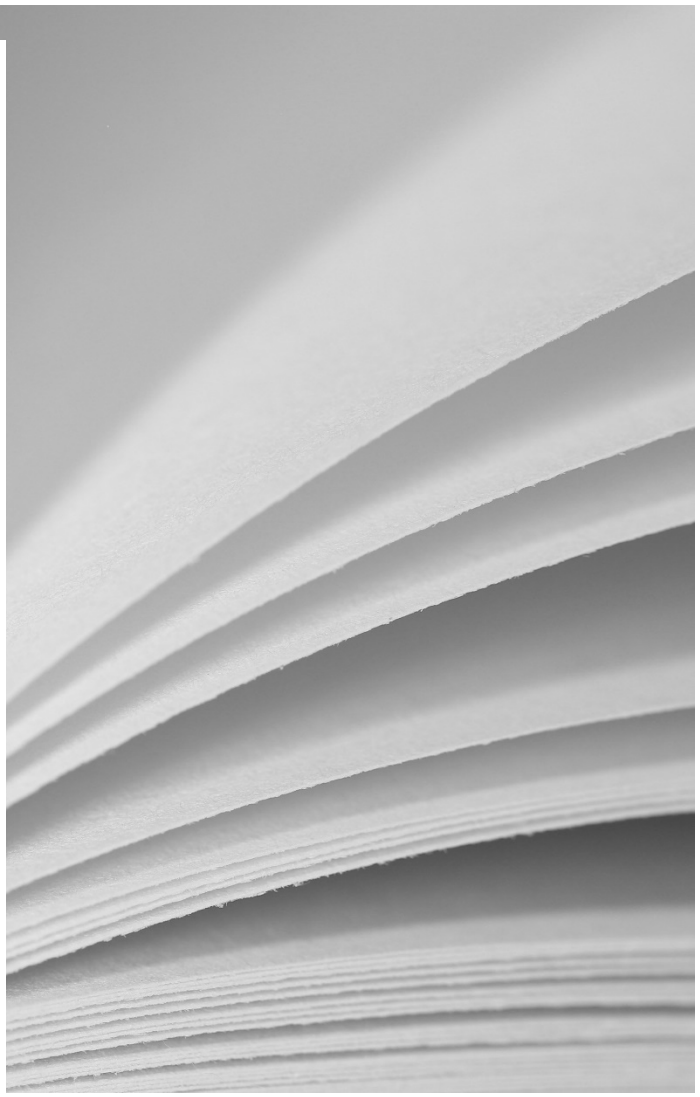
Evaluation Plan for Colorado Department of Education's Dyslexia Pilot Program



2021

Overview

During the 2019 legislative session, the Colorado General Assembly created a dyslexia pilot program through House Bill 19-1134. The purpose of the program is to pilot the use of READ Act assessment results and a research-based protocol to identify markers of dyslexia in K–3 students. During the 2021-2022 school year, pilot sites will receive training and coaching to provide support to young students who may demonstrate the early markers for dyslexia.

Following the implementation of the pilot program (at the end of the 2021-22 school year), the Colorado Department of Education is required to evaluate the implementation of the pilot program and the effectiveness of the strategies in identifying and supporting more students in the participating local education providers than were identified and supported in nonparticipating local education providers. The Department must submit a report concerning the implementation and evaluation of the pilot program to the State Board of Education and the education committees of the State Senate and the House of Representatives on or before December 31, 2022. This document describes the plan for the evaluation of the pilot program.

Table of Contents

Contents

Introduction

During the 2019 legislative session, the Colorado General Assembly created a dyslexia pilot program through House Bill 19-1134. According to the preamble of the Bill, parents of children identified as having dyslexia had voiced concerns related to the adequacy and effectiveness of the methods and tools for identifying students who have dyslexia and the adequacy of the educational supports for these students. Though there had been various efforts at both the state and school district levels to address the issues related to effective identification and support for students with dyslexia, these efforts had not resulted in significant progress in educating these students. Therefore, the General Assembly, recognizing the obligation of the state of Colorado to provide educational opportunities to all children that will enable them to lead fulfilling and productive lives, found it is necessary to create a working group of parents and educational experts to review the work of educational experts and local education providers in Colorado and in other states in the area of identification of and educational support for students with dyslexia, and to use their findings to inform future efforts by the state and local education providers to identify and effectively support students with dyslexia.

The General Assembly further found that a pilot program through which the Department of Education works with a group of volunteer local education providers to use early literacy assessment results to identify markers of dyslexia and provide support to young students who may demonstrate the early markers for dyslexia would strengthen the ability of local education providers throughout the state to identify and effectively support students with dyslexia. According to the Statement of Work through which this evaluation plan was ordered, the pilot program includes three schools that will participate in the program from June 2021 through June 2022. Through the pilot program, schools will receive no-cost training and support for Grade K-3 teachers in utilizing the pilot identification process for markers of dyslexia. All pilot program activities will be communicated and delivered by the University of Oregon pilot program team.

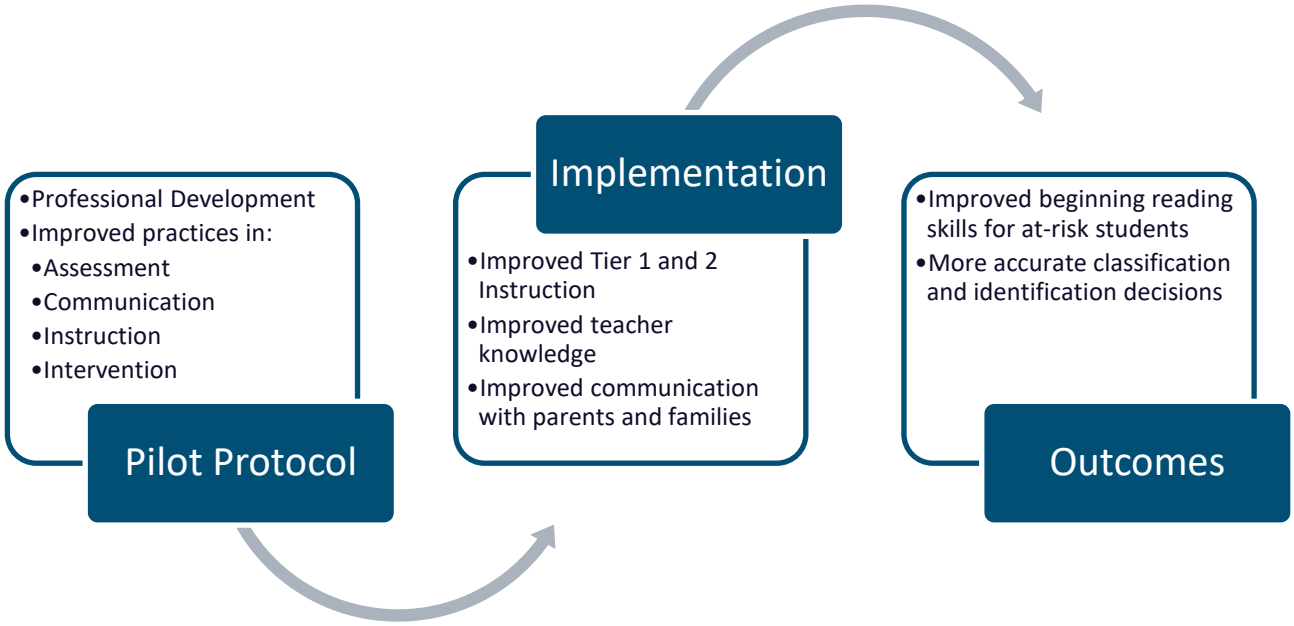
Following the implementation of the pilot program (at the end of the 2021-22 school year), the Department is required to evaluate the implementation of the pilot program and the effectiveness of the strategies in identifying and supporting more students in the participating local education providers than were identified and supported in nonparticipating local education providers. Based on the evaluation, the Department shall refine the resources for technical support, identification, and interventions, as necessary, and disseminate the resources to all local education providers in the state. Upon request, the Department shall also provide the technical support necessary to effectively use the resources. On or before December 31, 2022, the Department shall submit to the State Board and the Education Committees of the Senate and the House of Representatives, or any committees, a report concerning the implementation and evaluation of the pilot program. The Department may include in the report any recommendations for legislation that the department deems necessary based on the evaluation of the pilot program.

Evaluation Plan Overview

This document describes the plan for the evaluation that will inform the Department's report to the State Board and the Education Committees of the Senate and House. It is written under the assumption that the primary intended user of the evaluation results is Colorado Department of Education. Secondary users include the state's dyslexia work group; any relevant government bodies, such as the State Board and the Education Committees of the Senate and the House of Representatives; and the University of Oregon. The following sections describe the dyslexia pilot program's theory of change and the focus of the evaluation. Chapter 2 describes the Methods. Chapter 3 describes the Analysis and Interpretation Plan. Chapter 4 describes the Dissemination Plan.

Theory of Change

When conducting a program evaluation, it is important to have a clear theory of change to guide the formation of research questions, study design, and interpretation of results (CDC, 2011). Based on the House Bill 19-1134, the University of Oregon's application (RFP DAAA 2020000098), and prior research (e.g., Fien et al., 2021; Smith et al., 2016), the following theory of change is proposed: The pilot program will consist of professional development delivered by the University of Oregon on a dyslexia screening and intervention protocol that is intended to improve practice in the areas of assessment, communication, instruction, and intervention. If the professional development and protocol are successful in improving these practices, the improved practice may be observed through improved Tier 1 and Tier 2 instruction; improved teacher knowledge; and improved communication with parents and families. The improved practices should in turn lead to improved student outcomes, such as better beginning reading skills for at-risk students, and more accurate classification and identification decisions. The magnitude and direction of any effects will depend on baseline conditions and may therefore vary across schools.



It is important to document the stages of program development at every stage of a program's implementation and evaluation (CDC, 2011). Development tends to be described in terms of one of three states: program planning, implementation, and program maintenance (CDC, 2011). The dyslexia pilot is best described as being somewhere between the *planning* and *implementation* stage. Many aspects of the pilot and protocol are currently being implemented, but only for the first time. Therefore, the resources needed for successful implementation are not fully understood. An evaluation should therefore focus on inputs and activities, as well as outputs and short-term outcomes, with an understanding that further improvements to the program's inputs and activities may be necessary for achieving desired outputs and short-term outcomes (CDC, 2011).

It is also important to note for methodological reasons that many pilot activities are now in the implementation stage. At the time of this writing, the pilot protocol and a draft of the evaluation plan has been reviewed by the dyslexia work group, or a representative thereof. The University of Oregon has begun to deliver professional development to the pilot schools. An initial needs assessment has been conducted at the pilot schools. All schools reported that they were at an intermediate level of implementation. Documentation of these activities is critical because they are altering baseline conditions in ways that can mitigate the evaluation's ability to make causal inferences about the dyslexia pilot program. For example, any improvements that already have been made in teacher knowledge at pilot sites will not be observed in measures that have yet to be administered.

Evaluation Focus

Table 1

Evaluation Questions by Topic and Interested Stakeholders

Topic	Interested Groups
Usability	
1. To what extent did teachers in the pilot schools find the professional development on the protocol to be relevant and useful?	CDE, pilot schools, UO, State Board, DWG
2. To what extent did teachers find the intervention protocol easy to use?	CDE, pilot schools, UO, State Board, DWG, Acadience Inc
3. To what extent did teachers in the pilot program perceive that the intervention protocol would meet the needs of their students?	CDE, pilot schools, UO, State Board, DWG, Acadience Inc
4. To what extent did teachers have a positive perception of team meetings?	CDE, pilot schools, UO, State Board, DWG
Implementation	
5. To what extent did teachers receive training as intended?	CDE, UO, State Board
6. To what extent was the protocol administered by teachers?	CDE, UO, State Board
7. To what extent did implementation of a multitiered system of support in reading improve at the school level from beginning to end of year?	CDE, UO, State Board
Effectiveness	
8. How does pilot teacher knowledge of the science of reading and markers of dyslexia compare to the knowledge of teachers in comparison schools?	CDE, pilot schools, UO, State Board, DWG
9. To what extent did the implementation of evidence-based reading instruction improve from pre- to post-intervention?	CDE, pilot schools, UO, State Board, DWG
10. To what extent did the intervention protocol change student outcomes (e.g., reading assessment scores, lower rates of risk for dyslexia, number of students referred for comprehensive evaluation, number of students meeting SLD eligibility requirements and requiring special education services)?	CDE, pilot schools, UO, State Board, DWG, Acadience Inc

Program evaluations can serve three overlapping goals: rendering judgements about a program (i.e., accountability), facilitating improvements in the program (i.e., program development), and knowledge generation (i.e., transferability; CDC, 2011). To pursue these aims, evaluations must engage stakeholders, describe the program, focus the evaluation design, gather credible evidence, justify conclusions, and ensure use of results. While undertaking these actions, evaluators must balance considerations of accuracy, utility, feasibility, and ethics. The overarching purpose of the evaluation is to improve the Department's capacity to refine the resources for technical support, identification, and interventions; provide the technical support necessary to effectively use the resources; and make recommendations for legislation. Though the evaluation will examine program effectiveness, it should be emphasized for the sake of transparency that the small size of the study and methodological constraints described the in the Methods section, will limit the evaluation's ability to make valid generalizable inferences about the program's effectiveness. Therefore, based on input from the Department and other stakeholders, the evaluation will seek to answer questions concerning the pilot's usability, implementation, and effectiveness described in Table 1.

Stakeholder Interest

Stakeholders have an interest in understanding program evaluation results. However, different groups of stakeholders often have different interests, which call for different dissemination plans. Prior to the creation of this document, members of the state dyslexia work group were engaged on several occasions by Colorado Department of Education about (a) the nature and scope of the pilot program and protocol to be implemented by the University of Oregon, and (b) the nature and scope of the evaluation plan. In meetings on May 11, 2021 and August 10, 2021, a member of the dyslexia work group, Dr. Laura Santerre-Lemon, provided input into the evaluation plan, reviewed a draft of this document, and made suggestions for revisions.

Additional stakeholders who have directly or indirectly provided input into the dyslexia pilot and evaluation plan include Acadience Inc, the publisher of one of the assessments to be used in the pilot program, and school personnel. Although Acadience is not a primary audience for the evaluation results, they have an interest in the pilot program and evaluation insofar as the evaluation could be interpreted as providing information about the assessment's usability and appropriateness for dyslexia screening and identification. School personnel, meanwhile, have a critical interest in the study results insofar as they may affect their future work. However, a major component of this evaluation is designed to assess their needs and feedback through data collection and analysis. It may therefore be superfluous to seek their feedback on the evaluation plan itself.

Table 1 identifies which stakeholders are apt to have a particular interest in findings related to each research question. The Dissemination Section describes steps that will be taken to ensure the interests of all stakeholders are considered before findings are disseminated, including the mode and timing of communication.

COVID-19 and School Closures

It is also important to document that the COVID-19 has had a major impact on the planning and development of the dyslexia pilot program and evaluation plan, and it may continue to do so in the upcoming years. As described in the Methods section, COVID-19 has had a large, indirect effect on the pilot program and evaluation, especially through stay-at-home orders and the closure of schools to in-person instruction. One area of impact has been the program's initial start date, which had to be delayed due to travel restrictions and the infeasibility of recruiting schools during the spring of 2020. A second area of impact has been recruitment. Relatively few schools opted into this pilot program despite the lengthy solicitation window. The reluctance of schools to undertake the types of new initiatives promoted within the pilot is almost certainly related to the fact that pandemic has placed, and continues to place, enormous burdens on families and schools. A third area of impact concerns the pilot program activities. The University of Oregon was not allowed to carry out the pilot program activities as originally proposed due to health and safety measures. For example, the University of Oregon originally proposed to have their coaching personnel conduct classroom visits and observations as part of their professional development efforts. This component of the pilot had to be eliminated to reduce the risk of disease transmission. The extent to which the pandemic and related safety measures will continue to affect the dyslexia pilot study and evaluation is unclear. Suffice it to say, the actual implementation of the dyslexia pilot will differ from the intended implementation. There is also higher than typical potential for deviations from the evaluation plan. To the extent possible, the evaluation will follow best practice recommendations for adjusting evaluation plans during disruptions, such as COVID (Moore et al., 2021).

Methods

The research questions described in this evaluation plan (Table 1) will be answered using data gathered during the 2021-2022 study pilot study. The study sample, measures, design, and procedures are described in the sections that follow.

Sample

Prior to the creation of this research plan, the Department recruited three schools to participate in the dyslexia pilot program via a voluntary response to a solicitation. Recruitment took place from fall 2020 to winter 2021, resulting in a combined applicant pool from which schools were reviewed and selected. As an incentive to participate and to offset any costs associated with the pilot study, schools were offered $10,000 for their participation. To be considered, schools needed to complete an application and obtain district level support. Additionally, schools had to use Acadience Reading as a fall, winter, and spring interim assessment. Schools that had participated in previous Department projects (such as the Early Literacy Grant, Early Literacy Assessment Tool project, or the Structured Literacy project) were encouraged to apply. Schools that participated in these projects were given priority for selection, as were schools that have invested in other evidence-based reading initiatives aligned to the READ Act. Five schools applied. Applications were scored on the following criteria: Leadership, PD & Training, Implementation of Evidence Based Practices, and Commitment to the Pilot. All applicants met the criteria. However, one school was excluded because it was fully virtual, and another did not use the required assessment system.

School Characteristics

Table 2

School Characteristics of Pilot Sites According to the National Center for Education Statistics

School	Setting	Student Total	Title 1	Student:Teacher	FRPL Eligible	Per Pupil Spending
Ignacio Elementary (NCES 080477001444)	Rural-Distant	320	Yes	13.06:1	62%	$15,051
Singing Hills Elementary (NCES 080372001532)	Rural-Fringe	360	No	13.85:1	16%	$10,135
Academy for Advanced Creative Learning (NCES 080306006456)	Large City	293	No	13.91:1	15%	$6,948

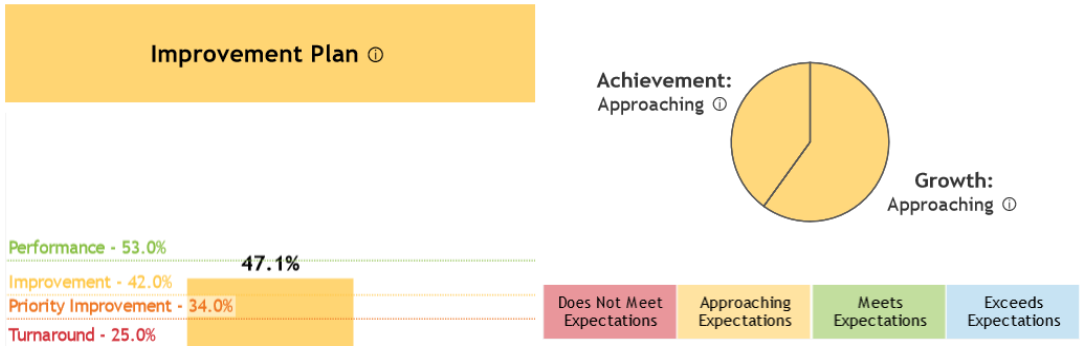
Table 2 describes the characteristics of the three schools that applied and were accepted into the pilot. The schools are located in rural and urban settings. One school is a Title 1 school with a high percentage of students qualifying for free and reduced priced lunch. The schools are of similar size and have similar student/teacher ratios. The schools serve roughly 970 students, about 540 of which can be expected to participate in the pilot study. Assuming 8 teachers participate per grade, there will be approximately 96 teachers in the pilot.

Figure 1 provides snapshots of school performance. In 2019, Ignacio Elementary was approaching improvement in academic achievement and growth. Singing Hills Elementary met academic achievement and growth goals. Academy for Advanced Creative Learning exceeded academic achievement goals and met growth goals.

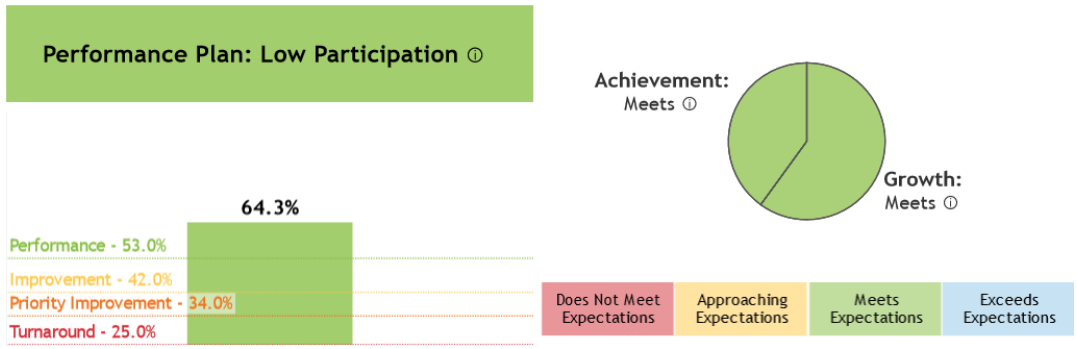
Figure 1

Snapshot of school performance in 2019

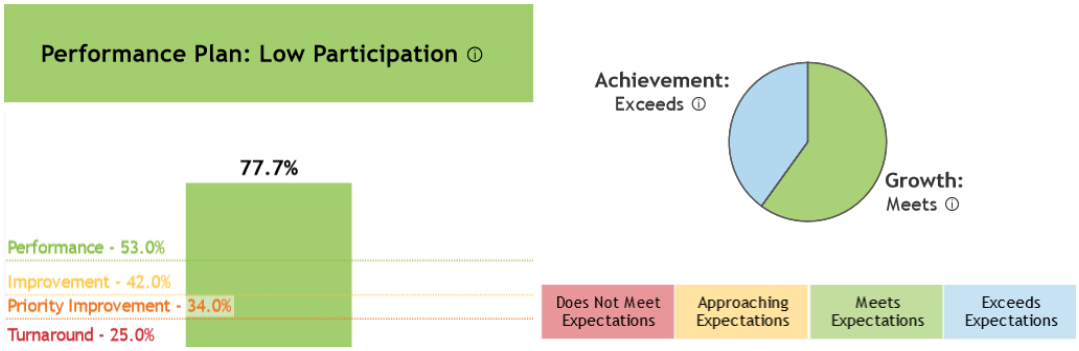
Ignacio Elementary (CDE 4252)



Singing Hills Elementary (CDE 7925)



Academy for Advanced Creative Learning (CDE 0517)



Teacher Characteristics

The 2020 Teaching and Learning Conditions Survey described educator experience levels for Ignacio Elementary and Academy for Advanced Creative Learning. At Ignacio Elementary, about 23% of educators were in their first year, 27% of had 2-3 years of experience, 10% had 4-5 years of experience, 17% had 6-10 years of experience, 10% had 5-10 years of experience, and 13% had more than 20 years of experience. At Academy for Advanced Creative Learning, about 35% of educators were in the first year, 12% had 2-3 years of experience, 23% had 4-5 years of experience, and 31% had 6-10 years of experience. Results were unavailable for Singing Hills due to a low response rate.

The Teaching and Learning Conditions Survey also yields a composite score that describes how favorably educators view their teaching and learning conditions based on ratings of school leadership, teacher leadership, student conduct, instructional practice and support, professional development, time, facilities and resources, community support and involvement, overall reflection, and district support. In 2020, teachers at Ignacio Elementary viewed their conditions less favorably (71.5%) than the state average (76.5%), down from a gap of only 1% in 2018. The gap was driven by low favorability ratings of the general approach to professional development and availability of time. Academy for Advanced Creative Learning viewed their conditions more favorably (85%) than the state average (76.5%) in 2020. The 2020 rating was 1% lower than the 2018 rating of 84%, which was also higher than the state average. Current and historical results were unavailable for Singing Hills.

Student Characteristics

Table 3

Student Characteristics in 2018-2019 by School

	SWDs	English Learners	Minority	Race and Ethnicity					
				White	AIAN	Black	Hispanic	Asian	Two or more races
Ignacio Elementary	10%	6%	75%	24%	33%	0%	38%	>1%	4%
Singing Hills Elementary	12%	4%	18%	79%	1%	>1%	14%	>1%	4%
Academy for Advanced Creative Learning	4%	0%	21%	75%	1%	1%	15%	>1%	8%

Table 3 describes the students in the pilot schools by disability status, English Learner status, racial/ethnic minority status, and race/ethnicity. Data are from the Colorado Department of Education "Performance Snapshots" except for Race and Ethnicity data, which are from NCES. Table 4 describes the total and percent of students by grand and schools that had significant reading deficiencies in 2018-2019.

Table 4

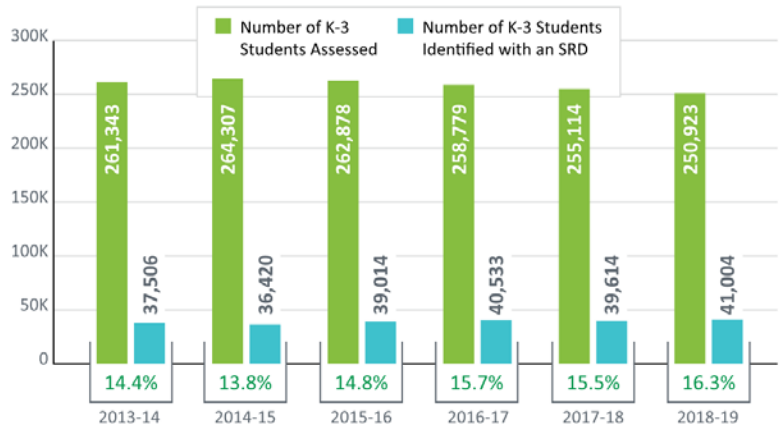
Total and Percent of Students with Significant Reading Deficiencies in 2020-2021 by Grade and School as Reported in Pilot Application

	Academy for Advanced Creative Learning		Ignacio Elementary		Singing Hills Elementary		Total	
	n	%	*n*	%	*n*	%	*N*	%
Grade K	6	21%	4	11%	5	8	15	12%
Grade 1	6	21%	18	32%	6	11	30	22%
Grade 2	5	14%	9	20%	15	25	29	20%
Grade 3	2	6%	14	38%	11	19	27	21%

For comparison, Figure 2 describes state-wide longitudinal trends in significant reading deficiency status (Colorado Department of Education, 2020). Compared to historical trends, pilot schools had higher than average identification rates for reading deficiency in 2020. In 2019, state-wide grade level averages were similar to those of the pilot school rates in that statewide identification rates in Grade K (11.5%) were about 6-8 points lower than the rates in Grades 1-3 (CDE, 2020).

Figure 2

Longitudinal Student Reading Deficiency Rates from 2013-2014 to 2018-2019



Adapted from Colorado Department of Education (2020)

Measures

Research questions for this evaluation are organized to address three overarching topics: protocol usability, implementation, and effectiveness. The sections that follow describe the measures that will be used to answer questions under each topic, including the conditions of administration.

Usability

End of Year feedback survey. In the spring of 2022, the University of Oregon will administer a survey to pilot sites that describes the extent to which teachers (a) found the professional development on the protocol to be relevant and useful, (b) the extent to which teachers found the intervention protocol easy to use; (c) the extent to which teachers perceived that the protocol met the needs of students within their classrooms; and (d) the extent to which team meetings were (i) of high quality, (ii) relevant, and (iii) useful, and (e) the extent to which participants perceived that their school administrator was engaged with and supportive of the pilot program. The survey will include the questions concerning level of protocol implementation that appeared in a survey administered by the University of Oregon in the summer of 2021, as well as a follow-up questions to be constructed by the evaluator regarding change in perceptions of the protocol. University of Oregon will provide a draft of the survey to the evaluator one month prior to administration date to confirm the adequacy of the survey questions. The evaluator may suggest revising or adding questions to ensure that sufficient information will be collected such that results can improve the capacity of the Department to provide and refine the resources for technical support, identification, and interventions. University of Oregon will provide survey results at the respondent level to the evaluator by March 31, 2022.

Monthly meeting agenda minutes. The University of Oregon is collecting monthly agenda minutes from team meetings that inquire about questions and concerns of pilot school team members. Records will be submitted to the evaluator in two waves. The first wave of data will be submitted by January 1, 2022. The second wave will be submitted by March 31, 2022.

Implementation

Project team activities. Records of project team activities will be submitted to the evaluator by the University of Oregon. Records to be submitted include the number and length of contacts; the number of trainings provided; and activity logs/checklists. Records will be submitted to the evaluator in two waves. The first wave of data will be submitted by January 1, 2022. The second wave will be submitted by March 31, 2022.

Pilot school activities. University of Oregon will gather and submit records of pilot school activities to the evaluator. Records to be submitted include documentation of (a) assessment provision for all protocol measures (i.e., screener, diagnostic assessment, family history questionnaire, teacher input

questionnaire, and intervention implementation data), including the date at which each assessment was administered; (b) teacher participation in pilot training (e.g., through attendance sheets); (c) use of the protocol; (d) data team meeting minutes; (e) MTSS-R Team and PLC Team meeting minutes; and (f) walk-through checklists describing instruction and intervention. Records will be submitted to the evaluator in two waves. The first wave of data will be submitted by January 1, 2022. The second wave will be submitted by March 31, 2022.

Multitiered System of Support for Reading Checklist (MTSS-R Checklist; University of Oregon, 2019)**.** The MTSS-R Checklist is an adaptation of the Planning and Evaluation Tool-Revised (Kame'enui & Simmons, 2003). It is similar to the Reading Tiered Fidelity Index (R-TFI; St. Martin et al., 2015) in scope, describing the implementation of standards priorities and goals; administration, organization, and communication; instruction and intervention; assessments; and professional development and job-embedded collaborative learning. Both pilot schools and comparison schools will administer the MTSS-R Checklist. Pre-pilot scores will be used as a check on the validity of the planned comparisons. Change in scores from pre- to post-pilot will also be used to gauge the extent to which the pilot was implemented. Records will be submitted to the evaluator by the University of Oregon and the Department in two waves: The first wave of data will be submitted by September 1, 2022. The second wave will be submitted by March 31, 2022.

Effectiveness

Teacher knowledge. Teacher knowledge will be measured post-intervention in pilot and comparison schools with a modified version of the Teaching Reading Essentials survey (Moats, 2006) as published in (Moats, 2009). The survey will include seven additional questions about dyslexia based on a National Center on Improving Literacy's fact sheet about dyslexia (National Center on Improving Literacy, 2020). A pre-test will not be administered because pilot activities began prior to the creation of this evaluation plan which would bias results against the pilot schools. There are no established cuts for minimally acceptable scores. However, scores will be compared to the scores reported in the validation study.

Acadience Reading (Good & Kaminski, 2018). Acadience Reading is a set of reading measures used to assess early literacy skills in Grades K-6. According to the technical manual, it can be used to identify students at risk for reading difficulties, help teachers identify areas to target instructional support, progress monitor during interventions, and examine the effectiveness of educational supports. Acadience includes six standardized subtests: First Sound Fluency (FSF), Letter Naming Fluency (LNF), Phoneme Segmentation Fluency (PSF), Nonsense Word Fluency (NWF), Oral Reading Fluency (ORF), and MAZE. The recommended administration schedule for benchmarking is described in Table 5. As described in the following section, both change in score and change in risk status will be described across schools in both conditions.

Table 5

Acadience Reading Assessment Schedule by Season and Grade

	Beginning	Middle	End
Kindergarten			
FSF	X	X	
LNF	X	X	X
PSF		X	X
NWF		X	X
Grade 1			
LNF	X		
PSF	X		
NWF	X	X	X
Grade 2			
NWF	X		
ORF	X	X	X
Grade 3			
ORF	X	X	X
MAZE	X	X	X

To examine change in beginning reading proficiency, the evaluation will examine growth on LNF in Grade K, NWF in Grade 1, and ORF in Grades 2-3 in pilot and comparison schools. Growth will be examined using the benchmark scores for students in each grade and school using (a) data from all students, and (b) data from students who were below benchmark in the fall.

Pilot program application. To be considered for the pilot program, schools had to complete an application that inquired about the number and percentage of students per grade with a significant reading deficiency, and the nature of the school's MTSS-R. Comparison schools included in the evaluation will complete a modified version of the application at the time they complete the MTSS-R checklist (i.e., Summer/Fall 2021) that inquires about the 2020-2021 school year.

Post-pilot evaluation survey. The evaluator will develop a post-pilot survey based on Colorado Department of Education's pilot program application. The Department will work with the evaluator to administer the survey to both pilot and comparison schools. In addition to the extant questions, the survey will collect information on (a) number and percentage of students who received Level 1 parent communication letters per grade per school, (b) the number and percentage of students who received Level 2 parent communication letters per grade per school (i.e., the number and percentage of students "flagged" with initial markers of dyslexia), (c) the number and percentage of students referred for comprehensive evaluation per grade per school, (d) the number and percentage of students meeting SLD eligibility criteria and requiring special education services per grade per school,

(e) the number of students who received evidence-based interventions the previous school year, and the names of the interventions. The survey will ask respondents to upload template versions of their parent notification letters. Additional questions will be completed by comparison schools that inquire about the extent to which the following measures are administered to students making inadequate progress in Tier II: re-administration of screening battery; family history questionnaire, teacher input, intervention and implementation data. Data will be submitted to the evaluator by March 31, 2022.

Design

Descriptive and explanatory nonexperimental analyses will be used to answer questions concerning the intervention's usability, implementation, and effectiveness. The decision to use descriptive analysis is informed by the study design and purpose. Under ideal circumstances, a rigorous evaluation of a program or policy would utilize either a randomized control trial (RCT) design, or a quasi-experimental method such as an interrupted time series or regression discontinuity design to promote causal inferences about the program's effects (Shadish et al., 2002). Prior to the creation of the research plan, however, the Department recruited three schools to participate in the pilot program via a voluntary response to a solicitation. The non-random method of assignment to treatment precludes the use of an RCT. Meanwhile, quasi-experimental methods are not feasible due to the limited number of schools participating in the study. Typically, a minimum of 30 clusters are recommended for analyses involving multiple levels because small sample analyses do not perform well in terms of estimation bias, Type I error rates, and relative power (Kreft, 1996). Multilevel models that perform better with small samples, meanwhile, cannot promote inferences about higher level predictors, such as treatment effects (McNeish & Stapleton, 2016).

A repeated single-case design methodology with schools at the unit of analysis would typically constitute a useful alternative to an RD when sample size is an issue (Kratochwill et al., 2020). However, implementing a single-case design requires control over treatment such that cases can serve as their own counterfactuals. The What Works Clearinghouse identifies three potential forms of experimental control: (1) treatment introduction and withdrawal, (2) iterative manipulation of treatment across different observational phases; and (3) staggered introduction of treatment across different points of time (e.g., multiple baseline; Kratochwill et al., 2020). It would have been possible to utilize one of these designs in a multi-year study. However, the COVID-19 pandemic reduced the length of this pilot to a single school year. It is neither technically feasible nor ethical to utilize one of these methods during the 2021-2022 school year. Technically, the treatment condition is too complex to make delay, manipulation, or withdrawal a viable approach within a single school year. Unlike an academic intervention, the pilot protocol cannot be simply stopped or halted at a predesignated time. Ethically, delaying, manipulating, or withdrawing the treatment mid-year is problematic because it would theoretically attenuate student and teacher learning (if not actively harm it) during what is expected to be a challenging academic year.

Given the limitations on the study design, descriptive and explanatory nonexperimental analyses will be used to describe the usability of the pilot, as well as longitudinal change in student and teacher outcomes from pre- to post-pilot. In accordance with House Bill 19-1134 which requires the Department to "evaluate the implementation of the pilot program and the effectiveness of the strategies in identifying and supporting more students in participating local education providers than were identified and supported in nonparticipating local education providers", the evaluation will describe longitudinal trends in the identification and support of students in the three pilot school, as well trends in three matched comparison schools. Comparison schools will be drawn from Colorado's Acadience users, and ideally matched on student and school characteristics (e.g., percent minority, percent EL, percent SWD, Title 1, annual progress rating) as feasible. Matching procedures will depend on data availability and suitability, but will follow IES best-practice recommendations (Van Dine et al., 2021). When possible, moving averages will be used in calculations to protect against annual fluctuations, but COVID-19 will likely complicate inferences about recent historical trends.

It should be emphasized that, even after matching schools, the sample size will not meet the minimum recommended size for the most common statistical comparisons, even when methods that ignore the clustering of students are included for consideration (VanVoorhis & Morgan, 2007). For comparison, it can be noted that Pennsylvania's state-authorized dyslexia pilot study used RD in a quasi-experiment in a multi-year, multi-cohort study of 42 schools across eight districts (Kuchle et al., 2018). Utah's state-authorized study of a pilot K-3 reading program utilized ITS with five districts and 14 schools over two years (Fratto et al., 2018). Though both of these studies included samples that were arguably small relative to their research aims, their substantially larger size and longer duration than the Colorado pilot afforded a different set of methodological options.

Because the small scope and duration of the Colorado pilot constrains design options, schools will be compared through data visualizations, potentially supplemented with repeated *t*-tests and non-parametric tests. *T*-tests and non-parametric tests also have the potential to mislead about an intervention's true effects for reasons discussed above in reference to the quasi-experimental methods (e.g., poor error rates), but are one of the few viable options because they are not dependent on sample size (Morgan, 2017). However, additional assumptions must be met, including assumptions about the normal distribution of data and the homoscedasticity of variance. The extent to which these methods will be viable is uncertain in light of research documenting (a) floor effects in some the measures to be examined (Catts et al., 2009), and (b) variable growth according to some of the student characteristics of interest, such as initial ability level (e.g., Fien et al., 2010) and race and socioeconomic status (e.g., Little, 2017; Willoughby et al., 2019). Overall, however, these methods are comparable to those used in other state dyslexia pilot studies. For instance, the state of Washington's dyslexia pilot study examined growth over one year in a variety of student reading outcomes in four districts with 10 schools through visual analysis (Young, 2009). Utah's pilot study report relied heavily on chi-squared difference testing (Fratto et al., 2018).

Analysis and Interpretation Plan

Table 6

Summary of Analysis and Interpretation Plan

Question	Data	Sites	Collection Timeline	Responsibility	Analysis	Expected Outcome
Usability						
1 2 3 4	EOY Survey Monthly minutes	Pilot Pilot	By March 31 2022 Monthly	• Evaluator to review/revise • UO to collect and deliver	Quantitative Description	• Higher ratings are better • Identify remaining TA needs
Implementation						
5	Project team activities	Pilot	January/March 2022	• UO to collect and deliver	Quantitative Description	• PD was delivered as planned
6	Pilot school activities	Pilot	January/March 2022	• CDE to collect and deliver	Quantitative Description	• Program and protocol implemented with fidelity
7	MTSS-R Checklist	Both	ASAP/March 2022	• UO and CDE to collect and deliver	Quantitative Description or Comparison	• Pilot schools will improve more than comparisons
Effectiveness						
8	Post-pilot survey	Both	March 31 2022	• Evaluator to create and disseminate with CDE	Quantitative Description or Comparison	• Pilot schools will score higher than comparison schools
9	EOY survey, teacher self-rating	Pilot	March 31 2022	• Evaluator to review/revise • UO to collect and deliver	Quantitative Description or Comparison	• Pilot schools will improve in self-rating
10	Acadience scores	Both	March 31 2022	• CDE to arrange data transfer	Quantitative Description or Comparison	• Pilot schools will have faster growth in ability and risk reduction
	EOY survey	Pilot	March 31 2022	• Evaluator to review/revise • UO to collect and deliver	Quantitative Description or Comparison	• Implementation levels in pilot schools increase
	Post pilot survey	Both	March 31 2022	• Evaluator to create and disseminate with CDE	Quantitative and Qualitative Description	• More improvement in pilot schools, but schools may not have data

This section describes the analysis and interpretation plan for the evaluation research questions. In the plan, which is summarized in Table 6, research questions that use the same measures and methods are grouped together. Because the pilot and evaluation is a one-year project, there are no plans to publish interim results, as might be done in a multi-year study. It is therefore recommended that all key stakeholders have a chance to review the plan prior to the dissemination of any study findings.

Usability

RQ1: To what extent did teachers in the pilot schools find the professional development on the protocol to be relevant and useful?

RQ 2: To what extent did teachers in the pilot program perceive that the intervention protocol would meet the needs of their students?

RQ 3: To what extent did teachers find the intervention protocol easy to use?

RQ 4: To what extent did teachers have a positive perception of the protocol?

Factors such as the perceived need for an intervention, the perceived benefits of an intervention, feelings self-efficacy and self-proficiency, and program compatibility and adaptability have all been observed to influence the program implementation (Durlak & DuPre, 2008). In general, practitioners who recognize a need for an intervention and are confident and knowledgeable enough to implement it will make a greater effort to do so. To promote a better understanding of program usability and to improve the capacity of the Department to provide and refine the resources for technical support, identification, and interventions, the evaluation will summarize trends in the End of Year feedback survey and monthly meeting minutes. The survey questions will be structured such that they correspond to the four research questions above. Given that comparison schools will not complete these measures and there are no pre-established cut-offs for rating usability for these measures, data will be used to improve the capacity of the Department to provide and refine the resources for technical support, identification, and interventions, rather than attempting to classify the pilot as usable or not usable. Results will be described at the group level (i.e., the three pilot schools) unless results warrant finer-grained reporting (e.g., cross-school heterogeneity in trends).

Implementation

RQ5. To what extent did teachers receive training as intended?

The purpose of RQ4 is to describe the extent to which pilot teachers received training in accordance with the protocol that University of Oregon developed with the Department, with the understanding that the pilot cannot be implemented in accordance with the University's original proposal due to COVID-related challenges, and that subsequent changes to the protocol may be necessary because of the pandemic. To answer the question, project team activities and pilot team activities (i.e., attendance sheets) will be quantitatively described. Where relevant, the total number

of activities will be described as portion of the total agreed upon number of activities (e.g., number of trainings provided number of trainings intended to be provided). As with the usability data, there are no pre-established cut-offs for determining whether an intervention training was provided at an acceptable level. Results will be interpreted with the general understanding that greater rates of activity are desirable. Results will be described at the group level (i.e., the three pilot schools) unless results warrant finer-grained reporting (e.g., cross-school heterogeneity in trends).

RQ6: To what extent was the protocol administered by teachers?

The purpose of RQ6 is to evaluate the extent to which pilot teachers implemented the protocol in accordance with the training the received from the University of Oregon. To answer the question, pilot school activities will be quantitively, and where appropriate, qualitatively described. The measures of pilot school activities to be collected provide different types of information about protocol implementation. The records of assessment provision, for example, provides a shallow but objective measure of protocol implementation insofar as there are clear expectations about what measures should be administered and under what circumstances. Meeting minutes provide deeper, but more subjective, insights into protocol administration. Trends in meeting minutes will be described to improve the capacity of the Department to provide and refine the resources for technical support, identification, and interventions (e.g., through a consideration of how time was spent across meetings). Results will be described at the group level (i.e., the three pilot schools) unless results warrant finer-grained reporting (e.g., cross-school heterogeneity in trends).

RQ7: To what extent did implementation of a multitiered system of support in reading improve at the school level from beginning to end of year?

RQ7 serves two key purposes. First, through a consideration of improvement on the MTSS-R checklist, the evaluation can determine whether the schoolwide reading model generally improved from pre- to post-intervention. Second, comparing MTSS-R levels at the pilot schools to the comparison schools can help validate the appropriateness of the comparison schools that were selected to promote inferences about pilot effectiveness. As part of the evaluation, pilot schools will be matched to comparators using school data that is routinely collected by the Colorado Department of Education. However, it is possible that comparator schools will differ from the pilot in schools in terms of factors that are more proximal to student reading outcomes despite attempts to find suitable matches. Observing the initial status and growth on the MTSS-R checklist can provide important context for understanding the pilot's effectiveness. That said, it should be noted that results from the MTSS-R checklist may biased against the pilot schools because (a) pilot schools began receiving professional development prior to the administration of the MTSS-R checklist, and (b) comparison schools will complete the measure later than the pilot schools (e.g., because they must be identified before they complete the measure). Furthermore, this approach does not account for the accuracy of the initial self-rating. It is possible that pilot schools may come to believe that they

overrated themselves on the pre-pilot measure because of the training they receive. Results should be interpreted accordingly.

Effectiveness

RQ8: How does pilot teacher knowledge of the science of reading and markers of dyslexia compare to that of teachers in comparison schools?

RQ 7 is meant to provide insight about the extent to which the pilot was successful in improving teacher knowledge about the science of reading and dyslexia. Ideally, this question would compare growth from pre- to post- intervention in the pilot and comparison schools. However, at the time of this document's creation, the pilot schools have already received professional development on these topics, which would bias the results against the pilot schools. As an alternative, end of year scores on the teacher knowledge measure will be compared across the three schools. The primary limitation of this approach is that it does not account for initial status, which prevents inferences about pilot effectiveness. For instance, if pilot schools begin the year with substantially lower levels of knowledge than the comparison schools, it is possible that they will also end the year with lower levels of knowledge, even if the pilot is successful in helping the teachers grow more rapidly than teachers in comparison schools.

RQ9: To what extent did the implementation of evidence-based practices improve from beginning to end of year?

RQ 9 is meant to describe the extent to which the implementation of evidence-based practices improved from pre- to post-pilot. It will not be possible to collect observational data that directly addresses this question because the University of Oregon will not be able to observe instruction. Therefore, pilot teacher self-report of level of implementation of the dyslexia protocol will be compared from pre- to post- intervention. At the time of this writing, pilot school teachers rated themselves at "intermediate" level. It is assumed that teacher self-ratings will improve from pre-post. However, it is also possible that teachers will come to believe that they over-estimated their ability in the pre-pilot survey. In this scenario, teachers may continue to rate themselves at the intermediate level or lower, but the pilot could be deemed successful insofar as it promoted more accurate self-assessment. Therefore, the evaluator will work with the University of Oregon to construct an open-ended question on the survey that probes for this possibility.

RQ10: To what extent did the pilot improve student outcomes?

RQ 10 will examine the extent to which the intervention protocol improved student outcomes. It is expected that pilot schools will increase their Tier 1 and Tier 2 Acadience scores from fall to spring at a faster rate than their comparators (i.e., steeper slope), assuming the matched comparisons are valid. Average slopes will be calculated for all students in the pilot in each school, and for students at-risk. It is expected that rates of risk for dyslexia (as operationalized with Acadience cut-scores for risk) will decline in the pilot schools (i.e., % of students at risk in the spring will be lower in the spring than

in the fall), and that the rate of reduction will be greater than that of the comparison schools (i.e., steeper negative average slope).

The evaluation will also describe the number of students referred for comprehensive evaluation and the number of students meeting SLD eligibility requirements and requiring special education services from pre- to post-pilot. However, it is difficult to project the direction of change because the percent of students with significant reading deficiencies substantially varies by year and grade in the pilot schools. For instance, Academy of Advanced and Creative Learning reported that about 20% of students had significant reading deficiencies in Kindergarten and Grade 1, which is 6% higher than the state average. However, they also reported that only 6% of students had significant reading deficiencies in Grade 3, which is 13% lower than the state average in 2019 (Colorado Department of Education, 2020). Given the heterogeneity in baseline conditions even within schools, it seems doubtful that the dyslexia pilot will have an effect with a singular direction and magnitude within or across schools. Moreover, it is unclear what "effectiveness" would mean in this context. For example, if the cross-grade heterogeneity in rates of reading deficiency at Academy of Advanced and Creative Learning is due successful intervention practices (e.g., school intervenes early; parents intervene more intensely outside of school), it would suggest a very different trajectory than if it is due to unreliable classifications. Therefore, change will be described with the intention to improve the capacity of the Department to provide and refine the resources for technical support, identification, and interventions rather than to evaluate program success or failure.

Dissemination Plan

This section describes the plan for disseminating products and information related to the evaluation. As described in Table 7, four distinct dissemination products will be developed during the evaluation: (1) the evaluation plan, (2) a summary of initial results and project status, (3) an in-depth summary of results, and (4) the Department's final report. Two documents (i.e., the evaluation plan and initial results and status report) will have drafts associated with them with distinct dissemination strategies. The section that follows describes the rationale for the dissemination strategy for each product.

Rationale by Product

Evaluation Plan. For the sake of transparency and public accountability, it is recommended that key stakeholders review the *final* version of this document (i.e., the evaluation plan). Stakeholders have a right to comment on decisions that might affect the likelihood of obtaining useful information (CDC, 2011). Obtaining stakeholder input will also increase the likelihood that evaluation results are used (CDC, 2011). It is important to seek input on the evaluation plan because there will not be another opportunity to receive input that could affect the evaluation's execution. Although an initial results and project status report will be made available to stakeholders in the Winter/Spring of 2022, it will not be feasible to make major adjustments to the evaluation at that time.

Initial results and status report. The purpose of the initial results and status report is primary to serve as a helpful "check-in." The report will provide an opportunity to describe data collection, as well as any logistical challenges originating from the pilot program, or from external factors, such as new stay-at-home orders. Descriptive information will be included in the status report to gauge data completeness. No evaluative inferences will be made about the pilot in the document. The draft of the report may include sensitive information, for instance, challenges within specific schools or with specific products. Therefore, the draft will not be publicly disseminated. The final version of the report will be identical to draft report, but with sensitive information removed and edits made based on feedback from the Department and the University of Oregon.

Table 7

Dissemination Plan Summary

Product	Method	Timing	Archiving
Evaluation Plan Draft	• Reviewed by CDE and University of Oregon • Not otherwise published or disseminated	August 2021	• Archived by CDE, evaluator, and University of Oregon
Evaluation Plan (this document)	• Reviewed by all key stakeholder groups prior to plan execution	ASAP	• Archived by CDE, evaluator, and University of Oregon • Disseminated after stakeholder review only upon request and at CDE discretion
Initial Results and Project Status Report Draft	• Short presentation or write-up reviewed by CDE and University of Oregon • Not otherwise published or disseminated	Winter/Spring 2022	• Archived by CDE, evaluator, and University of Oregon as desired
Initial Results and Project Status Report	• Short presentation or write-up reviewed by stakeholders	Winter/Spring 2022	• Archived by CDE, evaluator, and University of Oregon as desired • Disseminated after stakeholder review only upon request and at CDE discretion
Evaluator's Write-Up	• Detailed description of methods, results, and conclusions, with (a) a summary of the winter/spring status report if appropriate, (b) guidance on the interpretation of results, and (c) suggestions for final reporting, reviewed by CDE and University of Oregon	Spring/Summer 2022	• Archived by CDE, evaluator, and University of Oregon • Not disseminated after review
CDE's Report on Evaluation	• Description of methods, results, and conclusions as CDE deems appropriate	Winter 2022	• Publicly available • Disseminated in accordance with state/CDE policy

Evaluator's Write-Up. The evaluator will write up the results of the evaluation in accordance with their scope of work. The confidential write-up will include an in-depth description of methods, results, and conclusions, as well as (a) a summary of the winter/spring status report if appropriate, (b) guidance on the interpretation of results, and (c) advice and suggestions for final reporting. It will be reviewed and retained by the Department, the evaluator, and University of Oregon.

The evaluator's write-up will not be made publicly available because it is broad in purpose and will likely contain sensitive information that has the potential to be misused. According to House Bill 19-1134, the Department has an interest in understanding usability issues so that it can refine the resources for technical support, identification, and interventions, as necessary, and disseminate the resources to all local education providers in the state. The Department must therefore have access to sensitive information, such as the specific strengths and weaknesses of a product as described by teaching personnel; or information that has the potential to cause harm to personnel at the pilot or comparator school sites, such as non-compliance with the protocol. The public does not have an interest in the full scope of information that will be reported in the write-up. Furthermore, some of the information that will be included has the potential to be (a) be misused by proponents and critics of the pilot, and/or (b) misinterpreted by the general public. For example, the results from this study will have a limited ability to promote generalization about the pilot's effectiveness to other times or contexts. Similarly, the evaluation's usability findings will describe what was observed, but not what may have been observed if alternative protocols or products were employed. Given these caveats, the in-depth write-up should not be made publicly available or disseminated apart from the CDE and University of Oregon pilot personnel.

CDE's Report on Evaluation. The Department's Report may utilize text and information from the Evaluator's Write-Up as desired for reporting, barring the stipulations described above. An important consideration the Department will need to make prior to publication of the Report concerns data aggregation and reporting. Typically, state pilot reports identify participating schools by name, but are able to mask sensitive information by reporting results by condition or group. The extent to which this will be feasible in the proposed study is unclear. Aggregating results by condition is apt to be inappropriate for some measures because of differences in initial starting conditions. For example, rates of significant reading deficiency in Grade 3 could increase for Academy of Advanced and Creative Learning, but decrease in Ignacio Elementary, due to their very different starting points. Averaging these trends could make it appear that no change occurred, when in reality, both schools changed, but in opposite directions.

The Department should also consider dissemination plans for the final report and communicate them to the evaluator prior to the creation of the evaluator's write-up. Because the evaluation will be better able to provide information about how the Department can provide technical assistance than program effectiveness, it may be in the Department's interest to use research findings beyond the

final report as part of their technical assistance efforts. Planning in advance for this possibility will increase the likelihood that the evaluation can contribute to these efforts.

Finally, the Department, evaluator, and University of Oregon should establish follow-up and contingency plans that address the misuse of the evaluation's findings. Despite prevention efforts, misuse of research findings can occur for a variety of reasons (CDC, 2011). Motivated stakeholders may seek to undermine a program by emphasizing negative findings. Proponents may seek to generalize positive findings beyond what the research supports. To prevent such misuse, a point of contact should be designated who will be responsible for communicating with the public and interested parties, answering questions about the results, and addressing misuse of evaluation results when it occurs.

References

Catts, H. W., Petscher, Y., Schatschneider, C., Sittner Bridges, M., & Mendoza, K. (2009). Floor effects associated with universal screening and their impact on the early identification of reading disabilities. *Journal of Learning Disabilities*, *42*(2), 163–176. https://doi.org/10.1177/0022219408326219

Centers for Disease Control and Prevention. (2011). *Developing an Effective Evaluation Plan: Setting the Course for Effective Program Evaluation*. Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office of Smoking and Health; Division of Nutriion, Physical Activity, and Obesity. https://www.cdc.gov/obesity/downloads/cdc-evaluation-workbook-508.pdf

Colorado Department of Education. (2020). *READ Report: 2020 Annual Report to the Colorado Read Act*. Colorado Department of Education. https://www.cde.state.co.us/cdedepcom/readactreport

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, *41*(3–4), 327–350. https://doi.org/10.1007/s10464-008-9165-0

Fien, H., Nelson, N. J., Smolkowski, K., Kosty, D., Pilger, M., Baker, S. K., & Smith, J. L. M. (2021). A conceptual replication study of the Enhanced Core Reading Instruction MTSS-Reading Model. *Exceptional Children*, *87*(3), 265–288. https://doi.org/10.1177/0014402920953763

Fien, H., Park, Y., Baker, S. K., Stoolmiller, J. L. M., & Kame'enui, E. J. (2010). An examination of the relation of nonsense word fluency initial status and gains to reading outcomes for beginning readers. *School Psychology Review*, *39*(4), 631–653.

Fratto, K., Throndsen, J., Suddreth, D., & Nielsen, D. (2018). *Report to the Education Interim Commitee: Interventions for Reading Difficulties Pilot Program Report*. Utah Department of Education. https://le.utah.gov/interim/2018/pdf/00004483.pdf

Good, R. H., & Kaminski, R. A. (2018). *Acadience Reading*. Dynamic Measurement Group. https://acadiencelearning.org/wp-content/uploads/2020/01/AcadienceReading_AssessmentManual.pdf

Kame'enui, E. J., & Simmons, D. (2003). *Planning and evaluation tool-Revised*.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., & Rindskopf, D. M. (2020). *Single-case designs technical documentation.* What Works Clearinghouse. http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kreft, I. G. G. (1996). *Are multilevel techniques necessary? An overview, including simulation studies.* [Unpublished manuscript]. California State University.

Kuchle, L., Brown, S., & Coukoulis, N. (2018). *Annual Evaluation Report for the Pennsylvania Dyslexia Screening and Early Literacy Intervention Pilot Program: Pilot Year 2, 2016-2017 School Year*. American Institutes for Research. https://files.eric.ed.gov/fulltext/ED582923.pdf

Little, M. (2017). Racial and socioeconomic gaps in executive function skills in early elementary school: Nationally representative evidence From the ECLS-K:2011. *Educational Researcher*, *46*(2), 103–109. https://doi.org/10.3102/0013189X17698700

McNeish, D. M., & Stapleton, L. M. (2016). The Effect of Small Sample Size on Two-Level Model Estimates: A Review and Illustration. *Educational Psychology Review*, *28*(2), 295–314. https://doi.org/10.1007/s10648-014-9287-x

Moats, L. (2009). Knowledge foundations for teaching reading and spelling. *Reading and Writing*, *22*(4), 379–399. https://doi.org/10.1007/s11145-009-9162-1

Moats, L. C. (2006). *Teacher reading essentials*. Sopris West Educartional Services.

Moore, H., Heinemeier, S., Gillis, M., Shaver, D., Schaff, J., & Lammert, J. (2021). *Evaluations During Disruptions: Course Corrections and Other Considerations*. Center to Improve Program and Project Performance. https://osepideasthatwork.org/sites/default/files/CIPP-Evaluation-During-Disruptions-Report-Full.pdf

Morgan, C. J. (2017). Use of proper statistical techniques for research studies with small samples. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, *313*(5), L873–L877. https://doi.org/10.1152/ajplung.00238.2017

National Center on Improving Literacy. (2020). Myths vs. Facts: Breaking Down the Truth About Dyslexia. *Washington, D.C.: U.S. Department of Education, Office of Elementary and Secondary Education, Office of Special Education Programs, National Center on Improving Literacy*. https://improvingliteracy.org/brief/understanding-dyslexia-myth-vs-facts

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.

Smith, J. L. M., Nelson, N. J., Smolkowski, K., Baker, S. K., Fien, H., & Kosty, D. (2016). Examining the Efficacy of a Multitiered Intervention for At-Risk Readers in Grade 1. *The Elementary School Journal*, *116*(4), 549–573. https://doi.org/10.1086/686249

St. Martin, K., Nantais, M., Harms, A., & Huth, E. (2015). *Reading Tiered Fidelity Inventory (Elementary Level Edition)*. Michigan Department of Education, Michigan Integrated Behavior and Learning Support Initiative.

Van Dine, D., Randel, B., & Kulte, M. (2021). *A Guide to Identifying Similar Schools to Support School Improvement* (REL 2021–09). Institute of Education Sciences. https://files.eric.ed.gov/fulltext/ED613435.pdf

VanVoorhis, C. R., & Morgan, B. L. (2007). Understanding power and rules of thumb for determining sample sizes. *Tutorialsin Quantitative Methods for Psychology*, *3*(2). file:///C:/Users/BRIANG~1/AppData/Local/Temp/VanVoorhisMorgan-2007.pdf

Willoughby, M. T., Wylie, A. C., & Little, M. H. (2019). Testing longitudinal associations between executive function and academic achievement. *Developmental Psychology*, *55*(4), 767–779. https://doi.org/10.1037/dev0000664

Young, C. (2009). *Lorraine Wojahn Dyslexia Reading Pilot Program*. https://app.leg.wa.gov/ReportsToTheLegislature/Home/GetPDF?fileName=Lorraine%20Wojahn%20Dyslexia%20Pilot%20Reading%20Program_1f056327-cf4a-4c7e-a2bb-20b503b3af2f.pdf