SRI Education™

February 2018

Evaluation of Education Connections: Supporting Teachers with Standards-Based Instruction for English Learners in Mainstream Classrooms

Final Report

SRI Project P22361

Submitted to:

Sarah Catherine Moore
Lindsey Massoud
Joanna Duggan
Christopher Fredericks

Center for Applied Linguistics

Prepared by:

Deepa Patel
Xin Wei
Katrina Laguarda
Regie Stites
Hannah Cheever
Rebecca Goetz

 This material is based upon work supported by the U.S. Department of Education (DoED) under Grant Number U411C130013, through a subaward from the Center for Applied Linguistics (CAL). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessaril reflect the views of the DoED or of CAL.

Contents

Exhibits

Executive Summary

The Center for Applied Linguistics (CAL) received an Investing in Innovation (i3) development grant in 2014 to develop an online site, Education Connections (EdConx), to train and support teachers to integrate English language proficiency standards with content standards to deliver lessons that are both content-rich and accessible to ELs. CAL collaborated with the University of Oregon to design the EdConx online platform as a one-stop site for 6th- through 12th-grade teachers of mainstream classrooms to access teaching resources and other supports for standards-based instruction relevant to ELs that were consistent with sheltered instructional practices. CAL designed EdConx to meet the varying needs of secondary teachers across subjects and to offer a menu of supports from which teachers could pick and choose to create their own personalized professional development experience. The EdConx platform offered academic and teaching resources vetted by experts in EL instruction, online courses designed and facilitated by instructors from TESOL International, live events/webinars led by subject-matter experts identified by CAL, and forums to support teacher collaboration.

Theory of Action and Evaluation Design

CAL hypothesized that teachers' engagement with the site for 30 minutes per week over 1.5 years, in whatever way they chose to meet their professional needs, would lead to expanded teacher knowledge and increased use of standards-based academic content and EL-relevant instructional practices in classrooms that include ELs, as well as improvement to students' English language proficiency.

To test this theory, CAL collaborated with six districts and with SRI to evaluate the impact of EdConx and teachers' engagement with the platform between fall 2014 and spring 2017. The six rural, urban, and suburban districts were located in four states: California, Maryland, North Carolina, and Virginia. Within districts, CAL focused recruitment on 6th- through 12th-grade mainstream teachers of core subjects. The research team evaluated the implementation of EdConx and its impact on teacher practice and students' English proficiency using a randomized controlled trial (RCT) in three districts and a quasi-experimental design (QED) study in three other districts. SRI designed these two distinct studies because not all districts agreed to participate in a RCT.

Impact of EdConx on Students' English Proficiency

To measure EdConx's impact on students' English proficiency, SRI compared the English language proficiency of ELs taught by English language arts (ELA) teachers in the treatment group with the English language proficiency of ELs taught by ELA teachers in the control group in the RCT districts.[1] The research team narrowed the sample to students in classrooms of ELA teachers because CAL hypothesized that changes to instruction in ELA classrooms would have the most immediate impact on language proficiency. Students' English proficiency was measured by the test administered in their state—California English Language Development Test (CELDT) in California

[1] The student outcomes analysis only includes the RCT districts. The QED districts were not included in the student outcomes analysis because the process of collecting student data from District D—the largest QED district—would have been too burdensome. District D required parent consent and student assent for collecting any student-level extant data, even de-identified data, which would not have been feasible given evaluation resources.

and WIDA's ACCESS for ELs (Assessing Comprehension and Communication in English State-to-State for English Language Learners) in North Carolina and Virginia.

The student outcomes analysis included 455 ELs in 48 ELA teachers' classrooms in treatment and 469 ELs in 43 ELA teachers' classrooms in control. We used a multilevel model nesting students in teachers to estimate the impact of EdConx on improving ELs' English proficiency scores.[2] Because teachers were randomly assigned to the treatment or control condition a year before students were assigned to teachers during the second year of EdConx implementation, all students who participated in this study are considered "joiners."[3] To include joiners in the analysis for a cluster RCT, baseline equivalence on student outcome measures must be established. After controlling for baseline English proficiency, student demographics, and teacher demographic characteristics, we found no difference in English proficiency scores between treatment and control students at posttest.

Impact of EdConx on Teachers' EL-Relevant Instructional Practice

CAL's theory predicted that the EdConx site and related resources would impact EL's English proficiency by increasing teachers' knowledge, understanding, and use of standards-based, EL-relevant instructional practices. Thus, measuring these dimensions of teacher practice is the first step to understanding how EdConx might affect teachers' instruction and improve student learning. To assess teachers' adoption of the instructional practices supported by EdConx, we surveyed all teachers in the RCT and QED study twice. Survey items asked how often teachers used academic content standards and English language proficiency standards in instruction, how often they used sheltered instruction strategies to support ELs, and how often they used strategies to improve the engagement of ELs in class. Teachers in both the treatment and control/comparison groups completed a survey at baseline, near the beginning of treatment teachers' exposure to EdConx, and they completed the same survey as a follow-up approximately 1.5 years later, at the end of the study period. The survey contained 16 items designed to assess how frequently the teachers engaged in practices that support ELs in accessing academic content. Due to the two different study designs, the analyses for the RCT and QED studies were conducted separately.

RCT. The impact analysis included only teachers in the RCT districts who responded to items about instructional practice on the follow-up survey, regardless of whether they also responded to the baseline survey. Of the 424 teachers randomized at baseline, 215 responded to the follow-up survey. About half of the teachers at baseline attrited from the study, either because they did not respond to the follow-up survey or because they asked to be removed from the study. Overall teacher attrition was 49 percent, with a differential attrition rate of less than 1 percent between treatment and control. These attrition rates fall within the low attrition limits set by the What Works Clearinghouse (What Works Clearinghouse, 2014) and the National Evaluation of i3.[4] Because not

[2] We also ran a three-level model with students nested in teachers and teachers nested in schools. The 3-level model produced similar results as the two-level model.

[3] A "joiner" is defined as a student who joined teachers' classrooms after random assignment. Because we examine student impact of EdConx during the teacher's second year of participation, all students joined the teachers' classrooms after random assignment and are considered "joiners."

[4] When differential attrition is under 1.4 percent, overall attrition can be as high at 60 percent, so attrition here is within the attrition standards to be applied by the NEi3. Because attrition is within acceptable limits, there is a low probably that it systematically biased the teacher sample at follow-up, undermining the validity of the study's experimental design.

all teachers in the RCT who responded to the follow-up survey study had also completed a baseline survey, we used multiple imputation to estimate missing values at baseline.

To test the difference in teacher survey outcomes between treatment and control groups, we developed a multilevel model with teachers nested within schools, controlling for teacher-level demographics, a baseline survey score, and the district dummy variables.

QED study. Twenty treatment teachers and 34 comparison teachers completed the baseline survey, and 18 treatment teachers and 25 comparison teachers completed both baseline and follow-up surveys. Only those teachers with both a baseline and follow-up survey were included in the QED outcomes analysis.

We used propensity score weighting, a rigorous quasi-experimental approach, to examine differences in instructional practice between treatment and comparison teachers. By applying propensity score modeling methods to ensure treatment and comparison groups were equivalent at baseline, we can attribute the group difference in instructional practice at follow-up to EdConx participation. As with the models developed for the RCT, the propensity score-weighted multilevel model nested teachers within schools, controlling for teacher-level demographics, a baseline survey score, and the district dummy variables.

In both the RCT and QED samples, the estimated impact of EdConx on teachers' use of all EdConx-supported instructional practices was small and not statistically different from zero, meaning that on average EdConx had no impact on teacher practice, as measured by the survey.

Implementation of EdConx

To better understand why EdConx resulted in no effects on student and teacher outcomes, we examined the fidelity with which EdConx was implemented. Overall, CAL delivered EdConx components as designed. Despite this, teachers' use of and satisfaction with the site did not meet the benchmarks set by CAL for fidelity of implementation. Forty percent of teachers in the RCT districts (n = 225) and 36 percent of teachers in the QED districts (n = 36) failed to enroll in EdConx by logging on to the site and creating a user profile. Most of the teachers who logged on to the site spent fewer than 100 minutes on the platform in total, well below the intended 30 minutes on average per week. A closer look at treatment teachers' time on the site and their perceptions of the site's relevance, their familiarity with the site, and their satisfaction with EdConx as measured by follow-up survey responses and interviews helps to explain teachers' minimal use of the site.

Treatment teachers in QED districts reported greater familiarity with EdConx and saw it as more relevant than those in the RCT districts, possibly due to more robust recruitment practices and more frequent face-to-face meetings. CAL's recruitment efforts in the QED districts were more robust than those in the RCT districts and included several information sessions for teachers about the program and meetings with principals to discuss the benefits of EdConx. In addition, CAL held several face-to-face sessions in one QED district to address specific teacher needs. These sessions may have helped to attract teachers who judged that EdConx was relevant to their professional needs and priorities, and these sessions may have led them to be more familiar with the site.

Treatment teachers who reported familiarity with EdConx were also more likely to report satisfaction. When asked about their satisfaction with EdConx components on the survey, teachers who reported at least some familiarity with EdConx were also more likely to report

satisfaction with its core components. This finding suggests that teachers who used EdConx and were more familiar with it viewed it as a valuable resource.

The breadth of materials offered on the EdConx platform and the flexibility offered to teachers when navigating the site may have created an online experience with too many disparate components, making it difficult for teachers to identify specific resources relevant to their instructional needs. By design, a key feature of EdConx was to meet each teacher's unique interests and needs. During interviews teachers reported a variety of motivations for engagement with the site, from interest in accessing subject-specific resources for teaching ELs to enrolling in courses that offered university credit, suggesting several different user profiles based on interests and motivation. During interviews, treatment teachers stressed that they valued the content on the site and the flexibility of being able to access EdConx in the manner appropriate to their needs. Several teachers, however, also shared that EdConx, in trying to meet teachers' varying motivations and interests, provided too many different options, which made it difficult to identify the resources that were most appropriate for supporting their instructional needs.

Teachers faced a number of barriers to participating in EdConx, most notably, lack of support from school leaders and trouble with site navigation. Almost one-quarter of teachers in QED and RCT districts reported that they had difficultly taking what they learned from EdConx and applying it in their classroom. Similarly, teachers in the RCT districts QED districts reported that leaders at their schools were largely unaware of the resources available through EdConx. Because school leaders typically support teachers with classroom instruction on a daily basis, their lack of awareness of EdConx resources most likely contributed to teachers' challenges applying what they had learned. Although there was some variation on the survey between RCT and QED teachers' assessment of the EdConx site's ease of navigation, interviews revealed several challenges with the site's structure and difficulty navigating back to useful resources on subsequent visits to the site, likely contributors to teachers' low engagement.

Conclusion and Implications

Although the evaluation found that EdConx had no effect on teachers' use of EL-relevant instructional practices or improvements to students' English proficiency, the implementation findings suggest some key lessons for improving the design and implementation of future online professional development interventions, particularly with an eye to increasing levels of teacher participation. Among these lessons are the following:

- **Target recruitment.** Teachers have varying motivations and priorities for engaging with EdConx. Teachers' interests were often disparate, and EdConx was not able to meet them all simultaneously. Thus, before launching recruitment for professional development, specifying the user profile(s) that program developers theorize will most benefit from the program and targeting those teachers could better yield a group willing and ready to engage fully with the program.
- **Engage school leaders.** During recruitment, CAL worked directly with district officials and rarely had direct contact with school leaders. Conveying to school leaders the value of professional development for the teachers in their school could help to elevate the program to one school leaders and teachers prioritize as a resource for professional learning and growth.

- **Conduct user testing early and often.** Interviews with teachers revealed navigation issues with the EdConx site. Engaging in usability testing early as well as iteratively throughout the development of online professional development could help to identify these kinks and address them immediately to ensure a smooth online experience.
- **Be flexible.** After the first year of implementation, CAL recognized that a fully online program was not suitable for many of the participating treatment teachers. It adjusted the program to a blended model that incorporated face-to-face sessions to complement the online site.

Introduction

Classrooms across the United States have seen a 40 percent increase in English learners (ELs) over the last two decades (Batalova & McHugh, 2010; NCES, 2017). During this time, educators have struggled to close the achievement gap between ELs and their non-EL peers: On the 2015 National Assessment of Educational Progress (NAEP), fourth-grade ELs' performance lagged behind that of their non-EL peers by 36 points in reading and 25 points in math. Among eighth-graders, the gap was even wider, 44 points in reading and 38 points in math, comparable to results on the 2007 NAEP (OELA, 2016). The increase in the EL population and persistent achievement gap underscore the urgency for educators to develop structures and systems that better meet ELs' learning needs.

As the number of ELs in U.S. schools has increased, it has become common practice to include them in the classrooms of mainstream teachers—those who teach in a general education classroom and have been trained in a core subject (math, science, English, social studies)—rather than in the classrooms of bilingual or English as a second language teachers (Fregeau & Leier, 2015; Kaplan et al., 2011; Lucas & Villegas, 2010; Pettit, 2011). Language development experts argue that mainstream teachers can be effective in supporting ELs' academic achievement when they know how to simultaneously develop students' content and language skills in the context of their discipline (Gottlieb, 2013; Master et al., 2015; Turkan et al., 2014). However, mainstream teachers often are inadequately prepared to do this (Turkan et al., 2014). This problem is compounded by the Common Core State Standards and Next Generation Science Standards, which have greater language and content demands and place higher expectations on mainstream teachers to support ELs in accessing challenging content while developing their English language skills (Santos, Darling-Hammond, & Cheuk, n.d.).

The Center for Applied Linguistics (CAL) received an Investing in Innovation (i3) grant in 2014 to address this gap in mainstream teachers' knowledge and skill. It developed an online site, Education Connections (EdConx), to train and support teachers to integrate English language proficiency standards with content standards to deliver lessons that are both content-rich and accessible to ELs. The EdConx site consists of academic and teaching resources vetted by experts in EL instruction, online courses designed and facilitated by instructors from TESOL International, live events/webinars led by subject-matter experts, and forums to support teacher collaboration. In addition to the online resources, CAL held in-person information sessions with teachers to support their use of the EdConx site. From fall 2014 through spring 2017, CAL partnered with six school districts in four states to provide access to the EdConx site to 6th- through 12th-grade teachers of mainstream classrooms. By design, teachers could interact with the site and its available resources in the way that best met their professional needs.

SRI Education evaluated the implementation of EdConx and its impact on teacher practice and students' English language proficiency across the districts. This report presents the results of the evaluation, beginning with an overview of the EdConx site and theory of action and evaluation design, followed by the results of the impact and implementation studies. The report ends with a discussion of implications for future online professional development.

Overview of EdConx

CAL designed the EdConx online site as a one-stop site for 6th- through 12th-grade teachers of mainstream classrooms to access teaching resources and other supports for standards-based instruction relevant to ELs. CAL developed these resources and supports to be consistent with sheltered instructional practices, providing guidance in four key areas:

- **Defining** language and content objectives that are both aligned with English language proficiency and content standards.
- **Modifying** the original curriculum by differentiating, adapting, or scaffolding content to make it more accessible to ELs.
- **Cultivating** the capital ELs bring to the classroom and their learning by encouraging native language development, incorporating varied instructional approaches, and making connections beyond the classroom.
- **Applying** language and content learning in ways that promote target language and native language development.

CAL's design of EdConx was intended to meet the varying needs of teachers and to offer a menu of supports from which teachers could pick and choose to create their own personalized professional development experience. CAL worked with the University of Oregon (UO) to adapt its learning management system, ObaVerse, to deliver each of the elements of the EdConx site. The EdConx online site had the following four components, each providing guidance aligned with the four key areas of sheltered instruction (defining, modifying, cultivating, applying).

- **Forums** offered EdConx users a place to post and discuss ideas, questions, and challenges related to designing and implementing EL-relevant instruction. As part of the forum discussion board, CAL created and disseminated email newsletters ("Tuesday Tips" and "Fridays Facts") with links to online resources, advice from subject matter experts, and ideas for EL instruction.
- **Resources** contained links to articles from peer-reviewed journals, blog posts, lesson plans and other written resources on teaching ELs that had been vetted by experts in the field. CAL offered articles that would allow teachers to learn more about the theory behind best practices for EL instruction as well as practical resources that teachers could implement in their classrooms.
- **Live events/webinars** were hosted by subject matter experts online to discuss and share best practices in EL instruction. Live events/webinars included topics such as cultivating academic language and literacy, enhancing sheltering, formative language assessments for ELs, and a language-based approach to content instruction.
- **Online minicourses** designed and facilitated by staff from TESOL International provided structured training in instructional methods for secondary ELs, oral language development, content-based instruction for ELs, and family-school connections. Titles of courses included Content Area Assessment for Secondary English Learners, Overview of Methods of Instruction for Secondary ELs, and Vocabulary Development and Literacy Development for Secondary ELs.

In addition, core to EdConx implementation was regular communication between CAL and EdConx users and technical assistance from UO to support users with the EdConx site. CAL originally intended to provide EdConx professional development exclusively through the four components of the EdConx site. In Year 2 (2015–16) of program implementation, however, CAL decided to adopt a hybrid approach in response to teacher feedback. It kept the online program as the main source of training and other resources, but offered complementary in-person sessions as well.

As a first step to designing the evaluation, SRI worked with CAL to develop a visual representation of the logic behind the EdConx theory of action. The next section provides details about the theory of action and evaluation design.

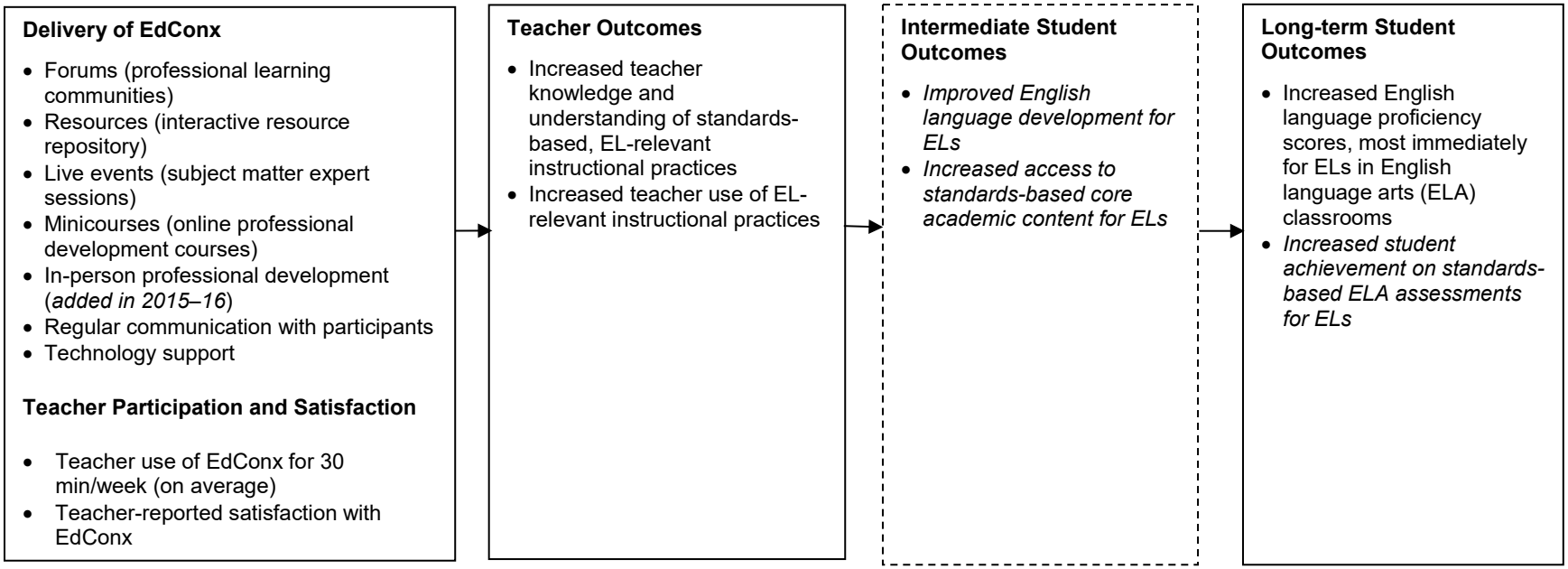
Theory of Action and Evaluation Design

The logic model shows the key components of the EdConx site and the theorized impacts of implementation of the core components of EdConx on teacher practice and student outcomes (Exhibit 1). CAL hypothesized that teachers' engagement with the site for 30 minutes per week over 1.5 years, in whatever way they chose to meet their professional needs, would lead to the following changes in teacher and student outcomes:

- **Teacher outcomes** included expanded teacher knowledge and increased use of standards-based academic content and EL-relevant instructional practices in classrooms that include ELs.
- **Student outcomes** included intermediate outcomes (improved language development and increased access to standards-based instruction in the core content areas) through which EdConx impacts long-term outcomes (increased English language proficiency and improved achievement on standards-based English language arts assessments). Although CAL theorized EdConx's impact on students' long-term outcomes in classrooms of all mainstream teachers, they hypothesized that the impact would be most immediate on students in classrooms of English language arts (ELA) teachers.

To test this theory, CAL collaborated with six districts across four states and partnered with SRI to examine the implementation and impact of EdConx. The following section details district and teacher recruitment and evaluation design.

Exhibit 1. EdConx Logic Model



Note. The names of the core components changed in 2015. The logic model depicted in this report includes these updates, retaining the original component names in parenthesis. Although the names have changed, the core function of each component has remained the same

Italicized text signifies outcomes that the study did not measure.

District and Teacher Recruitment

CAL conducted two rounds of district recruitment, starting with the recruitment of Districts A and B in fall 2014. When teacher recruitment efforts in those districts did not yield the number of teachers they had committed to serve with i3 funds, CAL conducted a second round of district recruitment in spring and fall 2015, which resulted in the participation of four additional districts. Districts A and B are large urban districts in North Carolina and California, respectively. District D is a suburban district in Maryland. Districts C, E and F are small rural districts in Virginia. These recruitment cycles led to three different timelines for participation in the evaluation (Exhibit 2). Within districts, CAL focused recruitment on 6th- through 12th-grade mainstream teachers of core subjects (English/language arts, mathematics, science, and social studies), although a small number of teachers of elective subjects (visual and performing arts, career-technical education, and digital literacy) also participated in the study.

Exhibit 2. Dates of Study Participation

	Start	End
Randomized Controlled Trial		
District A	1/1/2015	6/24/2016
District B	1/1/2015	6/24/2016
District C	6/8/2015	12/31/2016
Quasi-Experimental Design Study		
District D	1/1/2016	6/30/2017
District E	1/1/2016	6/30/2017
District F	1/1/2016	6/30/2017

Impact Studies

Because not all districts agreed to participate in a randomized controlled trial (RCT), the evaluation consisted of two separate studies—a RCT in Districts A, B, and C and a quasi-experimental design (QED) study using propensity score weighting in Districts D, E, and F. Both studies tested CAL's theory that EdConx would lead to increases in teacher knowledge and understanding of standards-based EL-relevant instruction and use of EL-relevant practices, and (in the RCT) students' English language proficiency. In the RCT districts, SRI assigned 6th- through 12th-grade mainstream teachers of core subjects who elected to participate in EdConx to treatment or delayed treatment at the start of the study. The treatment group received access to EdConx immediately, and the delayed treatment group received access after 1.5 years, serving as the control group during the study period. In the QED districts, the study team used propensity score weighting to compare 6th- through 12th-grade mainstream teachers who chose to participate in EdConx for 1.5 years with a group that elected to serve in a comparison group (and thus did not have access to EdConx). Once the treatment group was identified, treatment teachers received an email with login information to access the EdConx site.

To measure impact on student outcomes, we compared the English language proficiency of ELs taught by English language arts (ELA) teachers in the treatment group with the English language proficiency of ELs taught by ELA teachers in the control group. We narrowed the sample to students in classrooms of ELA teachers because CAL theorized that changes to instruction in ELA classrooms would have the most immediate impact on language proficiency. As a measure of English language proficiency, we used scores from state assessments of ELs.[5]

[5] The study team did not assess the impact of EdConx on student achievement as measured by the English (ELA) assessment because ELA assessments are administered to high school students in only one grade level in the four study districts, limiting sample size and the availability of baseline data.

To measure the impact of EdConx on teachers, the study team examined changes to teachers' behaviors and practices as reported on two surveys, one administered at baseline, before treatment teachers began participating in EdConx, and the other at follow-up, after 1.5 years.

Implementation Study

CAL's logic model predicted that teachers' use of any core component of the EdConx site (i.e., resources, live events/webinars, courses, forums) for 30 minutes on average per week over 1.5 years of participation in EdConx would lead to the intended teacher and student outcomes (see Exhibit 1). By design, CAL did not specify thresholds or define expected levels of engagement for each of the four core components of the site (forums, resources, live events/webinars, minicourses), believing that teachers should have the flexibility to choose how to use the EdConx site to address their individual needs. To measure the extent to which CAL implemented EdConx with fidelity, we examined CAL's delivery of EdConx components via the customized site as well as participants' engagement with the site for the intended 30 minutes on average per week and satisfaction with the program.

In the following sections of this report, we present the results of the student and teacher outcomes analyses, describe the fidelity with EdConx was implemented, and consider implications for the effectiveness of future development and implementation of online professional development systems.

Impact of EdConx on Students' English Proficiency

CAL theorized that teachers' participation in EdConx would ultimately lead to long-term student outcomes of improvements in ELs' English language proficiency as measured by state assessment scores (see Exhibit 1). To test this theory, we evaluated the impact of EdConx on ELs' English proficiency scores in Year 2 of the study in the RCT districts only.[6]

Measures of English Proficiency

Students' English proficiency was measured by the California English Language Development Test (CELDT) in District B and by WIDA's ACCESS for ELs (Assessing Comprehension and Communication in English State-to-State for English Language Learners) in Districts A (North Carolina) and C (Virginia). Test developers have established the reliability and validity of both tests, and both tests have been adopted statewide by the respective states. To pool students' English proficiency scores from North Carolina, California, and Virginia, we first standardized the scores of each test by grade within district(s) and then aggregated the data across tests.

Randomized Controlled Trial

Student sample. The study narrowed the sample of ELs in the analysis to only those in classrooms of ELA teachers participating in the RCT. Because teachers were randomly assigned to treatment or control condition a year before students were assigned to teachers during the second year of EdConx implementation, all students who participated in this study are considered "joiners."[7] To include joiners in the analysis for a cluster RCT, baseline equivalence on student outcome measures must be established. Exhibit 3 shows the student analytic sample. We define students in the analytic sample as those whose ELA teacher participated in the RCT and who had both baseline and posttest English proficiency scores.[8] Exhibit 4 shows baseline measure for English proficiency.

Exhibit 3. Schools, Teachers, and Students in the Student Outcomes Sample

	Schools			Teachers			Students		
	Treat-ment	Control	Total	Treat-ment	Control	Total	Treat-ment	Control	Total
District A	16	21	37	21	25	46	106	138	244
District B	9	6	15	15	9	24	214	141	355
District C	2	3	5	12	9	21	135	190	325
Total	27	30	57	48	43	91	455	469	924

[6] The student outcomes analysis only includes the RCT districts. The QED district were not included in the student outcomes analysis because the process of collecting student data from District D—the largest QED district—would have been too burdensome. District D required parent consent and student assent for collecting any student-level extant data, even de-identified data, which would not have been feasible given evaluation resources.

[7] A "joiner" is defined as a student who joined teachers' classrooms after random assignment. Because we examine student impact of EdConx during the teacher's second year of participation, all students joined the teachers' classrooms well after random assignment and are considered late joiners.

[8] For District A, we used WIDA ACCESS scores in spring 2015 as baseline and spring 2016 as posttest. For District B, we use CELDT scores in fall 2015 as baseline and fall 2016 as posttest. For District C, we use WIDA ACCESS in spring 2016 as baseline and spring 2017 as posttest.

Exhibit 4. EL Baseline Measures of English Proficiency

Baseline Measure of English Proficiency	Treatment Mean (*SD*)	Control Mean (*SD*)	*d* [a]
Composite score	0.06 (1.00)	-0.06 (1.00)	0.12

[a]*d* is standardized mean difference, which is the difference between the mean outcome of the intervention group and the mean outcome of the comparison group divided by the pooled within-group standard deviations (What Works Clearinghouse, 2014).

Student outcomes analysis. We estimated the impact of EdConx on improving ELs' English proficiency scores using a multilevel model nesting students in teachers.[9] To establish baseline equivalence, we first standardized English proficiency scores by grade within districts. Then we calculated baseline difference using standardized mean differences between the two groups. Last, the average impact of EdConx on English proficiency scores was evaluated using a multilevel model pooling students from the three RCT districts.

Treatment and control students' baseline scores differed by 0.12 standard deviation—within the What Works Clearinghouse baseline equivalence requirement. The size of the difference was small and not statistically different from zero, meaning that on average treatment and control students had very similar English proficiency at baseline. The main independent variable of interest in the model is the teacher-level dummy variable indicating treatment condition (treatment or control group). Covariates were added to the model to control for differences in students' baseline English proficiency scores and other student demographic characteristics such as race/ethnicity, gender, special education status, and grade level. In addition, because student assignment to teachers was not random, we also controlled for teacher-level background characteristics including gender, total years of experience, and education degree. District dummy variables were also included in the model.

Findings from the RCT Student Outcomes Analysis

After controlling for baseline English proficiency, student demographics, and teacher demographic characteristics, we found no difference in English proficiency scores between treatment and control students at posttest (Exhibit 5). The estimated impact of EdConx on improving students' English proficiency was small and not statistically different from zero, meaning that on average EdConx had no impact on students' English proficiency, as measured by CELDT and WIDA ACCESS exams.

Exhibit 5. EdConx Impact on ELs' English Proficiency Scores

English Proficiency Scores	Estimated Impact	Standard Error	Effect Size	Sample Size	
				Students	Teachers
Composite scores	0.05	0.06	0.05	924	91

Note. Two-level HLM was used where students were nested in teachers. A positive effect size favors the intervention group and a negative number favors the comparison group. Effect size measures the change (in standard deviations) in an average participant's outcome that can be expected if the participant was exposed to the intervention. Estimated impact and standard errors are the coefficient and standard errors associated with intervention variable from the two-level HLM model (students nested in teachers). Effect size = Estimated impact/ pooled standard deviations of the intervention and control group.
*p < .05, **p < .01, ***p < .001.

[9] We also ran a three-level model with students nested in teachers and teachers nested in schools. The 3-level model produced similar results as the two-level model.

Impact of EdConx on Teachers' EL-Relevant Instructional Practice

CAL's theory predicted that EdConx site and related resources would impact EL's English proficient by increasing teachers' knowledge, understanding, and use of standards-based, EL-relevant instructional practices (see Exhibit 1). Measuring these dimensions of teacher practice is the first step to understanding how EdConx might affect teachers' instruction and improve student learning. Here, we present the results of the RCT and QED studies assessing the impact of EdConx on study teachers' classroom instruction after 1.5 years of participation in EdConx.

Measures of Instructional Practice

To measure teachers' adoption of the instructional practices supported by EdConx, we surveyed all teachers in the study twice. Teachers in both the treatment and control/comparison groups completed a survey at baseline, near the beginning of treatment teachers' exposure to EdConx, and they completed the same survey as a follow-up approximately 1.5 years later, at the end of the study period.

The survey contained 16 items designed to assess how frequently the teachers engaged in practices that support ELs in accessing academic content. The items were replicated from two surveys on instructional practices for ELs: Grady and O'Dwyer (2014) and Council of Chief State School Officers (n.d.). In addition, SRI developed new items based on the eight components of EL instruction presented in *Making Content Comprehensible for English Learners: The SIOP Model* (4th Edition) (Echevarria, Vogt, & Short, 2012). Items asked how often teachers used academic content standards and English language proficiency standards in instruction, how often they used sheltered instruction strategies to support ELs, and how often they used strategies to improve the engagement of ELs in class. Teachers reported the frequency of these practices in their classrooms on a 6-point scale, ranging from *Almost never* to *Nearly daily*.

We performed an exploratory factor analysis using baseline survey data from respondents in all six districts to create a conceptually relevant outcome measure that would also be reliable and useful for analysis. The factor analysis demonstrated that all 16 survey items contributed to a single scale measuring teachers' EL-relevant instructional practices. The resulting scale was highly reliable, achieving an alpha of .85 (Exhibit 6). See Appendix A for the items in this scale.

After factor analysis, we created an instructional practice factor score by taking the average of all 16 items under this factor. This approach keeps the composite variable in the same scale as the original items, which makes it easier to interpret. Exhibit 6 shows the mean and standard deviation of the instructional practice survey scale.

Exhibit 6. Properties of Teacher Instructional Practice Scale

	Baseline Survey[a]		Follow-up Survey[b]			
	Mean	(SD)	Mean	SD	Alpha	Number of Items
Teachers' Use of All EdConx-Supported Instructional Practices	4.66	0.77	4.60	0.78	.85	16

[a] *n* = 257 combining teachers from the RCT and QED districts.

[b] *n* = 271 combining teachers from the RCT and QED districts.

Source: EdConx baseline and follow-up teacher survey.

Randomized Controlled Trial

The following section describes the RCT teacher sample and teacher outcomes impact analysis.

Teacher sample. The impact analysis included all teachers in the RCT districts who responded to items about instructional practice on the follow-up survey, regardless of whether they also responded to the baseline survey. Of the 424 teachers randomized at baseline, 215 responded to the follow-up survey. About half the teachers from baseline attrited from the study, either because they did not respond to the follow-up survey or because they asked to be removed from the study (Exhibit 7). Overall teacher attrition was 49 percent, with a differential attrition rate of less than 1 percent between treatment and control. These attrition rates fall within the low attrition limits set by the What Works Clearinghouse (What Works Clearinghouse, 2014) and the National Evaluation of i3.[10]

Exhibit 7. Teacher Attrition from Baseline to Follow-up, RCT

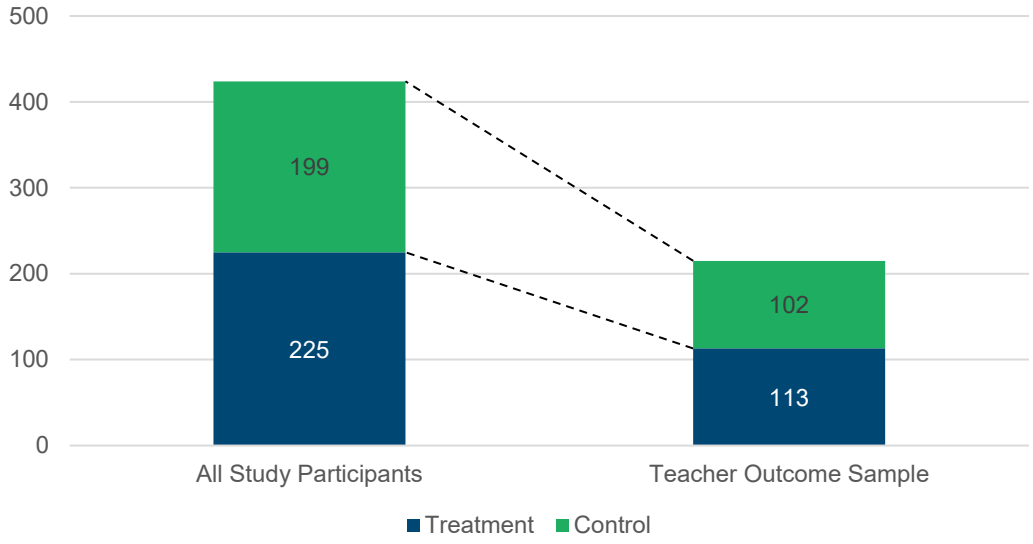


Exhibit 8 shows schools and teachers in the teacher analytic sample in the RCT by district.

[10] When differential attrition is under 1.4 percent, overall attrition can be as high at 60 percent, so attrition here is within the attrition standards to be applied by the NEi3. Because attrition is within acceptable limits, there is a low probably that it systematically biased the teacher sample at follow-up, undermining the validity of the study's experimental design.

Exhibit 8. Schools and Teachers in the Teacher Analytic Sample, RCT

	Schools[a]			Teachers		
	Treatment	Control	Total	Treatment	Control	Total
District A	32	30	42	59	57	116
District B	10	8	10	44	36	80
District C	2	3	3	10	9	19
Total	44	41	55	113	102	215

[a]In several cases, control and treatment teachers taught at the same schools. Thus, the total number of schools accounts for this duplication and is not a sum of the treatment and control columns.

Although all 6th- through 12th grade mainstream teachers were eligible for the study, most treatment and control teachers taught in high schools (74 percent of treatment teachers and 66 percent of control), with a smaller proportion teaching in middle schools. Nearly all study teachers taught core academic subjects (i.e., English language arts, math, science and social studies), most likely a result of CAL's focus on recruiting mainstream teachers of core content areas. A small number of teachers taught elective courses such as the visual and performing arts, career-technical education, and digital literacy (Exhibit 9).

Exhibit 9. Characteristics of Teachers at Follow-up, RCT

	Treatment (*n* = 113)	Control (*n* = 102)
Grade Level		
High school	74%	66%
Middle school	26%	34%
Subject[a]		
English language arts	39%	36%
Math	24%	25%
Science	30%	25%
Social science	11%	11%
Other (e.g., the arts, career-technical education, digital literacy)	2%	5%
	Treatment (*n* = 99)	**Control (*n* = 91)**
EL Focal Class[b]		
Class size	28	28
Number of ELs	8	7
Percentage of ELs	28%	29%

[a] Percentages do not sum to 100 because some teachers reported teaching more than one subject.
[b] The class with the highest proportion of ELs.
Source: EdConx follow-up teacher survey.

Because CAL developed EdConx for teachers who teach ELs in mainstream classes, we asked the teachers to report the number of ELs they taught in the class with the highest proportion of ELs. Both treatment and control teachers reported an average class size of 28 students (Exhibit 9). Of the students enrolled in these classes, treatment and control teachers reported that ELs accounted for nearly 30 percent of their focal classes, suggesting that EdConx succeeded in enrolling its intended audience in the project.

Teacher outcomes analysis. To evaluate the impact of EdConx participation on teachers, we compared treatment and control teachers' scores on a survey scale that captured dimensions of teachers' EL-relevant instructional practice. Because survey responses of teachers from the same school generally tend to be more alike than those of teachers chosen randomly from the population at large, it is necessary to account for the dependency in the data (Hedeker & Gibbons, 2006). To test the difference in teacher survey outcomes between treatment and control groups, we developed a multilevel model with teachers nested within schools, after controlling for teacher-level demographics, a baseline survey score, and the district dummy variables. The teacher-level demographic variables collected from district administrative data sets and included in these models were years of experience, gender, and highest degree attained. Multilevel modeling was selected for the analysis because of nesting of teachers within schools, which recognize the existence of such data hierarchies by allowing for residual components at each level. A two-level model that allows for grouping teacher survey outcomes within schools includes residuals at the teacher and school levels. Thus, the residual variance is partitioned into a between-school component (the variance of the school-level residuals) and a within-school component (the variance of the teacher-level residuals). The school residuals, often called *school effects*, represent unobserved school characteristics (such as school environment, leadership, and student composition) that affect teacher survey outcomes. It is these unobserved variables that lead to correlation between outcomes for teachers from the same school. More details on our approach to multilevel modeling are included in Appendix B.

Because not all teachers in the RCT who responded to the follow-up survey study had also completed a baseline survey, we used multiple imputation to estimate missing values on the baseline surveys. Imputation uses the information on nonmissing cases to make an educated guess about the values of the missing cases. Multiple imputation pools results from several iterations of imputation to obtain more reliable estimates than other imputation methods. With imputation of baseline scores, the sample in the impact estimates was 215 teachers and 55 schools (refer to Appendix B for a more detailed description of the analysis).

Quasi-Experimental Design Study

The following section describes the QED teacher sample and teacher outcomes impact analysis.

Teacher sample. Teachers from three districts participated in the QED study. Twenty treatment teachers and 34 comparison teachers completed the baseline survey, and 18 treatment teachers and 25 comparison teachers completed both baseline and follow-up surveys. Only those teachers with both a baseline and follow-up survey were included in the QED outcomes analysis. Although attrition is not relevant for QED studies in the same way as for RCTs, the teacher sample attrition from baseline to follow-up is shown in Exhibit 10 to provide readers with a complete description of our QED sample.

Exhibit 10. Teacher Attrition from Baseline to Follow-up, QED Study

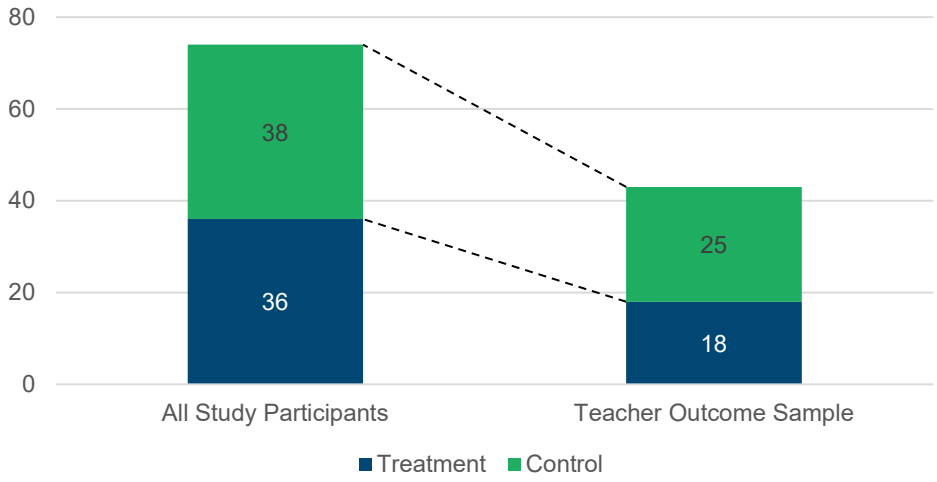


Exhibit 11 shows the number of schools and teachers included in the teacher outcomes sample, by district.

Exhibit 11. Schools and Teachers in the Teacher Analytic Sample, QED Study[a]

	Schools			Teachers		
	Treatment	Comparison	Total	Treatment	Comparison	Total
District D	11	9	13	14	21	35
District E	1	0	1	1	0	1
District F	2	1	2	3	4	7
Total	14	10	16	18	25	43

[a]Teachers completed both baseline and follow-up surveys.

Like the RCT, the QED study included 6th- through 12th- grade mainstream teachers, the majority of whom taught in high schools (72 percent of treatment teachers and 60 percent of control). Most teachers taught core academic subjects (i.e., English language arts, math, science, and social studies), consistent with CAL's targeting of core academic subject teachers during recruitment (Exhibit 12). A larger number of teachers taught elective subjects such as the arts, technical education, and digital literacy than in the RCT districts (17 percent of treatment and 24 percent of comparison teachers).

Treatment and comparison teachers taught similar numbers of students, reporting an average class size of 28 students. On average, treatment teachers reported that ELs made up larger proportions of their classes (52 percent) compared with comparison teachers (38 percent).

Exhibit 12. Characteristics of Teachers at Follow-up, QED Study

	Treatment (*n* = 18)	Comparison (*n* = 25)
Grade Level		
High school	72%	60%
Middle school	28%	40%
Subject[a]		
English language arts	28%	20%
Math	17%	24%
Science	6%	4%
Social science	33%	28%
Other (e.g., the arts, career-technical education, digital literacy)	17%	24%
EL Focal Class[b]		
Class size	28	28
Number of ELs	14	11
Percentage of ELs	52%	38%

[a] Percentages do not sum to 100 because some teachers reported teaching more than one subject.

[b] The class with the highest proportion of ELs.

Source: EdConx follow-up teacher survey.

Teacher outcomes analysis. In the QED study, we used propensity score weighting, a rigorous quasi-experimental approach, to examine differences in instructional practice between treatment and comparison teachers. By applying propensity score modeling methods to ensure treatment and comparison groups were equivalent at baseline, we can attribute the group difference in instructional practice at follow-up to EdConx participation (see Appendix B for details of the propensity score approach in.)

Before conducting propensity score weighting, we first examined the standardized mean score difference (the difference in means for the treatment and comparison groups, divided by a pooled standard deviation) using unweighted mean and standard deviations to identify how much treatment and comparison teachers differed on each baseline variable. After propensity score weighting for comparison teachers, we reexamined the standardized mean score difference to ensure that differences were less than 0.25, the standard recommended by the What Works Clearinghouse (What Works Clearinghouse, 2014), thereby demonstrating covariate balance on most baseline measures.

The balance on the covariates in the analysis improved after application of propensity score weighting (Exhibit 13). The standardized mean difference before propensity score weighting was greater than 0.25 standard deviation for years of teaching, BA degree, and baseline instructional practice. All of these large differences were reduced to less than 0.25 standard deviation between the two groups after propensity score weighting, creating two groups similar on all potentially confounding covariates included in the analyses.

Exhibit 13. Teacher Baseline Measures of Instructional Practice, QED Study

	Before Propensity Score Weighting			After Propensity Score Weighting	
Teacher Demographic Characteristics and Survey Score	Treatment Mean (*SD*)	Comparison Mean (*SD*)	*d* [a]	Weighted Comparison Mean (*SD*)	*d* [a]
Years of teaching	12.67 (6.15)	8.44 (5.42)	0.74	11.70 (4.52)	0.18
Female	0.83 (0.38)	0.80 (0.41)	0.08	0.87 (0.28)	-0.14
BA[b]	0.28 (0.46)	0.44 (0.51)	-0.33	0.24 (0.35)	0.11
Use of all EdConx-supported instructional practices at baseline	5.17 (0.64)	4.60 (0.91)	0.70	5.11 (0.55)	0.09

[a]*d* is standardized mean difference, which is the difference between the mean outcome of the intervention group and the mean outcome of the comparison group divided by the pooled within-group standard deviations (What Works Clearinghouse, 2014).

[b]BA = highest degree was a bachelor of arts.

To improve model precision, we adjusted for potential confounding baseline covariates as recommended by Rosenbaum and Rubin (1983). Specifically, the weight for treated teachers was 1.0, and the weight for comparison teachers was equal to $p_i/(1 - p_i)$, where p_i is the propensity score for the *i*-th comparison student. This approach weights the comparison group to create balance with the treatment group on observed covariates and thus estimates the effect of EdConx for the individuals who elected to be part of the treatment group. We selected weighting over other approaches, such as matching, because weighting successfully yields a comparison group equivalent to the treatment group at baseline using the available data (details below), provides flexibility with the distribution of the data, and retains all subjects in the analysis, maximizing statistical power to detect program effects.

As with the models developed for the RCT, the propensity score weighted multilevel model nested teachers within schools, controlling for teacher-level demographics, a baseline survey score, and the district dummy variables. The teacher-level demographic variables collected from district administrative data sets and included in these models were years of experience, gender, and highest degree attained a bachelor's.

Findings from the RCT and QED Study Teacher Outcomes Analyses

Exhibits 14 and 15 show the estimated difference in survey outcomes scores between treatment and control/comparison teachers controlling for baseline factors (baseline survey scores, teacher characteristics, and district) in the RCT and QED analyses, respectively. Effect size, measured in standard deviations of the underlying distribution of scale scores across sites, represents the impact of EdConx on teacher practice. In both RCT and QED samples, the estimated impact of EdConx on improving teachers' use of EdConx-supported instructional practices was not statistically different from zero, meaning that on average EdConx had no impact on teacher practice, as measured by the survey. In addition, the coefficients for treatment impact estimated here have

relatively large standard errors, which means that any differences between treatment and control are very close to zero and quite likely due to chance.

Exhibit 14. EdConx Impact on Teacher Measures of Instructional Practice, RCT

Teacher Outcome	Estimated Impact	Standard Error	Effect Size	Sample Size	
				Teachers	Schools
Use of all EdConx-supported instructional practices	-0.04	0.09	-0.05	215	55

Note. A two-level hierarchical linear model (HLM) was used where teachers were nested in schools. A positive effect size favors the treatment group and a negative number favors the comparison group. Effect size measures the change (in standard deviations) in an average participant's outcome that can be expected if the participant was exposed to the intervention. Estimated impact and standard errors are the coefficient and standard errors associated with the intervention variable from the two-level HLM model (teachers nested in schools). Effect size = Estimated impact/pooled standard deviations of the intervention and comparison groups. *p < .05, **p < .01, ***p < .001.

Exhibit 15. EdConx Impact on Teacher Measures of Instructional Practice, QED Study

Teacher Outcome	Estimated Impact	Standard Error	Effect Size	Sample Size	
				Teachers	Schools
Use of all EdConx-supported instructional practices	-0.25	0.16	-0.35	43	16

Note. A two-level propensity score weighted HLM was used where teachers were nested in schools. A positive effect size favors the intervention group and a negative number favors the comparison group. Effect size measures the change (in standard deviations) in an average participant's outcome that can be expected if the participant was exposed to the intervention. Estimated impact and standard errors are the coefficient and standard errors associated with intervention variable from the two-level HLM model (teachers nested in schools). Effect size = Estimated impact/ standard deviations of the intervention group (Stuart, Lee, & Leacy, 2013). *p < .05, **p < .01, ***p < .001.

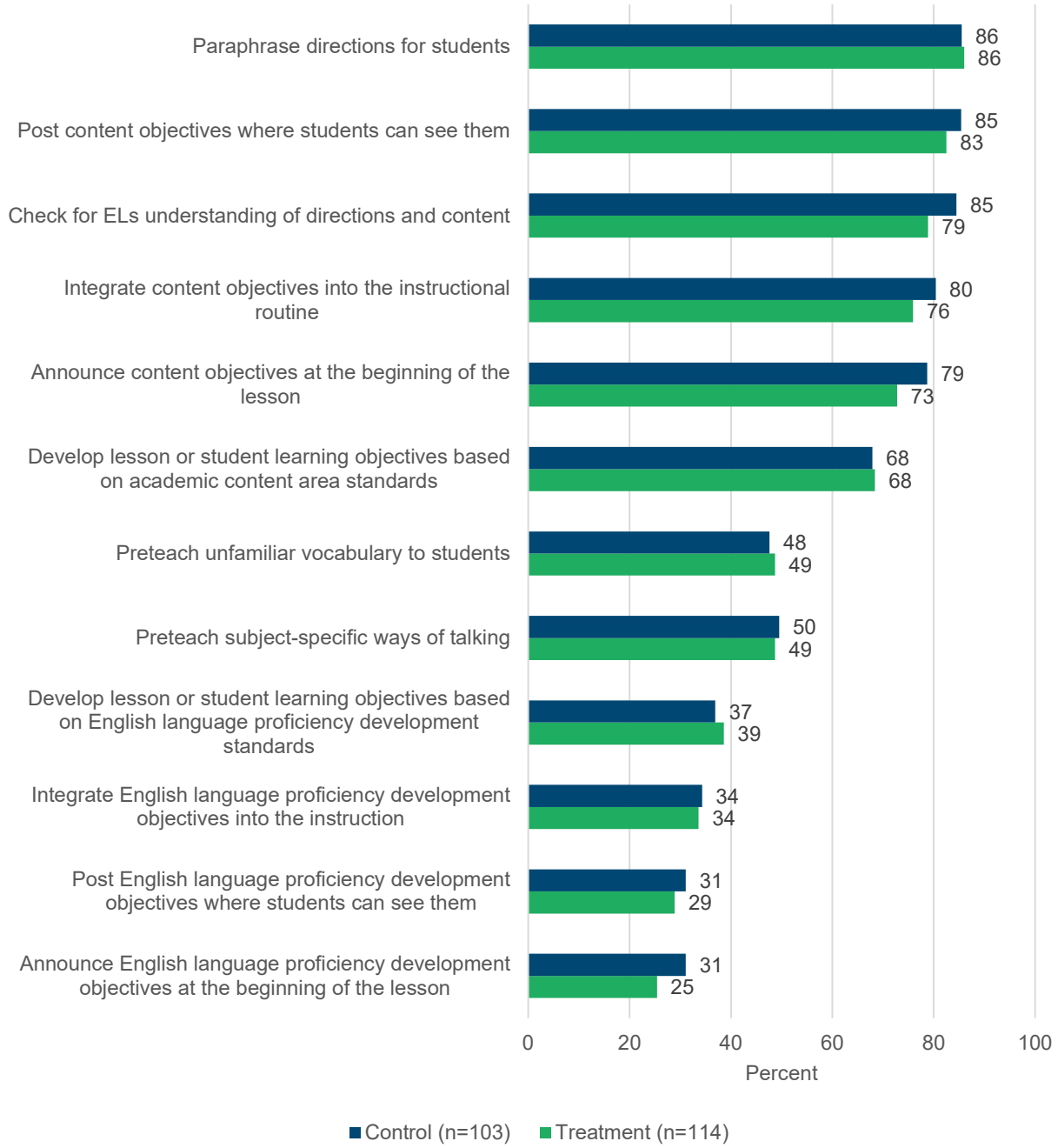
Teachers' Use of EL-Relevant Instructional Practices

The RCT and QED impact analyses on teacher outcomes showed no differences between treatment and control teachers' use of EL-relevant instructional practices after 1.5 years of EdConx participation (and, similarly, no difference between treatment and control students' English proficiency in the RCT districts). To better understand how the frequency of teachers' use of EL-relevant practices differed between treatment and control/comparison teachers, we examined teachers' reported use of these instructional practices on the follow-up survey. Consistent with the impact analysis reported above, treatment and control/comparison teachers used individual EdConx-supported practices with similar frequency. In addition, the majority of teachers in both groups reported using 10 of 16 EdConx-supported instructional practices at least twice a week at baseline, suggesting that these practices were more or less routine for most teachers even before engagement with EdConx.

A closer look at frequencies for individual survey items shows some room for growth in a few key areas. For example, teachers reported incorporating English language proficiency objectives into their instruction (a practice designed to benefit ELs in particular) much less often than content objectives (a practice that benefits all students). Between 60 and 90 percent of treatment and control/comparison teachers reported posting content objectives, announcing them at the beginning of class, incorporating content objectives into their instructional routine, and developing lesson or student learning objectives based on content standards at least twice per week (Exhibits 16 and 17).

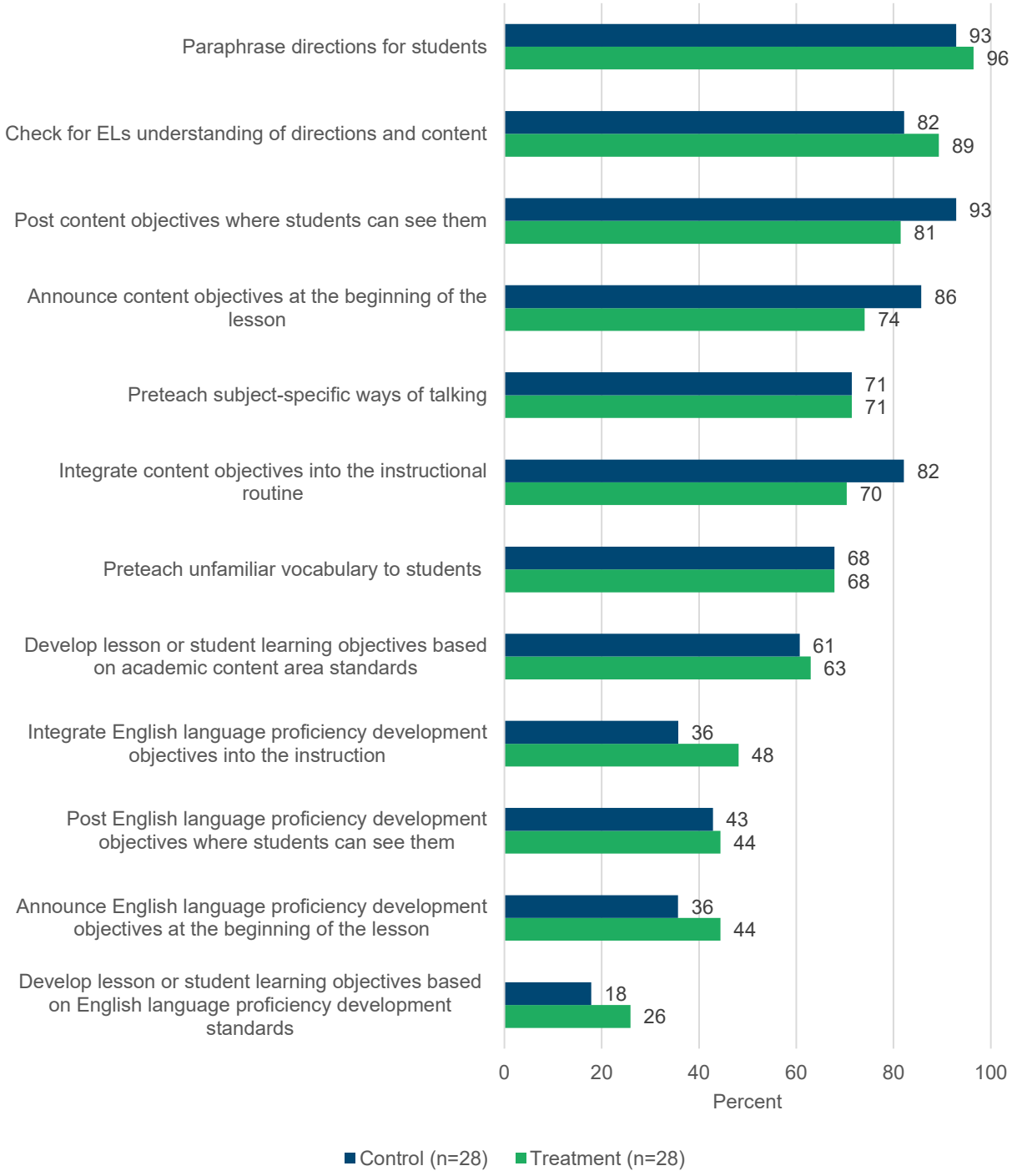
In comparison, between 20 and 50 percent of both treatment and control/comparison teachers reported developing lessons or student learning objectives based on English language proficiency development standards or incorporating English language proficiency objectives into their classroom instruction at least twice per week. Both treatment and control/comparison teachers appear to be less accustomed to planning lessons around objectives related to developing the language skills of ELs and/or require additional support to develop this practice.

Exhibit 16. Teachers' Use of EL-Relevant Instructional Practices at Least Twice per Week, RCT



Source: EdConx follow-up teacher survey.

Exhibit 17. Teachers' Use of EL-Relevant Instructional Practices at Least Twice per Week, QED Study

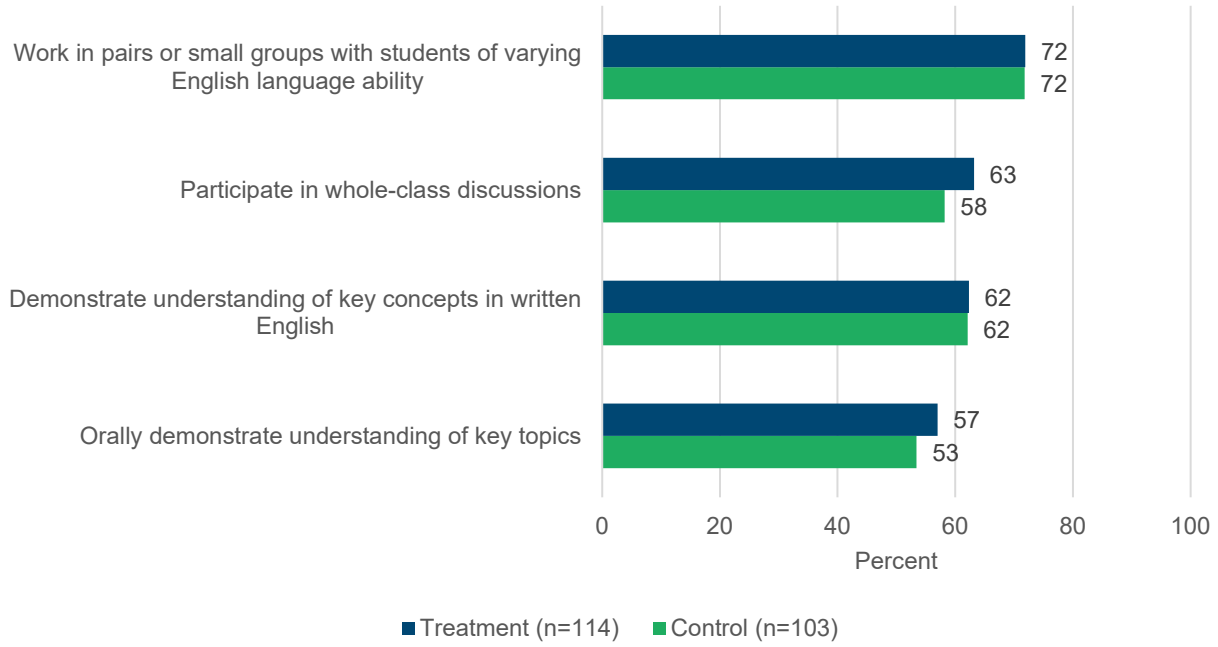


Source: EdConx follow-up teacher survey.

Similarly, both treatment and control/comparison teachers made less frequent use of instructional practices that specifically benefit ELs. Between 79 and 96 percent of treatment and control/comparison teachers reported checking for student understanding of directions and content and paraphrasing directions at least twice per week. In contrast, only 49 to 71 percent of treatment and control/comparison teachers reported using preteaching strategies, particularly relevant to ELs, at least twice a week. One possible explanation is that checking for understanding of directions and content and paraphrasing directions are strategies teachers use with all students regardless of English proficiency, and may thus, be a more routine part of their practice than preteaching unfamiliar vocabulary or subject-specific ways of talking.

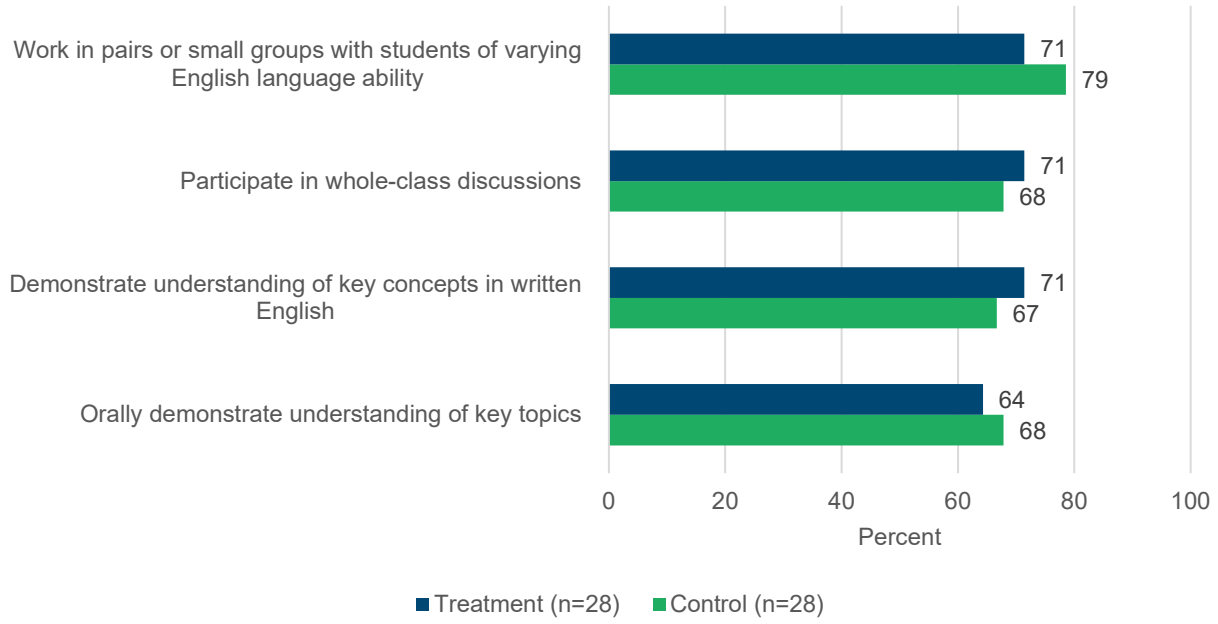
The frequency with which a majority of both treatment and control/comparison teachers reported grouping students of varying English language abilities, using whole-class discussions and requiring students to demonstrate understanding of key concepts orally and in written English suggests that they strived equally to create cultures of inclusiveness in which ELs are encouraged to participate (Exhibit 18 and 19).

Exhibit 18. Teachers Reporting EL Students' Participation in Class Activities at Least Twice per Week, RCT



Source: EdConx follow-up teacher survey.

Exhibit 19. Teachers Reporting EL Students' Participation in Class Activities at Least Twice per Week, QED Study



Source: EdConx follow-up teacher survey.

Both treatment and control/comparison teachers teach classes with large proportions of ELs. Thus, it is not surprising that both groups attempted to meet the needs of ELs by integrating content and language objectives into lessons, using instructional strategies that support ELs, and using strategies to improve the engagement of ELs in class. However, it is noteworthy that similarly small proportions of treatment and control/comparison teachers reported using English language proficiency standards and preteaching strategies given EdConx's specific focus on these areas. To better understand why there were no differences between treatment and control/comparison groups, we examined program implementation.

Implementation of EdConx

To complement the impact analyses reported above, we examined the extent to which EdConx was implemented as designed, with fidelity to the model as envisioned by CAL. We used data for the larger, baseline sample of teachers in the RCT and QED study districts rather than focusing narrowly on the sample of teachers who responded to the follow-up survey and thus were included in the outcomes sample, in order to capture fidelity of the program as offered to all teachers.[11] We examined two components of fidelity: CAL's delivery of the EdConx model (with 8 indicators chosen to assess implementation of the component, as shown in Appendix C) and treatment teachers' participation in and satisfaction with EdConx (with 2 indicators). In consultation with CAL, we set thresholds to indicate high, medium, or low fidelity of implementation on each indicator, and rules for rolling up indicators to the component level.[12]

Overall, CAL delivered EdConx as designed. CAL developed the EdConx site and delivered forums, minicourses, and live events/webinars as planned, meeting the threshold for high fidelity on most indicators measuring the delivery of the EdConx site components. The only indicator for which CAL did not meet the threshold for high fidelity was updating and maintaining EdConx resources. In Year 2, CAL met the standard for medium fidelity (posting 6–11 resources per month) in five districts and low fidelity (posting 5 or fewer resources per month) in one district (Exhibits C-1 and C-3). Despite high fidelity in the delivery of EdConx overall, teachers' use of and satisfaction with the site did not meet the fidelity benchmarks set by CAL.

According to the EdConx logic model (Exhibit 1), CAL theorized that if teachers participated in EdConx for 30 minutes on average per week in any of the core components over the course of the 1.5 years of implementation, the program would lead to an increase in teachers' use of EL-relevant instructional practices and improvements to ELs' English proficiency. Yet, of the 225 treatment teachers in the RCT baseline sample, 40 percent of teachers did not complete the first step of enrollment in EdConx by logging on to the site and creating a user profile (Exhibit 20). Similarly, of the 36 treatment teachers in the QED teacher outcomes sample, 36 percent had not logged on to the site even once during the study period. In addition, most of the teachers who logged on to the site spent less than 100 minutes total, well below the intended 30 minutes on average per week (Exhibit 21).[13] No treatment teachers attended any of the live events/webinars offered during the study period and only 25 teachers completed at least one minicourse across the RCT and QED districts.[14] As a result, only few treatment teachers engaged with EdConx as intended.

[11] The impact estimates reported in the teacher and student outcomes analysis represent an intent-to-treat (ITT) analysis, where all teachers were included in the sample without regard to how much they participated in EdConx. For the fidelity of implementation analysis, we included all teachers in the baseline sample on measures of engagement with EdConx, regardless of their inclusion in the teacher or student outcomes analyses. Survey exhibits include only treatment teachers who responded to the follow-up survey.

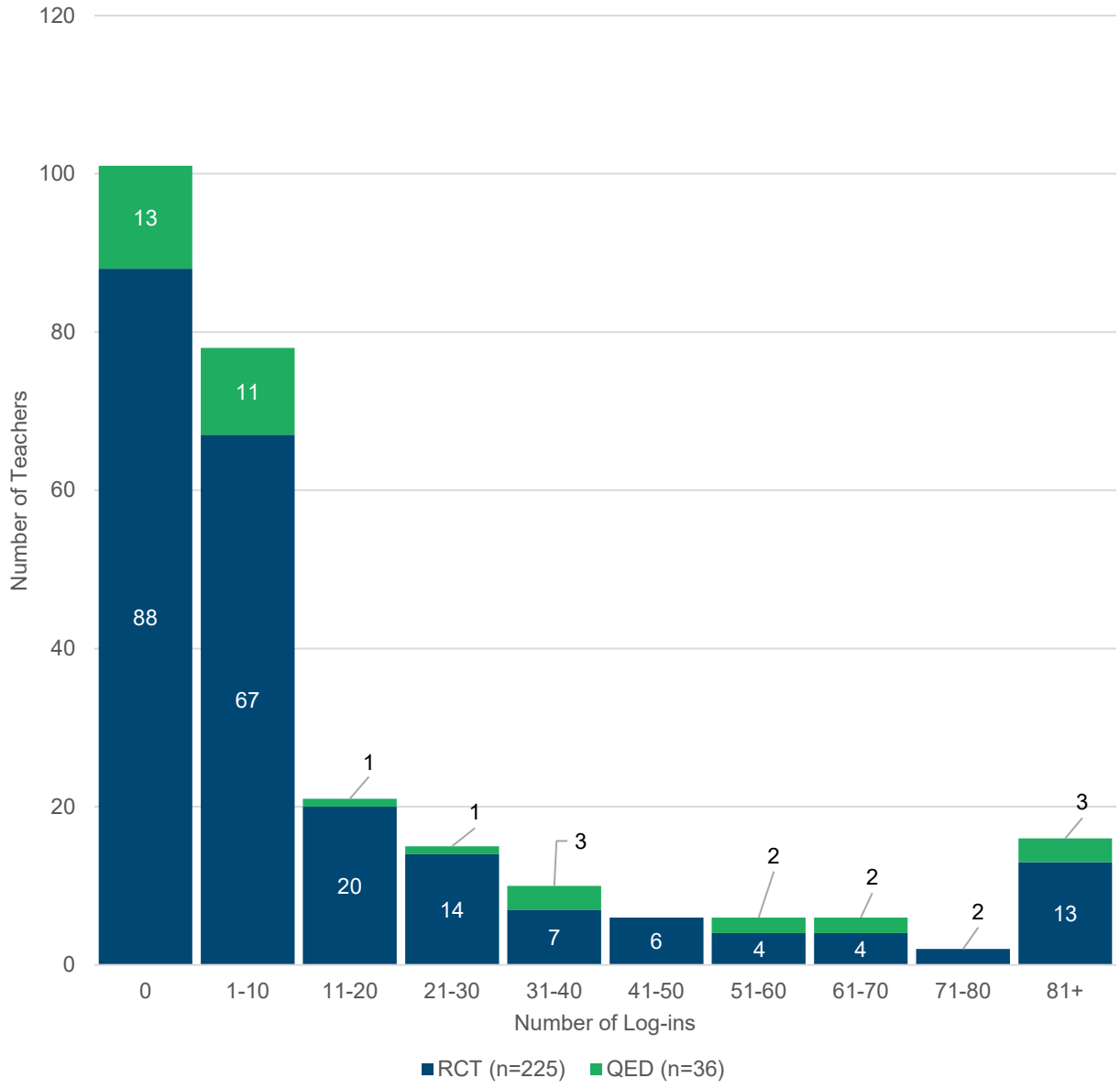
[12] Lacking findings from directly applicable prior studies, CAL used expert judgment to estimate thresholds for high-, medium-, and low-level implementation fidelity for EdConx.

[13] Exhibit 20 and 21 account only for log-ins to the EdConx site and does not include teachers' engagement in in-person events or live events/webinars.

[14] CAL opened the live events/webinars to district staff, EL coordinators, and teachers not participating in the study. Although no study participants attended the live events/webinars, non-study participants attended these events.

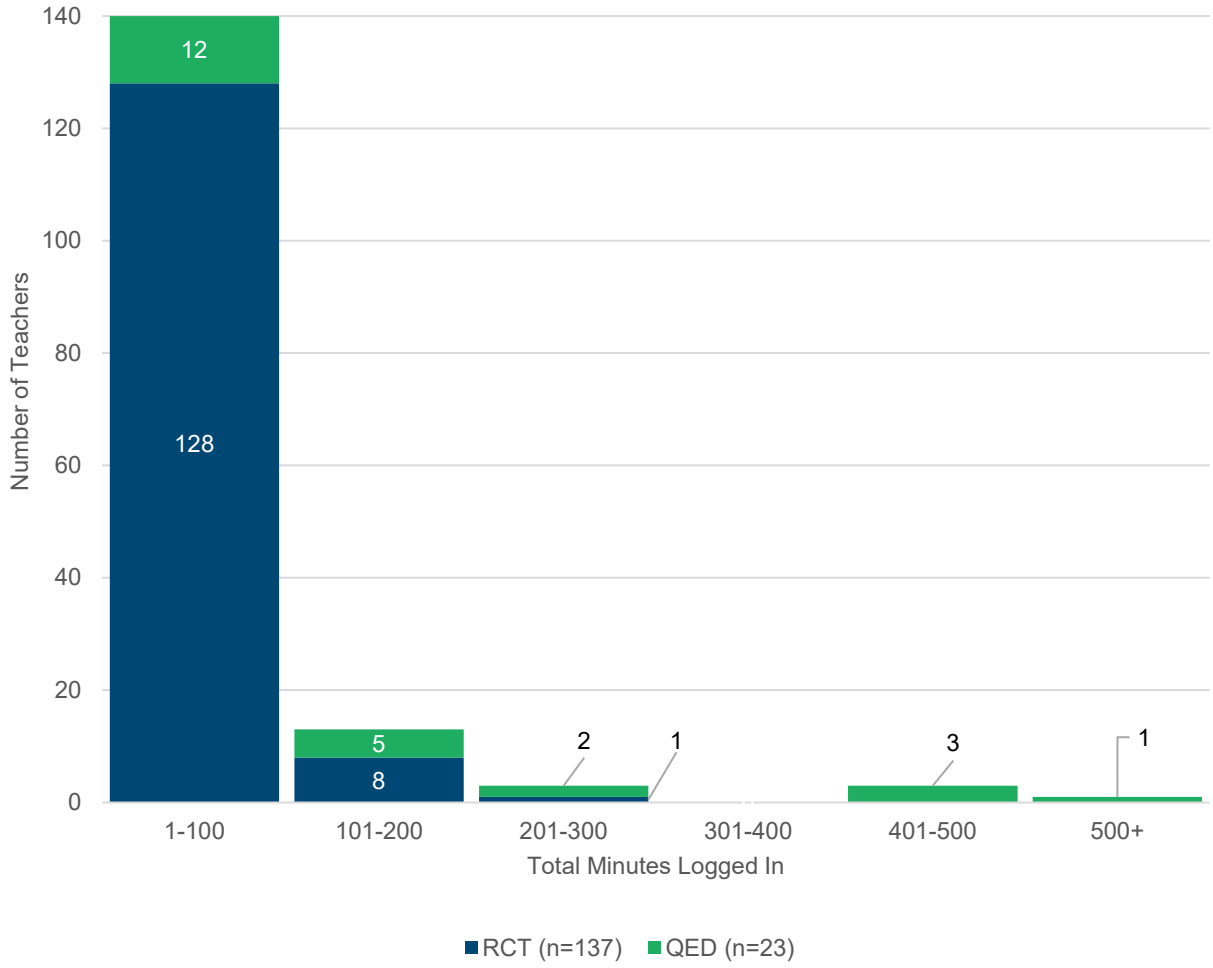
A closer look at treatment teachers' time on the site, follow-up survey responses, and interviews helps to explain their minimal use of the site. Survey responses and interviews provide more information about teachers' motivations for participating in EdConx, perceptions of the program's relevance to their needs, their experience participating in the program, and the factors that facilitated and hindered their participation in EdConx. The following section provides key findings from this analysis.

Exhibit 20. Total Number of Log-ins to EdConx, Among Teachers



Source: University of Oregon user statistics.

Exhibit 21. Total Number of Minutes Logged in to EdConx, Among Teachers Who Logged in at Least Once



Source: University of Oregon user statistics.

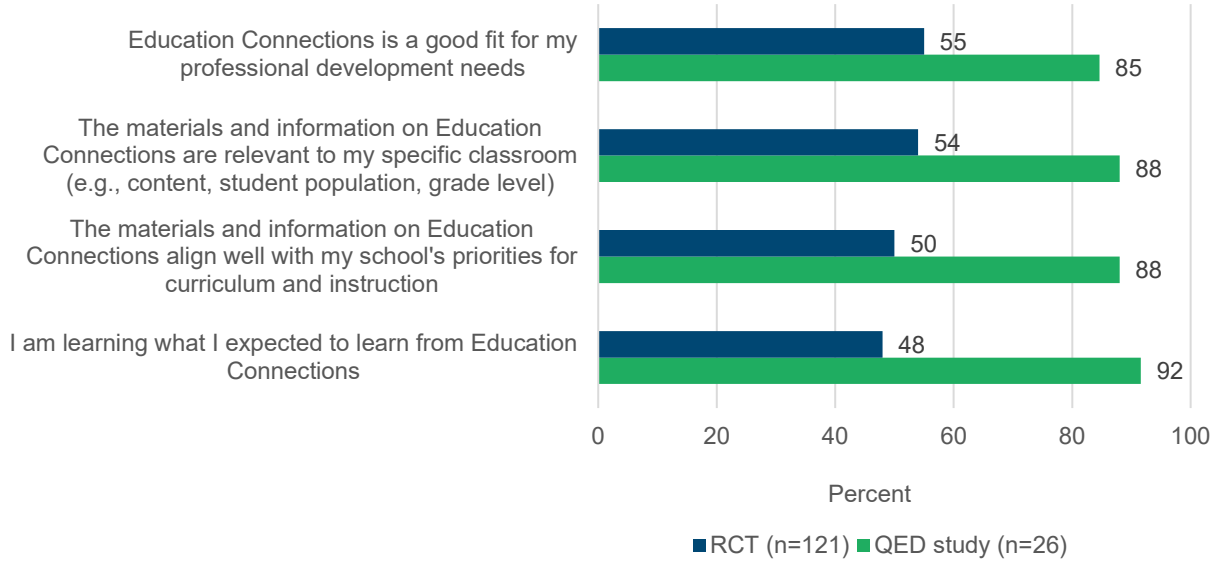
Treatment teachers in QED districts reported greater familiarity with EdConx and saw it as more relevant than those in the RCT districts, possibly due to more robust recruitment practices and more frequent face-to-face meetings.

Coherence with classroom context, school and district priorities, and teachers' professional needs are key characteristics of successful, high-quality professional development (Darling-Hammond, Hyler, & Gardner, 2007; Wei, et al., 2009; Desimone, 2009; Yoon, et al., 2007; and Porter, et al., 2000). Only about half of the treatment teachers in the RCT districts reported that EdConx was relevant to their professional needs, classroom assignment, and school priorities (Exhibit 22). In the QED districts, where teachers reported higher levels of familiarity with EdConx, more than 85 percent of treatment teachers reported that EdConx was relevant to their professional needs and school and district priorities.

In the QED districts, where teachers elected to participate in the treatment group, relatively few treatment teachers reported that they did not know enough about the EdConx site to rate their

satisfaction with each of its core components. This difference may be due to CAL's improved and more robust recruitment strategies in the QED districts. CAL structured its recruitment in the QED districts differently than in the RCT districts, holding in-person information sessions to convey the EdConx structure and resources to prospective participants and school leaders. These sessions may have helped to attract teachers who judged that EdConx was relevant to their professional needs and priorities, and these sessions may have led them to be more familiar with the site. In contrast, CAL's initial recruitment in the RCT districts relied on their district contact's support to encourage teachers to join the program. This reliance on district staff had varying success—in some instances teachers received misinformation or partial information about EdConx. The distance of Districts A and B from CAL's offices made it difficult to hold frequent in-person information sessions. In addition, school and district leaders spoke about EdConx as one of many valuable resources available to teachers but not as the primary resource for teaching ELs. Thus, teachers may not have viewed EdConx as a program the district prioritized. Greater familiarity with EdConx among teachers participating in QED districts could be attributed to CAL's recruitment practices building buy-in among principals and teacher, and combining online learning with face-to-face session to better meet teachers' needs.

Exhibit 22. Teachers Agreeing that EdConx is Relevant



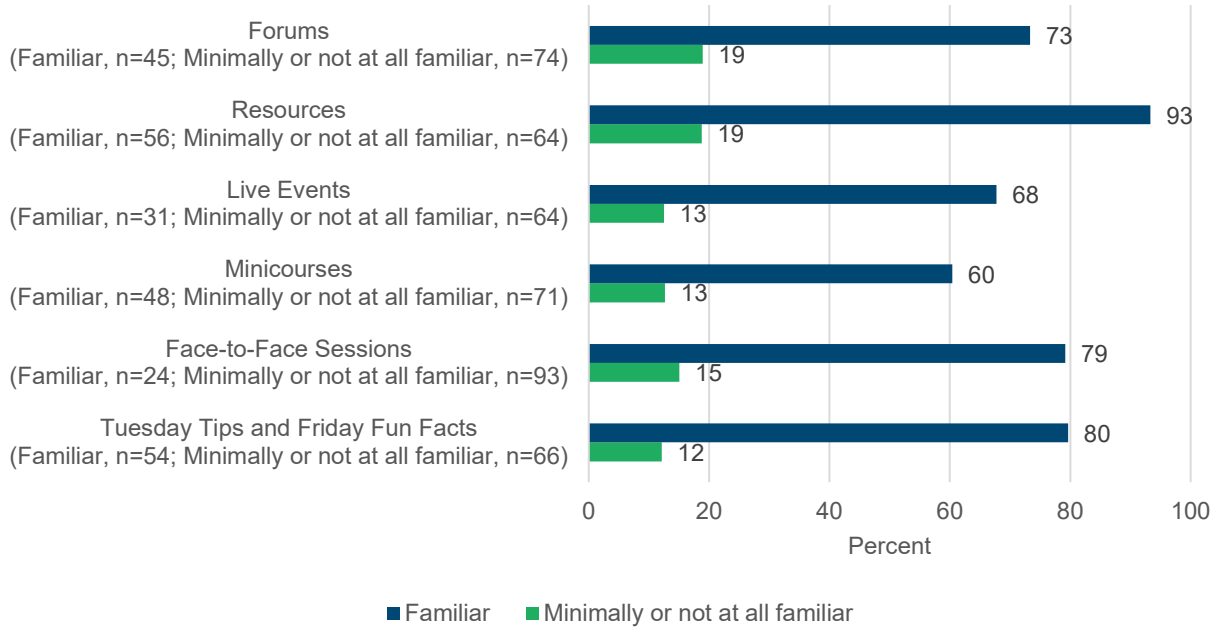
Source: EdConx teacher follow-up survey.

Treatment teachers who reported familiarity with EdConx were also more likely to report satisfaction.

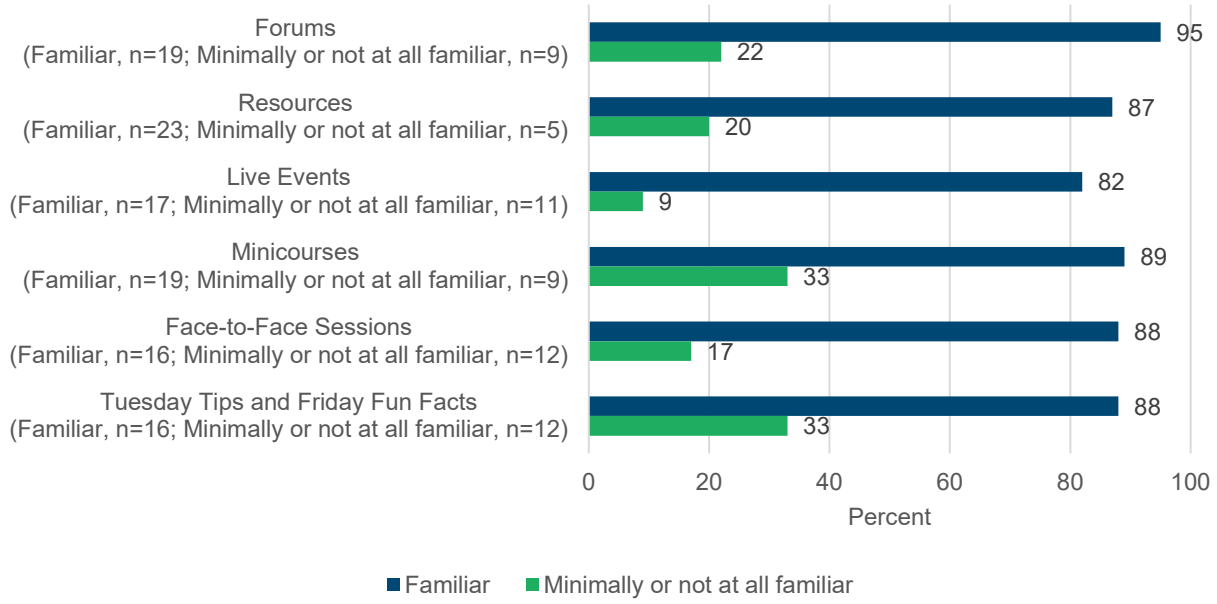
Exhibit 23 shows the proportion of teachers who reported they were satisfied with EdConx by their familiarity with each core component. More than 60 percent of treatment teachers in the RCT districts and more than 80 percent of treatment teachers in the QED districts who reported some familiarity with EdConx also reported satisfaction with the site. In contrast, less than one-third of treatment teachers who were not at all or minimally familiar with EdConx across RCT and QED districts reported satisfaction with EdConx's components, suggesting that teachers who used EdConx and were more familiar with it viewed it as a valuable resource.

Exhibit 23. Treatment Teachers Reporting Satisfaction with EdConx Components, by Familiarity

RCT



QED Study



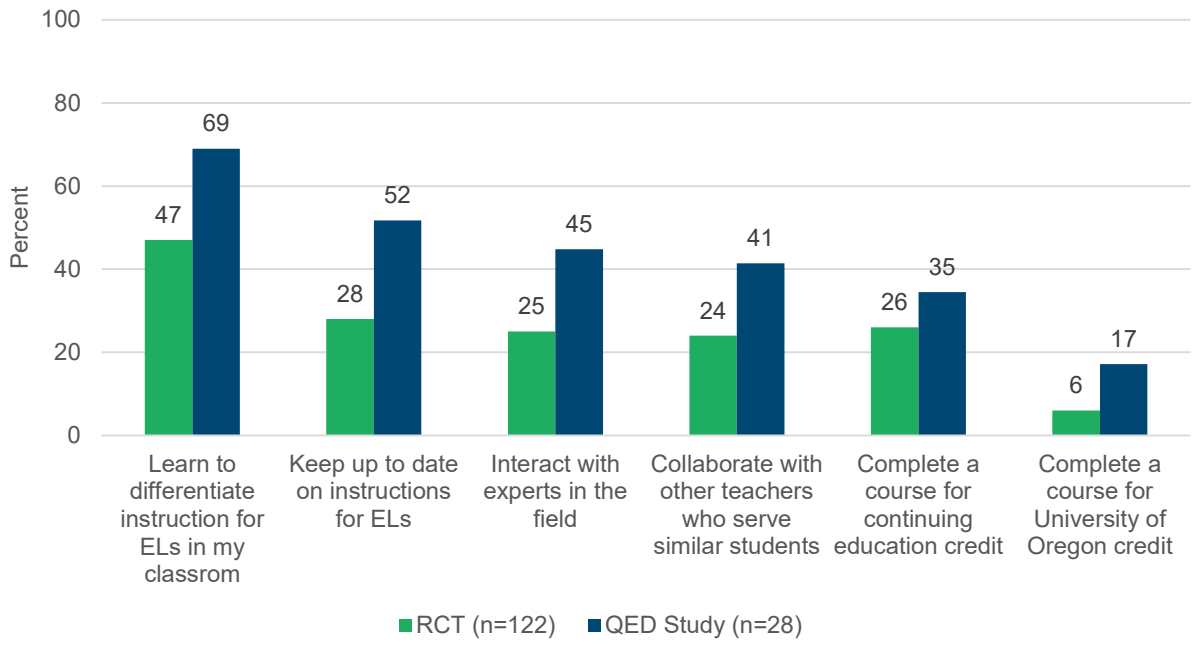
Source: EdConx teacher follow-up survey.

The breadth of materials offered on the EdConx platform and the flexibility offered to teachers when navigating the site may have created an online experience with too many disparate components, making it difficult for teachers to identify specific resources relevant to their instructional needs.

By design, a key feature of EdConx was to meet each teacher's unique interests and needs. During interviews teachers reported a variety of motivations for participation, from interest in accessing subject-specific resources for teaching ELs to enrolling in courses that offered university credit, suggesting there may be several different user profiles based on interests and motivation. Survey responses confirmed that teachers identified different priorities for joining EdConx, with a majority of treatment teachers reporting that a main motivator for their participation was to learn how to effectively differentiate instruction for ELs and keep up to date on instruction (Exhibit 24). In contrast, less than half of treatment teachers reported that they considered the mode of content delivery (i.e., completing courses, interacting with experts, collaborating with teachers who teach similar students) as a high priority for joining the program.

During interviews, treatment teachers stressed that they valued the content and flexibility of being able to access EdConx in the manner appropriate to their needs. Several teachers, however, also shared that EdConx, in trying to meet teachers' varying motivations and interests, provided too many disparate components, which made it difficult to identify the resources that were most appropriate for supporting their instructional needs.

Exhibit 24. Treatment Teachers' Motivation for Joining EdConx



Source: EdConx teacher follow-up survey.

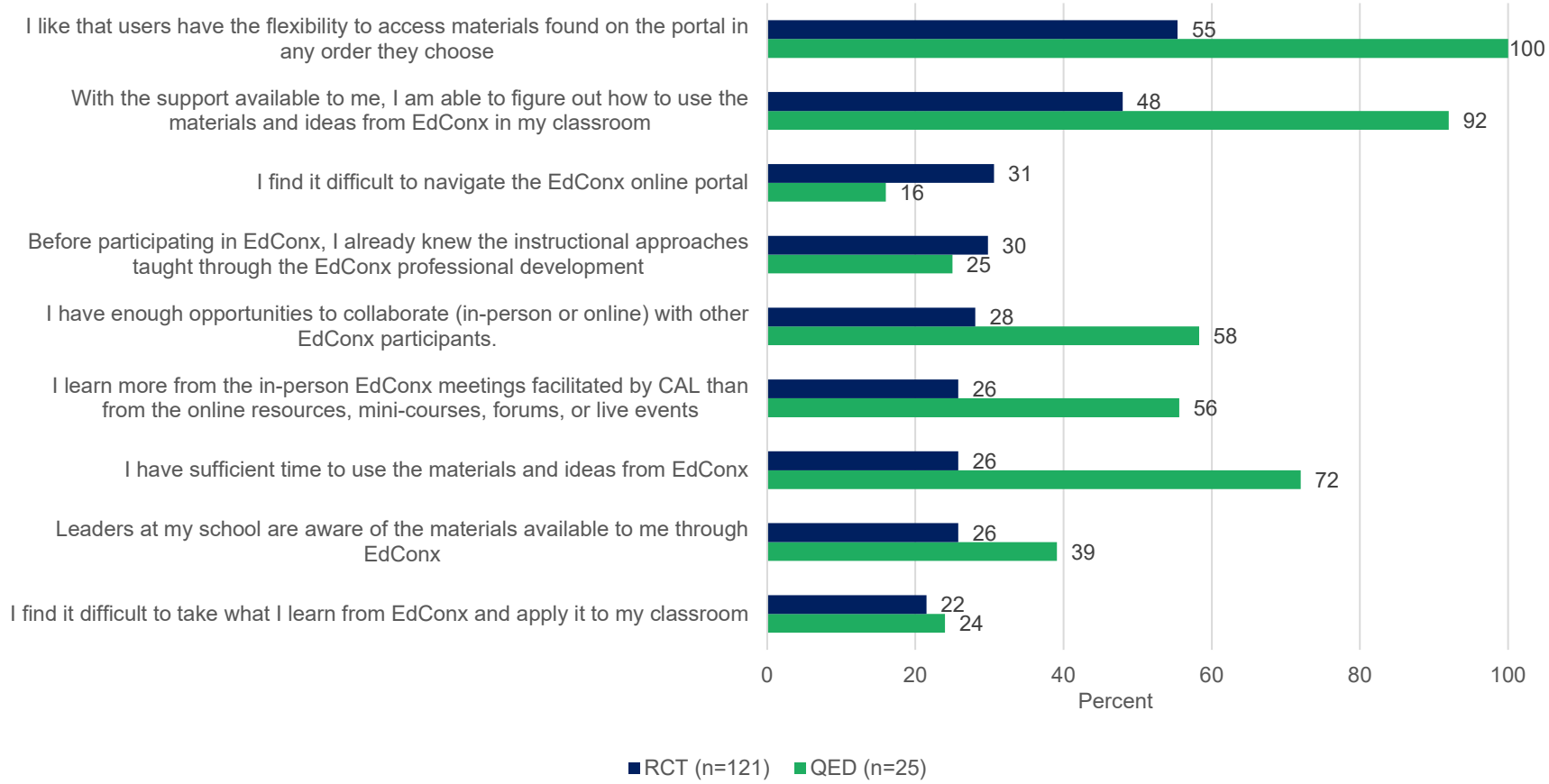
Teachers faced a number of barriers to participating in EdConx, most notably, lack of support from school leaders and trouble with site navigation.

A majority of treatment teachers in both the RCT and QED districts reported that they appreciated the flexibility that EdConx offers, and teachers in the QED districts in particular reported that they were able to figure out how to use the EdConx materials and ideas with the support available to them (Exhibit 25). However, teachers in both the RCT and QED districts also reported key challenges that may explain the low levels of use and familiarity with EdConx.

Treatment teachers in the RCT districts reported barriers such as insufficient time to use EdConx materials (with only 26 percent of teachers reporting that they had enough time), difficulty navigating the EdConx online site (31 percent), and difficulty taking what they learned from EdConx and applying it in their classroom (22 percent). In the QED districts, challenges with EdConx were different—fewer teachers reported difficulty navigating the site (16 percent) and a much larger proportion reported having sufficient time to use EdConx materials (72 percent). However, as in the RCT districts, about one-quarter of teachers in QED districts reported that they had difficultly taking what they learned from EdConx and applying it in their classroom. During interviews, teachers said that finding practical, easily implemented resources on the site was difficult. In addition, as reflected on the survey, only about 25 percent of teacher in the RCT districts and 40 percent of the teachers in the QED districts reported that the leaders at their school were aware of the resources available through EdConx. Because school leaders typically support teachers with classroom instruction on a daily basis, their lack of awareness of EdConx resources most likely contributed to teachers' challenges applying what they had learned. In addition, teachers typically focus their attention on school priorities for improvement, and if principals were largely unaware of the resources EdConx offered, this lack of awareness may have contributed to lower teacher engagement with EdConx.

Although there was some variation on the survey between RCT and QED teachers' assessment of the EdConx site's ease of navigation, interviews revealed several challenges with the site's structure and difficulty navigating back to useful resources on subsequent visits to the site, likely contributors to teachers' low engagement. Teachers complained that the site lacked a central home page with consistent links to the core components (i.e., resources, courses, live events, and forums) and that the site's primary search function yielded few results because limitations in searching all EdConx resources. Teachers expressed some frustration with the site's structure and difficulty navigating back to a useful resource on a subsequent visit to the site, likely contributors to teachers' low engagement.

Exhibit 25. Teachers Agreeing with Statements About Factors that Support or Hinder EdConx Use



Source: EdConx teacher follow-up survey.

Conclusion and Implications

The results of the intent-to-treat analysis of the impact of EdConx in the RCT and QED districts showed no statistically significant differences between treatment and control/comparison teachers in their adoption of EL-relevant instructional practices. Not surprisingly, given similarities in reported EL-relevant teaching practices there were also no statistically significant differences in EL student outcomes between treatment and control/comparison groups. The results of the EdConx implementation study showed high fidelity in the delivery of EdConx components but low fidelity in teacher participation and use of EdConx. Thus, the potential impact of EdConx on treatment teachers' instructional practices and student outcomes was constrained by extremely low levels of teacher participation.

Although the evaluation did not demonstrate that EdConx leads to increases in teachers' use of EL-relevant instructional practices or improvements to students' English proficiency, the implementation findings suggest some lessons for future work and for improving the design and implementation of future online professional development interventions, particularly with an eye to increasing levels of teacher participation. Among these lessons are the following:

- **Target recruitment.** Teachers had varying motivations and priorities for engaging with EdConx. Teachers' interests were often disparate, and EdConx was not able to meet them all simultaneously. Thus, before launching recruitment for professional development, specifying the user profile(s) that program developers theorize will most benefit from the program and targeting those teachers could better yield a group willing and ready to fully engage with the program.
- **Engage school leaders.** School leaders set priorities for teachers. During recruitment, CAL worked directly with district officials and rarely had direct contact with school leaders. Conveying to school leaders the value of professional development for the teachers in their school could help to elevate the program to one school leaders and teachers prioritize as a resource for professional learning and growth. This buy-in from school leaders could help to encourage teachers' use of the program.
- **Conduct user testing early and often.** Interviews with teachers revealed navigation issues with the EdConx site. Engaging in usability testing early as well as iteratively throughout the development of online professional development could help to identify these kinks and address them immediately to ensure a smooth online experience.
- **Be flexible.** After the first year of implementation, CAL recognized that a fully online program was not suitable for many of the participating treatment teachers. It adjusted the program to a blended model that incorporated face-to-face sessions to complement the online site.

References

Batalova, J. & McHugh, M. (2010). *Number and Growth of Students in U.S. Schools in Need of English Instruction*. Washington, DC: Migration Policy Institute. Retrieved from http://www.migrationpolicy.org/research/number-and-growth-students-us-schools-need-english-instruction-2009.

Becker, S. O., & Ichino, A. (2002). Estimation of average treatment effects based on propensity scores. *Stata Journal, 2*, 358-377.

Darling-Hammond, L., Hyler, M.E., and Gardner, M. (2017). *Effective teacher professional development*. Palo Alto, CA: Learning Policy Institute

Desimone, L.S. (2009). *Improving Impact Studies of Teachers' Professional Development: Towards Better Conceptualizations and Measures*. Education Researcher, 38(3), 181-199

Echevarria, J. J., Vogt, M., & Short, D. J. (2012) Making Content Comprehensible for English Learners: The SIOP Model (4th Edition). Boston, MA: Pearson/Allyn & Bacon.

Fregeau, L. & Leier, R. (2015). Language Policy and the Preparedness of Mainstream Teachers for Serving ELs in K-12 Classrooms. *TAPESTRY 7*(1). Retrieved from http://stars.library.ucf.edu/cgi/viewcontent.cgi?article=1050&context=tapestry.

Gottlieb, M. (2013). *Essential Actions: A Handbook for Implementing WIDA's Framework for English Language Development Standards*. Wisconsin: Board of Regents of the University of Wisconsin System. Retrieved from https://www.wida.us/standards/eld.aspx.

Grady, M.W., & O'Dwyer, L.M. (2014. The English Language Learner Program Survey for Principals (REL 2014-027). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands. Retrieved from http://ies.ed.gov/ncee/edlabs.

Hedeker, D., & Gibbons, R. (2006). Longitudinal data analysis. New York: Wiley.

Kaplan, K. C., Lavadenz, M., & Amras, E. G. (2011). Essential Elements of Effective Practices for English Learners. (1). Retrieved from http://www.ctdev.changeagentsproductions.org/wp-content/uploads/2015/01/PolicyBrief7blkCaliTogDec2011-1.pdf.

Lucas, T. & Villegas, A.M. 2010. The Missing Piece in Teacher Education: the Preparation of Linguistically Responsive Teachers. Yearbook of the National Society for the Study of Education (109)2. 297-318. Columbia University: Teachers College. Retrieved from https://www.researchgate.net/publication/292697945_The_missing_piece_in_teacher_education_The_preparation_of_linguistically_responsive_teachers.

Master, B., Loeb, S., Whitney, C., & Wyckoff, J. (2016). Different Skills? Identifying Differentially Effective Teachers of English Language Learners. The Elementary School Journal (117)2. Illinois: University of Chicago. Retrieved from http://files.eric.ed.gov/fulltext/ED529176.pdf.

National Center for Education Statistics. (2017). The Condition of Education 2017: English Language Learners in Public Schools. Washington, DC: U.S. Department of Education. Retrieved from https://nces.ed.gov/fastfacts/display.asp?id=96.

Office of English Language Acquisition (2016). *English Learners' (ELs') Results From the 2015 Nation's Report Card.* Washington, DC: U.S. Department of Education. Retrieved from http://www.ncela.us/files/fast_facts/OELA_FF_NAEP_2015_For_Grades48.pdf.

Pettit, S. K. (2011). Teachers' beliefs about English language learners in the mainstream classroom: A review of the literature. *International Multilingual Research Journal, 5*(2), 123-147. Retrieved from https://www.researchgate.net/publication/232948907_Teachers%27_Beliefs_About_English_Language_Learners_in_the_Mainstream_Classroom_A_Review_of_the_Literature.

Porter, A., Garet, M., Desimone, L., Yoon, K.S., & Birman, B. (2000). *Does professional development change teaching practice? Results from a three-year study.* Washington, DC: U.S. Department of Education.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rosenbaum, P., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41-55.

Rosenbaum, P., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association, 79*, 516-524.

Rosenbaum, P., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician, 39*(1), 33-38.

Santos, M., Darling-Hammond., L, & Cheuk, T. (n.d.). *Teacher development to Support English Language Learners in the Context of Common Core State Standards.* CA: Stanford University. Retrieved from http://edfs200ell.pbworks.com/w/file/fetch/54675861/Teacher%20Development.pdf

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology, 66*(8), 84-90. doi:10.1016/j.jclinepi.2013.01.013

Turkan, S., De Oliveira, L.C., Lee, O., & Phelps, G. (2014). Proposing a Knowledge Base for Teaching Academic Content to English Language Learners: Disciplinary Linguistic Knowledge. *Teacher College Record (116).* Columbia University: Teachers College. Retrieved from https://www.ets.org/research/policy_research_reports/publications/article/2014/jsel.

Wei, R.C., Darling-Hammond, L., Andree, A., Richardson, N., Orphanos, S. (2009) *Professional Learning in the learning profression: A status report on teacher development in the United States and abroad.* Dallas, TX. National Staff Development Council.

What Works Clearinghouse. (2014). *What Works Clearinghouse procuedures and standards handbook* (3 ed.). Princeton, NJ: U.S. Department of Education, Institute of Education Sciences.

Yoon, K.S., Duncan, T., Lee, S.W., Scarloss, B., & Shapley, K.L. (2007). *Reviewing the evidence on how teacher professional development affects student achievement* (Issues & Answers report, REL 2007-No. 033). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southwest.

Appendix A. Survey Items

Presented here is the survey scale used in the teacher outcomes analysis.

Exhibit A-1. Items on Instructional Practice Survey Scale

Items	Response Scale (1–6)
Teachers' Use of All EdConx-Supported Instructional Practices In {EL focal class}, approximately how frequently do you do the following? 1. Develop lesson or student learning objectives based on academic content area standards in English, math, science and social studies 2. Announce content objectives at the beginning of the lesson 3. Post content objectives where students can see them 4. Integrate content objectives into the instructional routine (i.e., students know to expect them) 5. Develop lesson or student learning objectives based on English language proficiency development standards 6. Announce English language proficiency development objectives at the beginning of the lesson 7. Post English language proficiency development objectives where students can see them 8. Integrate English language proficiency development objectives into the instructional routine (i.e., students know to expect them) 9. Pre-teach unfamiliar vocabulary to students 10. Pre-teach subject-specific ways of talking (e.g., justifying a claim with evidence in math or science) 11. Paraphrase directions for students 12. Check for ELs understanding of directions and content How frequently does an average EL student for {EL focal class} do the following? 13. Work in pairs or small groups with students of varying English language ability 14. Participate in whole-class discussions 15. Orally demonstrate understanding of key concepts 16. Demonstrate understanding of key concepts in written English	Almost never Once a quarter 1-2 times per month Weekly 2-4 times per week Nearly daily

Appendix B. Analysis Approach

Given that students were nested within teachers and teachers were nested within schools, the magnitude of the intervention effects on teachers was tested using multilevel analysis, or hierarchical linear modeling (HLM) (Raudenbush & Bryk, 2002). HLM adjusts standard errors to account for the dependence among students within teachers and/or teachers within schools, thus avoiding overestimation of statistical significance of the effect size. We conducted a series of multilevel analyses to examine the effect of the intervention on teacher and student outcomes, with and without adjusting for important covariates (e.g., pretest/resurvey scores and demographic characteristics).

RCT Attrition Analysis

Although randomizing teachers to conditions should result in statistically equivalent groups, a higher overall attrition and differential attrition between treatment and control groups may jeopardize the initial balance and the impact estimate may be biased (What Works Clearinghouse, 2014). The data analysis began with an attrition analysis. The treatment group attrition rate was 50 percent, the control group attrition rate was 49%, and the differential attrition rate was 1 percent. According to What Works Clearinghouse standards (2014), the overall and differential attrition rates are low for this study.

RCT Intent-to-treat HLM Analysis of EdConx Effect on Teacher Outcomes

Our ITT (intent-to-treat) analysis assessed the average effect of the treatment based on the initial treatment assignment regardless of how many of the teachers assigned to treatment actually received the treatment. The ITT impact estimate is the expected effect of the intervention when it was implemented in the real world, with less than perfect implementation and dosage. Two-level HLM was performed to take into account teachers nested in schools. Dependent variables were the teacher survey scale listed in Exhibit B-1. Independent variables for the simple model included a constant, treatment indicator, dummy variables for the districts, and a baseline score on the same outcome survey measure. Independent variables for the full model included all the independent variables from the simple model plus teacher demographic characteristics variables.

Level 1 is teacher level and level 2 is the school level.

$$Y_{is}^{postsurvey} = \beta_0 + \beta_1 Presurvey + \beta_2 Treatment + \beta_3 District + \beta_4 COV_{is} + \gamma_{is} + \mu_{os},$$

where i is the teacher; s is schools; $Y_{is}^{postsurvey}$ is the follow-up survey score; $Presurvey$ is the baseline survey score; Treatment = 1 for intervention teachers and Treatment = 0 for comparison teachers; District is two district dummy variables for District A and District C using District B as the reference group; COV_{is} is teacher background characteristics (years of teaching experience, female, and BA degree), γ_{is} and μ_{os} are individual and school random effects.

RCT Teacher Outcomes Analysis Treatment of Missing Data

For the RCT teacher outcome study, we used multiple imputation to mitigate bias resulting from missing data. We imputed for missing data on covariates and baseline measures using the EM (expectation-maximization) algorithm for multiple imputation using available scores from other

survey scores, teacher background information, and follow-up survey measures. However, we did not impute follow-up survey measures.

An SAS PROC MI procedure with EM statement was used for multiple imputation. Multiple imputation inference involved three distinct phases:

- The missing data were filled in five times to generate five complete data sets.
- The five complete data sets were analyzed using HLM.
- The results from the five complete data sets were combined for the inference using SAS PROC MIANALYZE.

RCT Teacher Outcome Results

The impact analysis models were multilevel, with teachers nested within schools. Each model controlled for teacher background characteristics (years of teaching experience, gender, and BA degree) and baseline survey score. Although many of these variables were not statistically significant, we included them in all models to increase the precision with which we could estimate the treatment effect and to minimize omitted variable bias. Results of the RCT full model teacher outcomes impact analysis are shown in Exhibit B-1.

Exhibit B-1. RCT Impact on Teachers' Use of EdConx-Supported Instructional Practices

	Coefficient	*SE*
Treatment indicator	0.001	0.10
Baseline survey score	0.64***	0.06
Teacher demographics		
Teacher's years of experience	0.01	0.01
Teacher is female	0.45***	0.12
Teacher has a BA	0.08	0.10
District		
District A	-0.05	0.11
District C	-0.02	0.17
Intercept	1.13	0.34
Number of observations		
Teachers	215	
Schools	86	

*p < .05, **p < .01, ***p < .001.

QED Study Teacher Outcome Analysis - Propensity Score Weighting

Propensity score techniques are quasi-experimental approaches developed to approximate findings obtained from randomized control trials (Becker & Ichino, 2002). They have been increasingly used in observational studies with cohort designs to reduce selection bias in estimating

treatment or intervention effects when randomized controlled trials are not feasible or ethical (Rosenbaum & Rubin, 1983, 1984, 1985). The propensity score is the predicted probability of participating in an intervention based on a set of potentially confounding covariates (e.g., years of teaching experience, gender, and education degree) using logistic regression. Propensity scoring attempts to equalize the mean values of potentially confounding covariates in the treatment and comparison groups and assures that differences in teacher outcomes are not the result of differences in mean values on those covariates. Although it aims to generate rigorous and unbiased estimates of the effects of a treatment on the outcome of interest, propensity scoring cannot account for unobserved confounders such as teacher motivation in participating in EdConx.

This study used propensity score methods to test the effect on EdConx on teacher survey outcomes. We estimated the average treatment effect on the treated (ATT). ATT estimates were adjusted for confounding using inverse propensity score estimators, as recommended by Rosenbaum and Rubin (1983). Specifically, the weight for EdConx was 1 and the weight for comparison teachers was equal to (pi/1-pi), where pi is the propensity score for the *i*-th comparison teacher. Weighting was selected over other approaches such as matching because of good performance in this data set (details below) and maximize the sample sizes.

QED Study HLM Analysis of EdConx Effect on Teacher Outcomes

The HLM models used for the QED study teacher outcomes were very similar to those specified above. The only difference was that propensity score weights were added to the QED HLM to eliminate the baseline difference between treatment and comparison teachers.

QED Study Teacher Outcomes Analysis Treatment of Missing Data

No imputation of missing data was used. The HLM analysis used an unimputed data set.

QED Study Teacher Outcome Results

The impact analysis models were multilevel with propensity score weights, with teachers nested within schools. Each model controlled for teacher background characteristics (years of teaching experience, gender, and BA degree) and baseline survey score. Although many of these variables were not statistically significant, we included them in all models to increase the precision with which we could estimate the treatment effect and to minimize omitted variable bias. Results of the QED teacher outcomes impact analysis are shown in Exhibit B-2.

Exhibit B-2. QED Impact on Teachers' Use of EdConx-Supported Instructional Practices

	Coefficient	*SE*
Treatment indicator	0.001	0.10
Baseline survey score	0.64***	0.06
Teacher demographics		
Teacher's years of experience	0.01	0.01
Teacher is female	0.45***	0.12
Teacher has a BA	0.08	0.10
District		
District A	-0.05	0.11
District C	-0.02	0.17
Intercept	1.13	0.34
Number of observations		
Teachers	215	
Schools	86	

*p < .05, **p < .01, ***p < .001.

HLM Analysis of EdConx Effect on Student Outcomes in the RCT Districts

A two-level HLM analysis was performed to compare EL's English proficiency score between treatment and control group taking into account that students were nested in teachers. Dependent variables were the English proficiency scores. Independent variables for the simple model included a constant treatment indicator at the teacher level, dummy variables for the districts, student demographic characteristics, and teacher demographic characteristics.

Level 1 is teacher level and level 2 is the school level.

$Y_{is}^{post} = \beta_0 + \beta_1 Pre + \beta_2 Treatment + \beta_3 District + \beta_4 COV_{is} + \gamma_{is} + \mu_{os}$, where i is the student; s is teachers; $Y_{is}^{postsurvey}$ is post English proficiency score; Pre is the baseline English proficiency score; Treatment = 1 for intervention teachers and Treatment = 0 for comparison teachers; District is two district dummy variables for District A and District C using District B as the reference group; COV_{is} is teacher background characteristics (years of teaching experience, female, and BA degree) and student characteristics (gender, race, grade level, and special education status), γ_{is} and μ_{os} are individual and teacher random effects.

Student Outcome Analysis Treatment of Missing Data

We did not impute any missing data. HLM analyses were conducted using unimputed data sets.

Student Outcome Results

The results of the student outcomes impact analysis are in Exhibit B-3.

Exhibit B-3. Impact on EL's English Proficiency Composite Score

	Coefficient	*SE*
Treatment indicator	0.05	0.06
Baseline English proficiency score	0.50***	0.03
Student demographics		
Black	-0.33	0.29
White	-0.26	0.28
Hispanic	-0.29	0.26
Asian	-0.18	0.27
Student is female	0.06	0.06
Special education	-0.31***	0.08
Grade 6	-0.28	0.17
Grade 7	-0.10	0.14
Grade 8	-0.15	0.13
Grade 9	-0.05	0.12
Grade 10	-0.01	0.11
Grade 11	0.04	0.12
Teacher demographics		
Teacher is female	0.08	0.07
Teacher has a BA	0.04	0.07
Teacher's years of experience	-0.001	0.01
District		
District A	0.57	0.08
District C	0.15	0.08
Intercept	0.04	0.29
Number of observations		
Students	924	
Teachers	91	

*p < .05, **p < .01, ***p < .001.

Appendix C. Implementation Fidelity

Exhibit C-1. Fidelity Matrix for the Evaluation of EdConx in the RCT Districts (n = 225)

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 1. Delivery of the EdConx Model *Do program developers deliver training as designed?* Decision rules for rolling up to the component level for each district: High: Scored high on at least 67% indicators and low on none. Med: The component does not reach the threshold for high fidelity, and fewer than 50% of the indicators scored as low. Low: Greater than or equal to 50% of the indicators scored as low.						**District A 2014–15[a]**	**District B 2014–15[a]**	**District C 2015–16[b]**	**District A 2015–16[b]**	**District B 2015–16[b]**	**District C 2016–17[c]**
1.1	Update and maintain the forum	Program developers post weekly contribution to the forum, including Tuesday Tips and Friday Fun Facts	Website analytics	Quarterly	High = 1 Forum posted per week, at least 36 weeks Med = 1 Forum posted per week, for 20–35 weeks Low = 1 Forum posted per week, for 19 or fewer weeks	High	High	High	High	High	High
1.2	Update and maintain resources	Program developers post 12 resources per month throughout program implementation	Website analytics	Quarterly	High = 12 resources posted per month, on average Med = 6–11 resources posted per month, on average Low = 5 or fewer resources posted per month, on average	High	High	Medium	Medium	Medium	Low

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 1. Delivery of the EdConx Model *Do program developers deliver training as designed?* Decision rules for rolling up to the component level for each district: High: Scored high on at least 67% indicators and low on none. Med: The component does not reach the threshold for high fidelity, and fewer than 50% of the indicators scored as low. Low: Greater than or equal to 50% of the indicators scored as low.						**District A 2014–15[a]**	**District B 2014–15[a]**	**District C 2015–16[b]**	**District A 2015–16[b]**	**District B 2015–16[b]**	**District C 2016–17[c]**
1.3	Host live events	Program developers host 9 live events (e.g., webinars) over the 1.5-year program	Report from CAL or EdConx archives	Once during the 1.5-year program	High = Hosted 9 live events Med = Hosted between 5 and 8 live events Low = Hosted fewer than 5 live events	N/A	N/A	N/A	High	High	High
1.4	Provide online minicourses	Program developers provide 9 courses over the 1.5-year program	Report from CAL	Once during the 1.5-year program	High = Provided 9 minicourses Med = Provided between 5 and 8 mini courses Low = Provided 4 or fewer minicourses	N/A	N/A	N/A	High	High	High

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 1. Delivery of the EdConx Model *Do program developers deliver training as designed?* Decision rules for rolling up to the component level for each district: High: Scored high on at least 67% indicators and low on none. Med: The component does not reach the threshold for high fidelity, and fewer than 50% of the indicators scored as low. Low: Greater than or equal to 50% of the indicators scored as low.						**District A 2014–15[a]**	**District B 2014–15[a]**	**District C 2015–16[b]**	**District A 2015–16[b]**	**District B 2015–16[b]**	**District C 2016–17[c]**
1.5	Online site maintenance	Site is accessible and available from 6 a.m. to 9 p.m. at least 90% of school days	Report from University of Oregon (UO)	Once during each year of implementation	High = Site is available and accessible at least 90% of school days Med = Site is available and accessible 50%-89% of school days Low = Site is available and accessible for 49% or fewer of school days	High	High	High	High	High	High
1.6	In-person professional development session (*measure added in 2015–16 for new sites)*	Program developers deliver in-person professional development session twice over the 1.5-year program	Report from CAL	Once during the 1.5-year program	High = Delivered at least two professional development sessions Low = Delivered fewer than one professional development sessions	N/A	N/A	N/A	N/A	N/A	High

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 1. Delivery of the EdConx Model *Do program developers deliver training as designed?* Decision rules for rolling up to the component level for each district: High: Scored high on at least 67% indicators and low on none. Med: The component does not reach the threshold for high fidelity, and fewer than 50% of the indicators scored as low. Low: Greater than or equal to 50% of the indicators scored as low.						**District A 2014–15[a]**	**District B 2014–15[a]**	**District C 2015–16[b]**	**District A 2015–16[b]**	**District B 2015–16[b]**	**District C 2016–17[c]**
1.7	Regular communica-tion with participants	Program developers communicate with participants through a MailChimp campaign for at least 90% of the live events and minicourses	Report from CAL	Once during the 1.5-year program	High = Sent MailChimp campaign for at least 90% of live events and minicourses Med = Sent MailChimp campaign for 50–89% of live events and minicourses Low = Sent MailChimp campaign for fewer than 49% of live events and minicourses	High	High	High	High	High	High

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 1. Delivery of the EdConx Model *Do program developers deliver training as designed?* Decision rules for rolling up to the component level for each district: High: Scored high on at least 67% indicators and low on none. Med: The component does not reach the threshold for high fidelity, and fewer than 50% of the indicators scored as low. Low: Greater than or equal to 50% of the indicators scored as low.						**District A 2014–15[a]**	**District B 2014–15[a]**	**District C 2015–16[b]**	**District A 2015–16[b]**	**District B 2015–16[b]**	**District C 2016–17[c]**
1.8	Technology Support	UO responds to at least 90% of help tickets submitted by participants within 24 hours.	Report from UO	Once during the 1.5-year program	High = UO responded to at least 90% of help tickets within 24 hours Med = UO responded to 50–89% of help tickets within 24 hours Low = UO responded to fewer than 49% of help tickets within 24 hours	High	High	High	High	High	High

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 2. Participation in and Satisfaction with EdConx *Do participants participate in EdConx at the intended levels? Are participants satisfied with EdConx?* Decision rules for rolling up to the component level for each district: High: Scored high on the one indicator and low on neither. Med: Scored medium on one indicator and low on neither. Low: Scored low on at least one indicator.						**District A 2014–15[a]**	**District B 2014–15[a]**	**District C 2015–16[b]**	**District A 2015–16[b]**	**District B 2015–16[b]**	**District C 2016–17[c]**
2.1	Participation	Enrolled teachers participate in EdConx for 30 minutes per week on average [f]	Website analytics and live event attendance list	Quarterly	High = 75%+ of participants participate for 30 minutes per week on average Med = 50–74% of participants participate for 30 minutes per week on average Low = Less than 50% of participants participate for 30 minutes per week on average	Low	Low	Low	Low	Low	Low

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 2. Participation in and Satisfaction with EdConx *Do participants participate in EdConx at the intended levels? Are participants satisfied with EdConx?* Decision rules for rolling up to the component level for each district: High: Scored high on the one indicator and low on neither. Med: Scored medium on one indicator and low on neither. Low: Scored low on at least one indicator.						**District A 2014–15[a]**	**District B 2014–15[a]**	**District C 2015–16[b]**	**District A 2015–16[b]**	**District B 2015–16[b]**	**District C 2016–17[c]**
2.2	Satisfaction	Survey respondents report being satisfied with EdConx[g]	Teacher post-survey	Post-survey given at the end of participation	High = 75%+ of participants are satisfied or very satisfied on at least 4/6 components Med = 50–74% of participants are satisfied or very satisfied on at least 4/6 components Low = 49% or fewer of participants are satisfied or very satisfied on at least 4/6 components	N/A	N/A	N/A	High	Medium	High

[a] 6 months, from Jan 2015 to June 2015.
[b] 10.5 months, from July 2015 to June 2016.
[c] 4.5 months, from July 2016 to Dec 2016.
[d] Discounting the monthly of July and 2 weeks in December.
[e] Did not have course completion surveys for spring 2015.
[f] Sample was every eligible study treatment teachers, not including the 64 teachers who attrited.
[g] Sample was every treatment teachers who completed the follow-up survey.

Exhibit C-2. Roll-up to the Component Level, RCT Districts

Component	Number of Indicators	Component Level Fidelity			Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity?
		District A	District B	District C			
Delivery of EdConx	8 (note: 2 indicators are assessed in Year 2 only)	Year 1			**High**: At least 2 of 3 districts meet "High" and none meet "Low" **Med:** Does not reach threshold for high implementation and no districts meet "Low" **Low:** One or more districts meets "Low" Implemented with fidelity if sample meets "High"	Year 1	Yes
		High	High	High		High	
		Year 2				Year 2	
		High	High	Medium		High	
Participation in EdConx	2 (note 1 indicator is assessed in Year 2 only)	Year 1			**High**: At least 2 of 3 districts meet "High" and none meet "Low" **Med:** Does not reach threshold for high implementation and no districts meet "Low" **Low:** One or more districts meets "Low" Implemented with fidelity if sample meets "High"	Year 1 Low	No
		Low	Low	Low			
		Year 2				Year 2	
		Low	Low	Low		Low	

Exhibit C-3. Fidelity Matrix for the Evaluation of EdConx in the QED Study Districts (n = 36)

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 1. Delivery of the EdConx Model *Do program developers deliver training as designed?* Decision rules for rolling up to the component level for each district: High: Scored high on at least 67% indicators and low on none. Med: The component does not reach the threshold for high fidelity, and fewer than 50% of the indicators scored as low. Low: Greater than or equal to 50% of the indicators scored as low.						**District D 2015–16[a]**	**District E 2015–16[a]**	**District F 2015–16[a]**	**District D 2016–17[b]**	**District E 2016–17[b]**	**District F 2016–17[b]**
1.1	Update and maintain the forum	Program developers post weekly contribution to the forum, including Tuesday Tips and Friday Fun Facts	Website analytics	Quarterly	High = 1 Forum posted per week, at least 36 weeks Med = 1 Forum posted per week, for 20–35 weeks Low = 1 Forum posted per week, for 19 or fewer weeks	High	High	High	High	High	High
1.2	Update and maintain resources	Program developers post 12 resources per month throughout program implementation[c]	Website analytics	Quarterly	High = 12 resources posted per month, on average Med = 6-11 resources posted per month, on average Low = 5 or fewer resources posted per month, on average	Medium	Medium	Medium	Medium	Medium	Medium

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
1.3	Host live events	Program developers host 9 live events (e.g., webinars) over the 1.5-year program	Report from CAL or EdConx archives	Once during the 1.5-year program	High = Hosted 9 live events Med = Hosted between 5 and 8 live events Low = Hosted fewer than 5 live events	N/A	N/A	N/A	High	High	High
1.4	Provide online minicourses	Program developers provide 9 courses over the 1.5-year program	Report from CAL	Once during the 1.5-year program	High = Provided 9 minicourses Med = Provided between 5 and 8 minicourses Low = Provided 4 or fewer minicourses	N/A	N/A	N/A	High	High	High
1.5	Online site maintenance	Site is accessible and available from 6 a.m. to 9 p.m. at least 90% of school days	Report from UO	Once during each year of implementation	High = Site is available and accessible at least 90% of school days Med = Site is available and accessible 50–89% of school days Low = Site is available and accessible for 49% or fewer of school days	High	High	High	High	High	High

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
1.6	In-person professional development session (*measure added in 2015–16 for new sites)*	Program developers deliver in-person professional development session twice over the 1.5-year program	Report from CAL	Once during the 1.5-year program	High = Delivered at least two professional development sessions Low = Delivered fewer than one professional development sessions	N/A	N/A	N/A	High	High	High
1.7	Regular communication with participants	Program developers communicate with participants through a MailChimp campaign for at least 90% of the live events and minicourses	Report from CAL	Once during the 1.5-year program	High = Sent MailChimp campaign for at least 90% of live events and minicourses Med = Sent MailChimp campaign for 50–89% of live events and minicourses Low = Sent MailChimp campaign for fewer than 49% of live events and minicourses	High	High	High	High	High	High

Key Elements of Component		Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
1.8	Technology support	UO responds to at least 90% of help tickets submitted by participants within 24 hours.	Report from UO	Once during the 1.5-year program	High = UO responded to at least 90% of help tickets within 24 hours Med = UO responded to 50%- 89% of help tickets within 24 hours Low = UO responded to fewer than 49% of help tickets within 24 hours	High	High	High	High	High	High

	Key Elements of Component	Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 2. Participation in and Satisfaction with EdConx *Do participants participate in EdConx at the intended levels? Are participants satisfied with EdConx?* Decision rules for rolling up to the component level for each district: High: Scored high on the one indicator and low on neither. Med: Scored medium on one indicator and low on neither. Low: Scored low on at least one indicator.						**District D 2016–17[a]**	**District E 2016–17[a]**	**District F 2016–17[a]**	**District D 2016–17[b]**	**District E 2016–17[b]**	**District F 2016–17[b]**
2.1	Participation	Enrolled teachers participate in EdConx for 30 minutes per week on average[d]	Website analytics and live event attendance list	Quarterly	High = 75%+ of participants participate for 30 minutes per week on average Med = 50–74% of participants participate for 30 minutes per week on average糎 Low = Less than 50% of participants participate for 30 minutes per week on average	Low	Low	Low	Low	Low	Low

	Key Elements of Component	Operational Definition of Indicator	Data Source(s) for Measuring Indicator	Schedule for Data Collection	Implementation Score	Year 1			Year 2		
Component 2. Participation in and Satisfaction with EdConx *Do participants participate in EdConx at the intended levels? Are participants satisfied with EdConx?* Decision rules for rolling up to the component level for each district: High: Scored high on the one indicator and low on neither. Med: Scored medium on one indicator and low on neither. Low: Scored low on at least one indicator.						**District D 2016–17[a]**	**District E 2016–17[a]**	**District F 2016–17[a]**	**District D 2016–17[b]**	**District E 2016–17[b]**	**District F 2016–17[b]**
2.2	Satisfaction	Survey respondents report being satisfied with EdConx[e]	Teacher post-survey	Post-survey given at the end of participation	High = 75%+ of participants are satisfied or very satisfied on at least 4/6 components Med = 50–74% of participants are satisfied or very satisfied on at least 4/6 components Low = 49% or fewer of participants are satisfied or very satisfied on at least 4/6 components	N/A	N/A	N/A	High	Medium	High

[a] 6 months, from Jan 2016 to June 2016.

[b] 10.5 months, from July 2016 to June 2017.

[c] Discounting the month of July and 2 weeks in December.

[d] Sample is every treatment teachers who completed the baseline survey, not including the 6 teachers who attrited.

[e] Sample is every treatment teachers who completed the follow-up survey.

Exhibit C-4. Roll-up to the Component Level, QED Districts

Component	Number of Indicators	Component Level Fidelity			Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity?
		District D	District E	District F			
Delivery of EdConx	8 (note: 2 indicators are assessed in Year 2 only)	Year 1			**High**: At least 2 of 3 districts meet "High" and none meet "Low" **Med:** Does not reach threshold for high implementation and no districts meet "Low" **Low:** One or more districts meets "Low" Implemented with fidelity if sample meets "High"	Year 1	Yes
		High	High	High		High	
		Year 2				Year 2	
		High	High	High		High	
Participation in EdConx	2 (note 1 indicator is assessed in Year 2 only)	Year 1			**High**: At least 2 of 3 districts meet "High" and none meet "Low" **Med:** Does not reach threshold for high implementation and no districts meet "Low" **Low:** One or more districts meets "Low" Implemented with fidelity if sample meets "High"	Year 1	No
		Low	Low	Low		Low	
		Year 2				Year 2	
		Low	Low	Low		Low	