WestEd
WestEd.org

Colorado READ Act Interim Assessment Comparability Analysis

Summary and Recommendations

Mariann Lemke, Angela Bowzer, Aaron Soo Ping Chow

June 2022

WestEd
WestEd.org

Colorado READ Act Interim Assessment Comparability Analysis

Summary and Recommendations

Mariann Lemke, Angela Bowzer, Aaron Soo Ping Chow

June 2022

Lemke, M., Bowzer, A., and Soo Ping Chow, A. (2022). *Colorado READ Act Interim Assessment Comparability Analysis: Summary and Recommendations*. 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. 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.

WestEd
WestEd.org

Table of Contents

LIST OF TABLES

Colorado READ Act Interim Assessment Comparability Analysis: Summary and Recommendations

The Colorado Reading to Ensure Academic Development Act (READ Act) requires districts to assess students in kindergarten through grade 3 to determine reading competency levels using an assessment from a state-approved list of commercial assessments. Students who do not meet minimum skill levels are identified with a significant reading deficiency (SRD). Identified students are then provided with services aimed at getting them on track to reading proficiently by the end of grade 3. Each assessment vendor provides a cut score to the Colorado Department of Education (CDE) that indicates to schools using their assessments which students should be classified as having an SRD, but vendors can use different approaches to set their cut scores. Additionally, the approved assessments vary in their modes of administration, content assessed, and other characteristics.

The Colorado Legislature first passed the Reading to Ensure Academic Development Act (READ Act) in 2012 and then revised it in 2019. The revised READ Act requires the Colorado Department of Education (CDE) to define "sufficient … growth to standard" over time for students identified as reading below grade level or identified with an SRD. Given differences among assessments used to identify students with an SRD, one of the recommendations from a first-year READ Act evaluation report (McCrary et al., 2021) was to convene a panel of

assessment experts to develop recommendations for measuring growth to standard and review the effectiveness of the assessment system for READ Act purposes.

To follow up on this recommendation, between November 2021 and May 2022, WestEd analyzed characteristics of approved READ Act interim assessments, along with score data from administration of the assessments, and met with an expert panel to discuss analyses and results. The purpose of the analyses was to examine the comparability of the assessments and their SRD cut scores and to test the feasibility of establishing a common growth scale across assessments; detailed findings from qualitative and quantitative analyses are available in separate reports. Overall, the results suggest that neither the content of the assessments nor student scores that identify students with an SRD are fully comparable. This document summarizes those findings and their implications, and offers recommendations related to developing a potential growth-to-standard approach and for the READ Act assessment system more broadly.

Findings

Nine READ Act interim assessments were reviewed as part of WestEd's analysis of assessment characteristics; two (Star Reading and Star Early Literacy) were excluded from analyses of score data because the reported scores did not match expected scale values and analysts could not distinguish between the assessments, which include different content. Several other approved assessments (aimwebPlus Spanish, IDEL, PALS Español) were excluded from both analyses because they were only part of pilot programs and therefore not used continuously or were used by very few schools (data for approximately 50 to 500 students across the state per year). Data from school years 2014/15 to 2018/19 were included in analysis of score data.

As shown in Table 1, two of the READ Act interim assessments are given to most students in the state, with 58 percent of students taking Acadience Reading and another 19 percent taking the i-Ready Diagnostic. The demographic characteristics of students taking different assessments varies. Table 1 provides percentages of students with different demographic characteristics by assessment across years (2014–19) and grades (K–3). As shown, all assessments have similar percentages of special education students. However, higher percentages of White students and lower percentages of students eligible for free- or reduced-price lunch take Acadience Reading, aimswebPlus, FastBridge aReading, and i-Ready Diagnostic compared to other assessments. Students taking these four assessments also typically have higher 3rd grade CMAS ELA scores than students using the other interim assessments. Additionally, higher percentages of English learner (EL) students take ISIP Español, ISIP ER, and PALS than the other assessments. Not

surprisingly, Spanish-language ISIP Español assessment test-takers are primarily Hispanic EL students. However, most EL students (92%) overall are tested using an assessment in English.

Table 1
Assessment Usage and Characteristics of Test-Takers

Assessment	Usage*	EL	Special Education	FRPL	Asian	Black	Hispanic	White	Average Grade 3 CMAS ELA Score & % Meeting or Exceeding Standards**
Overall Sample		16%	10%	43%	3%	4%	33%	54%	738 (39%)
Acadience Reading	62% (657,898)	12%	10%	45%	2%	3%	33%	56%	737 (38%)
aimswebPlus	0.6% (6,451)	7%	9%	29%	2%	1%	17%	70%	737 (44%)
FastBridge aReading	0.6% (5,911)	17%	12%	39%	0%	1%	27%	67%	743 (44%)
i-Ready Diagnostic	21% (220,862)	15%	12%	30%	5%	5%	24%	60%	736 (38%)
ISIP ER	6% (67,052)	24%	11%	58%	5%	13%	42%	34%	742 (45%)
ISIP Español	1% (15,552)	93%	10%	89%	0%	0%	96%	4%	733 (35%)
PALS	9% (94,862)	26%	11%	50%	4%	4%	43%	44%	732 (37%)

*** Percent of students across Grades K–3 from 2014–19 taking this assessment**
**** Average score and percent meeting or exceeding ELA proficiency for students taking this assessment**

While there are commonalities between the assessments used to identify students with an SRD, there are many differences that suggest that the meaning of "significant reading deficiency" is different across the approved interim READ assessments. Table 2 provides an overview of key characteristics of the assessments, including information about their stated purposes; content included; mode of administration; definition of SRD; technical characteristics; percentages of students classified as SRD on each assessment; and data on each assessment's relationship to CMAS.

Table 2
Summary of Findings

	Purpose	Content	Administration	SRD Score Definition	Technical Characteristics	Percent SRD	Relationship to CMAS		
Assessment	**Explicitly Claims to Identify Students "At Risk"?**	**Vendor-Identified Assessed Colorado READ Act Literacy Areas***	**Mode of Administration and Response Method**	**SRD Cut Score Interpretation**	**Validity, Reliability, Fairness**	**Ever Classified as SRD**	**Correlation of Grade 3 Interim Assessment Scores to CMAS**	**% SRD and not SRD in Grade 3 Meeting CMAS Standards**	**Equipercentile Equivalent CMAS Scores for Spring SRD Identification**
Acadience Reading	yes	PA, PH, F, C	paper-pencil with mostly verbal responses	10%-20% chance of meeting later benchmarks on this assessment	Validity: fully meets Reliability: fully meets Fairness: partially meets	20%	0.76	Not SRD: 45% SRD: 6%	696
aimswebPlus	yes	PA, PH, F	paper-pencil with mostly verbal responses	At or below 10th percentile	Validity: fully meets Reliability: fully meets Fairness: fully meets	17%	0.72	Not SRD: 50% SRD: 3%	699
FastBridge aReading	yes	PA, PH, V, C	online with selected responses	At or below 15th percentile	Validity: fully meets Reliability: fully meets Fairness: partially meets	16%	0.83	Not SRD: 47%: SRD: 0%	700
i-Ready Diagnostic	no	PA, PH, V, C	online with selected or created responses	More than one grade below grade level	Validity: fully meets Reliability: fully meets Fairness: fully meets	18%	0.83	Not SRD: 54% SRD: 0.4%	701

	Purpose	Content	Administration	SRD Score Definition	Technical Characteristics	Percent SRD	Relationship to CMAS		
Assessment	**Explicitly Claims to Identify Students "At Risk"?**	**Vendor-Identified Assessed Colorado READ Act Literacy Areas***	**Mode of Administration and Response Method**	**SRD Cut Score Interpretation**	**Validity, Reliability, Fairness**	**Ever Classified as SRD**	**Correlation of Grade 3 Interim Assessment Scores to CMAS**	**% SRD and not SRD in Grade 3 Meeting CMAS Standards**	**Equipercentile Equivalent CMAS Scores for Spring SRD Identification**
ISIP ER	yes	PA, PH, V, F, C	online with selected or created responses	At or below 20th percentile	Validity: partially meets Reliability: fully meets Fairness: does not meet	21%	0.78	Not SRD: 41% SRD: 0.3%	686
ISIP Español	yes	PA, V, F, C	online with selected or created responses	At or below 20th percentile	Validity: partially meets Reliability: fully meets Fairness: does not meet	19%	*	*	*
PALS	yes	PA	paper-pencil with mostly verbal responses	At or below 25th percentile	Validity: fully meets Reliability: fully meets Fairness: partially meets	24%	0.66	Not SRD: 43% SRD: 1%	698
Star Early Literacy	yes	PA, PH, V, C	online with selected responses	At or below 25th percentile	Validity: fully meets Reliability: fully meets Fairness: fully meets	**	**	**	**
Star Reading	no	V, C	online with selected responses	At or below 25th percentile	Validity: fully meets Reliability: fully meets Fairness: fully meets	**	**	**	**

*** A very small number of students taking ISIP Español took CMAS in English (*n*=149), about 4% of ISIP Español test-takers.**
****STAR assessments were excluded from analysis of score data.**

Purpose. As Table 2 shows, although all the READ Act interim assessments are used to identify students with significant reading deficiencies, not all were necessarily designed with this specific purpose in mind. For example, both i-Ready Diagnostic and Star Reading were originally intended to provide a measure of overall reading skill rather than focusing on students at risk of reading problems.

Content. READ Act assessments differ in the content they use to identify an SRD at each grade level and the emphases they place on different literacy areas. For example, five of the nine reviewed interim assessments (i-Ready Diagnostic, ISIP ER, ISIP Español, Star Early Literacy, and Star Reading) test vocabulary at kindergarten through grade 3, while other assessments test vocabulary only at a subset of grade levels. One assessment (Acadience Reading) does not assess vocabulary at any grade level. Five of the nine reviewed interim assessments (FastBridge aReading, i-Ready Diagnostic, ISIP ER, ISIP Español, and Star Reading) test skills associated with comprehension (listening or reading) across all grade levels. Four assessments (Acadience Reading, aimswebPlus, PALS, and CMAS ELA) use comprehension items only in the identification of an SRD at grade 2 and/or grade 3, and Star Early Literacy tests skills associated with comprehension at grade 1.

The assessments also differ in the ways they describe and assess the READ Act literacy areas. For example, the READ Act requires the assessment of phonemic awareness. Three assessments (aimswebPlus, i-Ready Diagnostic, and PALS) describe their assessments as measuring phonological awareness instead of, or alongside, phonemic awareness. Acadience Reading and aimswebPlus assess reading fluency directly by having students read aloud with an assessor counting the number of words read correctly, while ISIP ER and ISIP Español assess fluency online through a task in which every fifth or sixth word of a text is left blank and students choose from three options to fill in each blank. Similar fill-in-the-blank tasks are used in several other assessments as a measure of comprehension. As a result of these differences, the meaning of the SRD designation can vary across assessments.

Administration. Three of the reviewed READ Act interim assessments (Acadience Reading, aimswebPlus, and PALS) are administered individually or in small groups by an assessor, while the other assessments (FastBridge aReading, i-Ready Diagnostic, ISIP ER, ISIP Español, Star Early Literacy, Star Reading) are administered online. The assessments administered by an assessor have items presented orally or via a paper form, and students respond aloud or write, depending on the measure being administered. For the online interim assessments, students select answers or move objects on a screen to respond. For some items, students listen to instructions or other text being read aloud and select a response; for other items, students read text presented onscreen and select a response. The online interim assessments use computer adaptive testing (CAT) algorithms that assign different items to students based on their performance. FastBridge aReading assigns items to students considering only item difficulty (not content), while five assessments (i-Ready Diagnostic, ISIP ER, ISIP Español, Star Early Literacy, and Star Reading) administer items using item difficulty in conjunction with content

domains. Assessments administered by an assessor are fixed-form assessments in which all students are administered the same test items.

Technical characteristics. A prior review of interim assessment technical characteristics (Friedrich et al., 2020) found that all but two reviewed assessments (ISIP ER and ISIP Español) fully met validity criteria and all met reliability criteria. However, five of the nine assessments did not fully meet fairness criteria, indicating that evidence was not available about how well the assessments function across different groups of students.

SRD Score Definition. Seven of the nine interim assessments established the cut score used to identify SRD based on normative performance ranging from the 10th percentile to the 25th percentile. Acadience Reading set its cut score based on likelihood of future reading achievement within the Acadience Reading assessment, and i-Ready Diagnostic established its cut score using a criterion-based method. According to some vendors, performance at or below the provided SRD cut score is performance below or far-below grade-level expectations; to others, this performance is linked to a risk of end-of-year reading deficits; and still to others, this performance is indicative of a need for reading intervention.

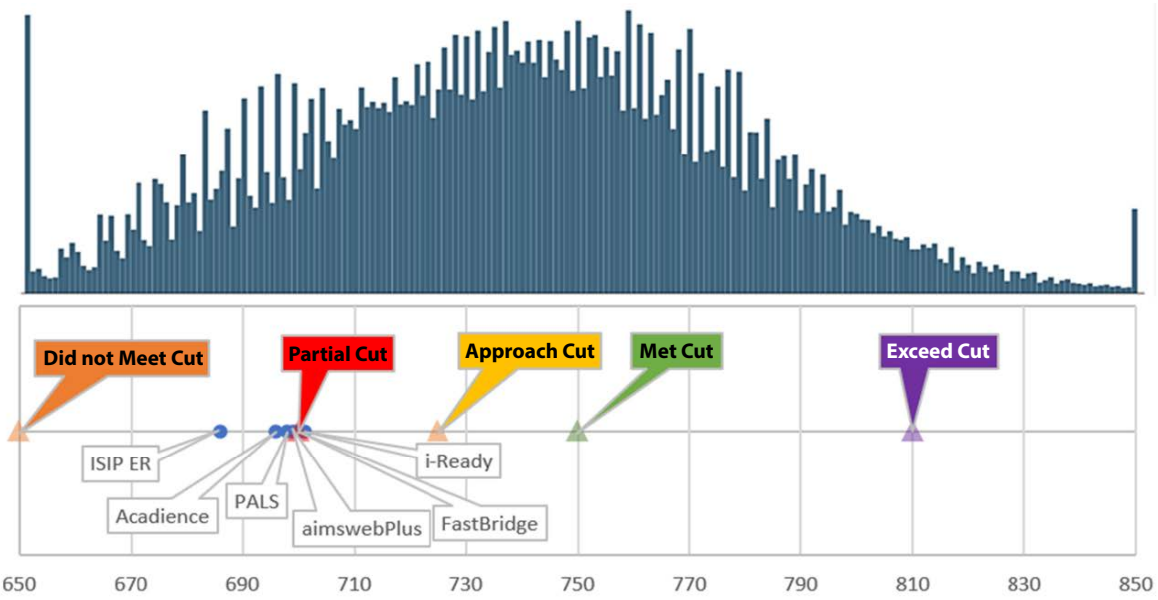
Relationship to CMAS ELA. Ultimately, CDE hopes to use information from the interim assessments to examine the extent to which students are growing towards becoming proficient readers (as measured by the CMAS ELA assessment in the 3rd grade). Therefore, additional analysis focused on how well the interim assessments map to CMAS and how their SRD cut scores compare to CMAS and to one another. All READ Act interim assessments reviewed were moderately or strongly correlated with CMAS ELA scores; meaning, performance on the interim assessments is at least somewhat predictive of performance on CMAS. The i-Ready Diagnostic and FastBridge aReading have the strongest relationships with CMAS ($r = 0.83$) and PALS has the weakest ($r = 0.66$). Of the assessments administered in English, the three assessments most highly correlated with CMAS are administered online (i-Ready Diagnostic, FastBridge aReading, and ISIP ER), while the three assessments with the weakest correlations with CMAS are administered via paper and pencil (Acadience Reading, aimswebPlus, and PALS).

Across all assessments, nearly all of the students (98%) classified as having an SRD in the 3rd grade score below proficiency on CMAS (i.e., failed to meet expectations). However, the majority of students classified as **not** having an SRD (53%) also failed to demonstrate proficiency on CMAS by the end of 3rd grade. This number includes a large percentage of students who score above the SRD cut score (and consequently may not receive READ Act services) but who also score below grade-level expectations. Data also show that classification consistency (i.e., how well being identified as SRD corresponds to CMAS performance) is generally lower for students in specific demographic subgroups. Additionally, different

assessments may classify students as SRD from some subgroups, such as students eligible for free- or reduced-price lunch, at different rates.

Although vendors used different approaches to setting their SRD cut scores, using an equipercentile approach to placing SRD cut scores on the CMAS scale shows that the SRD cut scores are relatively similar across assessments and generally cluster around a value of 700, the CMAS Partially Met Expectations cut score (see Figure 1).[1]

Figure 1. Equipercentile Linking of Interim Assessment Scores to CMAS Scores (3rd Grade)



That is, students identified as SRD are scoring at least 50 points and about two performance levels below the CMAS Met Expectations cut score of 750. A sample growth-to-standard projection model estimated to test the feasibility of such an approach shows that students who are classified as having an SRD at any grade level are unlikely to grow sufficiently to be proficient in reading by the end of 3rd grade, as measured by the CMAS exam. However, it does appear that students identified as SRD are experiencing growth as measured by increasing

[1] In the equipercentile approach, SRD cut scores are applied to the CMAS distribution (e.g. if the SRD cut score is the 15th percentile, the 15th percentile of the CMAS distribution is the equipercentile CMAS equivalent to that SRD cut score).

percentages of students exceeding predicted cut scores corresponding to the Partially Met Expectations and Approached Expectations performance levels.

Recommendations Based on Findings

This next section uses information from the analyses to suggest areas for CDE to consider in its future READ Act interim assessment–related work.

Assessment Selection

The READ Act requires that assessments approved for use in meeting READ Act requirements be vetted by an external organization and that a new approval process take place every few years. An initial set of assessments was approved in 2013 and a new approval process will take place in 2022. Based on analyses of currently approved assessments and their data, the following section provides recommendations related to potential future assessment selection criteria and processes.

Consistency in SRD Identification Across Assessments. The READ Act's approach to early reading is grounded in a multi-tiered system of support (MTSS) framework, in which scores are used to identify students with significant reading deficiencies, so that additional resources and supports can be provided for these students through a READ Act plan. However, results of the qualitative analysis of the interim assessments show that the ways in which students are identified as having an SRD across assessments differ. The content used to assess students to determine if they should be classified as having an SRD can differ, the manner in which students are assessed can differ, and the method used to establish a cut score that identifies a student as having an SRD can differ. Results of the quantitative analyses show that SRD cut scores, when mapped onto a common scale (grade 3 CMAS ELA) using equipercentile linking, appear to be similar. However, further analysis matching students across assessments shows that SRD cut scores may operate differently for different groups of students—specifically, they are less precise for students eligible for free- and reduced-price lunch, EL students, and American Indian, Hispanic. and Black students. If an important goal for CDE is to ensure that SRD determinations are consistent across students in the state, then a different approach to selecting interim assessments may be needed, such as selecting fewer assessments that are similar in their content or administration mode; selecting assessments that define SRD in similar ways; and/or considering evidence on how the assessments function for different subgroups of students. Additionally, some assessments (e.g., Fastbridge aReading) are used in very small numbers of schools, and the state may wish to consider current usage in conjunction with other criteria in its approval processes.

Situating all READ Act Assessments in an MTSS Framework. The READ Act's MTSS-type approach to early reading includes universal screening of all students through interim assessments, followed by additional diagnostic assessments to determine the specific needs of

students identified during initial screening as in need of additional support. Given that approach, CDE should consider reviewing and approving interim and diagnostic assessments together and approving "sets" of assessments that can meet both purposes. Some of the assessments approved for interim and diagnostic use, in fact, are the same (e.g., i-Ready Diagnostic). The state may wish to consider whether different criteria should be applied to assessments to be used for interim and diagnostic use. Rather than using a generic rubric about assessment quality, the state should consider a rubric more focused on the use of the assessments in an MTSS framework, such as the National Center on Intensive Intervention rubrics. In particular, the state may wish to keep in mind criteria for universal screening related to "practicality," which is about who can administer the tests and how long they take. Even more important than selection of assessments, according to MTSS framework criteria, is how data resulting from their administrations are used and how the assessments and their data align to approved instructional and intervention programs. The state's rules note that: "The list of evidence-based or scientifically-based instructional programming and supporting technologies, including software, for assessing and monitoring student progress must be aligned with the recommended reading assessments."

CDE could also consider an approval process that includes assessment and instructional materials. As with diagnostic assessments, some approved interim assessments are part of an instructional program to which student performance is connected in score reports (e.g., i-Ready Diagnostic, ISIP ER, ISIP Español). Assessment vendors could support this process by providing specific information about what scores they provide in each READ Act literacy area and describing how they could be used to determine next steps in those areas. This approach would also require application of a common definition of each literacy area by the vendors. The state could then consider developing guidance for districts on ways the instructional programs and assessments can best operate as a system. Guidance could address questions such as: what data do I get from a given diagnostic assessment and how might that lead to choice of a particular intervention, or implementation of which instructional program follows best from reported student performance on an approved interim assessment?

Meeting READ Act Goals From a Content Point of View. Because different assessment vendors define and measure literacy areas differently, it is not easy to compare them accurately to one another or to READ Act requirements. That is, content described as "phonics" for one assessment may be described as a different literacy area (or no READ Act literacy area) for another assessment. Further, how each literacy areas is assessed in practice can look very different across assessments, especially when student response mode varies (e.g., verbally responding to items versus selecting responses on a computer). To better facilitate consideration of how well assessments align to literacy areas required for assessment by the

READ Act and to allow comparisons across assessments, we suggest CDE consider requesting several specific types of information from vendors, including:

- **A map of assessment content to the READ Act minimum competencies for each grade level and administration time period** (fall, winter, spring)—more specifically, information about numbers of items and points associated with READ Act minimum competency areas, so the state can weigh how well each assessment's content represents the READ Act literacy areas, and a description or sample items that illustrate how the competency is addressed. As an example, a minimum competency standard at kindergarten is: "Identify and produce groups of words that begin with the same sound (alliteration)." Asking vendors to explain how their assessment addresses this minimum competency standard will allow CDE to consider not only the content assessed by each assessment, but also the ways in which the content is assessed.
- **A specific description of what SRD—connected to the READ Act definition of SRD—and "fall reading competency" means for each assessment.** Additionally, documentation about why each cut score is appropriate for its intended use, and a description of how and why cut scores changed from previous years for assessments that were previously approved would be useful. The state should also collect fall and spring grade-level benchmark scores from vendors (currently, only fall "competency" benchmarks are collected, and not all are available on the CDE website). Collecting this information from each vendor will allow the state to use the vendor's own information to assess student growth toward proficiency. Finally, the state may also wish to consider asking vendors to commit to developing predictive validity studies specific to Colorado's SRD classifications and performance on CMAS ELA.
- **Specific information about how any expected dyslexia identification criteria are met in each assessment.** Such information will facilitate the state's move toward incorporating consideration of identification with READ Act assessments.

Spanish-Language Assessments. The READ Act requires that some interim assessments be available in Spanish, and previous criteria for approval of Spanish-language assessments do include some additional criteria for these assessments. They sought documentation "that the test specifically identifies students with a '*significant reading deficiency*' in their native language (i.e., test developers consider what constitutes a proficient reader in the target language rather than directly translating the measures of a proficient reader in English into the target language." Yet, the READ Act does not define what minimum competency might look like in Spanish, so it is unclear how these criteria should be met. Review of currently approved Spanish-language assessments, for example, shows that they include additional content (e.g., use of accents) and necessarily address other content differently (e.g., letter sounds). The definition of a significant reading deficiency in Spanish may also need to be different than the READ Act definition, which was developed for literacy skills in English. Finally, many students who take Spanish-language READ Act assessments also take the Colorado Spanish Language

Arts assessment (CSLA) rather than the English language Colorado Measures of Academic Success (CMAS). CDE may, therefore, wish to consider how Spanish-language READ Act interim assessments and their SRD cut scores relate specifically to the CSLA. As a starting point, CDE should consider approval criteria that reflect how Spanish-language assessments would be used in Colorado. For example, it should be defined for vendors which students in which types of instructional programs should take the Spanish-language assessments, and then requiring evidence from vendors about the validity and reliability of their assessments when used with the types of students who would be taking them in Colorado. More broadly, CDE may wish to consider how READ Act definitions and competencies apply to reading skill development in Spanish.

Relationship of READ Act Assessments to READ Act Policy Goals

The purpose of the Colorado READ Act, as defined in legislation, is to "provide students with the necessary supports they need to be able to read with proficiency by third grade so that their academic growth and achievement is not hindered by low literacy skills in fourth grade and beyond." The legislation requires some specific actions, such as assessing students each year, and offers definitions of SRD and "reading competency." The legislation, however, does not describe how all the various required actions and definitions might work together to achieve intended outcomes. Better specifying the mechanisms by which required actions—such as assessing students—are intended to lead to desired outcomes and clarifying those desired outcomes themselves will be important to determining how READ Act assessments can best support policy goals. This section offers suggestions for areas of additional policy discussion and refinement related to READ Act assessments.

Identification of Students With SRD. While schools are permitted to use a body-of-evidence approach to determine SRD classifications, WestEd's analyses showed that school-provided SRD classifications for students nearly always matched the SRD classification that students would have received based solely on their READ Act interim assessment score. Therefore, should CDE wish to use the interim assessment scores themselves as a proxy for SRD classifications in future analyses (for example, in developing a growth-to-proficiency model), this method will work. However, if CDE's intent in allowing for the body-of-evidence approach is for schools to use multiple data points in making SRD determinations, schools may need more guidance or support in using additional data (or additional research into how schools apply a body-of-evidence approach may be needed).

Evaluating Effectiveness of READ Act. The revised READ Act of 2019 required CDE to engage an independent evaluator to describe "… effective processes, procedures, methods, and strategies used by local education providers … achieving significant academic growth to standard in reading for students identified as having significant reading deficiencies and as reading below grade level." Indeed, it is this requirement to develop an approach to measuring growth to standard for purposes of evaluating READ Act effectiveness that motivated WestEd's analyses

of READ Act assessments and scores. Results of that work show that a single growth-to-standard model may not be possible. However, collecting additional READ Act assessment data might support this goal. Currently, CDE collects only one interim assessment score per student per year. Collecting multiple scores within the year (at least fall and spring, and potentially fall, winter, and spring) would enable comparisons of student progress within schools and districts using the same READ Act interim assessments without necessitating direct comparisons across assessments.[2] Rigorous evaluation methods intended to provide causal evidence about which districts, schools, or programs are showing success could be carried out with such data.

Defining "Reading Competency" as Measured by CMAS. The READ Act defines "reading competency" as "a student meets the grade-level expectations in reading adopted by the state board." Given that the state's measure of whether students are meeting grade-level expectations is the CMAS, this definition would suggest that the READ Act aims for students to meet or exceed expectations on CMAS. However, CMAS is a measure of the state's academic standards in English Language Arts, not just reading. CMAS measures reading, writing, and use of language. READ Act interim assessments tend to focus (particularly at the early grades) on specific foundational reading skills and behaviors, as described in the READ Act itself. A question for CDE to consider is whether the outcomes of the READ Act should be measured in terms of overall CMAS performance or a narrower construct of reading represented by CMAS reading subscores. Reading subscore data were unavailable for WestEd's analyses; so, it was not possible to evaluate whether a stronger relationship between READ Act assessments and CMAS reading subscores might be observed than between READ Act assessments and CMAS scores overall, or whether a subscore might increase error because it was based on fewer items.

READ Act Activities and Goals. The READ Act, as implemented, includes some activities aimed at all students (e.g., requirements for training for teachers) and some activities aimed only at some students (e.g., additional funds and READ Act plans for students identified as having a SRD). Results of analysis show that there are many students who are not classified as SRD, and therefore not receiving additional READ Act resources or plans, but who **also** do not achieve proficiency on the CMAS ELA exam. Additionally, the equipercentile equivalent CMAS scores that corresponded to the SRD cut score of each assessment clustered around the Partially Met Expectations CMAS performance level (two levels below proficiency). This result may suggest a need for additional supports, not just for students identified as SRD, but for other students as well. More broadly, an expected outcome of proficiency on the CMAS grade 3 exam appears to be at odds with the use of assessments to identify only students most "at risk" (in this case, those identified as having a SRD). If the intended outcome of use of the READ Act assessments is that more students are proficient at grade 3, this goal should be operationalized through a

[2] Such an approach would also, however, require interim assessments whose scales allow for meaningful measurement of growth and a way to interpret that growth. Such criteria would need to be included in an assessment selection process.

theory of action that describes how use of the data and other activities will lead to this outcome for students at **all levels of interim assessment** performance.

Development of an Approach for Measuring Growth to Standard

As noted, the revised READ Act (2019) charged CDE with defining growth to standard for students reading below grade level or identified with an SRD. As part of the analysis of READ Act score data, WestEd tested several approaches to developing a model for measuring student progress toward meeting state standards. Results of this analysis form the basis of recommendations and considerations for future work in this area.

Utility of Growth-to-Standard Measures. The READ Act describes sufficient growth for students found to be "at risk" (either identified as having an SRD or reading below grade level) as putting them on a path to "adequately demonstrating proficiency by the end of third grade." Assuming that "adequately demonstrating proficiency by the end of third grade" means performing at least at the Met Expectations performance level on the grade 3 CMAS ELA assessment, a growth metric that uses this standard as the desired outcome is unlikely to provide much useful information. Results from analysis of data from 2014/15 through 2018/19 show that less than two percent of students identified as SRD in grade 3 met expectations on the grade 3 CMAS exam. Further, only 47 percent of students who are **not** classified as having an SRD in grade 3 met CMAS expectations. Therefore, while the results of analysis show that it is technically feasible to create a relatively accurate growth-to-standard proficiency model, creating such a model for students identified as SRD is unlikely to provide useful information. That is, measuring whether or not students identified as SRD make sufficient progress from year to year to get on a path towards proficiency will likely simply tell us that they do not. Nor do students "reading below grade level." Choosing a different outcome (such as reducing the numbers of students identified over time as SRD or having SRD students move toward proficiency without having to reach it by grade 3) or extending the timeframe to meet the target might create more feasible expectations without significant additional supports for students.

Furthermore, results in this report show that it will likely not be possible to create a single growth-to-standard model including all assessments together. Should CDE wish to measure growth toward proficiency for students identified as having a SRD using interim assessment scores, it will likely be necessary to carry out analyses separately for each assessment, since results of the matched sampling, along with the qualitative results, suggest it is not appropriate to put all assessments on a single scale.

If the goal of a growth-to-standard model is to help identify whether READ Act SRD interventions (or READ Act activities, more broadly) are "working" by showing students' learning progress toward proficiency, a single model may not be possible. However, collecting additional data might support this goal. Currently, CDE collects only one interim assessment score per student per year. Collecting multiple scores within the year would enable

comparisons of student progress within schools and districts using the same READ Act interim assessments without necessitating direct comparisons across assessments. Such an approach would, however, require interim assessments with scales that allow for meaningful measurement of growth and a way to interpret that growth, whether this is linked to CMAS performance or not.

Improving Quality and Utility of READ Act Assessment Data

Analysis of READ Act interim assessment scores revealed some challenges with the data itself, as currently collected. This section provides recommendations related to improving the quality and utility of future READ Act assessment data.

Consider Trying to Collect Data Directly from Vendors. CDE currently collects assessment score data from districts. Analysis suggests that in some cases, districts are reporting inaccurately (for example, including incorrect scales). To standardize the reporting of data, CDE could consider developing agreements with vendors and districts to collect READ Act assessment data directly from vendors themselves. Collecting data directly from vendors would necessitate working through issues related to student identifiers and privacy, but in the longer term, it might benefit both districts and CDE by reducing burden on districts and improving quality and consistency of data received by CDE. At minimum, CDE should create a template for vendors to report minimum and maximum scores for its assessments for each year they are approved for use and collect these data. This information will better enable CDE and other data users to verify the scores reported.

Clarify Data Collection Layouts and Other District Guidance. Assuming CDE is not able to collect assessment data from vendors, some changes to the data layouts provided to districts to guide their reporting could improve the consistency and quality of data. These changes include:

- **Dates and fall scores.** Specify that scores from fall assessments (for students who did not test in the spring because they showed grade-level competency in the fall) should be reported, along with dates reflecting a fall administration. Reported dates in current data appear to all be spring dates, which may be accurate if districts are testing all students in the spring, but adding guidance on how to handle cases of fall-only testing for students might ensure that all data reported are accurately tied to a time period.
- **Scores used to determine SRD**. If a composite score is used for SRD identification purposes, that score should be collected; where subscores are used, those should be collected. To ensure districts provide the score used to determine SRD, improve the

data collection layout documentation to make the description of scores needed more prominent.

- **Language of instruction.** Collect info (or map to info the state already has) on language of reading instruction to provide context to interpret performance of EL students.
- **Star Early Learning.** The currently approved Star Early Learning assessment consists of two separate assessments—Star Early Literacy and Star Reading (with Star Early Literacy targeting younger students who are beginning readers and Star Reading targeting more independent readers, typically with a transition between tests around grade 2). However, in the current data collection, results from these products are combined. Given that the assessments test different content, CDE should consider collecting scores separately for each test. In addition, there are different (and numerically overlapping) reporting scales available for Star assessments, and the scores in the current READ Act data may include different scales for the same assessments. Making information about which scale to include more prominent in data layout documentation might help prevent such situations in the future.
- **Information on SRD cut scores.** Each assessment provides SRD cut scores to CDE, which posts them on its website. However, for at least one assessment, scores on the CDE website do not specify exactly how to apply the cut score (i.e., should students at the cut score be considered as meeting the SRD criteria, or not?). Additionally, to enable analyses over time, CDE should maintain a list of historical cut scores for each year an assessment is approved.

Collect Multiple Scores Within a Year. As noted, CDE currently collects only a single score per year per student. For the purposes of measuring growth and potentially evaluating effectiveness of READ Act activities, CDE might wish to consider collecting fall and spring scores (and winter, where available). Collecting multiple scores might also improve the utility of the scores overall, as CDE would be able to conduct additional analyses examining trends in early literacy attainment; for example, CDE would be able to examine how proportions of students meeting (or not meeting) benchmarks compare across grades and time periods. These trends could then inform decisions.