Evaluation Plan for Colorado Department of Education's Dyslexia Pilot Program Year 2: 2022-2023

2022

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 was 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, three pilot sites received training and coaching to provide support to young students who may demonstrate the early markers for dyslexia. Following the pilot program's first year, the Colorado Department of Education evaluated 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. In 2022, the Colorado State Board of Education approved a second year of the Dyslexia Pilot Program. The Department recruited three new schools to take part in the pilot during the 2022-2023 school year. This document describes the plan to evaluate the second year of the Colorado State Dyslexia 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 Colorado General Assembly further created a dyslexia pilot program. The purpose of the program was 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, three pilot schools received training and coaching from the University of Oregon to provide support to young students who demonstrated early markers for dyslexia. The results of the pilot evaluation suggested that educators generally found the training to be high in quality, relevance, and usefulness. There was less agreement that the dyslexia protocol was easy to use, with most educators agreeing that additional training would make it easier to implement the protocol. Evaluation results suggested that the pilot was successful at improving levels of communication with parents about dyslexia, keeping records about the delivery and receipt of evidence-based interventions, and comprehensive evaluation practices. Results concerning the effect of the pilot on student reading development, rates of comprehensive evaluations for special education, and rates of special education eligibility were inconclusive.

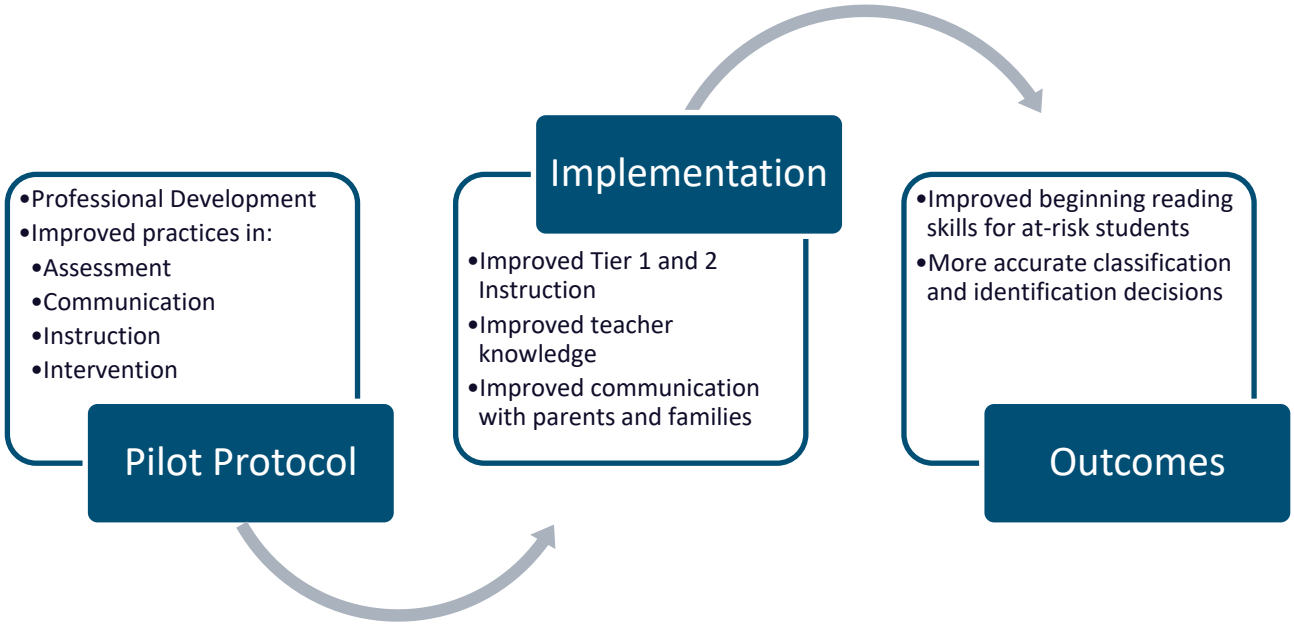
In the spring of 2022, a second year of the pilot was established. Three new schools were selected by the Department to participate in the pilot. Following the second year of the pilot program, the Department is required to evaluate the program's implementation. This document describes the plan for the evaluation that will inform the Department's report to the State Board and the Education Committees of the Senate and House.

Evaluation Plan Overview

This document is written under the assumption that the primary intended user of 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 in the *implementation* stage. An evaluation of the pilot's first year provided descriptive information about how the pilot program was implemented and the resources needed to support implementation. The Department and the University of Oregon will use this information to improve the implementation of the pilot program in its second year.

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.

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. To what extent do pilot teachers possess knowledge of the science of reading and markers of dyslexia?	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

Methods

The research questions described in this evaluation plan (Table 1) will be answered using data gathered during the 2022-2023 pilot evaluation. 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 in the spring of 2022. 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. Six schools applied. Applications were scored on the following criteria: Leadership, PD & Training, Implementation of Evidence Based Practices, and Commitment to the Pilot. One school was excluded because did not use the required assessment system. The three highest scoring schools were admitted to the pilot.

School Characteristics

Table 2

Estimated School Characteristics of Pilot Sites

School	Setting	Student Total	Title 1	Student:Teacher	FRPL	Minority	EL	SWDs
School 1	Rural: Remote	250	Yes	13.16	35%	13%	3%	21%
School 2	City: Large	264	No	19.06	42%	33%	2%	8%
School 3	City: Large	218	Yes	12.11	82%	67%	15%	14%

Table 2 describes estimated characteristics of the three schools that applied and were accepted into the pilot. Collectively, the schools serve roughly 732 students, about 480 of whom can be expected to participate in the pilot study. Assuming 8 teachers participate per grade, there will be approximately 96 teachers in the pilot schools. Figure 1 provides snapshots of school performance in 2019, the latest available data on school performance.

Figure 1

Snapshot of school performance in 2019

Teacher Characteristics

The 2022 Teaching and Learning Conditions Survey described educator experience levels for two of the pilot schools. At School 1, about 32% of educators were in their first year, 9% of had 2-3 years of experience, 9% had 4-5 years of experience, 14% had 6-10 years of experience, 18% had 5-10 years of experience, and 18% had more than 20 years of experience. At School 2, about 25% of educators were in the first year, 15% had 2-3 years of experience, 20% had 4-5 years of experience, and 25% had 6-10 years of experience. Results were not available for School 3.

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 2022, teachers at School 1 viewed their conditions more favorably (86%) than the state average (72%). Teachers at School 2 similarly viewed their school more favorably than the state average (82%). Results were unavailable for School 3.

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 Measures

End of Year survey. The usability of the pilot protocol and related materials will be assessed via a survey that was co-developed by the evaluator and the University of Oregon. The survey assesses (a) the quality, relevance and usefulness of the protocol, (b) the extent to which teachers find the intervention protocol easy to use; (c) the extent to which teachers perceive that the protocol meets the needs of students within their classrooms; (d) the quality, relevance and usefulness of team meetings; (e) the extent to which participants perceive that their school administrator is engaged with and supportive of the pilot program; and (f) teacher knowledge of dyslexia. The University of Oregon will administer the knowledge scale to pilot teachers in the summer/fall of 2022, and the full survey to the pilot schools in the spring of 2023.

Implementation Measures

Project team activities. To assess trends in implementation, the University of Oregon will submit records of project team activities to the evaluator. Records that may be submitted include pre-pilot needs assessment, monthly meeting minutes, the number and length of contacts; the number of trainings provided; and activity logs/checklists. Submitted documents will be summarized by the evaluator with the intention of understanding how implementation occurred and improving the Department's capacity to refine the resources for technical support.

Pilot school activities. The University of Oregon will submit records of pilot school activities to the evaluator to provide additional information about trends in implementation. Records that may be submitted include (a) records of assessment provision of all measures; (b) records of teacher participation in the pilot training; (c) records of the 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. Submitted documents will be summarized by the evaluator with the intention of understanding implementation to ultimately improve the Department's capacity to refine the resources for technical support.

Multitiered System of Support in Reading Implementation Checklist (MTSS-R Checklist; National Center on Improving Literacy, 2020)**.** MTSS-R implementation will be assessed via a self-assessment checklist to be completed by school leaders. The MTSS-R Implementation Checklist is a schoolwide measure that rates MTSS-R implementation across five elements: Core Reading Instruction and Intervention; Data Use; Professional Development and Coaching; School Leadership; and Mutual Support Involving Families and the School. The measure contains about 270 items, yielding a rich description of the school environment. Most items consist of a descriptive statement, such as, "All families receive our master schedule of reading instruction, which includes contact information for them to learn more and ask questions." Qualified school leaders rate each item by providing a score between 0 and 2, where 0 indicates the practice has not been implemented, a 1 indicates partial implementation, and a 2 indicates full implementation. Scores are then averaged within each element. For the purpose of the evaluation, element scores will also be summed to yield a total score that ranged from 0 to about 10, with higher scores indicating higher implementation levels. Pilot schools and comparison schools will complete the measure in the fall and spring. The Department will email the measure and instructions to an appropriate contact at all participating schools.

Effectiveness Measures

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. 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 evaluation survey. Apart from the Acadience scores, all effectiveness data will be collected via an end-of-year survey. The survey will ask school leaders to provide the 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 eligibility criteria for 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, and (f) a description of the data sources used in comprehensive evaluations for dyslexia. To administer the survey, the Department will email a link to an appropriate contact at all participating schools.

Design

Descriptive, nonexperimental analyses will be used to answer questions about the pilot's usability, implementation, and effectiveness. 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 Evaluation 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 precluded the use of an RCT. Meanwhile, quasi-experimental methods were not feasible due to the limited number of schools participating in the study (Kreft, 1996), the focus on higher level predictors, such as treatment effects (McNeish & Stapleton, 2016), and the short duration of the pilot. The sample size in this study does not meet minimum recommendations for relevant non-parametric analyses (Morgan, 2017), and for some key measures, is not sufficiently powered to detect group differences. Assuming an effect size of .25 on student outcomes (e.g., Smith et al., 2016), the use of demographic and pretest covariates with an R^2 of .74 and a level-2 ICC of .20 for Colorado Grade 3 reading (e.g., Hedges & Hedberg, 2014), an HLM analysis would only achieve a power of .12 (Tipton et al., 2022). Therefore, the effectiveness analyses will simply provide the means and standard deviations for each measure, grade, and school for pilot schools and a group of comparison schools. The comparison schools will be identified using Mahalanobis distance matching on five-year averages of the following school characteristics: total number of students, Title 1 status, charter status, student to teacher ratio, percent of minority students, percent of English learners, and percent of special education students.

Analysis and Interpretation Plan

Table 6

Summary of Analysis and Interpretation Plan

Question	Data	Sites	Collection Timeline	Responsibility	Analysis	Expected Outcome
Usability						
1 2	EOY Survey	Pilot	April 2023	• UO to administer	Quantitative Description	• Higher ratings are better • Identify remaining TA needs
3 4	Monthly minutes	Pilot	Monthly			
Implementation						
5	Project team activities	Pilot	January/April 2023	• UO to collect and deliver	Quantitative Description	• PD was delivered as planned
6	Pilot school activities	Pilot	January/April 2023	• UO to collect and deliver	Quantitative Description	• Program and protocol implemented with fidelity
7	MTSS-R Checklist	Both	ASAP/April 2023	• CDE will administer via email	Quantitative Description	• Pilot schools will improve more than comparisons provided a similar pre-test
Effectiveness						
8	EOY survey	Pilot	April 2023	• UO to administer	Quantitative Description	• Pilot schools will improve in knowledge
9	Post-pilot survey	Pilot	Spring 2022/ April 2023	• CDE will administer survey via email	Quantitative Description	• Pilot schools will show greater adherence to the protocol (e.g., more parent letters sent)
10	Acadience scores	Both	April 2023	• CDE to arrange data transfer	Quantitative Description or Comparison	• Pilot schools will have faster growth in ability and risk reduction
	Post-pilot survey	Both	April 2023	• CDE will administer survey via email	Quantitative Description	• No hypotheses concerning student outcomes (e.g., SWD rates)

This section describes the analysis and interpretation plan for the evaluation research questions. Table 6 lists the research questions, methods, and measures. 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

RQ 1: 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 of 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, 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

RQ 5. To what extent did teachers receive training as intended?

The purpose of RQ 5 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. 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 unless results warrant finer-grained reporting.

RQ 6: To what extent was the protocol administered by teachers?

The purpose of RQ 6 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. Results will be described at the group level unless results warrant finer-grained reporting.

RQ 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?

RQ 7 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.

Effectiveness

RQ 8: To what extent do teachers in the pilot schools possess knowledge of the science of reading and dyslexia?

RQ 8 is meant to provide insight about the extent to which teachers in the pilot schools possess knowledge of the science of reading and dyslexia following the pilot. Ideally, this question would compare growth from pre- to post- intervention in the pilot and comparison schools. However, it is not feasible to administer a knowledge assessment to teachers in the comparison school. Therefore,

this question will be answered with a simple pre/post description of teacher knowledge in the pilot schools.

RQ 9: 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, the evaluation will compare differences in markers of protocol implementation, such as a description of the measures used when making dyslexia-related classifications, and the number of letters send to parents indicating that their child has markers of dyslexia.

RQ 10: 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 assuming similar intercepts. If data are available, the evaluation will also compare rates of risk for reading difficulties using Acadience's risk measures. It is expected that rates of risk for dyslexia will decline in the pilot schools, and that the rate of reduction will be greater than that of the comparison schools. It should be noted that data availability is a concern because schools may not finish screening before the due date for the evaluation report. In such a case, the evaluator will work with the Department to develop contingency plans for analysis and reporting.

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. Moreover, it is unclear what "effectiveness" would mean in this context. 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 2023, 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 (this document)	• Reviewed by CDE and University of Oregon • Not otherwise published or disseminated	August 2022	• Archived by CDE, evaluator, and University of Oregon
Evaluation Plan	• Reviewed by all key stakeholder groups prior to plan execution	August 2022	• 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 2023	• 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 2023	• 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 2023	• 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 2023	• 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, averaging trends across units with different initial scores 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

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

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

Hedges, L., & Hedberg. (2014). Intraclass correlations and covariate outcome correlations for planning 2 and 3 level cluster randomized experiments in education. *Evaluation Review*.

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

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

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

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

Tipton, E., Spybrook, J., Miller, K., Weiss, M., & Hedges, L. (2022). ***The Generalizer***. https://thegeneralizer.org/about