An Investigation of Direct Instruction Spoken English for At-Risk English Learners

Grant Award Number: R305A150325

Methods and Measures Documentation

Oregon Research Institute

Investigators:
Barbara Gunn, Oregon Research Institute (Project Director)
Erin Chaparro, University of Oregon
Keith Smolkowski, Oregon Research Institute
Patricia Vadasy, Oregon Research Institute

March 10, 2021

Table of Contents

Purpose of the Documentation 3
Project Description and Research Design Overview 3
Participants 4
  Districts and Schools 4
  Teachers 4
  Students 5
Intervention 6
  Curriculum Description 6
  DISE Training and Coaching 6
Procedures 8
  Recruitment 8
  Permission and Consent 8
  Sample Selection 9
  Randomization 9
  Assessment 10
  Observations 10
Measures 11
  Student Measures 11
  Teacher and Classroom Measures 12
Data Structure 15
References 17

Purpose of the Documentation

This document describes methods used for *An Investigation of Direct Instruction Spoken English for At-Risk English Learners* funded by Institute of Education Sciences (IES) to Oregon Research Institute (ORI; Grant Award Number: R305A150325). This report supplments the published papers with additional details, such as variable names for measures inlcuded in archived data. This document may repeat text from manuscript without quotation marks or attribution, so readers should consult published papers (Chaparro et al. 2020; Cil et al., 2020).

Project Description and Research Design Overview

The proposal to the Institute of Education Sciences described this project as a four-year evaluation of the efficacy of DISE for developing English oral language proficiency and improving basic academic outcomes for ELs compared to the counterfactual language instruction typically provided in high-quality middle school programs (comparison). The study was designed to address the following research questions:

1. Main effects. Does the *DISE* program lead to significantly greater gains in English oral language proficiency and basic academic outcomes for ELs compared to the English language instruction typically provided to ELs in middle school classrooms? Do these effects maintain in 8th grade?
2. Moderation. What student characteristics (i.e., students' first language, baseline English oral language proficiency, years in the United States, cognitive ability, and IEP status), teacher characteristics (i.e., years of education, years teaching English learners), and other aspects of instruction (i.e., student teacher ratio, number of different languages spoken by EL students in one class, time spent on oral language instruction, one versus two years of instruction) moderate the relationship between program and outcomes on measures of language and academic achievement?
3. Mediation. Do key instructional variables, such as the rates of teacher demonstrations or opportunities for student independent practice, appear to mediate the relationship between program use and student outcomes?

During the course of the project, and for the impact manuscript, we refined the research questions to be more explicit and address certain realities of the project. For example, due to challenges with tracking and the coronavirus pandemic, we collected little data from 8th grade students. We also dropped the mediation question after determining that the planned analysis may have led to biased results (e.g., Imai et al., 2011; Maxwell & Cole, 2007; Mitchell & Maxwell, 2013). These issues led to the following updated research questions:

1. Do 6th and 7th grade English language development (ELD) teachers who use the DISE curriculum, compared to business as usual ELD instruction, improve English oral language proficiency among students identified as English Learners with limited, beginner to early intermediate English language proficiency across one school year and two school years?
2. Do middle schools with English language development (ELD) teachers who use *DISE*, compared to instruction as usual, improve distal language and reading outcomes? This exploratory question asks about distal language and far-transfer reading skills that extend beyond instruction typical of beginning-level students.

3. Does student initial skill moderate the relationship between condition and outcomes on measures of language performance?
4. Is there a relation between instructional variables, such as the rates of teacher demonstrations or opportunities for student independent practice, and student outcomes?

The study randomly assigned 29 schools in Texas, Washington, and Oregon, within district, to either implement *DISE* or conduct business as usual (BAU), a waitlist control condition, in their beginner ELD classrooms. We treated classrooms and schools as the same unit, as only one teacher participated in each school and taught all students in their classroom with either *DISE* or BAU. ELD teachers assigned to *DISE* received training from a certified *DISE* trainer. Teachers in the BAU condition taught the ELD curriculum approved by their district. BAU teachers were offered *DISE* and associated professional development at the end of their study participation. The study was designed with power (.80) to detect medium-small to medium effect sizes (0.35 to 0.47).

Thirty-one middle schools initially agreed to participate. After random assignment, we dropped two schools in a district with only three total students from the two schools, leaving 29 schools.[1] We tested all sixth and seventh grade students in their participating teachers' beginner ELD classrooms in the fall and spring of Years 1 and 2 (including students who joined their classrooms). We included students who moved out of study classrooms at subsequent testing periods if they could be located within the school or the district. Figure 1 provides an overview of the research design and sample of students, teachers, and schools.

Participants

For this project we use the term *wave* to refer to the groups of schools that join the study each year. The waves correspond to the years of the study. Wave 1 schools joined in academic year 2015-16, Wave 2 in 2016-17, Wave 3 in 2017-18, and Wave 4 in 2018-19.

Districts and Schools

As presented in Figure 1, we approached 115 districts to participate in the study. One hundred and five districts failed inclusion criteria or declined participation. (See Recruitment section below for more details.) We randomized 42 schools in 10 districts to condition. After randomization, we dropped the two schools in one Texas district, one in each condition, because they had too few students who met criteria for participation. We dropped the 11 schools in the other Texas school district because they had a change in administrators after randomization and the new administrator declined participation. These two districts were not included in the research activities or the analyses. Thus, eight districts participated in the DISE study. Of the 29 middle schools in the final sample, 14 were in the treatment condition and 15 in the BAU condition, with 25 schools in Texas, two in Washington, and two in Oregon. Schools joined the study in waves across the Grant Years 1, 2, 3, and 4. Table 1 presents the demographic information for the eight participating districts.

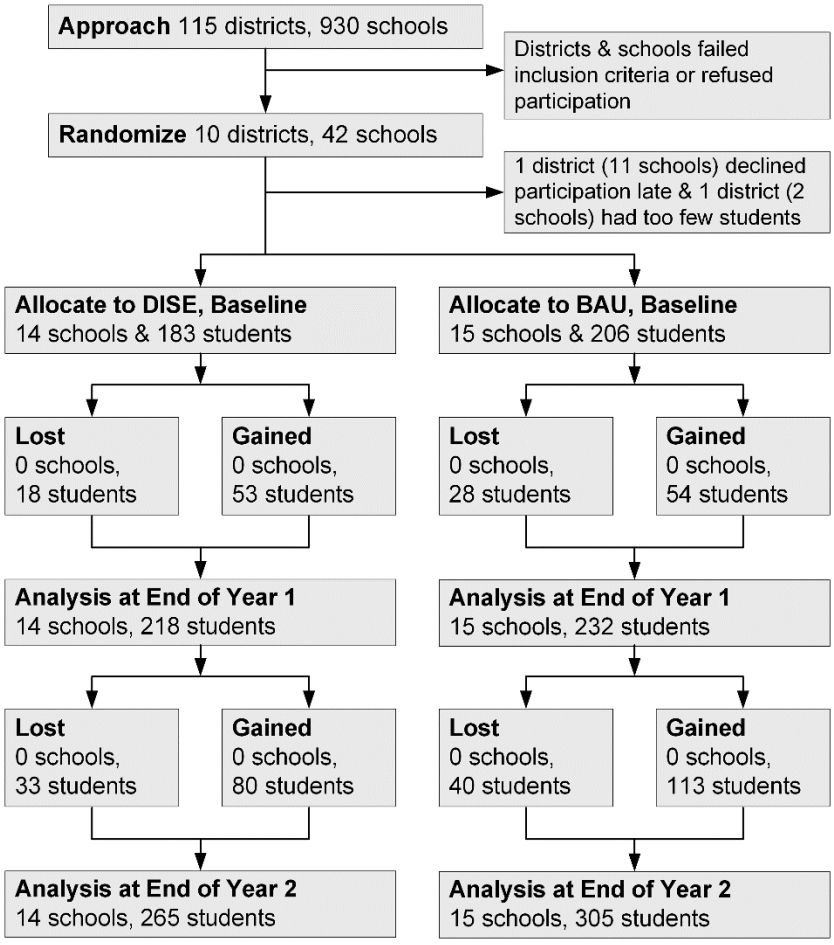
Teachers

Overall, 37 teachers participated over the course of the study. As described above, all

[1] In addition, one district agreed to participate but was unable to fulfill their agreed participation due to a change in district administration. We randomly assigned schools but conducted no other research activities in the district.

schools had only a single beginning-level ELD teacher at one time, except for one school with a teacher and an instructional assistant. In five schools, a participating ELD teacher left (e.g., for maternity leave) and was replaced by another teacher (sometimes more than once).

Figure 1. Participant Flow Diagram



Teachers delivered classroom instruction primarily in English, including practice speaking, listening, reading, and writing in English. Across the two years of participation in each school, a total of 18 intervention teachers[2] delivered DISE in their EL classrooms and 19 BAU teachers delivered their usual instruction. All teachers but one, an instructional assistant in a DISE school, were certified to teach English to nonnative English speakers (e.g., ESL, TESOL). Teachers taught an average of 11.6 years in total (11.9 in DISE schools, 11.3 in BAU) and 11.8 years of ESL (8.3 in DISE, 14.5 in BAU). In the DISE schools, 10 teachers reported that they had received their bachelor's degree, four a master's degree, and one an associate's. In BAU schools, nine teachers reported a bachelor's, seven a master's degree, one nearly completed a doctorate, one had a law degree, and one did not report a degree. One teacher in each condition had a special education certification.

Students

Within these teachers' classrooms, we assessed 144 sixth and seventh grade students in the fall and 130 in the spring of Grant Year 1; 366 students in the fall and 417 in the spring of Year 2; 324 students in the fall and 376 in the spring of Year 3; and 95 students in the fall and 107 in the spring of Year 4. A majority of students spoke Spanish as their first language. Table 2 provides additional information about the sample of students, teachers, and schools, such as the percent of students on an individualized education plan (IEP).

[2] Seventeen teachers delivered DISE, and one teacher chose to train his instructional assistant to deliver DISE to enable them to divide the classroom into two groups (lower and higher English language proficiency) for instruction.

Intervention

Curriculum Description

The authors of *DISE* designed a scope of instruction, derived from an analysis and selection of the content. They then developed the sequence, or order and amount, of content to be taught. The authors also considered the scheduling of the skills to be taught, and how to clearly present the skills, and then designed the lessons to systematically and sequentially develop increasing levels of oral language proficiency. DISE first teaches the most elementary skills across categories of language (i.e., speaking and pronunciation, listening, academic and social vocabulary, and syntax), and then proceeds systematically through additions and alterations that create more complicated applications. The work is cumulative, so that newly taught content recurs throughout the program. The development of the skills, the number of times skills appear in an activity, and the number of times they are reviewed, were mapped out during development. All the sub skills that are needed for more complex speaking tasks are taught earlier and integrated systematically, first bundled as simple tasks, and later embedded in progressively more elaborate conversation. The instruction also explicitly focuses students' attention on the language forms that are taught, a significant concept in instructed language learning (CCSSO, 2012b; Ellis, 2005b). The sequential, structured design of *DISE* make it easy for ELD teachers to follow, and increases the likelihood of implementation fidelity in a rigorous Goal 3 randomized trial comparing it to the ELD instruction provided in comparison school districts.

Table 1. *District Demographic Information*

		District				Students			Teachers	
District	State	*DISE* Middle Schools	BAU Middle Schools	Total Middle Schools	Locale	Total in All Grades	% with LEP	% with IEP	Total, FTE	Pupil/ Teacher Ratio
A	TX	4	5	9	City: Large	43,710	37%	8%	2,538	17.2
B	TX	4	4	8	City: Mid-size	34,792	35%	7%	2,283	15.2
C	WA	1	1	2	Town: Distant	6,805	31%	14%	352	19.3
D	TX	1	1	2	City: Large	6,064	27%	9%	390	15.6
E	TX	2	1	6	Suburb: Large	25,297	27%	11%	1,679	15.1
F	OR	0	1	2	Town: Remote	4,082	19%	13%	217	18.9
G	TX	1	2	3	Suburb: Large	12,625	11%	6%	764	16.5
H	OR	1	0	4	Suburb: Mid-size	10,854	6%	18%	518	20.9

DISE Training and Coaching

In Year 1 the 5 teachers assigned to the *DISE* condition (Wave 1) participated in two days of training from a certified National Institute for Direct Instruction (NIFDI) trainer. There were 4 teachers from District A and 1 teacher from District C. Drs. Chaparro and Gunn participated in the training along with 2 district administrators. The trainer provided a program overview and the rationale behind the design of the *DISE* lessons. The *DISE* teachers learned the instructional skills they need to teach the exercises as intended. They also learned presentation techniques (e.g., quick pacing, clear demonstrations) and monitoring and correction techniques. The training session also covered procedures for assessing and motivating students. Teachers received individual feedback from NIFDI trainers during the training.

Table 2. *Characteristics and Measure Descriptives for Student Sample (N = 746) by Condition*

		Control	DISE
Total Sample *N*		412	334
Participation by Assessment Period	Year 1 Fall (T_1) *N* (%)	212 (51%)	185 (55%)
	Year 1 Spring (T_2)	236 (57%)	223 (67%)
	Year 2 Fall (T_3)	298 (72%)	246 (74%)
	Year 2 Spring (T_4)	310 (75%)	265 (79%)
Female *N* (%),121 missing		171 (48%)	131 (48%)
Ethnicity, 132 missing	Hispanic *N* (%)	290 (83%)	221 (84%)
	Hispanic White	193 (55%)	127 (48%)
	Hispanic AIAN	67 (19%)	81 (31%)
	Non-Hispanic White	14 (4%)	19 (7%)
	Non-Hispanic Asian	25 (7%)	7 (3%)
Free or reduced-price lunch *N* (%),163 missing		327 (99%)	252 (100%)
Special education *N* (%),114 missing		9 (2%)	4 (1%)
Grade point average	Year 1 *M* (*SD*) *N*	2.9 (0.6) 194	2.8 (0.8) 179
	Year 2	2.5 (0.6) 181	2.5 (0.6) 161
Days in school *M* (*SD*) *N*	Year 1 *M* (*SD*) *N*	150 (35) 152	143 (38) 129
	Year 2	160 (22) 186	159 (24) 149
IPT *M* (*SD*) *N*	T_1 *M* (*SD*) *N*	8.6 (8.2) 206	7.3 (6.5) 183
	T_2	11.5 (10.7) 232	10.8 (7.6) 218
	T_3	13.6 (13.5) 298	12.2 (8.1) 245
	T_4	15.6 (12.9) 305	14.9 (9.6) 265
ORF WCPM	T_2 *M* (*SD*) *N*	61.1 (32.7) 233	58.1 (27.8) 218
	T_4	74.6 (33.6) 307	73.1 (28.5) 264
WJ Test 1 W Scores	T_2 *M* (*SD*) *N*	429.7 (19.8) 230	429.3 (17.7) 217
	T_4	437.4 (21.6) 306	436.5 (18.1) 264
WJ Test 2 W Scores	T_2 *M* (*SD*) *N*	433.9 (26.8) 231	429.8 (25.6) 218
	T_4	445.5 (28.7) 306	443.5 (26.5) 265
WJ Test 6 W Scores	T_2 *M* (*SD*) *N*	459.9 (22.2) 232	460.7 (19.8) 217
	T_4	466.3 (22.2) 294	469.2 (18.4) 258

Note. Percentages represent the proportion of the total (row 1); students with missing information not included in percentages. Ethnicity reported for only categories with more than 5% of cases in either condition. AIAN = American Indian and Alaska Native. Participating schools were in session for 172 to 177 days. Days in school calculated from days enrolled minus absences. IPT = Idea Proficiency Test; ORF WCPM = oral reading fluency words correct per minute; WJ = Woodcock-Johnson.

In Year 2 the 5 new teachers assigned to the *DISE* condition (Wave 2) received the same training and coaching as the Wave 1 teachers received. In Year 2 the trainer also provided a half day of refresher training, visited each Wave 1 *DISE* classroom, and provided up to 4 Skype or FaceTime coaching sessions with each teacher. These remote coaching sessions included watching the teacher in real time teaching the *DISE* lesson, and then speaking with the teacher after class to debrief the lesson, answer questions, and provide feedback.

In Year 3 the 12 teachers in Wave 1 completed their participation in the study. We offered the *DISE* curriculum and training to the 6 wait list Wave 1 teachers. In September 2017 the District C wait list teacher received the *DISE* curriculum and the training, but District A administration did not respond to offers to provide their wait list teachers with the *DISE* curriculum and training.

In Year 4 (September 2018), a national trainer from the National Institute for Direct Instruction provided *DISE* training to treatment teacher from District H, the 4 wait list teachers from District B and the one wait list teacher from District D. We provided them each with a class set of *DISE* as part of their participation in the study. We talked with and sent e-mails to the District A bilingual department administrator about training their 5 WL teachers, with no response, so we considered them a decline. In May 2019 the wait list teacher in District F received the curriculum and training from the NIFDI trainer. One treatment teacher from District E went on maternity leave. The teacher who replaced her for the 2018–2019 school year also received the *DISE* training.

In Year 5 (October 2019), the NIFDI national trainer provided DISE training to the one wait list teacher in District E. District G administration declined the training for their two wait list teachers.

Procedures

Recruitment

In the spring and summer before each year of this 4 year study, investigators contacted districts with concentrations of ELs in middle school. We focused our recruitment efforts on the West Coast including Washington, Oregon, California, Arizona, and, because of their large concentration of ELs, we also contacted Texas school districts. To be eligible to participate, the middle school ELD teacher in each building had to be using an English language curriculum other than *DISE* and be willing to continue using their current curriculum or to teach *DISE* daily for two years.

Across the four years of recruitment, we contacted 115 districts, with 103 districts either not responding to the initial contact or declining after reviewing the materials, eight districts accepting, and five withdrawing or discontinuing participation. The most common reasons for declining included participation in other research projects, insufficient numbers of ELs, new administrators, and concerns about using a new English oral language curriculum. A recruitment summary, with the names of the districts and reasons for accepting or declining participation is provided in the Appendix.

Permission and Consent

The study was reviewed by the ORI Institutional Review Board in April, 2015. The study was considered exempt under exemption category #1 (Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (a) research on regular and special education instructional strategies, or (b) research on the

effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods). Active consent was not required from parents or caregivers because of the ORI IRB approval and none of the districts required additional consent because all of the testing was considered typical practice in these settings.

In addition to IRB review and approval, we also submitted applications for permission to conduct research in Districts that required prior approval. Of the eight districts that participated, two Texas districts required applications before we began study activities.

Sample Selection

Teachers and students participated for two years, with the exception of Wave 4 teachers and students, who participated for one year. Throughout this document, we use *year* to refer their year of participation. We use *Grant Year* to refer to the year of the grant.

In Wave 1, for the 2015-2016 school year, we included all students in 6th or 7th grade in the participating teachers' beginner ELD classrooms. For the 2016-2017 school year we followed those cohorts of students in the participating teachers' classrooms. Between spring and fall 2016, approximately 50 students moved out of the study classrooms; many of them were "mainstreamed" out of EL instruction or moved out of the participating districts. Due to this level of attrition, we decided to include new 6th and 7th grade students who joined the study classrooms at both fall and spring testing in Year 2. We also located and tested students who had remained in the district but were no longer placed in those study classrooms (e.g. students who had been mainstreamed).

In Wave 2, most of the schools in District B grouped their classrooms by grade rather than EL level, so we included both the 6th and 7th grade classes. In District D, we included the 6th and 7th grade students in the mixed-grade beginner EL classrooms. Due to the very high rates of attrition we observed year to year in Wave 1, we continued the approach of (a) including new students who joined the focal "study classrooms" and (b) locating and testing students who were still in the district but no longer in those study classrooms. We decided to follow all students if they could be located within the school or district.

In Wave 3, we began by including the 6th and 7th grade students in the beginner EL class periods in District E and in the newcomer classes in District G. As with previous waves, new students in the study classrooms were incorporated into the sample at each testing period, and students who moved out of study classrooms were tested at subsequent testing periods if they could be located within the school or district.

In Wave 4, for the 2018-2019 school year we included 6th and 7th grade students in two Oregon middle schools. Because it was the last year of the study, these students only participated for one year.

Randomization

Schools were randomized to condition after they agreed to participate in the study. We randomly assigned schools with the uniform random number function in Excel after blocking on district.

In Wave 1, we randomized the two District C schools in June 2015 and the 9 District A schools in July 2015

In Wave 2, we randomized the eight District B schools in July 2016, and the two District

D schools in August 2016.

We randomly assigned two additional schools from another district in April 2016, but we dropped the schools from the study in September 2016. The district's Bilingual and ESL Coordinator told project directors that most of the beginning English learners in the 8 district middle schools attended the two middle schools, with typically enough beginners in those schools to meet the requirements of our research design. These two schools joined the study and were randomly assigned to condition, but after school started and class lists were finalized, we identified only three students in the two schools (combined) who met criteria for participation. Due to the likelihood of losing all three students during the study (attrition) as well as the travel expenses for testers, observers, and the *DISE* coach, we dropped the two schools from the study. We based the decision on the likelihood that the schools and their students would complete study activities and, if so, their limited contribution given the outsized expense.

In Wave 3 we randomized three middle schools in District E and three middle schools in District G to condition in September 2017. After randomization one of the intervention teachers in District E declined to teach *DISE* because it did not match her teaching philosophy. She remained in the study and participated in the testing and observation, and survey activities. She was included in the final intent-to-treat analyses as an intervention teacher.

In Wave 4 we randomized two Oregon middle schools, one in District H and one in District F, to condition in July 2018.

Assessment

Dr. Chaparro trained research assistants to collect student performance data prior to each assessment period. For the initial training, she provided a 4-hour interactive PowerPoint presentation. For returning assessors, 1½- to 2-hour refresher trainings were required. During both initial and refresher trainings, each assessor established reliability at or above 80% agreement with her on each test, and newer assessors verified their reliability through shadow scoring with experienced assessors on the first day of each assessment period.

All assessors were unaware of condition. All testing materials were labeled with school, district, and project student identification information except for their participation in the treatment or wait list condition, which was only accessible to key personnel. A testing team was deployed to each school in the fall and in the spring to conduct the individual students. Dr. Chaparro supervised the testing in all schools but did not conduct testing herself. All assessors were also required to maintain reliability through on-site shadow scoring. The primary assessor would administer the test to the student while the secondary assessor would sit next to the testing desk and score the student assessment while the student was being tested by the primary assessor. An over 80% average agreement was maintained for all shadow-scoring.

Project staff members attempted to follow all students for assessments as long as they remained in the same district. This included any students initially in a classroom assigned to condition even if the student no longer received English language instruction.

Observations

We conducted three observations per year of the ELD instruction provided by *DISE* and BAU teachers. In Year 1, one of the researchers trained three observers to use the observation tool, the Classroom *Observation of Student Teacher Interactions* (*COSTI*; Smolkowski & Gunn, 2012). The initial 2-hour training introduced the coding system and codebook, with guided

practice coding videos of ELD instruction. Observers practiced coding individually and as a group to compare reliabilities and to adjust decision rules as needed. They then visited a local middle school ESL classroom twice to practice coding in a real classroom and to establish final reliability. In Year 2, one of the researchers trained two additional observers using the same procedures as in the Year 1 training. All observers received a refresher training in the fall of each subsequent year of the study. A detailed description of the *COSTI* and the observation timeline is provided in the Measures section.

Measures

Student Measures

Student Demographic Characteristics. We collected student demographic information from school districts at the end of each project year. Student demographic variables provided by districts included biological sex, ethnicity, free or reduced price lunch status, special education status, grade point average, and number of days in school.

Outcome Measures and Baseline Cognitive Performance. Because school districts are rightly concerned with the amount of testing time, we opted to focus on the following battery of measures assessing English language development. We selected the IDEA Proficiency Test as our primary measure of oral English language. We also administered three Woodcock-Johnson (WJ) language measures and the Dynamic Indicators of Basic Early Literacy Skills (DIBELS 6th ed.) measure of oral reading fluency (ORF) as experimental measures of distal outcomes. We did not expect changes in ORF, for example, but included it to assess the potential for student oral language skills to transfer to reading skill. The WJ language measures are more proximal to DISE instruction in that they measure English language but nonetheless assess concepts and skills that beginning-level students would not likely master within a year or two. Therefore, we focused our investigation on the IPT and considered the WJ measures and ORF as exploratory because they measure skills likely beyond the level of DISE instruction provided to students in this project.

English Oral Language Proficiency. We established students' baseline English oral language proficiency using the IPT assessment suite (Ballard & Tighe, 2010a). We administered the IPT II English Oral Language Test at the beginning and end of the school year. The IPT II English Oral Language Test is an individually administered assessment that takes approximately 5 minutes for students with lower English language proficiency and 40 minutes for students with intermediate to early advanced levels of English Proficiency to complete (Ballard & Tighe, 2010a). IPT II English Oral Language has two alternate forms; we used one in the fall and one in the spring (Ballard & Tighe, 2010b; Ballard & Tighe, 2010c). According to the publisher, the internal consistency reliability of both forms of IPT II Oral English, E and F, is .90 (Ballard & Tighe, 2010a). The test-retest reliability was .95 (Ballard & Tighe, 2010a).

We also administered the Woodcock-Johnson IV Tests of Oral Language (WJ IV; Schrank et al., 2014), a well-established, reliable, and valid measure of oral language. The Oral Language cluster measures expressive and receptive language, vocabulary, listening comprehension, linguistic competency, reasoning, and memory. For this study, we used the picture vocabulary, oral comprehension, and understanding directions subtests as a distal measure of student growth in vocabulary and listening comprehension.

Oral Reading Fluency. Oral reading fluency (ORF) is a valid, reliable indicator of

overall reading competence (Fuchs et al., 2001; Slocum et al., 1995; Stanovich, 2000). We assessed students' oral reading fluency rate at spring data collection with one sixth-grade passage (DIBELS 6th ed.; Good & Kaminski, 2002). We calculated the number of words correctly read in 1 minute. DIBELS ORF has been demonstrated to be predictive of later reading proficiency (e.g., Baker et al., 2008; Good & Kaminski, 2002). Smolkowski and Cummings (2016) and Cummings et al. (2019) demonstrated the accuracy of DIBELS 6th edition ORF when predicting comprehensive tests.

Ravens Progressive Matrices. Because participating students spoke different languages, and typically spoke little to no English at baseline, we assessed them at baseline with a measure that requires no English language proficiency. We assessed their general (nonverbal) cognitive ability using the *Raven's Coloured Progressive Matrices (CPM)* (Raven, Raven, & Court, 1998) (internal consistency .80-.90). The test includes 60 non-verbal multiple-choice items that measure general intelligence. For each test item, students were asked to identify the missing element that completes a pattern.

Teacher and Classroom Measures

Teacher Demographic Survey. We collected demographic information from participating teachers for each year they participated in the study, The Participants section of this documentation provides information on teacher certification and educational background, years teaching middle school and English Learners, and use of instructional assistants.

DISE Lesson Progress Chart. We documented dosage and pacing of *DISE* instruction with a weekly Lesson Progress Chart (LPC) that asked teachers to report the lesson number and the DISE activities they completed each day. Teachers also reported the days and reasons why they did not teach DISE.

BAU Teacher Survey. We documented the ELD instruction in the BAU classrooms with a quarterly survey that asked teachers to report the ELD curricula they used, their instructional focus, and other considerations that affected instruction (e.g., substitute, testing). Teachers in BAU classrooms reported using commercially published ELD programs for their main English language instruction. Some of the instructional tools and assessments used included Milestones, Language Power, Florida Center for Reading Research Student Activities, Keys to Learning, Texas Primary Reading Inventory, and ESL Reading Smart, as well as software programs such as Rosetta Stone, Imagine Learning, and Read 180. Several teachers also created their own materials or activities, and many used books or shorter texts for silent and oral reading and comprehension discussions.

Classroom Observations of Student–Teacher Interactions (COSTI). In the fall, winter, and spring of each year we documented the rates of specific student-teacher interactions during *DISE* and BAU instruction that substantial research and theory suggests play a pivotal role in how well students learn and remember basic skills. We used the *COSTI* to record rates of (a) teacher demonstrations, (b) student independent practice, (c) student errors, and (d) corrective feedback from the teacher. Observers coded each time one of the four interactions occurred during 30 minutes of instruction. We audio recorded all observations for subsequent content coding. We conducted approximately 20% of the observations with two observers to establish reliability and maintained reliability at or above 80% for all observers. Observers demonstrated a very high level of interrater reliability; intraclass correlations (ICCs) ranged from .93 to .99. The stability of observed codes over time was lower, with ICCs from .43 to .80. ICCs for stability

document the extent to which teachers' rates of instruction practices remain the same over time and are analogous to test-retest correlations. With three observations, an ICC of .40 implies reliability of .67 for an aggregate mean; an ICC of .50 implies a reliability of .75 (Shoukri et al., 2004). Consistent with our theoretical model, we focused on the rates of demonstrations and independent practice, which produced stability ICCs of .57 and .80.

In Grant Year 1, observers conducted the *COSTI* observations in person. In Year 2, we pilot tested remote observation over video conferencing systems with four teachers to determine the feasibility and accuracy of remote coding. We compared in-person coding to coding online and averaged 80% agreement or above for teacher demonstration, group and individual practice. The real-time video and audio capabilities of the platforms enabled accurate *COSTI* coding and produced audio recordings for later coding. Based on these outcomes, we conducted the spring Year 2 observations remotely in all of the project classrooms. In the subsequent years of the study, we conducted, on average, two of the three yearly observations in person and one remotely per classroom.

Audio Content Coding of Classroom Observations. Whereas the *COSTI* observations document rates of student-teacher interactions, the *COSTI* is not designed to describe instructional content (focus) and instructional activities. For that reason, we also coded the audio recordings of the observations to describe the aspects of *DISE* and BAU English language instruction not captured with the *COSTI.* We chose the content codes by reviewing the teaching objectives for district-adopted English Learner language curricula and developed a set of codes that described the instructional content and the instructional format (i.e., of typical middle-school English language instruction. For each audio-recorded lesson, we coded 30 one-minute segments of the lesson for the following content areas and activities: (a) phonology; (b) morphology; (c) grammar; (d) syntax; (e) vocabulary; (f) school learning content and abstract concepts; (g) text reading; (h) other low-frequency content, such as discussions, literary forms, or reading aloud; (i) general directions and transitions; (j) teacher directed activities; (k) peer to peer activities without teacher facilitation; (l) teacher directed conversation; (m) independent seatwork; (n) teacher-directed writing; (o) teacher or student talk in Spanish; and (p) computer or audio assisted instruction. For most codes, however, we found insufficient variability for analysis. In addition, several were included to capture time in class but were deemed unrelated to DISE instruction (e.g., seatwork, writing, computer instruction). We examined only vocabulary, syntax, and teacher-directed activity. Coders demonstrated a very high level of interrater reliability; intraclass correlations (ICCs) ranged from .88 to .92. The stability of observed codes over time was lower, with ICCs of .15, .58, and .56 for vocabulary, syntax, and teacher-directed activity, respectively. Table 3 reports the codes by condition with frequencies higher than 2%.

Implementation of DISE Instruction. One ELD teacher in each treatment school delivered DISE instruction. This ensured that the intervention condition modeled the actual teaching conditions of typical end-users. The developers of DISE intended for teachers to teach 90 minutes daily. However, the teachers in this study had 90 minutes total to teach their beginners not only English oral language, but also English reading and writing. Due to this constraint, we asked the DISE teachers to provide 45–55 minutes of daily DISE instruction, at which pace they were expected to complete one DISE lesson every two days. The weekly Lesson Progress Chart (LPC) gave the DISE coach and the researchers' ongoing data on whether teachers met this goal, and barriers teachers faced to providing daily instruction. The number of lessons per week ranged from 0.7 to 1.5 in Year 1 of their participation and from 0.4 to 2.0 in

Year 2. Of the reasons teachers gave for not teaching DISE daily, 19% were for school closure, 22% for teacher absences, 28% for testing, 8% for school events, 4% for DISE unit tests, and 19% for other reasons.

Table 3. *Classroom Observation Instructional Content and Activity Codes*

Code	Definition	DISE % Time	BAU % Time
	Instructional Content		
Phonology	letter names and sounds, word pronunciation	2%	4%
Grammar	correct word usage, parts of speech, verb tense, contractions, use of words in the common *dialect*	13%	4%
Syntax	*sentence* structure, arrangement of words in sentences,	61%	7%
Vocabulary	explicit definitions, opposites, synonyms, academic, content area vocabulary	12%	20%
Text reading	discussion about text, comprehension questions, reading aloud,	5%	43%
	Instructional Activities		
General directions	teacher gives directions; transitions; students gather supplies	6%	17%
Teacher directed activities	teacher directs the instruction in a skill or activity	93%	47%
Peer to peer	students read or talk in pairs independently	0%	9%
Teacher directed conversation	extended conversational turns	1%	5%
Independent seatwork	students work independently without teacher leading	0%	11%
Teacher-directed writing	students write responses or answers in response to teacher directions	0%	6%

Note. The shaded rows represent content and instructional activities that were a focus of DISE lessons. We included the other categories as a description of the focus of instruction in both DISE and BAU classrooms during ELD instruction. Content and activities with less than 2% in both conditions were excluded from the table (e.g., morphology, school learning content and abstract concepts, teacher or student talk in Spanish), so the percents do not add to 100%.

Data Structure

The multilevel DISE data set includes assessments nested within students, students within classrooms, classrooms within teachers, teachers within schools, and schools within districts. The data set includes unique identifiers for each level of nesting. Table 4 describes the variables in the DISE data set including unique identifiers and design variables, student outcome measures, student and teacher sociodemographic characteristics, and classroom measures of instruction.

Table 4. *Variable Names and Descriptions for the DISE Data Set*

Variable Name	Description
Unique identifiers, design variables, and structural variables	
SeqID	Sequential ID
StudentID	Student ID
ClassID	Classroom ID
TeacherID	Teacher ID
SchoolID	School ID
DistrictID	District ID
Wave	Wave of school recruitment
Condition	Study condition (1=DISE, 2=BAU)
Tx	Study condition binary (1=DISE, 0=BAU)
Time	Assessment time (T1=1)
t	Assessment time (T1=0)
Student outcome measures	
iptCorrect	Student outcome: Idea Proficiency Test: items correct
orfCorrect	Student outcome: Oral Reading Fluency: words correct per minute
wj1W	Student outcome: Woodcock-Johnson: Test 1 W score
wj2W	Student outcome: Woodcock-Johnson: Test 2 W score
wj6W	Student outcome: Woodcock-Johnson: Test 6 W score
Student sociodemographic characteristics	
Grade	Student sociodemo: Grade level
RavensScore	Student sociodemo: Baseline cognitive performance (Raven's Progressive Matrices)
Gender	Student sociodemo: Gender
Ethnicity	Student sociodemo: Ethnicity
FRL	Student sociodemo: Free or reduced price lunch status
SPED	Student sociodemo: Special education status
GPA	Student sociodemo: Grade point average

Variable Name	Description
Teacher sociodemographic characteristics	
YearsTchMS	Teacher sociodemo: Years experience teaching middle school
YearsTchESL	Teacher sociodemo: Years experience teaching English as a second language
Education	Teacher sociodemo: Education
Certification	Teacher sociodemo: Certification or endorsement for teaching English learners
Classroom measures	
LessonStart	Classroom: First DISE lesson taught
LessonEnd	Classroom: Last DISE lesson taught
Lessons	Classroom: Total number of DISE lessons taught
LessonsPerWeek	Classroom: Number of DISE lessons taught per week
ObsMin	Classroom: Avg duration of COSTI observations
TDperM	Classroom: Avg teacher demos per min
PperM	Classroom: Avg student independent practice per min
EperM	Classroom: Avg student errors per minute
CperM	Classroom: Avg corrective feedback per min
ORGppn	Classroom: Instructional Activity % time observed: Organization
TDIRppn	Classroom: Instructional Activity % time observed: Teacher directed activity
PEERppn	Classroom: Instructional Activity % time observed: Student talk or read aloud
TCONppn	Classroom: Instructional Activity % time observed: Teacher-student conversation
SEATppn	Classroom: Instructional Activity % time observed: Seatwork
WRITEppn	Classroom: Instructional Activity % time observed: Student writing
COMPppn	Classroom: Instructional Activity % time observed: Computer or audio instruction
SPANppn	Classroom: Instructional Activity % time observed: Spanish talk
CCPhonoPpn	Classroom: Instructional Content % time observed: Phonology
CCGrammPpn	Classroom: Instructional Content % time observed: Grammar
CCSyntxPpn	Classroom: Instructional Content % time observed: Syntax
CCVocabPpn	Classroom: Instructional Content % time observed: Vocabulary
CCTextRPpn	Classroom: Instructional Content % time observed: Text reading
CCMorphPpn	Classroom: Instructional Content % time observed: Morphology
CCIdiomPpn	Classroom: Instructional Content % time observed: Idioms
CCCognaPpn	Classroom: Instructional Content % time observed: Cognates

Variable Name	Description
CCLearnPpn	Classroom: Instructional Content % time observed: School learning
CCOtherPpn	Classroom: Instructional Content % time observed: Other

Note. COSTI = Classroom Observation of Student-Teacher Interactions.

References

Baker, S. K., Smolkowski, K., Katz, R., Fien, H., Seeley, J. R., Kame'enui, E. J., & Thomas Beck, C. (2008). Reading fluency as a predictor of reading proficiency in low performing high poverty schools. *School Psychology Review*, *37*(1), 18–37.

Ballard & Tighe. (2010a). *IPT II-Oral English Technical Manual, Forms E & F*.

Ballard & Tighe. (2010b). *IPT II-Oral English Test, Form E*.

Ballard & Tighe. (2010c). *IPT II-Oral English Test, Form F*.

Chaparro, E., Smolkowski, K., Gunn, B., Dennis, C., & Vadasy, P. (2020). *Evaluating the efficacy of an English language development program for middle school English learners*. Manuscript submitted for publication.

Cil, G., Chaparro, E. A., Dennis, C., & Smolkowski, K. (2020). *The cost of adopting and implementing a new English language curriculum for middle school English learners*. Manuscript submitted for publication.

Cummings, K. D., Smolkowski, K., & Baker, D. L. (2019). Comparison of literacy screener risk selection between English proficient students and English learners. Learning Disability Quarterly. Advance online publication. https://doi.org/10.1177/0731948719864408

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific Studies of Reading*, *5*(3), 239–256. http://dx.doi.org/10.1207/S1532799XSSR0503_3

Good, R. H., & Kaminski, R. A. (Eds.). (2002). *Dynamic indicators of basic early literacy skills* (6th ed.). Institute for the Development of Educational Achievement. http://dibels.uoregon.edu/

Imai, K., Jo, B., & Stuart, E. A. (2011). Commentary: Using potential outcomes to understand causal mediation analysis. *Multivariate Behavioral Research*, *46*, 842-854. https://doi.org/10.1080/00273171.2011.606743

Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, *12*(1), 23-44.

Mitchell, M. A., & Maxwell, S. E. (2013). A comparison of the cross-sectional and sequential designs when assessing longitudinal mediation. *Multivariate Behavioral Research*, *48*(3), 301-339. https://doi.org/10.1080/00273171.2013.784696

Raven, J., Raven, J. C., & Court, J. H. (1998). *Raven manual: Section 4, Advanced Progressive Matrices, 1998 Edition*. Oxford, UK: Oxford Psychologists Press Ltd.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014). *Woodcock-Johnson IV Tests of Oral Language*. Riverside.

Shoukri, M. M., Asyali, M. H., & Donner, A. (2004). Sample size requirements for the design of reliability study: Review and new results. *Statistical Methods in Medical Research*, *13*, 251-271. https://doi.org/10.1191/0962280204sm365ra

Slocum, T. A., Street, E. M., & Gilberts, G. (1995). A review of research and theory on the relation between oral reading rate and reading comprehension. *Journal of Behavioral Education*, *5*, 377–398. https://doi.org/10.1007/BF02114539

Smolkowski, K., & Cummings, K. (2016). Evaluation of the DIBELS (sixth edition) diagnostic system for the selection of native and proficient English speakers at risk of reading difficulties. *Journal of Psychoeducational Assessment*, *34*(2), 103-118. https://doi.org/10.1177/0734282915589017

Smolkowski, K., & Gunn, B. (2012). Reliability and validity of the Classroom Observations of Student-Teacher Interactions (COSTI) for kindergarten reading instruction. *Early Childhood Research Quarterly*, *27*(2), 316-328. https://doi.org/10.1016/j.ecresq.2011.09.004

Stanovich, K. (2000). *Progress in understanding reading: Scientific foundations and new frontiers*. Guilford.