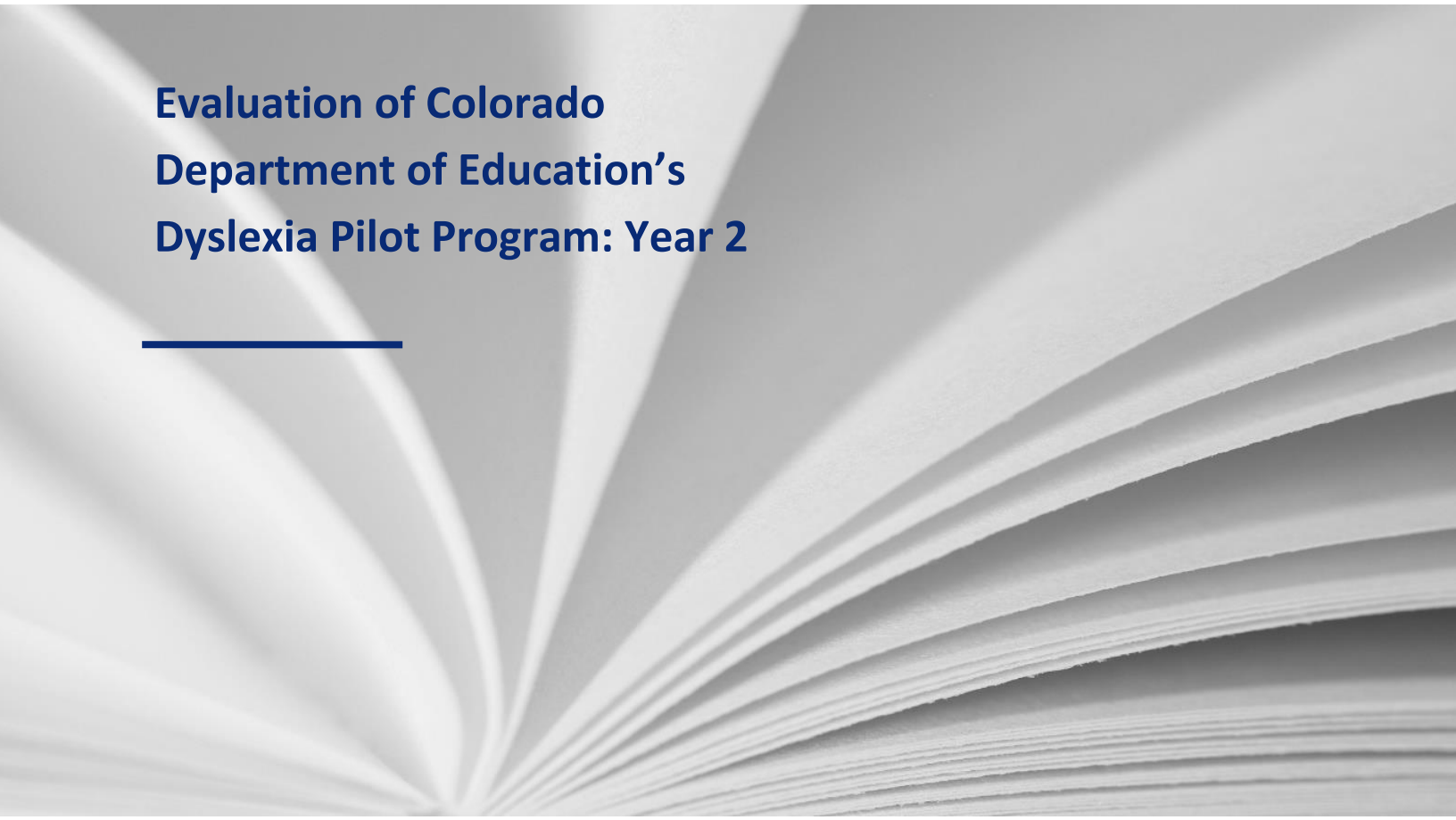
Evaluation of Colorado Department of Education's Dyslexia Pilot Program: Year 2



2022-2023

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, 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. This document describes the evaluation of the second year of the Colorado State Dyslexia Pilot Program.

Table of Contents

Introduction

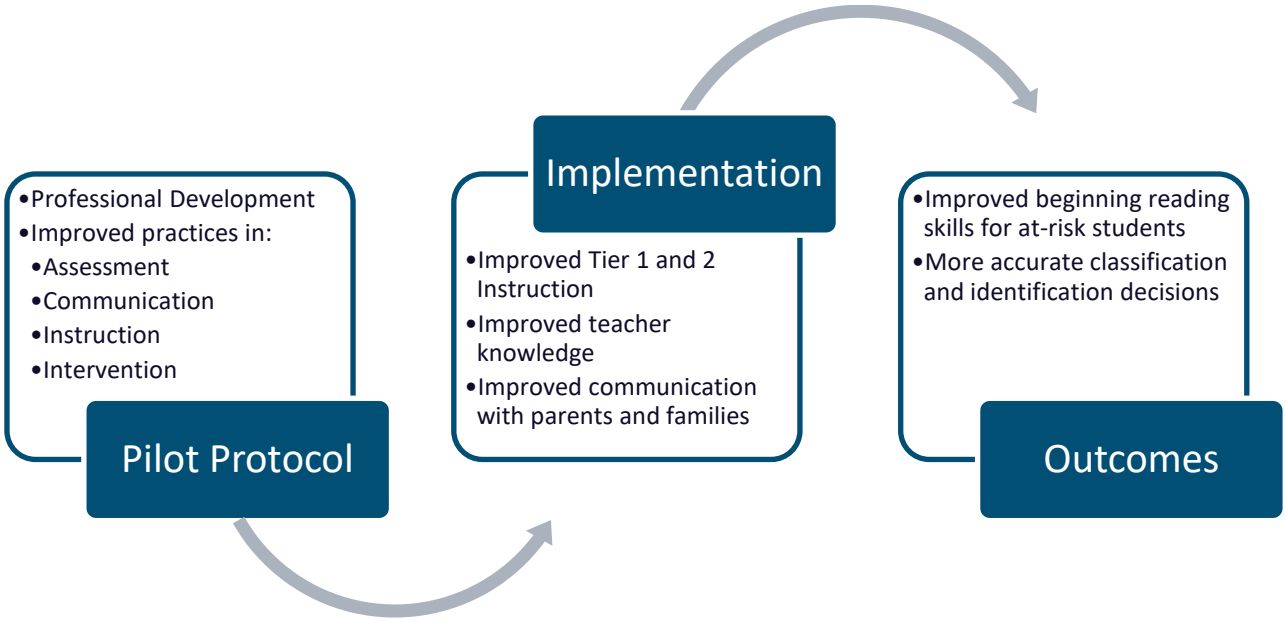
During the 2019 legislative session, the Colorado General Assembly created a dyslexia pilot program through House Bill 19-1134. According to the preamble of the bill, parents of children identified as having dyslexia had voiced concerns related to the adequacy and effectiveness of the methods and tools 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 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 established a pilot program through which the Department of Education would work 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 and strengthen the ability of local education providers throughout the state to identify and effectively support students with dyslexia.

The first year of the dyslexia pilot program ran from 2021 to 2022. Based on its promising results, the Department implemented a second year of the program, placing greater emphasis on understanding how to effectively implement and scale the program. This document describes the results of an evaluation aimed at helping the Department refine the resources needed for technical support, identification, and interventions; and disseminate the resources to local education providers.

Evaluation Report

In accordance with the bill that created the dyslexia pilot evaluation, this document was 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, which led the implementation of the pilot. The following sections describe: (a) the dyslexia pilot program's theory of change, (b) the dyslexia pilot protocol, and (c) the focus of the evaluation.

Theory of Change

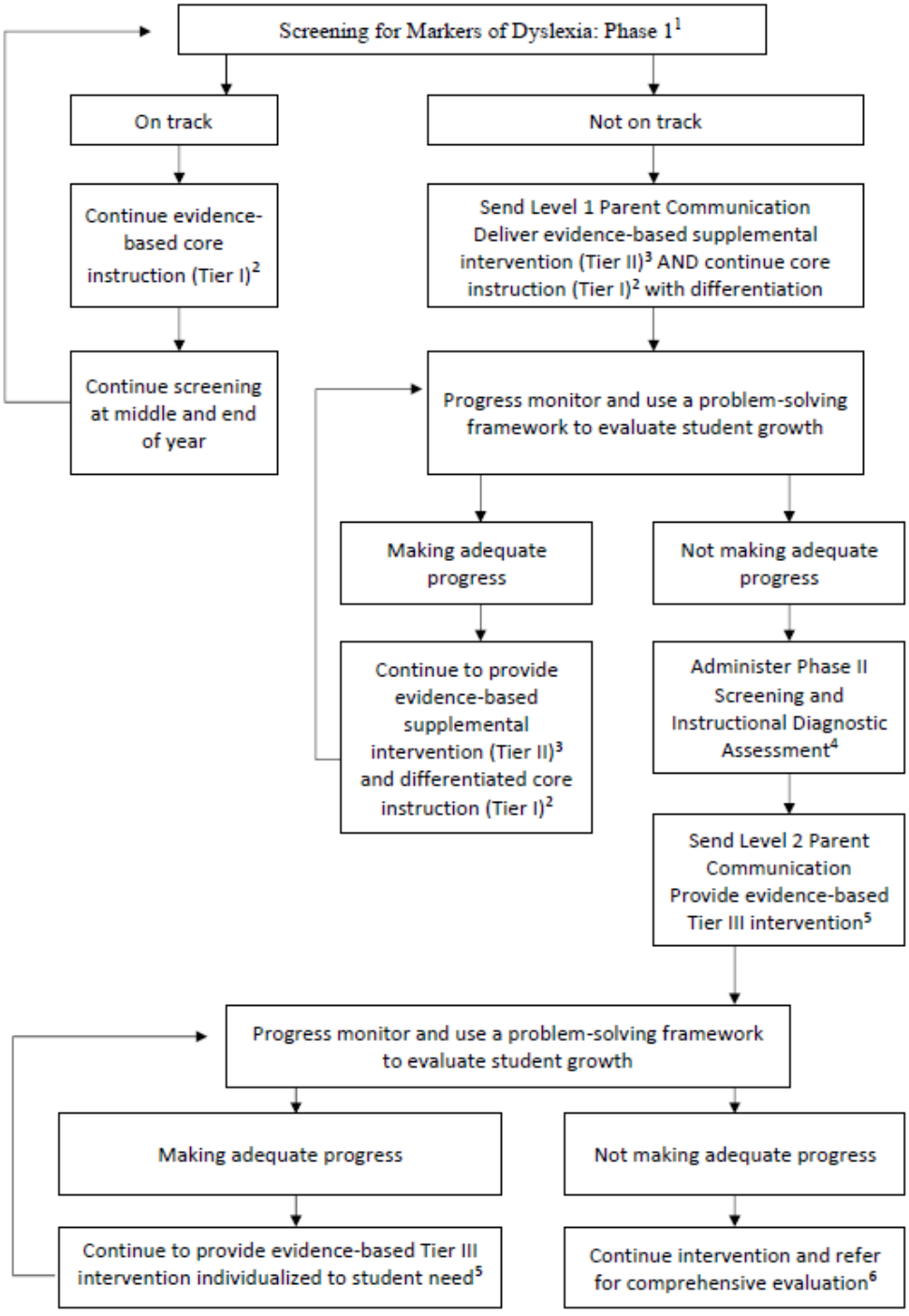


When conducting a program evaluation, it is important to have a clear theory of change to guide the formation of research questions, study design, and interpretation of results (CDC, 2011). Based on the House Bill 19-1134, the University of Oregon's application (RFP DAAA 2020000098), and prior research (e.g., Fien et al., 2021; Smith et al., 2016), the following theory of change was adopted for the pilot program: The pilot program consisted of professional development delivered by the University of Oregon on a dyslexia screening and intervention protocol that was intended to improve practice in the areas of assessment, communication, instruction, and intervention. To the extent that professional development was successfully delivered and implemented, the evaluation should observe improvements in Tier 1 and Tier 2 instruction; teacher knowledge; and improved communication with parents and families. These improved practices should in turn lead to improved student outcomes, such as higher 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.

Dyslexia Pilot Assessment and Intervention Protocol

A critical component of the dyslexia pilot program was the dyslexia assessment and intervention protocol that was developed by the University of Oregon with consultation from local stakeholders. The professional development provided by the University of Oregon to the pilot schools was expected to enhance school capacity to use this standardized process to support students with and at-risk for dyslexia. As illustrated in the figure on page 6, the protocol requires schools to screen students for markers of dyslexia in the fall to determine whether they are on track or at-risk for reading difficulties. Students who are not at-risk continue to receive typical core instruction (Tier 1) and are rescreened in the winter and spring. The families of students who are at-risk receive a letter

The Dyslexia Pilot Program Protocol



informing them of their child's assessment results and the school's intervention strategy. The student is then provided supplemental, evidence-based intervention and progress monitoring assessments (Tier 2). Students who make adequate progress continue to receive core and supplemental instruction. Students who do not make adequate progress are provided a diagnostic assessment and an intensive intervention (Tier 3), and their families are notified of the results. During intensive intervention, students are progress monitored using a problem-solving framework to evaluate their growth. Students who make adequate progress continue to be provided the intensive intervention that is individualized to their needs. Students who do not make adequate progress are referred for a comprehensive evaluation to determine special education eligibility.

Evaluation Goals

The purpose of this evaluation was to assess the usability, implementation, and effectiveness of the dyslexia pilot program. The findings from the evaluation are intended to improve the Department's capacity to refine resources for technical support, identification, and interventions; provide technical support necessary to effectively use the resources; and make recommendations for legislation. The adjacent table describes the specific research questions the evaluation sought to answer, organized by topic. The scope and focus of the evaluation were designed to replicate the evaluation of the pilot program's first year.

Evaluation Research Questions
Usability
1. To what extent did teachers in the pilot schools find the professional development on the protocol to be high quality, relevant, and useful?
2. To what extent did teachers find the intervention protocol easy to use?
3. What facilitated and hindered the implementation of the protocol?
4. To what extent did teachers have a positive perception of team meetings?
Implementation
5. To what extent did teachers receive training as intended?
6. To what extent was the protocol administered by teachers?
7. To what extent did implementation of a multitiered system of support in reading (MTSS-R) improve at the school level from beginning to end of year?
Effectiveness
8. To what extent do pilot teachers have knowledge of the science of reading and markers of dyslexia?
9. To what extent did the implementation of evidence-based reading instruction improve from pre- to post-intervention?
10. To what extent did the intervention protocol change student outcomes?

Methods

Sample

Following the design used in the first year of the pilot program, the Department recruited three schools to participate via a voluntary response to a solicitation. Recruitment took place during the summer of 2022, resulting in a combined applicant pool from which schools were reviewed and selected. To be considered for participation, schools needed to complete an application and obtain district level support. 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.

After the Evaluation Plan was developed, the Department recruited three comparison schools. In accordance with best-practice recommendations (Marcus et al., 2020), the evaluator used Mahalanobis distance matching to identify suitable comparisons using school characteristic data based on five-year averages to account for any COVID-related anomalies. Schools were matched on geographic locale, total number of students, Title 1 status, charter status, student to teacher ratio, percent of minority students, percent of English learners, and percent of students with reading deficiencies as data were available. An incentive of $1,000 was offered to all potential comparison schools. Recruitment efforts took place from September 2022 to January 2023. Table 1 describes select characteristics of the pilot and comparisons schools.

Table 1
Select Characteristics of Pilot and Comparison Schools

School	Setting	Student Total	Title 1	Student to Teacher Ratio	FRPL Eligible	TLCC	Teachers with 5 years' experience or less
Pilot Schools							
School 1	Rural	245	Yes	12:1	23%	86%	45%
School 2	City	258	No	13:1	32%	82%	20%
School 3	City	196	Yes	12:1	59%		
Comparison Schools							
School 4	City	274	Yes	12:1	70%		
School 5	Town	479	Yes	14:1	14%	72%	35%
School 6	Town	351	Yes	13:1	35%	84%	39%

Note. Figures for student total and Student to Teacher ratio were rounded to protect the confidentiality of participating schools. Missing values indicate data were unavailable. TLCC = Teaching and Learning Conditions Colorado survey score.

Measures

The sections that follow describe the measures that correspond to each topic area of the evaluation (i.e., Usability, Implementation, and Effectiveness).

Usability

End of Year survey. The usability of the pilot protocol and related materials was assessed via survey that was co-developed by the evaluator and the University of Oregon, with input from the Department. In the spring of 2023, the University of Oregon administered the survey to 26 practitioners in the pilot schools. They survey assessed (a) the quality, relevance and usefulness of the protocol, (b) the extent to which teachers found the intervention protocol easy to use; (c) perceived facilitators and barriers of protocol use, (d) the quality, relevance and usefulness of team

meetings; (e) the extent to which participants perceived that their school administrator was engaged with and supportive of the pilot program; and (f) teacher knowledge of dyslexia.

Implementation

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

Pilot school activities. The University of Oregon submitted records of pilot school activities to the evaluator to provide additional information about trends in implementation. Records that were identified as appropriate for review include (a) records of assessment provision of all measures; (b) records of teacher participation in pilot training; (c) records of 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 were 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 was assessed via a self-assessment checklist that was completed by school leaders. Pilot schools completed the measure in the fall and spring. Comparison schools completed the measures in the spring. 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 rated 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 were then averaged within each element. For the evaluation, element scores were also summed to yield a total score that ranged from 0 to about 10, with higher scores indicating higher implementation levels.

Effectiveness

Pilot evaluation survey. All effectiveness data was collected via an end-of-year survey. The survey asked 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 of students who received evidence-based interventions the previous school year, (e) a description of the data sources used in comprehensive evaluations for dyslexia, and (f) a description of the infrastructure that existed to support the identification of students with significant reading deficiencies in 2021-2022 and 2022-2023.

Design

Descriptive, nonexperimental analyses were 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 did not meet minimum recommendations for relevant non-parametric analyses (Morgan, 2017), and for some key measures, was not sufficiently powered to detect group differences.

Results

Usability

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

Twenty-six educators (15 teachers, 6 special educators, 2 administrators, 1 support staff, 1 instructional coach, and 1 other) completed the End of Year survey. In response to the statement, "The professional development provided by the University of Oregon for the dyslexia pilot was of **high quality**," 70% of respondents agreed, 17% slightly agreed, and 13% slightly disagreed. In response to the statement, "The professional development provided by the University of Oregon for the dyslexia pilot was **relevant** to my professional needs," 78% of respondents agreed, 13% slightly agreed, and 9% slightly disagreed. In response to the statement, "The professional development

provided by the University of Oregon for the dyslexia pilot was **useful**," 74% of respondents agreed, 17% slightly agreed, and 9% slightly disagreed.

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

Respondents were asked to rate the usability of the dyslexia pilot protocol. In response to the question, "How would you rate the usability of the dyslexia pilot protocol?" 30% of respondents described it as "very easy to use" and 70% of respondents described it as "easy to use." No respondents found it difficult to use.

***What facilitated and hindered use of the protocol*?**

To understand facilitators and barriers of protocol implementation, respondents were asked a series of constructed response and multiple-choice questions.

In response to the constructed-response question, "What factors and resources **facilitated implementation** of the dyslexia protocol?" respondents identified the following distinct factors: (1) high quality trainings and resources (e.g., professional development providers, videos, assessments) (2) usability of the protocol, (3) alignment with a preexisting multitiered system of support, (4) staff buy-in, (5) monthly check-ins, (6) on-site and in-person trainings), Enhanced Core Reading Instruction routines, (7) data summits, and (8) use of universal screeners.

In response to the constructed response question, "What factors and resources **hindered implementation** of the dyslexia protocol?" respondents identified the following distinct factors: (1) weather and other scheduling challenges, (2) lack of time, (3) competing initiatives and routines, (4) lack of fidelity of implementation by some staff, and (5) web-based delivery of content. Of these responses, **lack of time** was the most cited barrier. It can be noted that three respondents indicated that they did not face any barriers. One of these respondents further indicated that they expected implementation to continue to improve in the future.

Respondents were asked **how often they referenced the protocol training manual**. About 4% selected weekly, 54% selected monthly, 12% selected quarterly, 17% selected rarely or never, and 17% did not select a response.

Respondents were asked, "Apart from the pilot program, **what other professional development experiences did you have this year that helped** you implement the protocol?" Respondents could select as many experiences as applied from a pre-populated list of professional development programs that had previously been approved by the Department.

- 48% selected "Language Essentials for Teachers of Reading and Spelling."
- 17% selected "Orton Gillingham International."
- 17% selected "Other."

- 17% did not select a program.
- 4% selected “Strive.”

Of the respondents who selected “Other”, half wrote “Really Great Reading,” one wrote “PDSA protocol and Enhanced Core Reading Intervention,” and one did not provide a written response.

In response to the question, “Regardless of the professional development provider, **what topics have been the most useful** to you in implementing the protocol?” respondents were asked to select from a pre-populated list of relevant topics.

- 26% selected “Evidence-based instruction and intervention.”
- 17% selected “Databased decision making.”
- 17% selected “The science of reading.”
- 12% selected "Other.”
- 8% selected “Multitiered systems of support.”
- 8% selected “Universal screening.”
- 4% selected “The Definition of Dyslexia.”
- 4% selected “Comprehensive evaluations and identification.”
- 4% did not select a topic.

In response to the question, “**In what areas do you still need training**?” respondents were asked to select from the prepopulated list described in the previous question.

- 39% selected “Evidence-based instruction and intervention.”
- 17% selected “Comprehensive evaluations and identification.”
- 17% selected "Other.”
- 8% did not select a topic.
- 8% selected “Multitiered systems of support.”
- 4% selected “Disability policies.”
- 4% selected "Science of Reading.”
- 4% selected "Universal Screening.”

Of the respondents who selected “Other” to these questions, written responses varied, with some indicating that training in multitiered systems of support and components was already excessive, and others indicating that more training was needed both in general and in specific areas (i.e., universal screening and evidence-based instruction).

In response to the question, “**Do you think you need continued coaching** to implement the dyslexia protocol going forward?” 58% of respondents said “Yes,” 27% said “No,” and 17% did not respond.

To what extent did teachers have a positive perception of their team meetings?

In response to the statement, "My team meetings related to the pilot were generally of **high quality**," 58% of respondents agreed, 24% slightly agreed, and 8% slightly disagreed. In response to the statement, "My team meetings related to the pilot were generally **relevant** to my instructional needs," 54% of respondents agreed, 27% slightly agreed, 4% slightly disagreed, and 4% disagreed. In response to the statement, "My team meetings related to the pilot were generally **useful**," 57% of respondents agreed, 23% slightly agreed, and 8% slightly disagreed.

In response to the statement, "My school **leadership was invested** in the successful implementation of the dyslexia protocol," 69% of respondents agreed, 15% slightly agreed, and 4% slightly disagreed.

Implementation

To what extent did teachers receive training as intended?

To document trends in implementation, the University of Oregon submitted artifacts to the evaluator for a holistic review. These artifacts included **activity logs** describing the dates and descriptions of thirty meetings between the program providers and schools staff, and twenty-six training sessions variously focused on understanding dyslexia, student assessment, MTSS-R, databased decision making, intensification, and data summits. In addition to the activity logs, **attendance logs** for five professional development sessions and four data summits were submitted for one pilot school. These sessions were attended by about five to twenty individuals per school, with variation in attendance partially reflecting the different invited audiences (e.g., all staff vs school leaders only).

Also submitted were **agendas and notes** for seventeen pilot leadership check-ins (six for each of two pilot schools, and five for the third). The pilot leadership check-ins involved meeting with one to five school leaders within each school on a regular basis to identify and address any program or protocol implementation challenges. The challenges discussed at the meetings varied across schools and within schools over time. Some examples of challenges that were described include time constraints; issues of alignment with preexisting processes, resources, assessments, and/or practices; and questions and concerns about how to improve the use of student data. The notes also describe positive trends, such as a perceived lack of implementation challenges, perceived improvements in student learning; and perceived progress in addressing challenges noted in earlier meetings.

To what extent was the protocol administered by teachers?

To document pilot school implementation efforts, the University of Oregon submitted dyslexia protocol checklists for two of the pilot schools. These checklists describe twenty-five steps that should be taken by schools as part of the program, suggested completion dates, and fields to document the date of completion and evidence of completion. All steps were completed by the two schools that submitted checklists.

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

Table 2
Change in Pilot School MTSS-R Implementation from Beginning of the Year to the End of the Year

	Element 1		Element 2		Element 3		Element 4		Element 5		Total	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
School 1	1.57	1.75	1.32	1.42	1.75	1.79	1.68	1.71	0.67	1.00	6.99	7.67
School 2	1.69	1.76	1.10	0.86	1.33	1.58	1.39	1.49	0.83	1.25	6.34	6.94
School 3	0.91	1.48	1.12	1.43	1.00	1.41	1.32	1.66	0.83	1.69	4.06	7.67

MTSS-R implementation was measured at the beginning and end of the year in the pilot schools using the National Center on Improving Literacy's checklist. As illustrated in Table 2, all pilot schools improved their MTSS-R implementation from the beginning of the year to the end of the year.

Table 3
Comparison School End of Year MTSS-R Implementation

	Element 1	Element 2	Element 3	Element 4	Element 5	Total
School 4	1.04	0.00	0.08	0.15	0.42	1.69
School 5	1.19	1.23	1.33	0.50	0.75	5.00
School 6	1.55	0.91	0.96	1.24	1.25	5.91

MTSS-R implementation was also measured at the end of the year in the comparison schools (Table 3). It was not measured at the beginning of the year because comparison schools were being recruited at that time. At the end of the year, all pilot schools had higher MTSS-R implementation scores than the comparison schools.

Effectiveness

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

Post-pilot, respondents answered thirteen questions designed to assess their knowledge of dyslexia and evidence-based reading instruction. Most respondents answered most questions correctly, suggesting they had at least basic knowledge of dyslexia and evidence-based reading instruction. The three questions that produced error rates exceeding 20% were as follows:

- 41% of respondents described the statement, "Letter Naming Fluency is a useful predictor of low reading achievement" as false.
- 35% described the statement, "All students with dyslexia have phonological processing deficits" as true.
- 30% of respondents described the statement "All students with dyslexia required intensive intervention" as true.

To what extent did the implementation of evidence-based reading instruction improve from pre- to post-intervention?

Aspects of evidence-based reading instruction were assessed via a post-pilot survey. The survey included questions about each school's core curriculum, parent communication practices, comprehensive evaluation procedures, and students with significant reading deficiencies. Key findings are as follows:

Core Curricula and Supplemental Programs

Though core curricula were not expected to change because of the pilot, schools were asked to describe their core reading curricula to help contextualize evaluation findings. Two of the pilot schools used Wonders, one of which additionally used Enhanced Core Reading Instruction. One of the pilot schools used Houghton-Mifflin Harcourt's Into Reading. Supplemental programs used by the pilot schools included Really Great Reading, Enhanced Core Reading Instruction, Heggerty, SPIRE, Read Naturally, Take Flight, 6 Minute Solutions, 95% Group, and ReadWell.

The comparison schools used Into Reading, Collaborative Classroom, and Wonders. Supplemental programs included LLI, DIBELS/Burst, Estrellita, Fundations, Heggerty, Guided Reading, Systematic Instruction in Phonological Awareness, Lexia Core5, Really Great Reading (Blast, Countdown, HD word), Esperanza.

Change in Infrastructure to Support the Identification of Students with Significant Reading Deficiencies

Schools were asked to describe the infrastructure that existed in their schools to support the identification of students with significant reading deficiencies in the 2021-2022 and 2022-2023 school years. As illustrated in the table below, the pilot schools added a greater amount of new infrastructure than the comparison schools during the pilot.

Table 4
New and Continued Support Infrastructures by Condition

	Pilot Schools			Comparison Schools		
Infrastructure Type	**School 1**	**School 2**	**School 3**	**School 4**	**School 5**	**School 6**
Monthly MTSS meetings	**New**	Continued	Continued		**New**	Continued
Quarterly Databased Decision-making Meetings	Continued	**New**		Continued		
Universal Screening Data	Continued	Continued	**New**	Continued	Continued	
Progress Monitoring Data	Continued	Continued	Continued	Continued	Continued	Continued
Instructional Interventions	**New**	Continued		**New**	Continued	Continued
Records of Fidelity of Intervention		**New**				
Data Review Cycles	**New**	Continued		Continued		
Explicit Processes for Adjusting Intervention	**New**	Continued		**New**		

Parent and Family Communication

In all three pilot schools, the families of K-3 students received Level 1 letters from the school when their student was determined to be at-risk for dyslexia. The percentage of students who received such letters ranged from 0% to 100% (*M* = 56%), varying across grade level and school. In two of the pilot schools, families also received Level 2 letters when their student was determined to have markers of dyslexia. The percentage of students who received such letters ranged from 0% to 50%, varying across grade level and school (*M* = 21%). By contrast, comparison schools indicated that they did not send such communication letters to parents. These findings are consistent with the expectation that the pilot would increase levels of communication regarding dyslexia with parents and families.

To what extent did the intervention protocol change student outcomes?

Significant Reading Deficiencies

To explore the effect of the pilot on significant reading deficiencies, the evaluation examined year-to-year change in rates of significant reading deficiencies by grade and condition. In the pilot schools, the average rate of K-3 students classified has having significant reading deficiencies decreased from 2021-2022 to 2022-2023 (range = -2% to -14%). Only two of the comparison schools had relevant records for the 2021-2022 school year. For the two schools with relevant records, average rates of significant reading deficiencies also decreased (range = -1% to -8%), but to a lesser extent. Consistent with expectations, there was a large amount of heterogeneity in rates both within and across schools, which complicates attempts to interpret effectiveness.

Evidence-based Intervention

Schools were asked to indicate the percentage of students in each grade that received an evidence-based reading intervention for the 2021-2022 and 2022-2023 school years. No pilot school and one comparison school had records of receipt of intervention for 2021-2022. The average rate of receipt of intervention was 0% in the one comparison school with relevant records. One pilot school and one comparison school had records of receipt of intervention for 2022-2023. In 2022-2023, the average rate of receipt of intervention was 51% in the pilot schools (range = 49% to 54%) and 0% in the comparison school.

Comprehensive Evaluation Rates

To explore the effect of the pilot on comprehensive evaluations, the evaluation examined year-to-year change in the number of referrals for comprehensive evaluations for special education by grade and condition. Only two of the pilot schools and one of the comparison schools had records of the students who received comprehensive evaluation in both

years. The average number of students who received comprehensive evaluations decreased in both the pilot and comparison schools. Consistent with expectations, preexisting heterogeneity in comprehensive evaluation practices and the impact of COVID 19 make it difficult to interpret these changes.

Comprehensive Evaluation Procedures

To explore comprehensive evaluation procedures, schools were asked to select the data sources that are considered prior to a comprehensive evaluation for reading difficulties in their schools. As illustrated by the table below, pilot schools typically consulted a greater array of data sources prior to comprehensive evaluations, which was the expected outcome.

Table 5
Data Sources Considered Prior to Comprehensive Evaluation by School

	Pilot Schools			Comparison Schools		
School	**School 1**	**School 2**	**School 3**	**School 4**	**School 5**	**School 6**
Re-Administration of Screener	X	X	X		X	X
Family History	X	X	X			X
Teacher Feedback	X	X	X	X	X	X
Receipt of Intervention Data		X	X	X		X
Intervention Implementation Data	X	X	X	X		X
Progress Monitoring Data	X	X	X	X		X

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

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

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

Marcus, J., Peery, B., Klute, M., Pellerin, E., Mislevy, J., Wilkerson, S., & Schaeffer, V. (2020). *Regional Educational Laboratory Appalachia: Research-based strategies for effective remotelearning: Facilitators' handbook.* SRI International. https://ies.ed.gov/ncee/edlabs/regions/appalachia/events/materials/12-08-20_Workshop-1_student-engage_handbook_acc.pdf

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