . . measure students' reading skills in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and reading comprehension (22-7-1209 (2)(a)(II)(B))	<ul style="list-style-type: none"> <li>Content specifications for each grade level</li> <li>Evidence that the content of the assessment was developed and reviewed by experts, including teachers</li> </ul>





## Approved Assessments

Exhibit 2.1 READ Act Requirements for Approved Assessments and Example Evidence			
Criterion Number	Short Name	SB 19 – 199 Requirement	Example Evidence Reviewed
4	Reliability	Each of the recommended reading assessments is . . . reliable . . . proven to accurately . . . measure students' reading skills in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and reading comprehension (22-7-1209 (2)(a)(II)(B))	<ul style="list-style-type: none"> <li>• A model-based approach to reliability reported for each grade or one type of reliability were reported that are appropriate for the purpose of the assessment</li> <li>• Evidence that for each type of reliability reported, the lower bound of the confidence interval around the median estimate met or exceeded 0.70</li> </ul>
5	Diagnostic accuracy	Diagnostic reading assessments are proven to accurately identify students' specific reading skill deficiencies (22-7-1209 (2)(a)(II)(C))	<ul style="list-style-type: none"> <li>• Evidence of construct validity, such as convergent and discriminant analyses, demonstrating correlations that are reasonable for the grade and skills assessed</li> <li>• Consequential validity evidence is provided and is connected to a well-articulated theory of action about how the assessment(s) are intended to work</li> </ul>
6	Spanish language assessment	(At least one of the recommended) reading assessments (for kindergarten, and first, second, and third grades) is normed for the performance of students who speak Spanish as their native language, which assessment is available in both English and Spanish (22-7-1209 (2)(a)(II)(D))	<ul style="list-style-type: none"> <li>• Evidence that the student sample utilized for norming the assessment is representative of students who speak Spanish as their native language</li> <li>• Evidence that experts in Spanish language and literacy were included in the development of items</li> </ul>
7	Paper and pencil diagnostic assessment	(The list of recommended) reading assessments and reading diagnostics includes (at least one assessment and one diagnostic that a student can complete using) paper and pencil rather than using a computer (22-7-1209 (2)(a)(II)(E))	<ul style="list-style-type: none"> <li>• Evidence of comparability of all forms for each grade level, using a representative sample of students</li> </ul>

1. *Evidence-based.* The evaluation team assessed if there was evidence of the use of research and theory to inform the selection of assessment targets, the methods and measures used in the assessment, and the



## Approved Assessments

assessment process itself. Examples of evidence that the evaluation was seeking include:

- simulation evaluation results;
- cognitive laboratories reports;
- pilot study reports or pilot study evaluation reports;
- field test reports or field test evaluation reports;
- a well-articulated theory of action about how the assessment(s) are intended to work; and
- a stated assessment purpose and documentation of theoretical basis for assessment.

2. *Standards aligned.* The evaluation team assessed if there was evidence showing the degree of alignment of the assessment items to Colorado Academic Standards for reading. Examples of evidence that the evaluation was seeking include:

- independent (non-vendor) alignment study results;
- vendor-completed alignment study results; and
- a detailed description of the processes used for ensuring the alignment to standards.

3. *Validity.* The evaluation team assessed provided evidence to determine the extent to which the assessments are valid for their intended use. For this body of validity evidence, the evaluation was seeking:

- evidence that shows assessment of skills in one or more of the five areas of reading instruction (phonemic awareness, phonics, vocabulary development, reading fluency, and reading comprehension);
- content specifications for each grade level; evidence that the content of the assessment was developed and reviewed by experts, including teachers;



## Approved Assessments

- information on the students included in pilot and/or field testing to determine if they included English Learners and students with disabilities;
  - reported reading levels for passages, including how levels were established;
  - evidence that the internal structure of the assessment supports proposed score interpretations;
  - evidence that total test score and relevant sub-scores are related to external variables; evidence that the assessment(s) provide sub scores or information to identify specific reading deficiencies;
  - evidence to support classification procedures; and
  - evidence of classification accuracy analysis showing that the assessment appropriately identifies students' reading deficiencies.
4. *Reliability*. The evaluation team assessed provided evidence to determine the extent to which the assessments demonstrate reliability for the intended use. For this body of reliability evidence, the evaluation was seeking:
- evidence that shows assessment of skills in one or more of the five areas of reading instruction (phonemic awareness, phonics, vocabulary development, reading fluency, and reading comprehension);
  - a model-based approach to reliability reported for each grade or one type of reliability were reported that are appropriate for the purpose of the assessment;
  - evidence that for each type of reliability reported, the lower bound of the confidence interval around the median estimate met or exceeded 0.70;
  - that there are a sufficient number of alternate forms, and evidence is strong for comparability of alternate forms; and



## Approved Assessments

- SEM estimates are reported for score ranges and cut scores for each assessment.
5. *Diagnostic accuracy.* The evaluation team assessed provided evidence to determine the extent to which the diagnostic assessments demonstrate validity and reliability for the specific intended use. For this requirement, vendor-provided evidence was reviewed to determine the extent to which the assessment accurately identifies students' specific reading skill deficiencies including:
- evidence of construct validity, such as convergent and discriminant analyses, demonstrating correlations that are reasonable for the grade and skills assessed;
  - consequential validity evidence is provided and is connected to a well-articulated theory of action about how the assessment(s) are intended to work;
  - evidence of involvement of content experts in determining the score at which there is a high probability that a student does or does not require intervention;
  - evidence for classification accuracy for identifying students with a "significant reading deficiency," including evidence for adequate sensitivity and specificity;
  - evidence that experts in applicable content and the progression toward reading proficiency are significantly involved in the development process;
  - evidence that total test and relevant sub scores are related to external variables as expected;
  - evidence used to inform the setting of cut scores and a rationale for why certain forms of evidence are included and others are not; and
  - a clear description of the criterion or measure that was used to provide evidence for valid classifications.



## Approved Assessments

6. *Spanish language assessment.* The evaluation team assessed if a Spanish version of an assessment was offered and, if applicable, then the evidence that the assessments that were normed for the performance of students who speak Spanish as their native language. For this requirement, the evaluation sought:
  - evidence that the student sample utilized for norming the assessment is representative of students who speak Spanish as their native language;
  - evidence that experts in Spanish language and literacy were included in the development of items;
  - evidence from cognitive laboratories, pilot and/or field tests;
  - evidence of comparability of all forms for each grade level;
  - evidence that the reading constructs measured by the test are relevant to the target language;
  - evidence that assessment items have been reviewed to address cultural differences inherent to language and cultural stereotypes;
  - evidence that assessment items were back-translated to English by native Spanish speakers and reading content experts;
  - evidence that the translated test version does not privilege any dialect of the target language over others;
  - evidence of scaling is provided to ensure appropriate interpretability of scores across language versions of the test;
  - evidence that the form presentation is consistent with English version; and
  - evidence that cut points, confidence intervals, and indices of risk are consistent across both languages of the assessment.
7. *Paper and pencil diagnostic assessment.* The evaluation team assessed if an assessment was computer-based or non-computer-based (administer using paper and pencil) and, if applicable, that the various forms with



## Approved Assessments

demonstrated evidence of equivalence or comparability. For this requirement, vendor-provided evidence was reviewed for evidence of comparability of all forms for each grade level, using a representative sample of students.

### Description of Criteria Drawn from Additional Professional Standards

In addition to the above categories, the evaluation team included additional criteria to the evaluation and review rubric to provide supplemental information on the assessments. These additional criteria draw from established research on principles and characteristics for identifying high-quality assessments (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014; National Center on Intensive Intervention, n.d.). Exhibit 2.2 lists the criteria and includes sample evidence. The criteria are defined, and additional evidence is listed in the text that follows Exhibit 2.2.

Exhibit 2.2. Additional Criteria for Approved Assessments and Example Evidence			
Criterion Number	Short Name	Criterion	Example Evidence Reviewed
8	Removal of bias	The assessment development and review processes are designed and implemented to remove bias against all students	<ul style="list-style-type: none"> <li>Multiple-group confirmatory factor models for categorical item responses</li> <li>Bias reviews were conducted, and results were provided and include a representative panel composition</li> </ul>
9	Assessment administration guidance	The administration of the assessment is supported by appropriate guidance and resources	<ul style="list-style-type: none"> <li>An administration guide (or comparable set of resources) with a scripted administration protocol or guidelines for administration</li> <li>Standardized guidance scoring and interpreting scores</li> </ul>



## Approved Assessments

Exhibit 2.2. Additional Criteria for Approved Assessments and Example Evidence

Criterion Number	Short Name	Criterion	Example Evidence Reviewed
10	Appropriate accommodations	The assessment offers appropriate accommodations so all students can be fairly and accurately assessed	<ul style="list-style-type: none"><li>• Evidence that the assessment items and accessibility features permit all students to demonstrate their knowledge and abilities and do not contain features that unnecessarily prevent them from accessing the content of the item</li></ul>
11	Report usability	The assessment produces assessment data and information, such as student scores and score reports, that are usable for the intended audiences	<ul style="list-style-type: none"><li>• Evidence of user testing by a range of stakeholders is provided to demonstrate the utility of the reports for each intended audience</li><li>• evidence that reports provide a trajectory for student progress and that reports are designed for specific audiences, including districts, schools, classrooms, individual students, and parents/families</li></ul>
12	Data privacy	The assessment vendor has sufficient safeguards in place to protect student and teacher data privacy	<ul style="list-style-type: none"><li>• Evidence of how the district or school will receive all underlying data, in a timely and useable fashion, so that it can do further analysis as desired, including, for example, achievement, verification, forensic, and security analyses</li><li>• Evidence of how security safeguards have been tested and validated for computer-based tests and for paper-and-pencil tests, as relevant</li></ul>

8. *Removal of bias.* The evaluation team evaluated if there was evidence that the assessment was fair to all students and free of bias against all students. For this requirement, evidence was reviewed to determine the



## Approved Assessments

extent to which the development, review, and implementation processes were designed to eliminate bias against students. Evaluators reviewed for evidence such as one or more of the following types of analyses were conducted:

- multiple-group confirmatory factor models for categorical item responses;
- explanatory group models such as multiple-indicators, multiple-causes (MIMIC), or explanatory IRT with group decisions;
- Differential Item Functioning from Item Response Theory (DIF in IRT);
- that bias review(s) were conducted, and results were provided and include a representative panel composition; and
- evidence of the application of Universal Design for Learning principles throughout the assessment development process.

9. *Assessment administration guidance.* The administration of the assessment is supported by appropriate guidance and resources. For this requirement, submissions were reviewed for evidence of:

- an administration guide (or comparable set of resources) with a scripted administration protocol or guidelines for administration;
- a description of the appropriate testing environment; guidance on testing irregularities for before, during, and after the administration;
- standardized guidance scoring and interpreting scores;
- a description of how security safeguards have been tested and validated for computer-based tests and for paper-and-pencil tests (as relevant);
- evidence of average administration time is provided and is reasonable and balanced given the information provided by the assessment; and
- training for administration is offered, communicated, and readily available to teachers.





## Approved Assessments

10. *Appropriate accommodations.* The assessment offers appropriate accommodations so all students can be fairly and accurately assessed. The evaluation team reviewed submissions for:
- evidence that the assessment items and accessibility features permit all students to demonstrate their knowledge and abilities and do not contain features that unnecessarily prevent them from accessing the content of the item;
  - evidence that the assessment considers presentation, response, setting, and timing and scheduling; evidence that the assessment includes accommodations for students with disabilities;
  - information is provided to support the research and evidence base for allowable accommodations; evidence that potential accommodations do not compromise the interpretation or stated purpose of the test;
  - evidence that training materials include specific guidelines on the selection and implementation of any available administration or scoring accommodations; and
  - a description of the accessibility features that will be available is provided by the vendor.
11. *Report usability.* The assessment produces assessment data and information, such as student scores and score reports, that are usable for the intended audiences. The evaluation team assessed whether the reports generated from the assessment data provided useful information that supported instructional responses. The evaluation reviewed submissions for:
- evidence of user testing by a range of stakeholders is provided to demonstrate the utility of the reports for each intended audience;
  - training and examiner materials with clear instructions for the determination of whether the student potentially demonstrates a “significant reading deficiency”;



## Approved Assessments

- evidence is provided to demonstrate the utility of the reports for each intended audience; cut points, score ranges, and/or confidence intervals are clearly specified for specific age/grade ranges and administration windows; evidence that the estimated time for scoring is reasonable and balanced for the information provided;
  - evidence that reports provide a trajectory for student progress and that reports are designed for specific audiences, including districts, schools, classrooms, individual students, and parents/families;
  - evidence that reports are available in languages other than English;
  - scoring guidelines that are clear and easily interpreted; a description of the process and technology that will be used to issue reports in as timely a manner as possible is provided; and
  - a clear timeline is provided to show when assessment results will be available.
12. *Data privacy*. The assessment vendor has sufficient safeguards in place to protect student and teacher data privacy. For this requirement, the evaluation team reviewed submissions for:
- evidence of how the district or school will receive all underlying data, in a timely and useable fashion, so that it can do further analysis as desired, including, for example, achievement, verification, forensic, and security analyses;
  - evidence of how security safeguards have been tested and validated for computer-based tests and for paper-and-pencil tests, as relevant;
  - evidence of secure management of assessments and assessment data for all administration modes, so that no individual gains access to unauthorized information;
  - evidence of student privacy protection for all administration modes, reflecting compliance with all applicable federal and state laws and requirements.



### Data Collection and Methods

#### Information Used to Review Programs

We reviewed all materials submitted by the assessment vendor application in response to our request for evidence and any evidence submitted by the vendor prior to the conclusion of our review window.

#### Process Used for Collecting Information from Vendors

Each vendor that had an approved READ Act assessment was emailed by our internal vendor point of contact. The point of contact sent follow up emails and called nonresponsive vendors (see Appendix B.2). Of the 9 vendors that had approved assessments, all but 2 vendors responded with evidence for 15 assessments.<sup>8</sup>

#### Training for Independent Evaluators

The lead assessment evaluator conducted the training for the assessment reviewers. As part of the extended training, they reviewed the rubric, conducted a pilot review, calibrated on an initial sample, clarified definitions, documented decisions rules, and went through a formal training on process and recording scores.

#### Review of Rubric and Pilot Review

Upon approval of the rubric by CDE (see Appendix A.1), the assessment review team was assembled for an initial training session where they were walked through the entire rubric and allowed to ask clarifying questions. After all questions were answered, the assessment review team independently reviewed a common set of vendor evidence, applying the rubric as they went through the evidence. Assessment team staff were tasked with not only documenting their decisions based on the rubric, but also writing down notes where they encountered any questions, concerns, instances where they were not able to clearly apply the rubric criteria, or where evidence was provided that did not align

---

<sup>8</sup> The vendors for Terra Nova and Woodcock-Munoz LS did not submit information to the independent evaluation.



## Approved Assessments

with the rubric criteria. After all staff had completed their independent reviews, the assessment review team reconvened and compared and shared ratings and notes from their reviews.

### **Formal Reviewer Training**

After completion of the pilot review, the assessment review team was convened for formal training on the process that was used for the assessment review. The training included an overview of the materials that were to be used in the review (vendor evidence sets, rubric, decision rules, etc.), information on how to apply the rubric during reviews, how to record responses and notes, and specific review assignments.

### **Ratings and Resolution of Discrepancies in Ratings**

Each assessment was reviewed by at least two assessment reviewers. They met to resolve queries and identify and resolve any discrepancies in the ratings. After a review for an assessment was completed by both reviewers the ratings were compared and any differences were discussed and there was an attempt to reconcile and come to an agreement by the reviewers. If the reviewers had had different ratings for the same evidence a third rater reviewed the evidence set and made the ultimate determination of the rating. Reviewers were able to come to agreement for all assessment reviews so this step did not occur. Training leaders were available throughout the review process to support reviewers who had questions or concerns.

After raters had determined ratings for each criterion, criterion level ratings were aggregated into two summary ratings, one for compliance with SB 19-199 requirements and a second for additional professional criterion. The evaluation team applied the following decision-rules.

#### **Summary Rating: Compliance with all SB 19-199 Requirements**

- Fully meets: Received a rating of “Fully meets” on all indicators.



## Approved Assessments

- Largely meets: Received a rating of at least “Partially meets” on all indicators.
- Partially meets: Received a rating of “Does not meet” on at least one, but not all, indicators.
- Does not meet: Received a rating of “Does not meet” on all indicators.

Summary Rating: Compliance with Additional Technical and Quality

### Criteria

- Fully meets: Received a rating of “Fully meets” on all indicators.
- Largely meets: Received a rating of at least “Partially meets” on all indicators.
- Partially meets: Received a rating of “Does not meet” on at least one additional indicator.
- Does not meet: Received a rating of “Does not meet” on all indicators.

## Results and Discussion

The following summary table shows the number of assessments that fully, partially, and do not meet key READ Act Requirements and overall summary ratings.

Exhibit 2.3. Summary of Ratings for READ Act Requirements for Approved Assessments				
SB 19 – 199 Requirement	Rating			
	Fully Meets	Partially Meets	Does Not Meet	Not Applicable
Is evidence-based (22-7-1209 (2)(b)(I))	10	5	0	0
Is aligned with the preschool through elementary and secondary education standards for reading adopted by the state board pursuant to section 22-7-1005 (22-7-1209 (2) (a) (II) (A))	10	3	2	0



## Approved Assessments

Exhibit 2.3. Summary of Ratings for READ Act Requirements for Approved Assessments				
SB 19 – 199 Requirement	Rating			
Each of the recommended reading assessments is <b>valid</b> [and reliable] and proven to effectively and accurately measure students' reading skills in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and reading comprehension (22-7-1209 (2) (a) (II) (B))	13	2	0	0
Each of the recommended reading assessments is [valid and] <b>reliable</b> and proven to effectively and accurately measure students' reading skills in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and reading comprehension (22-7-1209 (2) (a) (II) (B))	14	1	0	0
Evidence that shows assessment of skills addressing the following targeted areas of scientifically based reading instruction	15	0	0	0
Each of the recommended reading diagnostics is proven to accurately identify students' specific reading skill deficiencies (22-7-1209 (2) (a) (II) (C))	4	2	0	9
At least one of the recommended reading assessments for kindergarten and first, second, and third grades is normed for the performance of students who speak Spanish as their native language, which assessment is available in both English and Spanish (22-7-1209 (2) (a) (II))	6	1	0	8
The list of recommended reading assessments and reading diagnostics includes at least one assessment and one diagnostic that a student can complete using paper and pencil rather than using a computer (22-7-1209 (2) (a) (II) (E))	7	0	0	8
	Fully	Largely	Partially	Does Not Meet
Summary Rating Compliance SB 19-199	6	7	2	0

Exhibit 2.4. Summary of Ratings for Additional Technical and Quality Criteria for Approved Assessments				
Additional Technical & Quality Criteria	Rating			
	Fully Meets	Partially Meets	Does Not Meet	Not Applicable



## Approved Assessments

Exhibit 2.4. Summary of Ratings for Additional Technical and Quality Criteria for Approved Assessments				
Additional Technical & Quality Criteria	Rating			
The assessment development and review processes are designed and implemented to remove bias against all students	9	4	2	0
The administration of the assessment is supported by appropriate guidance and resources	15	0	0	0
The assessment offers appropriate accommodations so all students can be fairly and accurately assessed	10	5	0	0
The assessment produces assessment data and information, such as student scores and score reports, that are usable for the intended audiences	14	1	0	0
The assessment vendor has sufficient safeguards in place to protect student and teacher data privacy	9	6	0	0
	Fully	Largely	Partially	Does Not Meet
Summary Rating: Additional Technical & Quality Criteria	4	9	2	0

### Discussion of cross cutting themes and issues that emerged from the assessment reviews

Overall, the approved assessments either fully or largely met the requirements outlined in the READ Act. Six of the 15 assessments reviewed fully met all READ Act requirements. However, for each required element, vendors provided evidence of varying levels of quality as detailed below. Only two assessments did not meet one requirement of the READ Act – alignment with Colorado’s state standards. Of the READ Act’s required elements for approved assessments, this was an area of weakness. Fewer approved assessments fully met additional technical and quality criteria. Only four fully met these criteria. Of the additional criteria, two assessments did not meet the bias removal criterion.

In the following text, we present lessons learned for each of the categories evaluated.



## Approved Assessments

1. *Evidence-based.*
  - a. During the evaluation of vendor-submitted evidence, it was clear that vendors approached the premise of “evidence-based or scientifically based” differently. Some vendors provided a brief statement that a study was conducted and offered limited reliability and/or validity data from said study. Other vendors provided a narrative about the theory of action behind the assessment supported by a wide range of supporting qualitative and quantitative data. These variations in responses may be explained by the availability of evidence that the vendors have with regard to their assessments.
  - b. Of all the READ Act criteria, this criterion is one of the strongest levers CDE has to differentiate between assessments submitted for READ Act approval, distinguishing between vendors that have continued to invest in their product through research and evaluation and those that have not.
2. *Standards aligned.*
  - a. Alignment evidence varied by vendor and assessment. Evidence ranged from highlighted copies of the Colorado Academic Standards to robust 3<sup>rd</sup>-party external evaluations. Alignment was shown at the item level and at the measure level. This result indicates that vendors approach their understanding of appropriate alignment evidence differently.
  - b. Some vendors were comfortable claiming alignment to the Common Core State Standards as being sufficient evidence for alignment to the Colorado Academic Standards without substantiating how the two sets of standards are comparable.
3. *Validity.*
  - a. Proving validity is an ongoing process, beginning at the initial conceptualization of the measure’s construct, continuing throughout the entire testing process, and extending into the interpretation of test





## Approved Assessments

scores. Thus, most of the vendors were able to produce a substantial amount of documentation related to the validity of their assessments. Much of the review time was spent reviewing this documentation and determining the quality of the validity evidence provided.

- b. Validity speaks to purpose of the assessment. The range of purposes of the assessments on the CDE-approved list varied. There were some diagnostic assessments that had a clear purpose (e.g., measuring student's understanding of vocabulary) while others claimed to accomplish multiple purposes (e.g., provide diagnostic data, benchmark data, and summative data). This variation in assessment purpose creates challenges in comparing assessments to one another as they are not an "apples to apples" comparison. This variation makes providing overall guidance around a set of assessments challenging as appropriate guidance may vary from assessment to assessment.

#### 4. *Reliability.*

- a. Reliability refers to the consistency of assessment results, or the degree to which student's results are the same when they take the same test on different occasions and when different but equivalent tests are taken at the same time or at different times. Vendors take different approaches to proving the reliability of their assessments depending on the type of assessment they have developed. Further, vendors are able to provide different levels of quality evidence, thus much of the review time was spent reviewing this documentation and determining the quality of the reliability evidence provided.
- b. Reliability is another strong lever CDE has to differentiate between assessments submitted for READ-Act approval. By increasing the rigor for the reliability evidence requirements CDE will ensure that only those assessments that are administered to students (at least within the same assessment) are being tested in an equitable manner.



## Approved Assessments

### 5. *Diagnostic accuracy.*

- a. Vendors were very clear as to which areas of reading their assessments covered and were able to readily produce a substantial amount of evidence regarding how they signaled to educators if a student has a Significant Reading Deficiency.
- b. The classification accuracy of the approved assessments varied. Classification accuracy indicates how well scores on a screening assessment correctly identify students at risk versus those not at risk. Assessments with strong classification accuracy maximize rates of true positive (correctly classifying a student as at risk) and true negative (correctly classifying a student as not at risk) classifications and they minimize rates of false positive (incorrectly classifying a student as at risk when he is not at risk) and false negative (failing to classify a student as at risk when he is at risk) classifications. For the READ Act, classification accuracy speaks to how accurately each assessment identifies students as having a serious reading deficiency. Given the number of assessments and the data provided by vendors to support their ability to identify students' specific reading skill deficiencies, it can be said that all the assessments can accurately identify students' specific reading skill deficiencies (in the areas the assessment addresses) but some do so more accurately than others.

### 6. *Spanish language assessment.*

- a. Of the 15 assessments reviewed, 7 included versions normed for the performance of students who speak Spanish. Three of these were diagnostic assessments and 4 were interim assessments.
- b. There were 12 evidence areas that we reviewed for within this criterion and all of the reviewed assessments performed fairly well, meeting at least 7. All vendors used experts in Spanish language and literacy in the development of items, but only one provided evidence that the translated test version does not privilege any specific dialect.



## Approved Assessments

- c. Vendors have utilized different approaches to the development of their assessments in Spanish (e.g., translation, translation and adaptation, etc.) and the rigor of the review of the assessments after development.
7. *Paper and pencil diagnostic assessment.*
    - a. Of the 15 assessments, 7 (4 diagnostic assessments and 3 interim assessments) offered paper and pencil options for administration. Of these 7 assessments, 4 were only paper and pencil, not offering a computer-based administration option.
  8. *Bias removal.*
    - a. Across most assessments, there was an overall lack of evidence to show that the assessments were free of bias, as determined by non-vendor staff. Most vendors demonstrated that their assessments were free of bias through internal reviews.
    - b. There was also a lack of evidence to show that the assessments were constructed in a way that all students were considered during the development process (Universal Design considerations).
  9. *Assessment administration.*
    - a. The approved assessments are all supported by administration information that guides the administrator through the assessment process. The clarity and ease of use of these materials varied by assessment.
    - b. Most vendors offer training on their assessments, though this is usually at an additional cost.
  10. *Appropriate accommodations.*
    - a. This was an area of weakness for vendor evidence. Summative and interim assessments from those vendors with a sizeable presence in large-scale testing were more likely to include evidence regarding accommodations. While the case may be made that some assessments are not able to allow for many accommodations due to



## Approved Assessments

the design of the assessment, there was a noticeable lack of content about accommodations and accessibility.

### 11. *Report usability.*

- a. Vendors produce a good number of reports, most for multiple audiences, and these reports vary in their presentation and language across assessments.
- b. No vendors offered reports in languages other than English.
- c. Few vendors provided evidence that spoke to engaging in activities to ensure that the reports they were producing were useful to stakeholders (e.g., user testing, cognitive labs).

### 12. *Data privacy.*

- a. Most assessment vendors provided adequate evidence that they complied with all expected data privacy requirements. Few communicated how districts and schools could obtain data for further analyses.

## **CDE's Processes for Identifying Items for Approved and Advisory List**

Our understanding of the process CDE used to inform the selection of the approved assessments during 2019-2020 reflected the READ Act components and intent. Most, if not all, of the CDE staff involved had just joined the department and there was already an effort to solicit bids from outside vendors to complete a review of the assessments and make recommendations. The process was transparent, and though we did not have the opportunity to review the RFP, our understanding is that it contained the expectations for both the qualifications of the vendor and the minimum criteria for the review of assessments. Once a vendor was secured through the competitive bidding process, the vendor, a team from the University of Houston led by Dr. Jeremy Miciak, developed the specific criteria that would be used to review the assessments.



## Approved Assessments

These criteria were shared with the public for general feedback and were then reviewed by a group of Colorado stakeholders during a convening hosted by CDE and facilitated by the assessment vendor. These stakeholders went through an application process run by CDE and selections were made to ensure representation in terms of state geography and experience with various student populations (e.g., English Learners, students with disabilities, Title 1 schools, turnaround and priority schools). The vendor incorporated the collected feedback from stakeholders and incorporated the feedback to the extent possible to finalize the review rubric. Information on the feedback that was collected and the process for incorporating feedback into the review rubric were not available for our review. CDE staff report that the majority of the feedback received from the field and the stakeholder meetings was focused on useability and bias/fairness.

Only new assessments (newly submitted assessments or new editions of approved assessments) were targeted for review; assessments that were already on the approved list were not reviewed. Assessment evidence for this review was collected by CDE through a request for information (RFI) process that followed Colorado's procurement process. Vendor submissions were submitted to the University of Houston, who then used them for their review. CDE was not part of the assessment review process outside of the elements focused on collecting stakeholder feedback.

The independent evaluation and CDE's evaluation had very similar approaches. Both processes developed criteria using foundational requirements from policy and/or CDE, utilized an independent entity to further develop review criteria, and included steps in the process for review and feedback from individuals outside of the vendor's organization. In our case, we used independent literacy and assessment experts and scholars as outside reviewers rather than Colorado stakeholders and input from the field. In both cases, the outside reviewers provided most of their feedback on the criterion focused on bias, useability, and support for English Learners and students who may require accommodations. Due to this, in addition to criteria represented by the READ Act



## Approved Assessments

---

requirements, our evaluation included criteria that reflects the aforementioned interests of the field and stakeholders and our outside reviewers. These criteria, more focused on the non-technical aspects of assessment, included a focus on bias/fairness, Spanish version development, administration support, accommodations, and useability of data.



# 3

## Instructional Programs on the Advisory List

The review of instructional programming considered three types of materials: core programs, supplemental programs, and intervention programs. Core programs are those that are used in general instruction. Supplemental programs are used in classrooms where more support beyond the core program is needed to supplement reading instruction. Intervention programs are used to support individual students who need intervention support for their reading development.

### Key findings

- **Nearly all of the approved instructional materials (core, supplemental, and intervention programs in English and Spanish) either fully meet or largely meet SB-19-199 requirements.**
- **Just over 10% of the programs fully met the criterion of being “evidence-based” or “scientifically-based.”**
- **The majority of core academic programs include academic content as required, and 87 percent of programs fully met this criterion.**
- **All but one of the instructional programs embedded some form of assessment.**



### Types of Instructional Programming Materials

The review of instructional programming considered three types of materials: core programs, supplemental programs, and intervention programs. *Core programs* are those that are used in general instruction and must target all five areas of scientifically based reading: phonemic awareness, phonics, reading fluency, vocabulary development, and reading comprehension. *Supplemental programs* are used in classrooms where more support beyond the core program is needed to supplement reading instruction. *Intervention programs* are used to support individual students who need intervention support for their reading development. Supplemental and intervention programs were subject to fewer review criteria, under the assumption that all students have access to a core instructional program.

### Description of Review Categories

The instructional materials review followed a rubric (see Appendix A.2) consisting of four main categories: (1) whether the program is evidence-based; (2) whether the program provides explicit and systematic skill development in the areas of phonemic awareness, phonics, vocabulary development, reading fluency, and reading comprehension, and is aligned with preschool through elementary and secondary state standards for reading adopted by the State Board; (3) whether the program includes texts on core academic content to assist the student in maintaining or meeting grade-appropriate proficiency levels in academic subjects in addition to reading; and (4) whether the program includes evidence-based or scientifically based, valid, and reliable assessments. The following describes each of the four areas with their criteria:





1. *Is evidence-based (22-7-1209 (2) (b) (I)).*

Vendors were invited to submit up to three research studies or a logic model or theory of action as evidence. The independent evaluators then evaluated the evidence provided using the ESSA evidence levels. As independent evaluators read through each study, they documented key findings, effect sizes, effects on students with reading deficiencies, and effects on ELs.

Independent evaluators also reviewed study designs for sample attrition, bias reduction, and baseline equivalence; the What Works Clearinghouse (WWC) reviews include these as important study design characteristics. The reviews presented here, then, approximate but are not as in-depth as WWC reviews.

In the instances where programs had been reviewed by the What Works Clearinghouse – a leading federal source of evidence-based information about education – the outcome domains and associated effectiveness ratings were recorded. After reviewing the available evidence, researchers assigned each program an evidence rating ranging from 1 to 4. Evidence ratings were guided by the ESSA levels of Evidence (see Exhibit 3.1). An ESSA level 1 or 2 earned a rating of fully meets, and an ESSA level of 3 or 4 earned a rating of partially meets. If a program could not demonstrate an ESSA evidence level of 4, then it would fail to meet. For the summary rating, a program could fully meet if it partially met on this indicator.

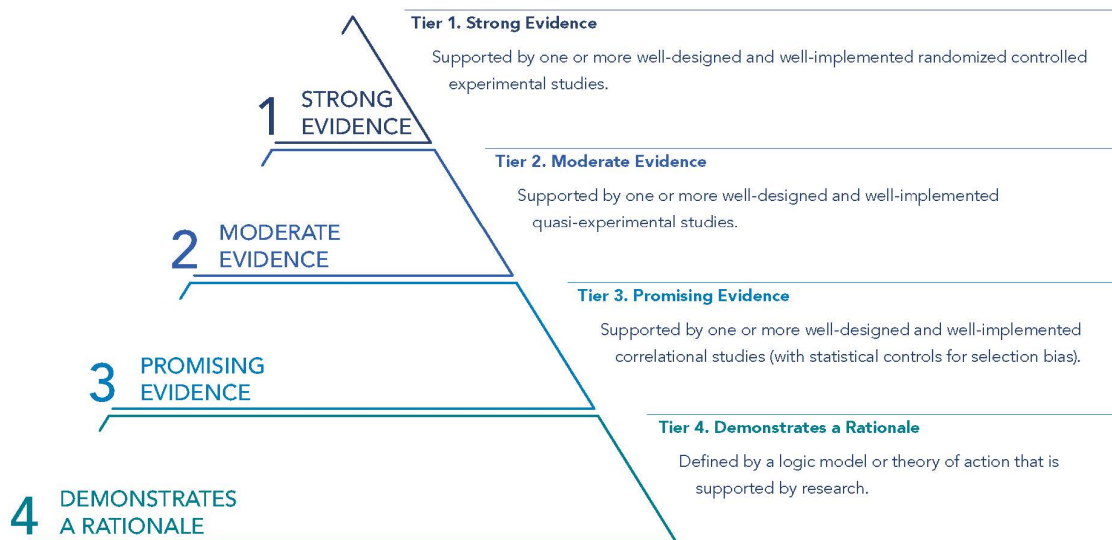
### ESSA Evidence Levels

The Every Student Succeeds Act (ESSA; 2015) establishes a four-tiered method of evaluating evidence. This framework is designed to ensure that states, districts, and schools can identify programs that work. Stronger research methods provide stronger evidence for a program, resulting in higher tiers of ESSA evidence levels. When a program has a higher tier rating, we can be more confident that it works. See Exhibit 3.1 below.



### Exhibit 3.1. ESSA Four Tiers of Evidence

The Every Student Succeeds Act (ESSA; 2015) establishes a four-tiered method of evaluating evidence. This framework is designed to ensure that states, districts, and schools can identify programs that work. Stronger research methods provide stronger evidence for a program, resulting in higher tiers of ESSA evidence levels. When a program has a higher tier rating, we can be more confident that it works.



- Provides explicit and systematic skill development in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and comprehension (22-7-1209 (2) (b) (II)), and is aligned with the preschool through elementary and secondary state standards for reading adopted by the State Board (22-7-1209 (2) (b) (II.5)). We evaluated whether skill development across reading areas were present, explicit, and systematic, using vendor-supplied information and EdReports, when available. Core programs were evaluated for all five areas of reading (phonemic awareness, phonics, reading fluency,*

#### Two Types of Reading Comprehension

Close reading: *Approach to comprehension focused on the text itself.*

Interactive reading: *Approach to comprehension focused on text and outside information related to text content.*



vocabulary development, and reading comprehension), whereas supplemental and intervention programs were evaluated only for the areas that vendors claimed to specifically target. Reading comprehension was evaluated along two dimensions in order to ensure that programs were compliant with both the READ Act and the Colorado State Standards. The two dimensions were close reading, which is an approach to comprehension focused on the text itself, and interactive reading, which is an approach to comprehension focused on text and outside information related to text content. The former is highly emphasized in the READ Act minimum competencies, whereas both are emphasized in the state standards.

A core program received a rating of fully meets if all elements fully met (i.e., demonstrated that they were explicitly and systematically taught). A program received a rating of partially meets if all elements at least partially met (meaning the element was present but did not suggest that it was presented both explicitly or systematically). Note, though, that we considered comprehension partially met if at least one of the two elements partially met.<sup>9</sup> Finally, a core program received a rating of does not meet if at least one of the elements of reading would have to fail to meet.<sup>10</sup> These decision rules were the same for supplemental and intervention programs, except that these programs were not required to address all the elements of reading. These decision rules applied only to those elements that the vendor claimed to address.

3. Includes texts on core academic content to assist the student in maintaining or meeting grade-appropriate proficiency levels in academic subjects in addition to reading (22-7-1209 (2) (b) (V)). The evaluators operationalized this requirement into four criteria: (a) grade-appropriate

---

<sup>9</sup> For example, if close reading partially met and interactive reading failed to meet while all other elements at least partially met, then this indicator would still fully meet.

<sup>10</sup> For reading comprehension, both elements would have to fail to meet.



text complexity, (b) a range of content areas (e.g., history, science) and genres (e.g., fiction, nonfiction), (c) support for students with disabilities, and (d) support for English Learners. The evaluators used vendor-supplied documentation and relevant EdReports indicators when available. The last two criteria were considered to ascertain grade-level access for students with disabilities and English Learners. These two criteria were not considered when determining summary ratings, as those were based on professional standards and not explicitly named in the READ Act. Supports for English Learners is the focus of Chapter 5. To fully meet the criteria of this indicator, both (a) and (b) had to fully meet. If both (a) and (b) partially met or if one fully met and the other did not meet, then this indicator would partially meet. If neither (a) nor (b) fully met and at least (a) or (b) did not meet, then this indicator would receive a rating of does not meet.

4. Includes evidence-based or scientifically based, valid, and reliable assessments (22-7-1209 (2) (b) (IV)). In place of a rating, the evaluators provided key information about embedded assessments. The key information listed whether assessments serve the purposes of providing formative, summative, or other information. We also noted whether assessments address the targeted areas of scientifically based reading instruction (i.e., phonemic awareness, phonics, reading fluency, vocabulary development, and reading comprehension). When available, we also summarized information regarding reliability and validity of embedded assessments.

Finally, the following decision rules were used to create an overall instructional program rating:

- Fully meets: Received a rating of at least “Partially meets” on the evidence-based indicator and received a rating of “Fully meets” on all other indicators.



## Instructional Programs on the Advisory List

- Largely meets: Received a rating of at least “Partially meets” on all indicators.
- Partially meets: Received a rating of “Does not meet” on at least one but not all indicators.
- Does not meet: Received a rating of “Does not meet” on all indicators.

### Rationale for additional professional standards

CDE asked the evaluation team to consider broadly accepted professional standards in addition to the READ Act requirements when reviewing materials. The professional standards we evaluated were supports for students with disabilities and supports for English Learners, described above and in more detail in Chapter 5. We considered supports for these two groups to be of particular importance given the prevalence of students who are English Learners (11.63%) and who are classified with a disability under IDEA in Colorado (14.69%).<sup>11</sup>

## Data Collection and Methods

### Description of Information Used to Review Instructional Programs

Information used to review programs include the following: For a program’s evidence base, we requested that vendors submit up to three research articles or reports. In case those were unavailable, we also requested a logic model or theoretical rationale for the program. To supplement this vendor-supplied information, we checked to see whether a *What Works Clearinghouse* review had been conducted. For information regarding program content, vendors were asked to answer specific questions and to provide examples that supported their response. When available, we

#### Why EdReports?

These high-quality independent reviews provide insight into the quality and complexity of texts included in the curriculum. They also consider whether curriculum-embedded tasks support grade-

also considered specific *EdReports* indicators that aligned with our criteria. If an *EdReports* review was available for a program,

#### Why What Works Clearinghouse?

The What Works Clearinghouse is an investment of the Institute of Education Sciences (IES) within the U.S. Department of Education. It reviews research on different programs, with the goal of providing educators with the information they need to make evidence-based decisions. It focuses on the results from high-quality research to answer the question “What works in education?” When they are available, we use these high-quality reports to supplement our own investigation of the evidence supporting the reading

<sup>11</sup> <https://www.cde.state.co.us/cdereval/2019-2020pupilmembership>



we first reviewed the *EdReports* ratings. If a program fully met one of our criteria based on the *EdReports* rating, then the vendor response and examples for that criterion were not reviewed. In instances where a criterion was not fully met based on an *EdReports* rating, then the vendor response and examples were reviewed and the rating was amended, if appropriate. *EdReports* ratings were only available for six of the reviewed programs.

### Process Used for Collection Information from Vendors

In order to gather information from vendors, vendors were contacted via an email that explained the purpose of the external evaluation. The email provided vendors with a form to complete with questions about specific aspects of their program. The form also solicited examples. (See Appendix B.3 for the Instructional Programming Vendor Request Form.) All but one vendor<sup>12</sup> responded with the requested information.

### Training for Independent Evaluators

The evaluation team created a rubric that specified criteria and rating options for each of the four categories described above. The rubric was developed in partnership with our external expert advisory panel and was approved by CDE (see Appendix A.2).

Evaluators were members of the advisory panel and WestEd staff with expertise in research, curriculum and/or English Learners. Evaluator training entailed three virtual sessions, during which evaluators learned about the CO READ Act, the review criteria, and how to use the rubric. As part of the training, evaluators asynchronously reviewed two programs, the final ratings for which were discussed during the training sessions.

### Ratings and Resolution of Discrepancies in Ratings

Teams of evaluators reviewed 69 CDE-approved programs. Reviewers with expertise in research were responsible for rating the evidence base, reviewers with expertise in curricular supports for English Learners were responsible for rating supports for English Learners, and reviewers with expertise in curriculum more broadly were responsible for rating the remainder of the criteria. Ten percent of the programs were reviewed by two sets of reviewers who came together to address any discrepancies and submit a final rating. In instances where reviewers were uncertain or could not resolve

---

<sup>12</sup> Amplify



discrepancies, an additional review was requested and performed by a member of the team who developed the rubric. This occurred for ratings in 5 programs.

### Results and Discussion

Nearly all of the approved instructional materials (core, supplemental, and intervention programs in English and Spanish) either fully meet or largely meet SB-19-199 requirements (Exhibit 3.2). Below, we describe the results for the four review categories.

Exhibit 3.2 – Summary Rating			
Summary Rating: Compliance with SB 19-199 requirements	Rating		
	Fully meets	Largely meets	Partially meets
Core programs in English*	10	1	0
Supplemental programs in English	18	7	2
Intervention programs in English	16	9	0
Programs in Spanish (all)	2	4	0
OVERALL	46	21	2

\*One core program did not receive a summary rating because it did not submit evidence to the external evaluators.

#### Is Evidence-based or Scientifically Based (22-7-1209 (2)(a)(II)(A))

While all instructional programs met the minimum standard for being evidence-based or scientifically based, just over 10% of the programs fully met this criterion (see Exhibit 3.3). The programs that fully met the evidence-based standard demonstrated impact on students' reading outcomes using rigorous research designs.

Exhibit 3.3 – Evidence-Based Programs			
Is evidence-based (22-7-1209 (2) (b) (I))	Rating		
	Fully meets	Partially meets	Does not meet
Core programs in English	4	8	0
Supplemental programs in English	1	25	0
Intervention programs in English	2	23	0
Programs in Spanish (all)	0	6	0
OVERALL	7	62	0



## Instructional Programs on the Advisory List

### **Provides Explicit and Systematic Skill Development in the Elements of Scientifically Based Reading Instruction (22-7-1209 (2) (b) (II)) and is Aligned with the Preschool Through Elementary and Secondary State Standards for Reading Adopted by the State Board (22-7-1209 (2) (b) (II.5)).**

Overall, the materials on CDE’s advisory lists for core, supplemental, and intervention programs offer explicit and systematic instruction in the elements of scientifically based reading instruction (Exhibit 3.4). Notably, 11 of the 12 English-language, core programs and two of three Spanish-language core programs fully met the independent evaluation’s criteria on all elements of scientifically based reading instruction. This is important because core programs are used to provide instruction to all students, including those who struggle with reading. The most common reason for ratings of partially meet was that the evidence provided by vendors to the independent evaluation did not fully demonstrate how the skill was systematically taught over the course of an entire school year. Vendors were asked to provide examples of systematic instruction. Vendors who received a rating of partially meet did not provide clear, sufficient evidence to demonstrate that the program had systematic instruction continued over the year.

Exhibit 3.4 – Skill Development			
Provides explicit and systematic skill development in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and comprehension (22-7-1209 (2) (b) (II)), and is aligned with the preschool through elementary and secondary state standards for reading adopted by the State Board (22-7-1209 (2) (b) (II.5))	Rating		
	Fully meets	Partially meets	Does not meet
Core programs in English	11	1	0
Supplemental programs in English	19	6	1
Intervention programs in English	16	9	0
Programs in Spanish (all)	3	1	2
<b>OVERALL</b>	<b>49</b>	<b>17</b>	<b>3</b>





## Instructional Programs on the Advisory List

Across all of 69 instructional programs, the independent evaluation found only three programs that had one or more elements of scientifically based reading instruction that did not meet the independent evaluation’s criteria for explicit and systematic instruction (Exhibit 3.5). The Writing Road to Reading, an English-language program approved for core, supplemental, and intervention use, did not meet the interactive reading dimension of reading comprehension. The differences in evaluation about reading comprehension stem from differences in CDE’s and the independent evaluation’s review criteria. CDE’s criteria focus on close reading of text, while the independent evaluation focuses on both close reading and interactive reading, which we define as making connections beyond a specific text. The evidence submitted to the independent evaluator did not show how instruction supported students in making connections beyond the text. Istation Espanol Lectura Temprana, a Spanish-language program on both CDE’s supplemental and intervention advisory lists, did not meet the independent evaluation’s criteria for phonemic awareness and the interactive reading dimension of reading comprehension. It is possible that the discrepancies in ratings stem from differences in the evidence base reviewed. CDE’s team had access to the full content of the program, whereas the independent evaluation team relied on a supplemental survey provided by the vendor.

**Exhibit 3.5. Numbers of Programs that Fully, Partially or Do Not Meet for Each Component of Scientifically Based Reading Instruction, by Program Type**

Component of Scientifically Based Reading	Program Type	Fully Meets	Partially Meets	Does Not Meet
Phonemic Awareness	Core - English	12	0	0
	Supplemental - English	20	2	0
	Intervention - English	20	2	0
	All - Spanish	3	0	2
Phonics	Core - English	12	0	0
	Supplemental - English	22	0	0



## Instructional Programs on the Advisory List

Exhibit 3.5. Numbers of Programs that Fully, Partially or Do Not Meet for Each Component of Scientifically Based Reading Instruction, by Program Type

Component of Scientifically Based Reading	Program Type	Fully Meets	Partially Meets	Does Not Meet
	Intervention - English	21	2	0
	All - Spanish	3	3	0
Fluency	Core - English	12	0	0
	Supplemental - English	15	3	1
	Intervention - English	14	4	0
	All - Spanish	5	1	0
Vocabulary	Core - English	12	0	0
	Supplemental - English	14	2	0
	Intervention - English	15	1	0
	All - Spanish	3	3	0
Reading Comprehension: Close Reading	Core - English	11	1	0
	Supplemental - English	11	4	0
	Intervention - English	11	4	0
	All - Spanish	6	0	0
Reading Comprehension: Interactive Reading	Core - English	11	0	1
	Supplemental - English	10	4	1
	Intervention - English	7	5	4
	All - Spanish	4	0	2

**Includes Texts on Core Academic Content to Assist the Student in Maintaining or Meeting Grade-appropriate Proficiency Levels in Academic Subjects in Addition to Reading (22-7-1209 (2) (b) (V)).**

The majority of core academic programs include academic content as required (Exhibit 3.6); 87 percent of programs fully met this criterion.



## Instructional Programs on the Advisory List

Exhibit 3.6 – Core Academic			
Includes texts on core academic content to assist the student in maintaining or meeting grade-appropriate proficiency levels in academic subjects in addition to reading (22-7-1209 (2) (b) (V))	Rating		
	Fully meets	Partially meets	Does not meet
Core programs in English	11	1	0
Core programs in Spanish	2	1	0
OVERALL	13	2	0

*Text complexity and quality.* This element was only evaluated for core programs, as our aim was to assess grade appropriateness of texts, and the evaluation team felt that an appropriate supplemental or intervention text would likely not be at grade level. All core instructional programs, in both English and Spanish, included texts written at grade level as evidenced by vendor-supplied quantitative and qualitative measures of text complexity (Exhibit 3.7). One core Spanish program was rated as “Partially met” due to a lack of qualitative evidence of grade-appropriate text complexity<sup>13</sup>. Having opportunities to read appropriately complex text is a pre-requisite for maintaining or meeting grade level proficiency standards. If students only access texts at easier proficiency levels, it would be impossible for them to meet grade level standards.

Exhibit 3.7. Number of programs that fully, partially, or do not meet for text complexity and quality, by program type			
Program Type	Rating		
	Fully Meets	Partially Meets	Does Not Meet
Core - English	11	0	0
Core - Spanish	2	1	0

*Range of texts.* The range of texts element was also only evaluated for core programs. Supplemental and intervention programs are intended to complement the use of a core program, and appropriate supplemental and

<sup>13</sup> McGraw Hill: Maravillas (2020)



intervention programs are likely to have a limited range of genres and content. To determine whether core programs included a sufficient range of texts, we considered the range of content as well as the genres included in the programs. All 11 English-language and all three Spanish-language core programs fully met expectations for the range of genres and content. Engaging students with such a range is important in building a foundation for comprehending, interpreting, and using a range of texts as students move into upper elementary where the focus of reading shifts to reading to learn.

*Cultural Representativeness.* Texts that are authored by individuals and with characters who reflect the ethnicities and cultures of students positively contribute to child development (Hughes-Hassell & Cox, 2010). Children interact with text features while reading, and when books are diverse, they engage in acceptance, exposure to different cultures, and challenge their worldviews (Okoye-Johnson, 2011; Shachar, 2012). There have been movements in children’s literature to increase diverse authorship, character, and story, but overall, most children primarily interact with texts that reflect a monolithic view of society and humanity, namely Eurocentric and middle class. To understand how culturally representative texts were, vendors were also asked to describe how their program is culturally representative, according to the following definition: (1) Program includes texts that offer (respectful) representations of people from cultural, socio-economic, and linguistic groups represented in Colorado schools (or, at least, some forms of cultural and linguistic diversity); and (2) Program includes texts that are written and illustrated by individuals who share the aspects of diversity represented in the texts. While 80% of vendors who provided a response in this section verbalized a commitment to diversity, only 43% provided evidence that the characters in their programs represent diverse populations and 3% of vendors provided quantitative evidence of diverse representation within their programs. Furthermore, the majority of vendors (56%) did not answer the question of diverse authorship of texts within their programs. Because of the limited information provided by vendors, evaluators did not a



sufficient empirical basis for rating whether programs used culturally representative texts.

Vendors' responses are reflective of the children's book industry as a whole. Children's books lack a range of diversity in book characters and book authors. Character diversity has been shown as the percentage of books with main characters from race/ethnicities as well as character representation in any roles. In a study of main characters in 455 picture books published in 2012, main characters were 75% white, 39 percent Black, 8 percent Asian, 7 percent Latino, and 2% either Native American or Middle Eastern (Koss, 2015). A review of overall character diversity in 3,682 children's books that the Cooperative Children's Book Center received in 2018 showed that minoritized children were represented in under 10% of picture books whereas white children were shown in 50% of picture books and animals in 27 percent (*Data on books about Black, Indigenous and People of Color published for children and teens*, 2018).

*Differentiation for students with disabilities.* All programs, regardless of program type, were analyzed to determine whether they provide differentiated instructional supports for students with disabilities, which the independent evaluation team defined as students with specific reading disabilities. To fully meet, a program had to demonstrate that supports were differentiated for students with specific reading disabilities. Only seven supplemental and three intervention programs fully met the evaluation's criteria (Exhibit 3.8). The differences between the evaluation's findings and the Advisory list result from differently worded and applied criteria. CDE defined differentiation as "linking assessment data with flexible grouping based on students' needs and progress", whereas the evaluation defines differentiation as the presence of differentiated instructional strategies specific to the needs of students with disabilities. CDE's definition is both reasonable and consistent with the overall data-based approach of the READ Act. One of CDE's requirements for intervention programs is that they link instruction to assessment for each component of scientifically based reading instruction. Although the vendors' submissions to the independent



## Instructional Programs on the Advisory List

evaluation team did not explicate how instruction would be differentiated for students with specific reading disabilities, individualized instruction based on formative assessments has the potential to address students' needs.

Exhibit 3.8. Number of programs that fully, partially, or do not meet for differentiation for students with disabilities, by program type			
Program Type	Rating		
	Fully Meets	Partially Meets	Does Not Meet
Core - English	5	6	1
Supplemental - English	7	15	4
Intervention - English	3	19	3
All - Spanish	2	2	2

### Includes Evidence-based or Scientifically Based and Reliable Assessments [22-7-1209 (2) (b) (V)]

All but one<sup>14</sup> of the instructional programs embedded some form of assessment. All but two of the instructional programs on CDE's advisory list that submitted information to the evaluation team include embedded formative assessments. Forty-six programs include summative assessments, while 23 programs include assessments for other purposes including placement in reading groups, recall, or informal purposes.

*Constructs addressed.* The core instructional programs on CDE's Advisory list for the most part address all five components of scientifically based reading instruction. Of the 11 English-language core programs which responded to the evaluation team's request for data, 9 offer embedded assessments on all five components of scientifically based reading instruction. The remaining two English-language core programs assess three components (one assesses phonics, vocabulary, and comprehension, while the other assesses phonics fluency, and reading comprehension). The three Spanish-language programs offer embedded assessments of all grade-level relevant components of scientifically based reading instruction.

<sup>14</sup> The 95% Group: *Vocabulary Surge: Unleashing the Power of Word Parts (Level A & B)*



*Reliability and validity evidence.* While all but one<sup>15</sup> of the instructional programs embedded some form of assessment, only four programs (some of which are categorized as both supplemental and intervention programs) submitted formal evidence of validity and reliability. (Many vendors noted that they did not have such evidence, while others submitted informal information about how they administer embedded assessments or feedback about their usefulness from teachers.) None of the core instructional programs submitted formal validity or reliability evidence. Lexia Learning System's *Core 5 Reading*, a supplemental and intervention program, submitted evidence of its assessment's alignment with several commonly used standardized measures of reading, including information related to the predictive validity of its assessments. Gander Publishing's *Seeing Stars*, a supplemental program submitted reliability information. Mind Play's *Virtual Reading Coach*, an intervention program submitted findings from a reliability and validity study. Curriculum Associates submitted the type of technical evidence for *iReady* that would qualify for a full assessment review, including information about test-retest reliability, standard error measurement, and the like. This vendor's assessments, which are on Colorado's approved list, were reviewed by the assessment team and fully meet all of the READ Act's requirements for approved assessments and largely meet the evaluation team's additional requirements. The types of assessments embedded in curriculum, especially core curriculum, are typically designed to guide in-the-moment instructional decision-making not to provide formal assessment determinations.

### **CDE Processes for Identifying Items for Approved and Advisory Lists**

The process CDE used to review instructional programs during 2019-2020 reflected the READ Act components and intent. CDE hired Dr. Stephanie Stoller, an educational consultant in the early literacy sphere and former Vice President

---

<sup>15</sup> The 95% Group: *Vocabulary Surge: Unleashing the Power of Word Parts (Level A & B)*



## Instructional Programs on the Advisory List

for Professional Learning at Acadience Learning Inc., to assist in developing an evidence-based rubric for evaluating instructional programs. Dr. Stoller has expertise on the mechanics of early literacy development and is deeply familiar with research on early reading intervention and success. Nearly 100 empirical studies, reports, and scientific articles were referenced in the rubric design process. The resulting rubric was comprised of elements that research has shown are central to learning to read. Additionally, best practices were derived from rubrics used by other states in successful material vetting processes.

The review process consisted of two phases. The first phase used a rubric primarily focused on how programs aligned with the science of reading; how instruction was explicit, sequential, systematic, and cumulative; and how the programs were supported by research. Programs that were reviewed favorably on the first phase of the rubric were invited to submit additional information for the second phase, which included individual academic components by grade level, closely examining all five components of scientifically based reading instruction (phonemic awareness, phonics, reading fluency, vocabulary, and reading comprehension). The rubrics were made publicly available, and CDE hosted an accompanying webinar and solicited public feedback. The contractor adjusted the rubrics based on the feedback received, and the CDE team provided final approval.

Reviewers were selected via a competitive application process. Selected reviewers were teachers representing districts that ranged in size and urbanicity. Many of the reviewers were instructional coaches or special education teachers. Some were bilingual in Spanish. Reviewers were not paid, and they were required to sign a conflict-of-interest statement.

CDE staff cast a wide net to reach instructional programming vendors. They contacted all vendors on the existing approved list, posted the information on their website, discussed the process during several monthly READ Act webinars, and encouraged districts to reach out to any vendors they wanted to





## Instructional Programs on the Advisory List

be considered. CDE also provided a technical assistance webinar for vendors to explain the process and answer questions.

Of 123 instructional programs reviewed, 89 passed Phase I of the rubric. Of those, 69 passed Phase 2 of the rubric. In total, 55% of reviewed programs were ultimately approved or partially approved. Exhibit 3.9 shows the number of each type of program that passed each phase of the rubric.

Exhibit 3.9. Number of programs that passed each phase of the rubric, by program type						
Program Type	Phase 1			Phase 2		
	Pass	No pass	Pass rate	Pass	No pass	Pass rate
Core	21	4	84%	15	6	71%
Supplemental	31	21	60%	25	6	81%
Intervention	37	9	80%	27	10	73%
Spanish*	5	1	83%	4	1	80%

\*Of the six Spanish programs that were reviewed, three were core programs, two were supplemental programs, and one was an intervention program.

In our professional judgment the instructional program review process reflects the goals of READ Act legislation. Both phases reflect an emphasis on the components of scientifically based reading instruction. The Phase 2 rubric criteria for each grade level are rooted in evidence, are clearly specified, and reflect both historical and current understandings of how the science of reading can be applied effectively in classroom practice. The review was executed in a thoughtful, systematic way that produced consistent ratings and allowed program vendors to appeal and clarify program content and approaches as needed.

The independent evaluation and CDE’s evaluation had very similar approaches. Both processes developed criteria using foundational requirements from policy and included steps in the process for review and feedback from individuals outside of the vendor’s organization. The independent evaluation solicited feedback on its rubric from an expert advisory panel of independent literacy experts and scholars, who also supported the review process.

Some operational criteria differed between CDE’s and the independent evaluation. Our expert advisory panel emphasized support for English Learners and students with disabilities and the need to investigate whether programs



## Instructional Programs on the Advisory List

---

demonstrated appropriate cultural representativeness. Due to this, our evaluation also included criteria that reflects the aforementioned interests of the field. In addition, we set a very high bar for research evidence to fully meet, to support stakeholders in understanding the range of evidence across programs.



# 4

## Professional Development Programs on the Advisory List

The professional development (PD) program review focused on four categories: whether the PD program (1) is evidence-based; (2) provides for explicit and systematic skill development in the five reading elements; (3) includes rigorous evaluations of teacher learning throughout and at the end of the course; and (4) has support for English Learners.

### Key findings

- Overall, all professional development programs fully met SB-19-199 requirements.
- One professional development program fully met the evidence base requirement.
- All six professional development programs the team reviewed met requirements around **explicit and systematic skill development and including evaluations of learning.**



### Description of Review Categories

The professional development (PD) program review focused on four categories: whether the PD program (1) is evidence-based; (2) provides for explicit and systematic skill development in the five reading elements; (3) includes rigorous evaluations of teacher learning throughout and at the end of the course; and (4) has support for English Learners. The evaluation team created a rubric for each of these four areas (see Appendix A.3 for the initial Professional Development evaluation rubric). The description of support for English Learners is reported in Chapter 5 of this document.

The following are the three areas with their criteria.

1. *Is evidence-based (22-7-1209 (2) (c))*. The team evaluated in what way the program evidence reflected one of the four ESSA evidence levels either by (a) formal research studies that demonstrate impact on teacher practice and student outcomes or (b) a logic model or theory of action that outlines how and why the program expects to have impact based on four research-based PD criteria (content focus, models of effective practice, feedback and direction, and ongoing support with sufficient duration of at least 45 hours). If vendors had formal research studies, then the evaluators reviewed the reports to determine whether the evidence was in alignment with ESSA evidence levels 1 or 2. If the vendors did not have formal research studies, then the team evaluated evidence of all four research-based professional development elements for alignment with ESSA evidence levels 3 or 4.
2. *Provides explicit and systematic skill development in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and comprehension (22-7-1209 (2) (c) (I))*. The team evaluated whether skill development across reading areas was present, explicit, and systematic using vendor-supplied information. Reading areas were phonemic awareness, phonics, vocabulary development, reading fluency, and reading comprehension. For reading comprehension,



## Professional Development Programs on the Advisory List

evaluators focused on “close reading” which was emphasized by READ Act minimum skill competencies and required in the evaluation of PD programs. This is different from the instructional programs evaluation which included both close reading and interactive reading (See Chapter 3), which are represented in the READ Act and the Colorado State Standards, each required in the evaluation of instructional programs.

3. *Includes rigorous evaluations of learning throughout and at the end of the course that a person taking the course must pass to successfully complete the course (22-7-1209 (2) (c) (II))*. The team evaluated the presence of (a) evaluation of teacher knowledge of program content, (b) evaluations both during and at the end of the course, (c) an indication of rigor, and (d) specific criteria and indicators of successful course completion (e.g., a certificate). For the indication of rigor, the team initially created two criteria: an indication within the assessment item and the presence of classroom evaluation. Measurement of the indication within the assessment item was based on Webb’s Depth of Knowledge levels 3 and 4 because of the use of Webb’s model in creating and assessing rigor in state standards (e.g., in the Common Core State Standards, 2010). The team noted that some vendors may instead be familiar with Bloom’s Taxonomy when assessing item rigor, so vendors viewed a comparison between the two when submitting information on our vendor request form. For classroom practice, the team expected that vendors would have some criteria to evaluate the application of program content to teaching.

### **Rationale for Inclusion of Additional Professional Standards**

CDE’s vendor solicitation referenced research-based professional development elements described by Joyce & Showers (2002). These elements included the presentation or theory and strategy as a rationale for active engagement, demonstration of new learning, practice with feedback, and ongoing support. The evaluation team aligned these research review elements with two other more recent reviews by Desimone (2009) and Darling-Hammond



and colleagues (2017) to determine a final set of evidence-based criteria. Specifically, Desimone's five research-based areas were: a focus on content with modeling, active learning with feedback, coherence, duration, and collective participation. Darling-Hammond and Gardner research review yielded six areas of PD: content focused that incorporates active learning, uses models of effective practice, offers feedback and direction, provides coaching and expert support, is of sustained duration, and supports collaboration. Both the Desimone (2009) and Darling-Hammond and colleagues (2017) professional development research syntheses base their findings on research that shows impact on changes in teacher practice and positive impact on student learning outcomes. The final set of criteria from these reviews for this evaluation were (a) content focus (focuses on five components of reading and incorporates active professional learning); (b) uses models of effective practice (e.g., demonstration); (c) offers feedback and direction; (d) ongoing support with sufficient duration (coaching; 45 hours).

## Data Collection and Methods

### Description of Information Used to Review Professional Development Programs

The evaluation team reviewed three documents or sets of documents to review the programs: (a) the vendor response to the evaluation's vendor request form, (b) the vendor application to CDE, and (c) any associated documents the vendor submitted with the application. The sequence of review started with the response to the evaluation's vendor request form because it included key categories for the evaluation. The evaluators then used the vendor application to CDE for supplemental information and to corroborate findings. (See Appendix B.4 for the Professional Development Programs Vendor Request Form.)

### Process Used for Collecting Information from Vendors

After creating the vendor request form, the team's vendor point-of-contact sent the request to each vendor by email. The point-of-contact sent follow up



emails and called nonresponsive vendors. All six vendors submitted responses to the vendor request form.

### **Training for Independent Evaluators**

Two of the project's lead evaluators conducted the professional development evaluations. They reviewed the rubric, calibrated on an initial sample, and further clarified definitions.

### **Ratings and Resolution of Discrepancies in Ratings**

One team of two lead evaluators reviewed all six professional development programs. The evaluators rated the four program criteria, including evidence to support English Learners which was not included in the final rating (See Chapter 5). If programs had ratings of partially meets or fully meets for being evidence-based; fully meets for skill development; and fully meets for assessment, then the summary rating was calculated as "fully meets." The evidence-based criteria allowed for "partially meets" because this meant the programs met at least ESSA evidence-levels 3 or 4.

The evaluators first reviewed one program together and agreed on ratings as a sample program. They then reviewed the final five programs independently and met to resolve queries and identify and resolve any discrepancies in the ratings. Two discrepancies occurred and were not more than one category different. In these cases, the evaluators discussed then concurred on the final rating.

## **Results and Discussion**

Overall, all professional development programs fully met SB-19-199 requirements (Exhibit 4.1). One fully met the evidence base requirement, and all six met requirements around explicit and systematic skill development and including evaluations of learning. Below we explain our findings in detail.

*Is evidence-based (22-7-1209 (2) (c)).* Five PD programs partially met the standard of evidence-based and one professional development program (CORE



## Professional Development Programs on the Advisory List

Elementary Reading Academy, face-to-face) fully met the standard of evidence based. Those that “partially met” did so because they met ESSA evidence levels 3 or 4. One fully met because it met ESSA evidence level 1 or 2. The five PD programs that partially met the standard provided a logic model or theory of action that outlined how and why the program expected to have impact. The program that fully met the standard provided three formal research studies that demonstrated impact on teacher practice or student outcomes. These formal research studies were conducted by researchers external to the vendor organization, were found on the vendor’s website, and did not have information on the funding source. The article findings were positive with overall effect sizes ranging from .17 to .18 for two of the three studies.

Exhibit 4.1

Professional Development Program Summary SB-19 – 199 Requirement	Rating			
	Fully meets	Partially meets	Does not meet	
Is evidence-based (22-7-1209 (2) (c)).	1	5	0	
Provides explicit and systematic skill development in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and comprehension (22-7-1209 (2) (c) (I)).	6	0	0	
Includes rigorous evaluations of learning throughout and at the end of the course that a person taking the course must pass to successfully complete the course (22-7-1209 (2) (c) (II)).	6	0	0	
	Fully meets	Largely meets	Partially meets	Does not meet
Summary Rating: Compliance with SB 19-199 requirements	6	0	0	0

*Provides explicit and systematic skill development in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and comprehension (22-7-1209 (2) (c) (I)). All six programs fully met the criteria for explicit and systematic skill development in all five*





## Professional Development Programs on the Advisory List

reading areas. The programs all included a heavy emphasis on building teachers' knowledge of the five core components of scientifically based reading instruction. For comprehension, the programs focused on close reading within a single text. We did not find consistent evidence of the professional development focusing on interactive reading, which is comprehension using information outside of the text itself (e.g., comparisons with other texts or use of reader background knowledge). Close reading is the primary focus of the READ Act and Colorado's minimum standards, while interactive reading is also included in the Colorado Academic Standards for Reading. Because the READ Act focuses on close reading, the evaluation did include ratings for interactive reading in the professional development review.

*Includes rigorous evaluations of learning throughout and at the end of the course that a person taking the course must pass to successfully complete the course (22-7-1209 (2) (c) (II)).* All six programs fully met the criteria for inclusion of rigorous evaluations of learning throughout and at the end of the course with criteria and an indication of completion.

All programs included evidence for rigor within the evaluation items, specifically with verbs like “analyze” and “apply” as related to Webb's Depth of Knowledge. However, most item types in the evaluations were recognition (e.g., true, false, multiple choice) and focused on knowledge and understanding of content knowledge learned in the program rather than the application or analysis of content (e.g., pedagogical knowledge, classroom implementation). The evaluators were unable to ascertain the number of items that might fall in either category because vendors provided samples of their evaluation items in accordance with the evaluation team's request.

Knowledge of scientifically based reading alone is unlikely to result in shifts in practice. Therefore, the evaluation team initially proposed to evaluate whether vendors' assessments included measures of classroom practice in their evaluation, even though this was not part of CDE's selection criteria. We found that none of the vendors included direct measures of impact on classroom



practice in their assessments. We decided to not include this in determining whether vendors met expectations for rigorous performance evaluations. Although analysis of classroom practice was not part of vendors' formal evaluations, five of the six vendors had classroom practice tied to coaching and feedback from coaches, suggesting that classroom practice was included as a formative assessment.

### **CDE Processes for Identifying Items for Approved and Advisory Lists**

CDE's process for identifying items for the professional development advisory list reflected the READ Act components and intent. Two CDE staff, the director and supervisor of PK-3 literacy led the processes. The staff used four criteria to guide the selection of PD programs: the standards for literacy, the new legislation (statute that was also written into the rule), historical information about previous rubrics, and feedback from previous contracts. CDE communicated the selection criteria through a detailed request for applications sent to existing advisory list professional development vendors on March 19, 2020 and provided a technical assistance webinar. CDE chose existing vendors because of the adjusted review timeline (due to the pandemic) and because the vendors had experience with the criteria in Colorado and would be able to meet the additional teacher training requirements.

CDE K-3 staff created a rubric for reviewing vendor proposals and invited stakeholders to comment on the draft. CDE sent out communication, posted the request to respond on website, and provided a TA webinar about how to provide rubric feedback. CDE leadership approved the rubrics in April, 2020.

Application reviewers are typically geographically diverse in Colorado. However, CDE was unable to use their standard recruitment process for reviewers due to the school shutdowns in March of 2020 caused by the pandemic. Instead, CDE invited reviewers from the instructional programs proposals, the CDE literacy team, and CDE preschool staff who had knowledge



## Professional Development Programs on the Advisory List

---

of scientifically based reading. The reviewers conducted their first review as a team on April 22, 2020 on a virtual meeting. The reviewers then reviewed programs independently and reconvened to come to consensus. Reviewers did not receive compensation under CDE policy.

CDE reported four successes about the selection process related to the review criteria. The criteria:

1. Increased the expectations about professional development for reading in the districts. The criteria gave districts and teachers insight into what CDE means by scientifically based reading.
2. Increased CDE's confidence in the advisory list. The criteria allowed them to confirm that the programs are scientifically based and aligned to the teacher standards and with statute and rules.
3. Allowed for transparency with the districts. Districts can review the rubrics and the comments and CDE can answer district questions based on the criteria and process.
4. Describe professional development specific to reading. In the past all professional development was combined with everyone on the same list.

CDE reported two lessons learned for improving the future process. First, the virtual reviews allowed reviewers more time for understanding of the rubric. Second, there is a need for clarity about what is meant for professional development by evidence-based reading so that the solicitation can include new possible vendors.

This independent evaluation and CDE's evaluation had very similar approaches. Each team developed processes using foundational requirements, used external stakeholders to provide feedback, and had multiple reviewers who came to consensus on ratings. However, the evaluation operationalized the criteria somewhat differently. The independent evaluation used a threshold of partially met for evidence-based professional development and fully met for the other criteria rather than a point threshold (CDE had an 80% threshold).



## Professional Development Programs on the Advisory List

---

The CDE selection processes in 2020 differed from the process in 2015, with the 2020 rubric being more specific due to the changes in SB-19-199 and emphasis on the new requirement for teacher training. Therefore, the 2020 CDE solicitation for professional development programs included more explicit guidance than in 2015. A second difference in the 2020 processes was the lack of external reviewer recruitment because of the pandemic.

In the future, CDE may want to consider (a) continuing the virtual review process due to the time it allowed for reviewers to understand the rubric independently rather than presenting all of the information in a one-day in-person session, and (b) allowing for time to continue the traditional recruitment so that CDE can have diverse stakeholders involved in the selection process.



# 5

## Supporting English Learners

English Learners bring immense potential to the classroom, including intellectual, linguistic, and creative resources to be built upon. Realizing English Learners' potential development – the gap between what a learner can accomplish and understand independently and that which a learner is capable of but has yet to achieve – relies on educators providing students with appropriate supports to help them realize their potential.

### Key findings

- **Five of six programs showed evidence that the professional development program helps teachers support multilingual or English Learner students in learning to read.**
- **All six programs referred to oral language support for English Learners.**
- **All six programs referred to Spanish or Spanish and other languages or dialects and translation to English.**
- **Some programs specifically referenced phonics or comprehension support for English Learners, and all programs referenced differentiation or vocabulary assistance for English Learners specifically for Tier 1 vocabulary words**



To achieve educational equity, English Learners must be offered opportunities to engage with and learn the same rigorous academic content as native English speakers. That goal is best achieved through differentiated instruction that takes into account English Learners' full academic and linguistic potential; this includes careful consideration of students' English language proficiency, as well as the many other factors that can impact learning (Fairbairn & Jones-Vo, 2010). In this chapter we discuss the evaluation team's specific considerations for English Learners for instructional programming and professional development.

Differentiated instruction is instruction purposefully designed to support individual students' learning; students come to the classroom with varied backgrounds and needs (Ford, n.d.). In developing the instructional programming and professional development programs rubrics, the evaluation team specified that a program would fully meet criteria for supporting English Learners if supports existed for English Learners of varying proficiency levels, and if language supports were provided for English Learners to access grade-level content.

### **Providing Language Supports**

A foundational theory of reading, called the Simple View of Reading (Gough & Tunmer, 1986), suggests that skilled reading is the product of foundational word reading skills (such as phonics) and language comprehension. Learning to read includes mastering foundational skills, but those foundational skills do not result in skilled reading in the absence of language comprehension. Similarly, strong oral language comprehension, in the absence of foundational decoding skills, will never lead to skilled reading.

For English Learners, this science of reading still applies. That is, the foundational word reading skills that any young learner must master are the same, such as understanding that letters represent sounds and can be combined in rule-governed ways to represent comprehensible words. However, focusing on



these foundational skills is not enough if the words being read are not – in fact – comprehensible (Goldenberg, 2020). Supporting English Learners in early reading requires providing additional English language instruction and scaffolding to support language comprehension.

### **Realizing English Learners’ Potential by Providing Access to Grade-Level Content**

English Learners bring immense potential to the classroom, including intellectual, linguistic, and creative resources to be built upon. Realizing English Learners’ potential development – the gap between what a learner can accomplish and understand independently and that which a learner is capable of but has yet to achieve – relies on educators providing students with appropriate supports to help them realize their potential (Billings & Walqui, 2017). Effective instruction of English Learners requires a pedagogical balance of high challenge and high support via the scaffolds designed and offered English Learners that support their engagement in a lesson, thus facilitating their development from the space of potential development - the Zone of Proximal Development (ZPD) – to one of autonomy; the goal of instruction (Billings & Walqui, 2017). Scaffolding is particularly important for English Learners who are simultaneously learning language and content in a new language.

### **What Do We Mean by Scaffolding?**

Scaffolding consists of two elements: structure and process. The structure of scaffolding refers to the “constant, but flexible” (Billings & Walqui, 2017) supports that are built into a lesson (Walqui & VanLier, 2010). An example of the structural aspect of scaffolding is the set of step-by-step guidelines that accompany the interactive task Think-Pair-Share. Another example is the supports embedded in a focal text, such as the meaningful chunking of a text with added subtitles and guiding questions that point readers to the main topic and support their understanding of key ideas.

The procedural aspect of scaffolding emerges in the moment, in response to something new the learner brings to classroom interactions with the teacher,



peer/s, and/or text (e.g., posing a question, co-constructing a new understanding, connecting ideas). The on-the-spot characteristic of the process of scaffolding makes it conditional to the particular learner and the situation in which it occurs.

In both examples above, the structure of the scaffold makes possible the process of scaffolding – those in-the-moment classroom interactions between teacher-student, student-student, and student-text, as teachers support students' participation in the activity, engagement with the curriculum, and construction of understanding. In these ways, scaffolding – both the structure and the process – supports students to simultaneously develop conceptual understandings, analytic processes, and the language needed to enact them.

### **Supports for English Learners of Varying Proficiency Levels**

With this understanding of scaffolding, it is clear that in order to truly meet the needs of English Learners, curriculum must include a *variety of differentiated scaffolds* that respond to the wide range of English language abilities English Learners bring to the classroom. To this point, a scaffold that might typically work well with a student who is identified as having proficiency at the “Bridging” level, is ineffective in supporting students at the “Emerging” or “Expanding” levels to access the curriculum, and therefore the related conceptual understandings, analytic practices, and language. In other words, scaffolds are only effective in as much as they account for the range of skills and needs of the learners they are intended to support. In the case of English Learners, a range of differentiated scaffolds must be designed that address their specific language abilities, in addition to the cognitive and developmental skill, in order to support English Learners' progression through their ZPDs and promote their autonomy as learners.

### **Differentiation for English Learners in Instructional Programming**

All programs, regardless of program type, were analyzed to determine whether supports existed for English Learners of varying proficiency levels, and if language supports were provided for English Learners to access grade-level content. To fully meet, a program had to demonstrate that scaffolding was





## Supporting English Learners

---

available that was specifically designed for English Learners, that the supports varied by English proficiency level, and that the language supports provided access to grade-level content, rather than simply reducing text complexity. More general approaches to differentiation, such as Universal Design, were rated as partially meets. The review criteria for programs in Spanish were different: In order to fully meet, the text could not be a direct translation of English, and it had to be representative of the varieties of Spanish spoken in Colorado. The programs did not meet if they did not meet either of these criteria.

The evidence that vendors provided yielded only two supplemental and no intervention programs that fully met the evaluation's criteria (see Exhibit 5.1 for number of programs that fully, partially, or do not meet criteria for supports for English Learners, by program type). Fewer supplemental and intervention programs met this evaluation's criteria because the CDE rubrics were more general and the independent evaluator's criteria were more specific. Specifically, CDE rubrics called for differentiation and support provided for English Learners but did not specify in what ways differentiation or support may be applied. The independent evaluators further defined these ways by specifying scaffolding, or strategies, specifically designed for English Learners, varied support related to English proficiency level, and supports that provide access to grade level content. Some vendors responded to the vendor request that their understanding of differentiation for English Learners was the simple fact that their program was an intervention that struggling students needed according to assessments. Therefore, they did not include specific differentiation or support for English Learners.



**Exhibit 5.1. Number of programs that fully, partially, or do not meet for differentiation for English Learners students, by program type**

Program Type	Rating		
	Fully Meets	Partially Meets	Does Not Meet
Core - English	6	5	1
Supplemental - English	2	19	5
Intervention - English	0	20	4
All - Spanish	6	0	0

### Differentiation for English Learners in Professional Development

Support for English Learners was not a requirement for vendors, and therefore was not included in the final rating of each professional development program. The Evaluation Team added the review criterion as it was of interest to CDE due to the needs of English Learners in the state.

Learners whose native languages are not English and who are learning to read in English have needs that are both similar and different to native English speakers learning to read. Researchers have pointed out that English Learners need a focus on Tier 1, or general, English vocabulary due to limited exposure in early childhood (August et al., 2005) and Spanish speakers would benefit from the integration of Spanish knowledge in learning and understanding English words and sentences (Pearson et al., 2007). Most recently, Goldenberg (2020) reviewed the science of reading knowledge base and research on effective instruction in reading for English Learners and proposed an emerging science of reading for English Learners, prioritizing English literacy and oral language proficiency.

The team examined vendor-supplied information to examine this criterion. Vendors submitted limited information to CDE about how professional development demonstrates how to provide support for English Learners, and the evaluation did not require vendors to submit additional information. Therefore,



this year, this evaluation only shows whether vendors signaled that they provide professional development around how to support English Learners.

Five of six programs showed evidence that the professional development program helps teachers support multilingual or English Learner students in learning to read.

The current evaluation only reviewed whether the PD program mentions the inclusion of support strategies for English Learners. However, adding this criterion allowed for a description of what approaches the programs included for future investigation. Specifically, this review points out three themes of EL support across programs: oral language, vocabulary, and a combination across a continuum of limited supports to more comprehensive.

The evaluators learned that programs had a continuum of how they described supports for English Learners. Some only mentioned that differentiation for English Learners is included in the program and described particular supports in one or more reading areas. The evaluators also learned that programs referred to second language learning in three primary ways: (a) support through oral language, (b) translation between Spanish and English, and (c) specific English Learner help in reading foundations, vocabulary, and comprehension. The following describe these three common ways that programs referred to differentiation for English Learners.

1. *Language support.* All six programs referred to oral language support for English Learners. For example, in the Teachers Top 10 tips there is a tool ("Tool 2 – Language Skills") where teachers learn about the strong contribution of language to literacy skills, the importance of modeling the use of language and academic terms, multiple ways to engage students in dialogue, and how to respond to students with weak language skills to help them develop into stronger skills. The English learner is highlighted throughout the Tool and teachers learn instructional strategies for oral language such as the 30 second conversation; how to engage in



## Supporting English Learners

---

conversation by modeling the use of language; repeating and expanding student responses; asking questions; repeating and clarifying; and providing the vocabulary terms and modeling syntax.

2. *Spanish/English translation.* All six programs referred to Spanish or Spanish and other languages or dialects and translation to English. For example, CORE online and face-to-face programs include an optional section describing differences between Spanish and English, such as the structure of each language, Spanish sound spellings and cognates, and a Spanish phonics survey.
3. *Reading Foundations, Vocabulary and / or Comprehension.* Some programs specifically referenced phonics or comprehension support for English Learners, and all programs referenced differentiation or vocabulary assistance for English Learners specifically for Tier 1 vocabulary words. For example, the Keys to Beginning Literacy manual describes English Learners as enrolling in school with fewer vocabulary words than their English-speaking peers, Tier 1 words may be challenging for English Learners, and visualization is particularly effective in teaching Tier 1 vocabulary to English Learners.

Recommendations related to the inclusion of explicit supports English Learners within instructional programs and professional development programs are described in Chapter 6.



# 6

## Conclusions and Recommendations

The Colorado State Legislature passed the Reading to Ensure Academic Development (READ) Act in 2012 and updated the Act in 2019. The revised Act requires an independent evaluation to identify and assess strategies that the state and local districts and schools have taken to support Colorado students in achieving proficiency in reading.

This report focuses on the findings related to Approved Assessments, advisory lists for Instructional Programming and Professional Development, and the Colorado Department of Education's (CDE) processes for

### Key findings

- **The materials the Colorado Department of Education approved for use with READ Act funds meet the minimum requirements in SB 19-199.**
- **Forty-four instructional programs “fully met” the core requirements outlined in the READ Act, fourteen “largely met” those requirements, and five “partially met” those requirements.**
- **The six professional development programs on the CDE’s advisory list were in compliance with all SB 19-199 required elements.**



## Conclusions and Recommendations

The overall conclusion from the review of assessments, instructional materials, and professional development programs is **that the materials CDE approved for use with READ Act funds meet the minimum requirements in SB 19-199** (see Exhibit 6.1).

Exhibit 6.1. Summary of Ratings				
	Fully Met	Largely Met	Partially Met	Did Not Meet / Not Rated
Assessments	6	7	2	0
Instructional Programs	46	18	3	0 / 1
Professional Development Programs	6	0	0	0

### Summary of Findings by Type of Material

All **approved assessments** met the minimum summary threshold for compliance with the SB 19-199 required elements. All 15 assessments either fully (10 assessments) or partially (5 assessments) met the requirement that the assessment be evidence-based. For the requirement of alignment to the Colorado Academic Standards, 13 of the 15 assessments included sufficient (10 assessments fully met, 3 assessments partially met) evidence, and 2 assessments did not meet this requirement. All vendors submitted sufficient validity evidence to either fully meet (13 assessments) or partially meet (2 assessments) the review criteria. Assessments performed similarly well (14 assessments fully met, 1 assessment partially met) in terms of reliability. Of the 6 measures on the approved list for diagnostic assessments, 4 assessments fully met and 2 assessments partially met the criteria for evidence that the diagnostic assessment accurately identifies students' specific reading skill deficiencies. Of the 7 assessments that offered administration in both English and Spanish, 6



## Conclusions and Recommendations

---

assessments fully met and 1 assessment partially met the criteria. All 7 assessments that offered administration using a paper and pencil rather than a computer fully met the evaluation's criteria.

For the additional technical and quality criteria that went beyond the scope of the requirements contained within SB 19-199, all approved assessments met the minimum summary threshold for compliance. For the evaluation's overall summary rating, 4 assessments received "fully met," 9 assessments received "largely met," and 2 assessments received "partially met." After reviewing submitted evidence for absence of bias against any particular types of students, 9 assessments fully met the criteria, with 4 assessments partially meeting and 2 assessments receiving a rating of "does not meet." All 15 assessments fully met the criteria for evidence that the assessment is supported by appropriate guidance and resources. For the criteria that the assessment includes appropriate accommodations, all of the 15 assessments included sufficient (10 assessments fully met, 5 assessments partially met) evidence. All assessment vendors submitted sufficient useability evidence to either fully meet (14 assessments) or partially meet (1 assessment) the review criteria. Assessment also performed well (9 assessments fully met, 6 assessments partially met) in terms of protecting student and teacher data privacy.

By and large, the **instructional programs** on CDE's advisory list met the core requirements outlined in the READ Act. CDE's two-phase process for reviewing instructional programs is grounded in the science of reading and provides a rigorous and reliable way to vet these programs. WestEd's review of approved instructional programs showed that all programs met the minimum threshold for evidence – a clear logic model rooted in the science of reading; this suggests that all programs have the potential to make a positive impact on students' reading outcomes. All but one program demonstrated the presence of skill development in the areas of phonemic awareness, phonics, reading fluency, and reading comprehension (as applicable), with 38 meeting criteria for explicit



## Conclusions and Recommendations

---

and systematic skill development. All core programs met the minimum requirements for including texts on core academic content to assist students in maintaining or meeting grade-appropriate proficiency in academic subjects in addition to reading. While all instructional programs included some form of embedded assessment, only four programs submitted formal evidence related to their assessments' reliability and validity.

All **professional development** programs on CDE's advisory list met the minimum summary threshold for compliance with the SB 19-199 required elements. Like the approved instructional programs, all six professional development programs met the minimum threshold for evidence – a clear logic model rooted in the science of reading; one program fully met the standard of evidence-based by providing rigorous research studies that demonstrate impact on teacher practice and student outcomes. All six professional development programs fully met the criteria for explicit and systematic skill development in all five reading areas of phonemic awareness, phonics, fluency, vocabulary, and comprehension, namely close reading within reading comprehension. All six programs fully met the criteria for inclusion of rigorous evaluations of teachers' learning throughout and at the end of the professional development course with specified criteria and an indication of completion. Although not included in the final ratings, five of six programs showed evidence that the professional development program helps teachers support multilingual or English learner students in learning to read.

### Summary Findings

In applying the evaluation's to CDE's approved assessments, instructional programs, and professional development programs, the evaluation team identified several cross-cutting findings.

First, CDE's advisory lists for instructional and professional development programs offer clear guidance about explicit and systematic instruction in the





## Conclusions and Recommendations

---

elements of scientifically based reading instruction. The approved assessments offer ways to measure students' growth in these areas. Together the emphasis on these five elements builds coherence among curriculum, instruction, and assessment. Maintaining a clear and consistent focus on these elements over time, while continuing to integrate new empirical research about K-3 reading development, promises to support educators in creating classroom environments that reflect the most up-to-date science of reading.

Second, while both assessment and instructional program vendors claimed alignment with the Colorado Academic Standards or Common Core State Standards, evidence of that alignment varied. Assessment vendors provided evidence that ranged from well-run alignment studies to highlighted copies of the state standards. Instructional programs, by and large, reflected the state's minimum standards for K-3 reading. However, not all instructional programs fully reflected the dimensions of reading that engage students in drawing inferences and making connections beyond the text as required by Colorado's reading standards.

Third, we found significant variation in the type and quality of evidence submitted. Only a small handful of instructional program and professional development vendors submitted high quality randomized controlled trials and quasi-experimental studies. In contrast, other vendors submitted and met evidence requirements with logic models that reflect reading theory but are not supported with rigorous empirical research, contain data tables that lack contextual information or interpretation, include case studies from a small number of observations, and rely on poorly designed empirical studies. Similarly, some assessment vendors provided a narrative about the theory of action behind the assessment supported by a wide range of supporting qualitative and quantitative data. Some vendors provided a brief statement that a study was conducted and offered limited reliability and/or validity data from said study.



## Conclusions and Recommendations

Fourth, CDE has been successful identifying Spanish-language materials. Nearly half of the approved assessments are normed for Spanish-speaking students. All but one instructional program in Spanish included the required components of scientifically based reading that meet the evaluation's thresholds. However, CDE was less successful in identifying instructional materials in English that offer differentiated support for English Learners; this was especially true of intervention and supplemental programs.

Fifth, the evaluation team found that a relatively small number of assessments, instructional programs, and professional development programs supported students with disabilities. Summative and interim assessments from those vendors with a sizeable presence in large-scale testing were more likely to include evidence regarding accommodations than other assessment vendors. Only seven supplemental and three intervention programs provided evidence that their materials could be differentiated for students with disabilities.

### Specific Recommendations

#### Recommendations for Approved Assessments

1. Is evidence-based or scientifically based (22-7-1209 (2)(a)(II)(A)).
  - a. Develop a baseline for what CDE considers to be the appropriate evidence-base for a READ Act-approved assessment. These criteria should differ based on the type (diagnostic, interim, summative) of assessment and should align to the purposes of these assessments as they are defined by CDE in the context of the states K-3 literacy program. The criteria should also include a consideration for when the evidence was produced and how relevant it may be given the duration since that evidence was collected.
2. Is aligned with the preschool through elementary and secondary education standards for reading adopted by the state board (22-7-1209 (2)(a)(II)(A)).
  - a. Set the expectation that vendors have a third party conduct an evaluation of their assessments to determine the alignment of their



## Conclusions and Recommendations

- assessment items to the Colorado Academic Standards. Provided evidence should include results from the study and the resulting action taken by the vendor to address any findings that showed weak or no alignment to the standards.
3. Diagnostic reading assessments are proven to accurately identify students' specific reading skill deficiencies (22-7-1209 (2)(a)(II)(C)).
    - a. Consider a higher threshold for classification accuracy. Most assessments were able to achieve the set benchmark (.70) for the area under the curve (AUC) statistic, which is an overall indication of the diagnostic accuracy. AUC values closer to 1 indicate the screening measure reliably distinguishes among students with satisfactory and unsatisfactory reading performance, whereas values at 0.50 indicate the predictor is no better than chance.
  4. (At least one of the recommended) reading assessments (for kindergarten, and first, second, and third grades) is normed for the performance of students who speak Spanish as their native language, which assessment is available in both English and Spanish (22-7-1209 (2)(a)(II)(D)).
    - a. Collect information from teachers and parents about the K-3 literacy program and their experiences with the available assessments in Spanish. This information can be used to develop more specific and rigorous criteria that may result in assessments that more closely adhere to the expectations of stakeholders and offer a more equitable experience to students being administered the assessment.
  5. The assessment development and review processes are designed and implemented to remove bias against all students.
    - a. Ensuring that assessments do not favor any gender, demographic, or ability is key to an equitable assessment process. Consider establishing common, specific, and clear criteria for what types of evidence vendors must be able to produce to demonstrate that a



## Conclusions and Recommendations

- representative group of outside stakeholders have reviewed their assessments and agree that they are without bias.
6. The assessment offers appropriate accommodations so all students can be fairly and accurately assessed.
    - a. Collect information from districts and schools about the K-3 literacy program and the impact on students who require testing accommodations or increased accessibility options to determine if the current assessments are serving them well.
    - b. Require vendors to provide specific evidence about how they have determined that accommodations are allowed or not and the impact that using accommodations would have on the validity and reliability of the resulting score interpretations.

### Recommendations for Instructional Programs

1. *Is evidence-based (22-7-1209 (2)(b)(I))*
  - a. Meeting ESSA Evidence Tier 1 and 2 levels requires vendors to conduct rigorous research to test whether their theories of action and logic models actually result in impact on student outcomes. We recommend that CDE clearly identify the degree to which evidence meets ESSA evidence tiers, make studies available, and provide guidance for how districts and schools might consider this research when selecting instructional programs.
2. *Provides explicit and systematic skill development in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and comprehension (22-7-1209 (2) (b) (II)), and is aligned with the preschool through elementary and secondary state standards for reading adopted by the State Board (22-7-1209 (2) (b) (II.5))*
  - a. Although the READ Act focuses primarily on the close reading aspects of comprehension, the Colorado Academic Standards for Reading, Writing, and Communicating also include interactive reading practices. CDE should consider guiding instructional programming



## Conclusions and Recommendations

- vendors to provide information about how they support interactive reading and further emphasizing selection criteria related to interactive reading when reviewing instructional programs.
3. *Includes texts on core academic content to assist the student in maintaining or meeting grade-appropriate proficiency levels in academic subjects in addition to reading (22-7-1209 (2)(b)(V))*
    - a. English Learners need additional instructional supports to meet grade-level proficiency levels in reading as well as their academic subjects; few programs offered the kinds of differentiated support necessary to help English Learners successfully learn to read. The evaluation team recommends that CDE convene an expert panel with dual expertise in English language acquisition and early reading development to create criteria for selecting instructional programs that are most likely to meet English learners' strengths and needs. CDE may want to establish a process in which vendors apply to receive a designation of including fully differentiated instruction for English learners.
    - b. The evaluation team further recommends that CDE consider guidance for vendors that instructional materials have diverse authorship and that characters are intentionally diverse, minimally at or above the current research-based percentages discussed in chapter 3.
  4. *Includes evidence-based or scientifically based and reliable assessments (22-7-1209 (2)(b)(V))*
    - a. Given that few instructional program vendors were able to provide systematic evidence about the validity and reliability of their embedded assessments, CDE should consider developing additional guidance for how districts and schools can leverage approved assessments to support instructional decision-making. This includes offering additional guidance about how to use data from approved assessments to select supplemental and intervention programs that will best meet students' needs. To the extent that approved



## Conclusions and Recommendations

- assessments' evidence warrants this, support schools and districts in using these assessment results to determine the intensity and duration of students' participation in supplemental and intervention instruction.
- b. Formal validity and reliability evidence for assessments embedded in instructional programs is rare. If this requirement stands, almost all instructional programs would be subject to potential removal from the advisory list. Given that these informal assessments are primarily used to guide instructional decision-making rather than make consequential decisions about student designations for READ Plans or grade level retention, the legislature may want to consider removing this requirement for assessments embedded in instructional programs to meet rigorous reliability and validity standards. Rather, educators should be guided to use State Board of Education approved assessments in conjunction with instructional programs.

### Recommendations for Professional Development Programs

1. *Is evidence-based (22-7-1209 (2) (c)).*
  - a. The evaluation recommends incorporating revised professional development criteria in future RFPs based on updated research. Only one professional development program presented rigorous, empirical research that demonstrated impact on student outcomes. Therefore, the evaluation team recommends that vendors' theories of action and practice need to clearly reflect the components of professional learning demonstrated by research to have an impact on teacher practice and student learning outcomes. Specifically, the evaluation recommends that CDE consider using the four criteria developed for this evaluation in future reviews: (a) includes content focused on the five components of scientifically based reading and incorporates active professional learning; (b) uses models of effective practice; (c) offers feedback and direction; and (d) provides ongoing support of sufficient duration.



## Conclusions and Recommendations

- Future evaluations of professional development programs related to reading could include the updated four professional development evidence-based themes so that the most updated, comprehensive research base is presented to vendors and used for program evaluation.
2. *Provides explicit and systematic skill development in the areas of phonemic awareness; phonics; vocabulary development; reading fluency, including oral skills; and comprehension (22-7-1209 (2) (c) (I).* Although the READ Act focuses primarily on close reading practices in descriptions of reading comprehension, the Colorado Academic Standards also includes interactive reading practices. CDE may consider whether close reading is the most important area of reading comprehension for professional development vendors to include in their program. If both types of reading comprehension are important to CDE, then guide professional development vendors in the distinction between these two types of reading comprehension.
  3. *Includes rigorous teacher evaluation throughout the program.* CDE may want to consider providing guidance for vendors to (a) submit copies of all evaluations to improve review of all program evaluation materials, (b) increase the number of assessment items related to pedagogical knowledge, and (c) add classroom performance to evaluation. One way to add performance evaluation may be to add a checklist, rating scale, or rubric to coaching practices to enhance the coaching relationship with explicit formative feedback.
  4. *Support for English Learners:* The evaluators have two primary recommendations for CDE regarding providing support for English Learners.
    - a. The approved professional development programs had a continuum of how they refer to differentiation of English Learners: from suggested teaching adaptations to a comprehensive approach. Because the



## Conclusions and Recommendations

---

- professional development programs are at or just over 45 hours, it may not be feasible to suggest that all programs include a focus on comprehensive differentiation for English Learners. Therefore, CDE may investigate ways that professional development programs could maximize differentiation for English Learners by and calling on experts in English language acquisition and early reading from a range of language development perspectives to inform broader professional development about English learners and early reading.
- b. Most or all of the professional development programs included at least three areas of differentiation for English learners: oral language support, English/Spanish translation, and vocabulary. CDE might consider suggesting that vendors include and describe these three areas for supporting English Learners in their professional development programs.





## References

- Adams, M. J. (1990). *Beginning to read: Thinking and learning about print*. MIT Press.
- August, D., Carlo, M., Dressler, C., & Snow, C. (2005). The critical role of vocabulary development for English Language Learners. *Learning Disabilities Research and Practice, 20*(1), 50–57. <https://doi.org/10.1111/j.1540-5826.2005.00120.x>
- Baker, S., Mercier Smith, J., Fien, H., Otterstedt, J., Katz, R., Luft Baker, D., & Thomas Beck, C. (2007). *Three-year report on Oregon Reading First: Impact and implementation*.
- Berkeley, S., Scruggs, T. E., & Mastropieri, M. A. (2010). Reading comprehension instruction for students with learning disabilities, 1995—2006: A meta-analysis. *Remedial and Special Education, 31*(6), 423–436. <https://doi.org/10.1177/0741932509355988>
- Billings, E., & Walqui, A. (2017). *The zone of proximal development: An affirmative perspective in teaching ELLs/MLLs*. <https://www.wested.org/resources/zone-of-proximal-development/>
- Borman, G. D., & D'Agostino, J. V. (1996). Title I and student achievement: A meta-analysis of federal evaluation results. *Educational Evaluation and Policy Analysis, 18*(4), 309–326. <https://doi.org/10.3102/01623737018004309>
- Bus, A. G., & Van Ijzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of Educational Psychology, 91*(3), 403–414. <https://doi.org/10.1037/0022-0663.91.3.403>
- Center on Enhancing Early Learning Outcomes. (2019). *Third grade reading laws: Implementation and impact*. <http://creativecommons.org/licenses/by/4.0itisavailableatwww.ccsso.org>. Phone
- Chard, D. J., Vaughn, S., & Tyler, B. J. (2002). A synthesis of research on effective interventions for building reading fluency with elementary students with learning disabilities. *Journal of Learning Disabilities, 35*(5), 386–406. <https://doi.org/10.1177/00222194020350050101>
- Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*.
- Data on books about Black, Indigenous and People of Color published for children and teens*. (2018). Cooperative Children's Book Center. <https://ccbc.education.wisc.edu/literature-resources/ccbc-diversity-statistics/books-by-and-or-about-poc-2018/> .
- Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. <https://doi.org/10.3102/0013189X08331140>



## References

- Ehri, L. C., Nunes, S. R., Stahl, S. A., & Willows, D. M. (2001). Systematic phonics instruction helps students learn to read: evidence from the National Reading Panel's meta-analysis. *Review of Educational Research, 71*, 393–447. <https://doi.org/10.2307/3516004>
- Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic Awareness Instruction Helps Children Learn to Read: Evidence From the National Reading Panel's Meta-Analysis. *Reading Research Quarterly, 36*(3), 250–287. <https://doi.org/10.1598/rrq.36.3.2>
- Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*(1), 1–44. <https://doi.org/10.1080/19345740802539200>
- Fairbairn, S., & Jones-Vo, S. (2010). *Differentiation instruction and assessment for English Learners: A guide for K-12 Teachers*. Caslon Publishing.
- Ford, K. (n.d.). *Differentiated instruction for English language learners*. <https://www.colorincolorado.org/article/differentiated-instruction-english-language-learners>
- Gamse, B. C., Jacob, R. T., Horst, M., Boulay, B., & Unlu, F. (2008). *Reading first impact study final report*. [www.edpubs.org](http://www.edpubs.org).
- García, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. In *Review of Educational Research* (Vol. 84, pp. 74–111). American Educational Research Association. <https://doi.org/10.2307/24434229>
- Goldenberg, C. (2020). Reading Wars, reading science, and English Learners. *Reading Research Quarterly, 55*(S1), S131–S144. <https://doi.org/10.1002/rrq.340>
- Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1).
- Greene, J. P. (1998). *A meta-analysis of the effectiveness of bilingual education*.
- Herlihy, C., Kemple, J., Bloom, H., Zhu, P., & Berlin, G. (2009). *Understanding Reading First: What we know, what we don't, and what's next*. <https://doi.org/10.4324/9780203142165>
- Hernandez, D. J. (2012). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation*.
- Hughes-Hassell, S., & Cox, E. J. (2010). Inside board books: Representation of people of color. *Library Quarterly, 80*(3), 211–230.
- Jeynes, W. H. (2008). A meta-analysis of the relationship between phonics instruction



## References

- and minority elementary school student academic achievement. *Education and Urban Society*, 40(2), 151–166. <https://doi.org/10.1177/0013124507304128>
- Joshi, R. M., & Aaron, P. G. (2000). The component model of reading: Simple view of reading made a little more complex. *Reading Psychology*, 21(2), 85–97. <https://doi.org/10.1080/02702710050084428>
- Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development* (3rd ed.). Associated for Supervision & Curriculum Dev (ASCD). <http://books.google.com/books?hl=en&lr=&id=MtBj2sJXHGwC&pgis=1>
- Kamil, M. (2004). Vocabulary and comprehension instruction: Summary and implications from the National Reading Panel findings. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 213–235). Brookes.
- Koss, M. D. (2015). Diversity in contemporary picture books: A content analysis. *Journal of Children's Literature*, 41(1), 32–42.
- Langenberg, D. N., Correro, G., Ferguson, G., Kamil, M. L., Samuels, S. J., Shaywitz, S. E., Williams, J., Yatvin, J., Ehri, L., Garza, N., Marrett, C. B., Shanahan, T., Trabasso, T., & Willows, D. (2000). *Report of the National Reading Panel Teaching Children to Read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. <https://doi.org/10.1002/ppul.1950070418>
- Lieberman, I., Shankweiler, D., Fischer, F., & Carter, B. (1974). Explicit syllable and phoneme segmentation in the young child. *Experimental Child Psychology*, 18, 201–212.
- Lloyd, D. N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement*, 38(4), 1193–1200. <https://doi.org/10.1177/001316447803800442>
- Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning. *Review of Educational Research*, 80(3), 300–335. <https://doi.org/10.3102/0034654310377087>
- Mcarthur, G., Sheehan, Y., Badcock, N. A., Francis, D. A., Wang, H. C., Kohnen, S., Banales, E., Anandakumar, T., Marinus, E., & Castles, A. (2018). Phonics training for English-speaking poor readers. In *Cochrane Database of Systematic Reviews* (Vol. 2018, Issue 11). John Wiley and Sons Ltd. <https://doi.org/10.1002/14651858.CD009115.pub3>
- National Center for Education Statistics. (2019). National 4th grade reading scores. In *National Assessment of Educational Progress*.
- Okoye-Johnson, O. (2011). Does multicultural education improve students' racial attitudes? Implications for closing the achievement gap. *Journal of Black Studies*, 42(8), 1252–1274. <https://doi.org/10.1177/0021934711408901>



## References

- Pearson, P. D., Hiebert, E. H., & Kamil, M. L. (2007). Vocabulary assessment: What we know and what we need to learn. *Reading Research Quarterly, 42*(2), 282–296. <https://doi.org/10.1598/rrq.42.2.4>
- Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis)abilities: Evidence, theory, and practice. In S. B. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97–110). Guilford.
- Shachar, R. (2012). Structuring of gender in Israeli society through children's reading and textbooks: where is mom's apron? *Journal of Research on Childhood Education, 26*, 249–263.
- Share, D., Jorm, A., Maclean, R., & Matthews, R. (1984). Sources of individual differences in reading acquisition. *Journal of Educational Psychology, 76*, 1309–1324.
- Slavin, R. E., & Cheung, A. (2005). A synthesis of research on language of reading instruction for English language learners. In *Review of Educational Research* (Vol. 75, Issue 2, pp. 247–284). SAGE Publications Inc. <https://doi.org/10.3102/00346543075002247>
- Snider, V. E. (1997). The relationship between phonemic awareness and later reading achievement. *The Journal of Educational Research, 90*(4), 203–211. <https://doi.org/https://doi.org/10.1080/00220671.1997.10544574>
- Solari, E. J., Grimm, R. P., McIntyre, N. S., & Denton, C. A. (2018). Reading comprehension development in at-risk vs. not at-risk first grade readers: The differential roles of listening comprehension, decoding, and fluency. *Learning and Individual Differences, 65*, 195–206. <https://doi.org/10.1016/j.lindif.2018.06.005>
- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*(1), 72–110. <https://doi.org/10.3102/00346543056001072>
- Stevens, E. A., Walker, M. A., & Vaughn, S. (2017). The effects of reading fluency interventions on the reading fluency and reading comprehension performance of elementary students with learning disabilities: A synthesis of the research from 2001 to 2014. *Journal of Learning Disabilities, 50*(5), 576–590. <https://doi.org/10.1177/0022219416638028>.The
- Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934–947. <https://doi.org/https://doi.org/10.1037/0012-1649.38.6.934>
- Suggate, S. P. (2016). A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities, 49*(1), 77–96. <https://doi.org/10.1177/0022219414528540>
- US Department of Education. (2015). *Reading First funding status*.



## References

---

<https://www2.ed.gov/programs/readingfirst/funding.html>

Vaughn, S., & Wanzek, J. (2014). Intensive interventions in reading for students with reading disabilities: Meaningful impacts. *Learning Disabilities Research & Practice*, 29(2), 46–53. <https://doi.org/10.1111/ldrp.12031>

Walqui, A., & VanLier, L. (2010). *Scaffolding the academic success of adolescent English Language Learners: A pedagogy of promise*.

Weiss, S., Stallings, D. T., & Porter, S. (2018). *Is Read to Achieve making the grade? An assessment of North Carolina's elementary reading proficiency initiative*.

Wolf, M., & Katzir-Cohen, T. (2009). Scientific studies of reading reading fluency and its intervention. *Scientific Studies of Reading*, 5(3), 211–239. <https://doi.org/10.1207/S1532799XSSR0503>

Wright, T. S., & Cervetti, G. N. (2017). A systematic review of the research on vocabulary instruction that impacts text comprehension. *Reading Research Quarterly*, 52(2), 203–226.



---

## List of Appendices

### A. Independent evaluation rubrics

A.1. Approved Assessments

A.2. Instructional Programs on CDE's Advisory List

A.3. Professional Development Programs on CDE's Advisory List

### B. Vendor request

B.1. Vendor submission window

B.2. Approved Assessments

B.3. Instructional Programs on CDE's Advisory List

B.4. Professional Development Programs on CDE's Advisory List

### C. Summary of individual program ratings

C.1. Definitions for Summary Ratings

C.2. Approved Assessments

C.3. Instructional Programs on CDE's Advisory List

C.4. Professional Development Programs on CDE's Advisory List