English Language and Literacy Acquisition-Validation (ELLA-V)
i3 Evaluation (Valid 22)
Final Report

Rebecca Wolf, Ph.D.
Gavin Latham, B.A.
Clayton Armstrong, B.S.
Steven Ross, Ph.D.
Mary Laurenzano, M.L.A.
Cecilia Daniels, M.S.
Jane Eisinger, M.S.
Joseph Reilly, Ed.D.

School of Education
Center for Research and Reform in Education

August 2018

Rafael Lara-Alecio, Ph.D., Principal Investigator
Beverly Irby, Ph.D., Co-Principal Investigator
Fuhui Tong, Ph.D., Co-Principal Investigator
Cindy Guerrero, Ph.D., Lead Coordinator
Laura Cajiao-Wingenbach, Ph.D., Lead Coordinator

Texas A&M University
College of Education and Human Development

About the Center for Research and Reform in Education (CRRE) at Johns Hopkins University

Established in 2004, the Center for Research and Reform in Education (CRRE) works to improve the quality of education for children in grades pre-K to 12 through its research into program effectiveness. Specifically, CRRE's work focuses on evaluating approaches to teaching and school organization, with studies completed on teacher professional development, comprehensive school reform, data-driven reform, cooperative learning, bilingual education, reading and math strategies, and after-school and summer learning programs.

Affiliated with Johns Hopkins University's School of Education, the CRRE research team currently includes numerous Johns Hopkins University professors and research staff with backgrounds including quantitative, qualitative, and evaluative research. The research team has published over 200 research documents, and within the past five years CRRE has conducted over 45 program evaluations nearing $10 million.

Specializing in education program evaluation, CRRE is able to conduct a wide range of educational research services. Past program evaluations have focused on both large-scale (national and state-level interventions) as well as small-scale program evaluations (school-level interventions), and include both published and private evaluations. CRRE frequently co-constructs measures consistent with an organization's logic model or theory of action, as well as goals of the initiative implemented, to effectively evaluate the desired objectives. To address the unique needs of each organization, CRRE offers the following range of evaluation studies:

- **Design and Implementation Quality:** Smaller "formative evaluation" case studies, often using observations, interviews, and surveys that focus on a program's design components and how they are received and used by target consumers (e.g., teachers, students, parents, etc.). Program improvements are a direct result of study outcomes.
- **Efficacy:** Medium-scale students that focus on how the program operates and affects educational outcomes in try-outs in pilot schools of small treatment-group vs. control-group comparisons.
- **Effectiveness:** Larger-scale "summative evaluation" studies that focus on the success of the program in improving outcomes in rigorous non-randomized ("quasi") experimental studies or randomized controlled trials.

Contents

Executive Summary

Overview

The English Language and Literacy Acquisition–Validation (ELLA-V) study was a five-year evaluation of a program that provided professional development, coaching, and curricula that targeted English-as-a-second-language (ESL) instruction for teachers of K–3 English learners (ELs). ELLA-V was implemented in 10 school districts in Texas in the 2013–14 through 2016–17 school years.

The project was federally funded by a grant from the U.S. Department of Education's Investing in Innovation (i3) Fund (PR/Award Number U411B120047). Professors at Texas A&M University were the recipients of the grant and developed the professional development, the coaching program, and the curricula. Researchers at the Center for Research and Reform in Education (CRRE) at Johns Hopkins University were contracted to conduct the independent evaluation.

The evaluation of ELLA-V was a multisite cluster randomized trial designed to meet the What Works Clearinghouse (WWC) standards for rigorous education research (WWC, 2017). The study used a mixed method design to estimate program impacts on student and teacher outcomes and document the fidelity of implementation and perceived quality of the program.

Program Description

ELLA-V provided ongoing virtual professional development and coaching and curricula to teachers of EL students. ELLA-V was implemented in grade 3 in 2013–14, grade 2 in 2014–15, grade 1 in 2015–16, and kindergarten in 2016–17. Teachers received the intervention for a single year, dependent on grade-level implementation.

Each school year, treatment teachers in one grade level received bimonthly virtual professional training for 18 sessions between September and May. Treatment teachers were also supported by coaches and observed, up to three times a year, depending on teacher need. Coaches provided feedback to teachers that was specific to teaching ELs. Finally, teachers were provided with EL-relevant curricula that reflected pedagogical best practices and was aligned with content-area standards and the instructional models used in the teacher professional development.

The ELLA-V professional development and curricula focused on literacy and science content, as well as cognitive-academic language proficiency to progress EL students' English language acquisition. Treatment 1 and Treatment 2 received equivalent professional development and coaching, but curricula materials differed across the two treatments. The curricula also differed across grade levels, according to student development.

Research Questions

1. What was the one-year impact of each ELLA-V intervention (T1 and T2) on K–3 students' performance in science, oral language, phonological awareness, English language development, reading, and writing, compared with the business-as-usual condition?

2. What was the one-year impact of each ELLA-V intervention (T1 and T2) on improving K–3 students' self-esteem, compared with the business-as-usual condition?

3. What was the one-year impact of each ELLA-V intervention (T1 and T2) on increasing K–3 teachers' quality of instruction, compared with the business-as-usual condition?

4. Was each component of ELLA-V implemented with fidelity?

5. How did principals and teachers perceive the effectiveness of each ELLA-V intervention (T1 and T2)?

Sample

Districts and schools in Texas were recruited to participate in the study if they served a majority EL and Spanish-speaking student population. The study sample included 79 schools in 10 districts in Texas across urban, suburban, small town, and rural sites. Schools were randomly assigned to one of the three treatment conditions: Treatment 1, Treatment 2, and Business-as-Usual. At least two ESL teachers per school and grade volunteered to participate in the study each year. Students in grades K–3 were recruited to participate in the study if they were in the classroom of the participating teachers, were an EL, and did not qualify for special education services.

Measures & Instruments

The evaluation estimated the impact of the ELLA-V interventions on student performance in science, oral language, phonological awareness, English language development, content-area reading, reading comprehension, reading fluency, English proficiency in reading and writing, and on students' self-esteem using the following measures:

- Iowa Test of Basic Skills (ITBS) science subtest
- Woodcock-Muñoz Language Survey-Revised (WMLS-R) reading and oral language subtests
- Test of Phonological Awareness 2nd Edition Plus (TOPA 2+)
- Texas English Language Proficiency Assessment (TELPAS) reading, writing, listening, and speaking subtests
- State of Texas Assessments of Academic Readiness (STAAR) reading subtest
- Dynamic Indicators of Basic English Literacy Skills (DIBELS) Oral Reading Fluency (ORF)

- The Hispanic EL Self-Esteem Inventory (SEI)

Teacher outcomes for this impact study were improved quality of instruction per pedagogical transitional bilingual theory. Teacher outcomes were assessed using the following instruments:

- Teacher Observation Record (TOR), which was developed by researchers at Texas A&M University to document the extent to which teachers implemented ESL-relevant instruction.
- Transitional Bilingual Observation Protocol (TBOP), which was also developed by researchers at Texas A&M University to capture certain pedagogical behaviors with ELs during classroom instruction.

Fidelity of implementation was measured using teacher attendance for professional development, coach observation reports, and shipment receipts for curricula materials. Teacher and principal perceptions about the professional development, curriculum materials, and coaching were captured via multiple data sources:

- Treatment and control teacher open-ended surveys
- Focus groups for treatment teachers
- Treatment teacher ePortfolios
- Treatment and control principal surveys
- Treatment principal interviews

Analysis

The impact of ELLA-V on student and teacher outcomes was estimated using hierarchical linear modeling. Propensity score weighting was also used to estimate program impact on teacher outcomes and some student outcomes due to large differences at baseline. To determine whether each of the key ELLA-V components was implemented with fidelity, at least 90% of schools in the fidelity sample had to achieve high levels of fidelity to the component.

Findings

The ELLA-V curricula targeted different content areas across treatments and grades. ELLA-V positively impacted student achievement in a few content areas when the ELLA-V curricula targeted those content areas. ELLA-V resulted in average improvements in science achievement for third-grade students who were exposed to intensive science-infused literacy ELLA-V curriculum (Treatment 1) compared with business-as-usual students. Yet third-grade students who were exposed to a science-infused oral language curriculum (Treatment 2) had similar gains in science as their business-as-usual peers.

ELLA-V also positively impacted oral language development in younger grades where the ELLA-V curricula had the strongest emphasis on oral language (grade K in Treatment 1 and grades

K–1 in Treatment 2). Similarly, kindergarten students who were exposed to ELLA-V curriculum that emphasized phonemic awareness (Treatment 1) outperformed business-as-usual students in phonemic awareness. Conversely, ELLA-V produced negative average effects on EL students' oral language for first-grade students in Treatment 1. Findings showed no difference between treatment and business-as-usual students in oral language or phonemic awareness in other treatment-grade combinations.

There were no observed impacts of ELLA-V on EL students' English language development or reading (measured in multiple ways) for any treatment or grade. Similarly, student writing was mostly unaffected by the intervention, though student writing was not a substantial focus of ELLA-V.

Student survey results also showed no differences in treatment and business-as-usual students' self-esteem in their classes taught in both English and Spanish, with the exception of first-grade students in Treatment 1 exhibiting greater self-esteem in using the Spanish language. However, the majority of teachers reported that one of the main benefits to students of ELLA-V was increased confidence and self-esteem in speaking English. ELLA-V helped teachers create classroom structures that enabled a risk-free, supportive environment for students to experiment using the English language. As a result, students practiced their English to a greater extent, which led to increased confidence in using the English language.

Treatment teachers were observed implementing research-proven ESL strategies to a greater extent than business-as-usual teachers. Strategies that treatment teachers reported using more frequently as a result of ELLA-V were grouping activities, differentiated instruction, visuals for learning new vocabulary, and sentence stems. Treatment teachers also spent a greater proportion of their instructional time targeting EL students' cognitive-academic language proficiency skills in English than did business-as-usual teachers.

Qualitative findings showed that the vast majority of treatment teachers and principals believed that the ELLA-V professional development, coaching, and curricula were effective in supporting them to meet the needs of their EL students. Teachers benefitted from the professional development, and even veteran teachers reported that they had learned something new. Teachers also appreciated the constructive criticism they received from the coaches. Teacher feedback about the curricula was more mixed, with teachers in grades K–1 overwhelmingly liking the curricula, while about half of teachers in grades 2–3 liked the curricula.

ELLA-V was mostly implemented with fidelity across treatments and grades, defined as at least 90% of schools in the fidelity sample fully participating in the intervention. The two exceptions were that only 43% of schools fully participated in the virtual professional development in the third-grade implementation, and 88% of schools fully participated in the virtual professional development in the kindergarten implementation. These percentages were less than the required 90%, but all other program components were implemented with fidelity for these schools.

Conclusion

ELLA-V improved EL teachers' quality of instruction, which led to improvements in oral language and phonological awareness for younger students and in science for third-grade students who were exposed to a literacy-infused science curriculum. Higher quality of instruction for treatment teachers was evident in increased use of ESL strategies (e.g., grouping activities, differentiated instruction, visuals for learning new vocabulary, and sentence stems) and a greater emphasis on cognitive-academic language proficiency compared with business-as-usual teachers.

With one exception, ELLA-V did not impact EL students' English language development, reading, writing, or self-esteem. Texas A&M researchers have found that ELs learn academic language incrementally, starting with oral language, and then pre-reading skills, and finally reading and writing (Tong, Irby, Lara-Alecio, & Koch, 2014). Given the backwards research design where students in each grade were exposed to the intervention for only one school year, EL students in older grades may not have reached their maximum potential under this intervention because they did not benefit from the cumulative effect of this intervention.

Another limitation is that treatment teachers were exposed to the intervention for only one school year, which may not have been adequate time for teachers to fully implement or students to fully benefit from the program. The professional development started in September, leaving ELLA-V teachers essentially 6–7 months to improve their instruction before EL student academic performance was re-assessed. Research has shown that practitioners may experience an "implementation dip," which is a short-term decrease in performance and confidence while new reforms are initiated (Fullan, 2004). Teachers in the treatment groups were asked to implement new instructional techniques, whereas teachers in the business-as-usual group could work to improve what they were already doing.

Different assessments may also help to explain some of the seemingly contradictory findings of program impacts on student outcomes. It is generally more difficult to identify program impacts on state or district tests as opposed to low-stakes assessments (Irby, Tong, Lara-Alecio, Mathes, Acosta, & Guerrero, 2010; Tong, Irby, Lara-Alecio, Mathes, 2008). In this study, there were positive impacts of ELLA-V on EL students' oral language using a low-stakes assessment but no observed effects on EL students' English language development using a high-stakes assessment. Moreover, some instruments were normed for monolingual English speakers, whereas other instruments were designed specifically for ELs. Therefore, tests normed for different student populations may measure different constructs even within the same domain (Bedore & Peña, 2008). Finally, the instrument used to measure EL students' self-esteem in this study may not have been adequately precise, given that study teachers overwhelming attributed ELs' improved confidence in speaking English to the intervention.

This report concludes that the ELLA-V was mostly implemented with fidelity and yielded improved outcomes for EL students in some content areas. More research is needed to identify the cumulative effects across multiple grade levels of the ELLA-V approach (oral language to pre-

reading to reading and writing) on EL students' academic performance and English language proficiency. The report also highlights the ongoing need for a system of supports for teachers of ELs. Professional development and coaching together positively impacted teacher quality, yet student outcomes were impacted only when curricula also targeted the content area.

Introduction

The English Language and Literacy Acquisition–Validation (ELLA-V) study was a five-year evaluation of a program that provided professional development, coaching, and curricula that targeted English-as-a-second-language (ESL) instruction for teachers of K–3 English learners (ELs). The project was federally funded by a grant from the U.S. Department of Education's Investing in Innovation (i3) Fund (PR/Award Number U411B120047). Professors at Texas A&M University were the recipients of the grant and developed the professional development, the coaching program, and the curricula. Researchers at the Center for Research and Reform in Education (CRRE) at Johns Hopkins University were contracted to conduct the independent evaluation. This report describes the methods and findings of the evaluation study.

Background

In the 2016–17 school year, ELs accounted for approximately 19% of the K–12 student population in Texas, a 38% increase from the 2006–07 school year (Texas Education Agency [TEA], 2017a). As of 2017, students classified as ELs were the lowest achieving student subgroup on Texas state assessments. For example, across grades 3–8, only 23% of EL students were on grade level according to the 2017 State of Texas Assessments of Academic Readiness (STAAR) reading test, as compared with 48% of all grade 3–8 students in Texas (TEA, 2017b).

Teachers in Texas and across the nation need more training and support to meet the academic needs of their EL students (Samson & Collins, 2012). In fact, the National Clearinghouse for English Language Acquisition found that only 30% of teachers of EL students had the necessary training to instruct ELs (Ballantyne, Sanderman, & Levy, 2008). Further, the achievement gap between ELs and mainstream students across content areas, such as science, has been a major concern of professional development reform (Irby et al., 2010; Lara-Alecio, Tong, Irby, & Mathes, 2009; Lee & Buxton, 2013; Tong, Irby, Lara-Alecio, & Mathes, 2008; Tong, Lara-Alecio, Irby, Mathes, & Kwok, 2008; Tong, Luo, Irby, Lara-Alecio, & Rivera, 2017). Teachers have not been sufficiently prepared or equipped to teach academic content and English language acquisition simultaneously (Bryan & Atwater, 2002; Correll, 2016; Lee, Hart, Cuevas, & Enders, 2004; Tong et al., 2017). Therefore, additional teacher professional development and supports are needed to help teachers meet the learning needs of their EL students (Buxton & Allexsaht-Snider, 2016; Tong et al., 2017).

Research has shown that teacher professional development can increase teacher effectiveness and positively impact student achievement when professional development is (a) sustained over time, (b) linked with curricula, and (c) focused on both pedagogy and academic content (Darling-Hammond & Richardson, 2009; Yoon, Duncan, Lee, Scarloss, & Shapley, 2007). Additionally, professional development has been shown to positively impact both teacher practice and student achievement for ELs specifically when it targets cognitive-academic language proficiency within an academic content area (Irby et al., 2010; Lara-Alecio et al., 2009; Tong, Irby, Lara-Alecio, & Mathes, 2008; Tong, Lara-Alecio, Irby, Mathes, & Kwok, 2008; Tong et al., 2017).

Curricula can also be leveraged to improve student outcomes, to the extent there is consistency between curricula and instruction (Tarr, Reys, Reys, Chávez, Shih, & Osterlind, 2008).

The ELLA-V project builds on these research-proven strategies and is a validation study of a previous project—English Language and Literacy Acquisition (ELLA)—developed by researchers at Texas A&M (Irby et al., 2010; Lara-Alecio et al., 2009; Tong, Irby, Lara-Alecio, & Mathes, 2008; Tong, Lara-Alecio, Irby, Mathes, & Kwok, 2008; Tong et al., 2017). The ELLA project was a randomized controlled trial implemented in one school district in Texas. Over the course of four school years (2004–08), ELLA provided teachers of ELs in grades K–3 bimonthly in-person professional development, which prepared teachers to implement an enhanced English as a Second Language (ESL) curricula. The professional development was aligned to ESL and content-area standards in both literacy and science, and it used research-based ESL strategies to optimize ELs' academic oral language and literacy development, or cognitive-academic language proficiency (CALP). Teachers also received curricular materials to implement with their students during the expanded ESL blocks. Students in both structured English immersion and transitional bilingual programs received the intervention during a 75-minute ESL block in kindergarten and 90-minute ESL block in grades 1–3, while the typical state-mandated ESL block was 45 minutes. Students were exposed to the intervention over four school years, depending on whether they remained in the same school, and teachers in each grade level were exposed to the intervention for one school year.

Researchers at Texas A&M evaluated effects of the ELLA program and found gains in oral language, phonological processing, and reading in English for EL students in grades K–2, relative to the business-as-usual condition (i.e., non-enhanced 45-minute ESL learning block) (Tong et al., 2008). A later study also identified positive impacts of ELLA on third-grade students' expressive vocabulary, oral reading fluency, and retell fluency (Tong et al., 2017). Yet there was no difference in reading achievement on the state reading test between third-grade ELLA and business-as-usual students (Irby et al., 2010).

With regard to program impacts on teacher outcomes, ELLA teachers spent more time developing EL students' CALP than control teachers (Lara-Alecio et al., 2009; Tong et al., 2017). For example, when teachers spoke in English, ELLA teachers spent more instructional time presenting or reviewing academic content than control teachers, whereas control teachers spent more time on social and academic routines than ELLA teachers (Lara-Alecio et al., 2009). Spending more instructional time targeting EL students' CALP is important, given the finding that ELs' academic English is the "prominent determinant in students' overall comprehension in language arts and content area classrooms" (Cummins, 2000; Tong et al., 2017, p. 294; Valdés, 2004).

This evaluation analyzes the impact of the English Language and Literacy Acquisition Validation (ELLA-V) on student and teacher outcomes. ELLA-V is designed to improve teacher effectiveness and student outcomes for ELs through ongoing virtual professional development

(VPD), virtual mentoring and coaching (VMC), and EL-relevant curricula. Therefore, ELLA-V contains the same programmatic elements as the earlier ELLA program and adds a teacher coaching component.

Teacher coaching and mentoring have been shown to positively impact academic outcomes for ELs, as well as teacher-student interactions and overall educational climate (Casteel & Ballantyne, 2010; Delaney, 2012; Pruitt & Wallace, 2012). Effective teacher mentoring and coaching provide teachers with content and pedagogical expertise, modeling of instructional strategies, and feedback on teacher practice (Pruitt & Wallace, 2012). Teacher coaching in ELLA-V followed these best practices.

ELLA-V also differs from ELLA in that the curricular components were redesigned to fit into a typical 45-minute ESL block. Program components from ELLA were separated into two interventions that were each evaluated as a different treatment in ELLA-V. The research design for ELLA-V was a multisite cluster randomized controlled trial, and schools within each school district were randomly assigned to one of three conditions: Treatment 1, Treatment 2, or Business-as-Usual. ELLA-V was implemented with a backwards design—grade 3 in year 1, grade 2 in year 2, grade 1 in year 3, and grade K in year 4—to examine program impacts after one year of treatment, as opposed to a multi-year effect on students as identified in the evaluation of ELLA. The next section provides more details about the ELLA-V program.

ELLA-V Program Description

ELLA-V provided ongoing virtual professional development and coaching and curricula to teachers of EL students. Each school year, treatment teachers in the target grade level received bimonthly virtual professional training for 18 sessions between September and May. Treatment teachers were also supported by coaches and observed up to three times a year depending on teacher need. Coaches provided feedback to teachers that was specific to teaching ELs. Finally, teachers were provided with EL-relevant curricula that reflected pedagogical best practices and were aligned with content-area standards and the instructional models used in the teacher professional development.

The ELLA-V professional development and curricula focused on literacy and science content, as well as cognitive-academic language proficiency to enhance EL students' English language proficiency. Treatment 1 and Treatment 2 received equivalent professional development and coaching, but curricula materials differed across the two treatments. The curricula also differed across grade levels, according to student development. Each program component is described in more detail in the following sections.

Virtual Professional Development (VPD). Treatment teachers received approximately 90 minutes of virtual training every two weeks from September to May, on average totaling three hours per month. During the professional development, teachers reviewed and practiced upcoming

lessons, reflected on and discussed student learning, and assessed pedagogical progress. The professional development focused on developing teachers' knowledge and use of ESL strategies. Topics focused on supporting ELs while teaching academic content and developing EL students' academic language skills, and thus included vocabulary building and fluency, oral language development, literacy development, reading comprehension, and disciplinary content knowledge. The professional development also featured ESL pedagogical strategies, such as structured opportunities for students to converse, less talking by the teacher, instruction clarifications, student engagement questioning strategies, structured activities, use of multiple forms of communication, and appropriate time spent on various instructional activities.

Virtual Mentoring and Coaching (VMC). Teachers received regular coaching on EL strategies from trained coaches provided by Texas A&M University. Each year, they received up to three rounds of lesson feedback, depending on teacher need, which occurred between January and May. Coaches provided feedback through field notes and observation records that assessed class routines, pacing, preparation, material usage, teacher talk vs. student talk, questioning strategies, and corrective feedback. All coaching was done via virtual tools such as LogMeIn. Coaches were also able to provide real-time direct feedback to teachers during instruction via Iris cameras and earpieces. Some teachers were also supported with additional live coaching sessions.

Curricula. Teachers also received curricular materials, which included lesson plans, lesson scripts, activity supplies, and formative assessments. All curricula materials were appropriate for a daily 45-minute ESL block. The curricula for Treatment 1 differed across grade levels and focused on oral language and phonemic awareness in grade K, oral language and learning to read in grade 1, learning to read in grade 2, and reading to learn (or content-area reading) in grade 3. Reading was therefore a focus in Treatment 1 across grades 1–3. Across grades, the curricula for Treatment 1 included Santillana Intensive English (SEI), Early Interventions in Reading (EIR-I and EIR-II), and Content Reading Integrating Science for English Language and Literacy Acquisition (CRISELLA).

The curricula for Treatment 2 largely focused on students' oral language development, and slightly varied across grade levels according to students' development. The curricula for grade 3 in Treatment 2 also contained a writing component. Across grades, the curricula for Treatment 2 included Story Re-Telling and Higher-Order Thinking for English Language and Literacy Acquisition (STELLA), Academic Oral Language in Science (AOLS), and Academic Oral and Written Language in Science (AOWLS). Table 1 outlines the differences in the ELLA-V curricula across treatments and grades.

Table 1. *Foci of ELLA-V curricula by treatment and grade.*

Grade	Treatment 1	Treatment 2
3	Content-area reading *(CRISELLA)*	Oral language + writing (*STELLA & AOWLS*)
2	Reading *(EIR-II)*	Oral language + writing *(STELLA & AOWLS)*
1	Oral language + reading *(SEI & EIR-I)*	Oral language *(STELLA & AOLS)*
K	Oral language + phonological awareness *(SEI)*	Oral language *(STELLA & AOLS)*

Science content was infused throughout all curricula to varying degrees. Treatment 1 grade 3 students were exposed to an intensive science-infused literacy curriculum, while science was less of a focus in Treatment 1 grades K–2. All grades in Treatment 2 were exposed to oral language curricula that also infused science vocabulary.

For Treatment 1, the focus of the curricula was oral language in grades K–1, reading in grade 2, and reading to access science content in grade 3. For Treatment 2, the major focus of the curricula for all grades was oral language and on having academic conversations. Appendix A provides more detail about the curricula by grade and treatment condition.

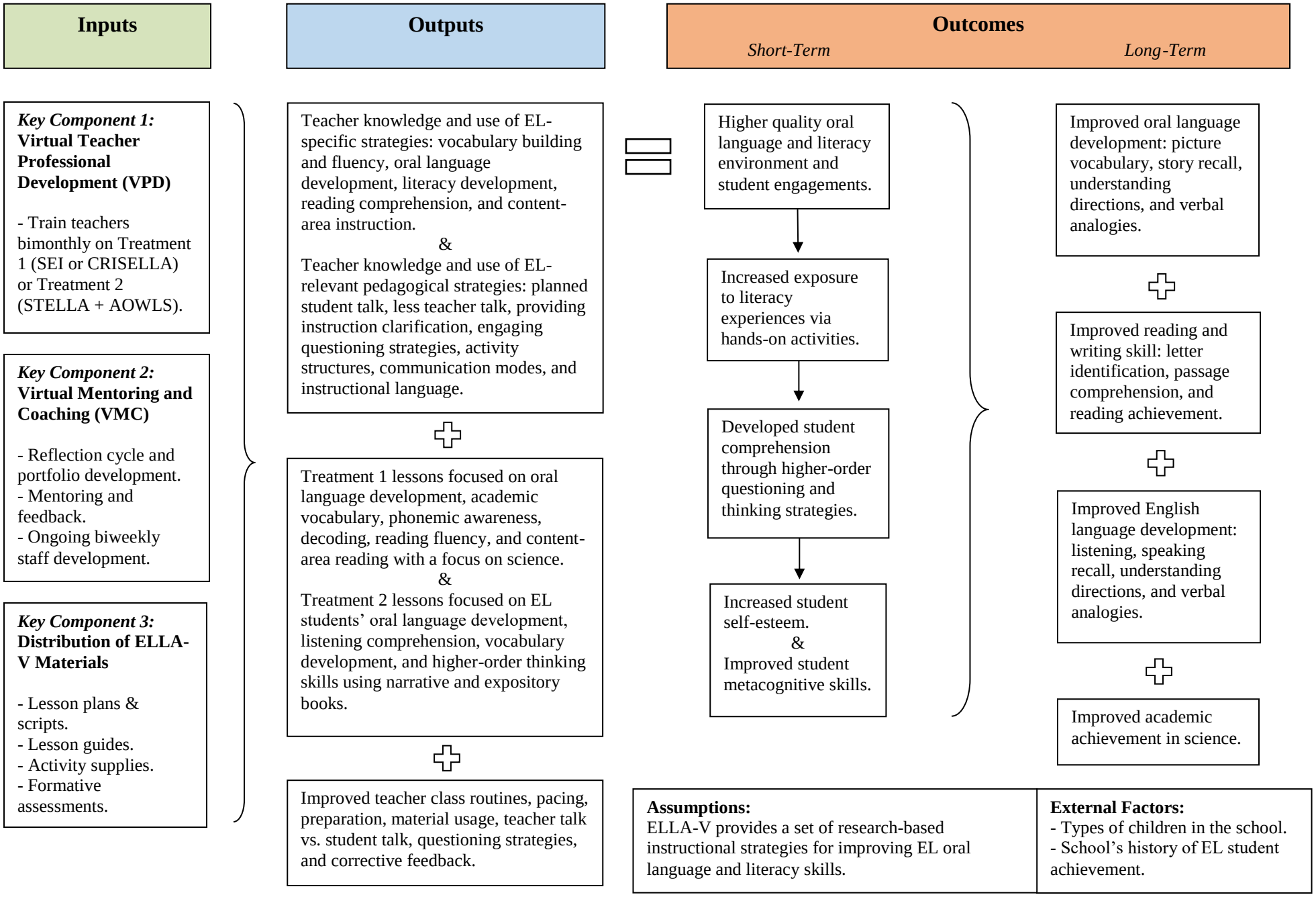
Evaluation Design

The evaluation of ELLA-V was a multisite cluster randomized trial designed to meet the What Works Clearinghouse (WWC) standards for rigorous education research (WWC, 2017). The study used a mixed method design to estimate program impacts on student and teacher outcomes and document the fidelity of implementation and perceived quality of the program.

Schools within each school district were randomly assigned by the independent evaluator to one of three conditions: Treatment 1, Treatment 2, or Business-as-Usual. ELLA-V was implemented with a backwards design—grade 3 in 2013–14, grade 2 in 2014–15, grade 1 in 2015–16, and grade K in 2016–17. Thus, students and teachers each participated in ELLA-V for only one school year, and program impacts were assessed after one year of participation. Moreover, because the intervention components of ELLA-V varied across grade levels, program impacts were estimated separately for each grade level.

The evaluation estimated the impact of ELLA-V on student performance in science, oral language, phonological awareness, English language development, reading, writing, and on students' self-esteem. The evaluation also examined program impact on the quality of teacher instruction. Finally, the evaluation documented whether each component of ELLA-V was implemented with fidelity. The key components and outcomes of implementation are detailed in the logic model, as shown in Figure 1.

Figure 1. Logic model for project ELLA-V.



Research Questions

1. What was the one-year impact of each ELLA-V intervention (T1 and T2) on K–3 students' performance in science, oral language, phonological awareness, English language development, reading, and writing, compared with the business-as-usual condition?

1. What was the one-year impact of each ELLA-V intervention (T1 and T2) on improving K–3 students' self-esteem, compared with the business-as-usual condition?

2. What was the one-year impact of each ELLA-V intervention (T1 and T2) on increasing K–3 teachers' quality of instruction, compared with the business-as-usual condition?

3. Was each component of ELLA-V implemented with fidelity?

4. How did principals and teachers perceive the effectiveness of each ELLA-V intervention (T1 and T2)?

Method

Sample

Districts and schools in Texas were recruited to participate in the study if they served a majority EL and Spanish-speaking student population. To be eligible for the study, a school needed to have an estimated 40 EL students in the third grade in the 2013–14 school year. Schools were first blocked into triads on the basis of district and TELPAS rating (e.g., beginning, intermediate, or advanced), whenever possible, and then randomly assigned to one of three treatment conditions (e.g., Treatment 1, Treatment 2, or Business-as-Usual).[1] Three cohorts of schools were randomized (see Table 3), for a total sample size of 79 schools in 10 districts in Texas across urban, suburban, small town, and rural sites. As shown in Table 2, district and schools in the study sample served a predominantly low-income and EL student population.

[1] All but one of the randomization blocks was comprised of three schools; one block was comprised of four schools. The analyses controlled for district and TELPAS dummy variables, as opposed to block, given school attrition from the study.

Table 2. *District and school sample characteristics.*

District Level	**Overall**			
Urbanicity				
Urban	10%			
Suburban	50%			
Town	10%			
Rural	30%			
EL	33%			
Low-income	82%			
School Level	**Overall**	**Treatment 1**	**Treatment 2**	**Business-as-Usual**
EL	62%	63%	61%	61%
Low-income	91%	92%	91%	90%
TELPAS Rating				
Beginning	24%	23%	23%	26%
Intermediate	68%	69%	69%	67%
Advanced	8%	8%	8%	7%
TELPAS Average Composite Score	1.9	1.9	2.0	2.0

In the spring of 2013, 63 schools were recruited to participate in the study and randomly assigned in summer 2013 to one of the three treatment conditions. In the spring of 2014, and prior to program implementation in grade 2, an additional 10 schools were recruited to participate in the study and randomly assigned in summer 2014 to one of the three treatment conditions. In the spring of 2016, and prior to program implementation in grade K, an additional 6 schools were recruited to participate in the study and randomly assigned in summer 2016. Table 3 outlines the number of schools that were recruited and randomly assigned.

Table 3. *Number of schools randomly assigned by treatment condition.*

	Overall	**Treatment 1**	**Treatment 2**	**Business-as-Usual**
3rd Grade	63	21	21	21
2nd Grade	+10	+3	+3	+4
1st Grade				
Grade K	+6	+2	+2	+2
Total	79	26	26	27

NOTE—All but one school participated in the study, but a few schools did not begin participation until one year following random assignment.

Each year, teachers were recruited to participate in the study prior to the start of the school year. At least two ESL teachers per school and grade volunteered to participate in the study each year. Treatment teachers were offered $3,250 for their participation, and business-as-usual teachers were offered $1,000 each school year. Third-grade teachers participated in the 2013–14 school year; second-grade teachers participated in the 2014–15 school year; first-grade teachers

participated in the 2015–16 school year; and kindergarten teachers participated in the 2016–17 school year.

Students in grades K–3 were recruited to participate in the study if they were in the classroom of the participating teachers, provided their parents or guardians consented to study participation. The majority of students in the study were in transitional bilingual classrooms. Students were recruited only if they were ELs and did not qualify for special education services. Each year of the study, students in the relevant grade level were recruited within the first six weeks of school.

Students and teachers in grade 3 were recruited in early fall 2013 and therefore shortly following school random assignment in the spring of 2013. Students and teachers in grades K–2 were recruited later, at least one year after school random assignment. Given potential bias due to non-random selection of participating teachers from study schools, the analytic teacher and student samples were restricted to those teachers and students with non-missing pretest and posttest scores so that baseline equivalence for each analytic sample could be established (WWC, 2017). Table 4 shows the teacher and student sample sizes across all treatment conditions and by grade.

Table 4. *Teacher and student sample sizes by grade.*

Grade Level	Teacher Sample Size	Student Sample Size
Third grade	112	2,000
Second grade	132	2,000
First grade	118	1,786
Kindergarten	126	1,857

NOTE—These sample sizes reflect the numbers of teachers and students who were included in any impact analysis.

Measures and Instruments

Student outcomes. The evaluation estimated the impact of the ELLA-V interventions on student performance in science, oral language, phonological awareness, English language development, reading achievement, English language development in reading, reading fluency, writing, and self-esteem using the following measures:

- ***Science.*** Iowa Test of Basic Skills (ITBS) (Dunbar & Welch, 2015) science subtest measures students' knowledge of concepts relating to life science, earth and space science, and physical science. This test was individually administered to students in grade 3 by trained testers[2] both prior to program implementation and after one year of treatment.

[2] Testers hired by the CRRE and trained by project personnel individually administered the following student assessments for the evaluation: ITBS, WMLS-R, TOPA 2+, DIBELS, and Hispanic EL Self-Esteem Inventory. All other student assessments (i.e., TELPAS, STAAR) were routinely administered to students by the school districts for purposes other than the study.

- ***Oral language.*** Woodcock-Muñoz Language Survey-Revised (WMLS-R) (Woodcock, Muñoz-Sandoval, Ruef, & Alvarado, 2005) oral language subtest measures students' listening and speaking skills, including language development and verbal reasoning. This test was individually administered to students in grades K–3 by trained testers both prior to program implementation and after one year of treatment.
- ***Phonological awareness.*** Test of Phonological Awareness 2nd Edition Plus (TOPA 2+) (Torgesen & Bryant, 2004) measures students' ability to isolate individual phonemes in spoken words and understand the relationships between letters and phonemes. This test was individually administered to students in grades K–1 by trained testers both prior to program implementation and after one year of treatment.
- ***English language development.*** Texas English Language Proficiency Assessment (TELPAS) (TEA, 2018) listening and speaking subtests measure EL students' ability to understand and use the spoken English language. Each year, teachers administer TELPAS to all ELs in Texas in grades K–12. TELPAS uses a 4-point scale.
- ***Reading achievement.*** State of Texas Assessments of Academic Readiness (STAAR) (TEA, 2013b) reading subtest measures grade-level reading expectations, including students' critical thinking, inferencing, making connections, understanding, and application in different genres of reading. STAAR is administered to all students in Texas each year beginning in grade 3.
- ***English language development in reading.*** Woodcock-Muñoz Language Survey-Revised (WMLS-R) (Woodcock, Muñoz-Sandoval, Ruef, & Alvarado, 2005) reading subtest provides a measure of reading skills, including letter and word identification skills and reading comprehension. WMLS-R was not designed to assess second language acquisition because the norming was based on monolingual English speakers. Unlike STAAR reading, passages in WMLS-R reading comprehension do not necessarily cover content area subjects. This test was individually administered to students in grades K–3 by trained testers both prior to program implementation and after one year of treatment.
- ***English language development in reading.***[3] TELPAS (TEA, 2018) reading subtest measures ELs' ability to read in content area subjects including mathematics, science, and social studies. This reading test was designed to detect progress in second language acquisition and uses a 4-point scale. Each year, teachers administer TELPAS to all ELs in Texas beginning in grade K.
- ***Reading fluency.*** Dynamic Indicators of Basic English Literacy Skills (DIBELS) Oral Reading Fluency (ORF) (Good & Kaminski, 2002) measures students' literacy skill in accuracy and fluency with connected text. This test was individually administered to students in grades 1–2 by trained testers both prior to program implementation and after one year of treatment.

[3] Note that adjustments for multiple comparisons were not applied because there was only one outcome measure per domain for confirmatory contrasts.

- ***Writing.*** TELPAS (TEA, 2018) writing subtest measures EL students' ability to produce written text with content at a grade-appropriate level. Each year, teachers administer TELPAS to all ELs in Texas in grades K–12. TELPAS uses a 4-point scale.
- ***Self-esteem in English and Spanish.*** The Hispanic EL Self-Esteem Inventory (SEI) was developed by researchers at Texas A&M University and was adapted from an earlier project (Irby, Tong, Nichter, Lara-Alecio, Hassey, & Guerrero, 2011). This inventory gauged students' self-esteem in using the English and Spanish language (separately), and assessed perceived efficacy to learn new words, read, listen to stories, comprehend language, converse, write, and answer questions. The inventory contained 24 items, 12 gauging self-esteem in using English and the other 12 gauging self-esteem in using Spanish. The survey used a 3-point scale (all the time, sometimes, never), and scores were created by averaging student responses across the 12 items for each language.[4] The inventory was orally administered to individual students in grades K–2 by trained testers and administered in writing to students in grade 3, both prior to program implementation and after one year of treatment.

For nearly all student outcomes, the same instrument was used both for the pretest and posttest. There were only two exceptions. The pretest for the TELPAS outcomes for kindergarten students was the Test de Vocabulario en Imagenes Peabody (TVIP) because no prior TELPAS scores were available for this grade. The TVIP was individually administered to kindergarten students by trained testers prior to program implementation. The second exception was that third-grade students' WMLS-R reading subtest pretest score was used as the pretest for the STAAR reading outcome because STAAR was administered to students starting in the spring of grade 3, as required by the state for all students.

ELLA-V project personnel at Texas A&M University were responsible for data collection, processing, and scoring. Project personnel also collected district data. Data were then transferred to the CRRE evaluation team, and the evaluation team checked, merged, and analyzed the data.

Teacher outcomes. Teacher outcomes for this impact study were improved quality of instruction per pedagogical transitional bilingual theory. Teacher outcomes were assessed using the following instruments:

- ***Teacher Observation Record (TOR).*** The Teacher Observation Record (TOR) was developed by researchers at Texas A&M University to document the extent to which teachers implemented ESL strategies (Tong, Irby, Lara-Alecio, Yoon, & Mathes, 2010). The TOR asked raters to rate teachers on approximately 10 items that gauged teachers' preparation for and delivery of ESL instruction. Topics included: appropriate materials and physical environment; lesson pacing; student engagement; teacher talking vs. student talking; use of leveled questioning; and cognitive feedback. The

[4] Internal consistency was achieved with a Cronbach alpha of 0.89 for the self-esteem in English items and with a Cronbach alpha of 0.90 for the self-esteem in Spanish items.

TOR used a 4-point scale, and scores were created by the CRRE using item response theory.[5]

- ***Transitional Bilingual Observation Protocol (TBOP).*** The Transitional Bilingual Observation Protocol (TBOP) was previously developed and validated from the four-dimensional bilingual pedagogical classroom theory (Lara-Alecio & Parker, 1994). TBOP captured certain pedagogical behaviors (e.g., language of instruction, language content, activity structure, communication mode, ESL strategies, etc.) during classroom instruction (Lara-Alecio et al., 2009; Tong et al., 2017). The TBOP asked raters to record the frequency of such behaviors; therefore, the TBOP score denoted the proportion of instructional time the teacher demonstrated the particular behavior.[6] TBOP scores were used to document both adherence to the intervention model as well as changes in teacher practices over time. The domain of interest for this study was the proportion of time the teacher spent presenting new academic content in English.

Both treatment and business-as-usual teachers were observed by trained observers three times annually and rated on both the TOR and TBOP instruments. Project personnel were extensively trained on the instruments by Texas A&M researchers and then observed and scored teachers virtually using videos of classroom practice. The first round of observations occurred in October/November, approximately 1–2 months after program implementation began. The second round of observations occurred in January, and the final round occurred in April/May. The scores from the initial observation were used as the pretest, and the scores from the final observation were used as the posttest. Teachers in all grades were observed using the TBOP instrument, and only teachers in grades K–1 were observed using the TOR instrument.

Fidelity of implementation. Fidelity of implementation was measured using teacher attendance for professional development, coach observation reports, and shipment receipts for curricula materials. Teacher and principal perceptions about the professional development, curriculum materials, and coaching were captured via multiple data sources:

- ***Treatment and control open-ended teacher surveys.*** Each school year, treatment and control teachers in the targeted grade level completed surveys administered by Texas A&M through an online platform. Treatment teachers were surveyed in both the fall and spring, while control teachers were surveyed only in the fall. Both treatment and control teacher surveys asked teachers to describe their standard ESL instructional block and use of curricula and pedagogical strategies. Additionally, treatment teachers were asked to describe the impact of the ELLA-V intervention on their instruction and professional growth and on students' academic language and self-esteem. Treatment

[5] Internal consistency was achieved for the TOR with a Cronbach alpha of 0.60 using pretest data only.

[6] Frequency data were provided to the CRRE by Texas A&M, and the CRRE calculated teachers' TBOP scores. Additionally, prior studies have found inter-rater agreement using the TBOP ranging from 0.65 to 0.98 in Kappa values (Bruce, Lara-Alecio, Parker, Hasbrouck, Weaver, & Irby, 1997; Breunig, 1998; Irby, Tong, Lara-Alecio, Meyer, & Rodriguez, 2007; Irby et al., 2010). However, given the multi-dimension-multi-rater nature of the instrument, a more rigorous process was developed to establish inter-rater reliability (IRR) using Gwet's (2012) AC_1 coefficient; the IRR using this approach ranged from .724 to .945 (Tong et al., 2017).

teachers were also asked to report specific ELLA-V pedagogical strategies they had implemented in their classrooms, as well as their reasoning for using (or not using) various strategies.

- ***Focus groups.*** Texas A&M researchers conducted focus groups for treatment teachers in the targeted grade level once per school year, either in person or virtually. The focus group protocols asked teachers to provide their perceptions of ELLA-V on student engagement and academic development, as well as the quality of program curricula, professional development, and coaching.
- ***Treatment teacher ePortfolio.*** Treatment teachers in the targeted grade level were asked to provide ePortfolio artifacts twice per year, and artifacts were proof of student progress using ELLA-V strategies. Teachers were also asked to provide artifacts that demonstrated how they implemented an ELLA-V lesson and documented the underlying educational philosophy and strategy behind the lesson.
- ***Treatment and control principal survey.*** Treatment and control principals in the targeted grade level were surveyed once per school year. The survey was administered by Texas A&M through the Survey Monkey online platform. The survey asked principals to provide details about their EL instructional models and curricula; the number and type of staff dedicated to ELs; educational challenges facing ELs; and general context of school leadership and community. Treatment principals were also surveyed about the perceived effectiveness of ELLA-V components, specifically curricula, professional development, and communication practices of Texas A&M.
- ***Treatment principal interview.*** Principal interviews were conducted once per school year by Texas A&M over the phone. Treatment principals were asked about the structure of their ESL and bilingual programs, their knowledge of the ELLA-V intervention, and their perception of ELLA-V efficacy, in regard to improving teacher quality and EL students' academic language development.

Analytic Approach

Impact study. The impact of ELLA-V on student and teacher outcomes was estimated using hierarchical linear modeling. Propensity score weighting was also used to estimate program impacts on teacher outcomes as well as some student outcomes due to large differences at baseline.

Hierarchical linear modeling. The impacts of the two ELLA-V interventions (T1 and T2) on student and teacher outcomes were estimated separately to understand the impact of each relative to the business-as-usual condition. Because the treatments and samples varied across grade levels, the effects of ELLA-V were also estimated separately by grade. Program effects were estimated using a hierarchical linear model with students or teachers nested within schools (Raudenbush & Bryk, 2002). The model to estimate program effects on student outcomes for a particular treatment and grade was as follows:

$$Y_{ij} = \gamma_{00} + \gamma_{01} treatment_j + \gamma_{10} grand\_pretest_{ij} + \gamma_{02} grand\_school\_EL_j + \gamma_{03} grand\_school\_beginning_j + \gamma_{04} grand\_school\_advanced_j + \gamma_{0k} \sum grand\_district\_dummy_j + u_{0j} + r_{ij}$$

Where:

Y_{ij}: Test score for student i in school j

γ_{00}: Grand mean for students in business-as-usual condition

γ_{01}: Treatment effect (model run separately for T1 and T2)

$Treatment_j$: Treatment indicator for school j

γ_{10}: Regression coefficient for the pretest

$grand\_pretest_{ij}$: Pretest score for student i in school j (grand-mean centered)

γ_{02}: Regression coefficient for the school-level proportion EL

$grand\_school\_EL$: Proportion EL in school j (grand-mean centered)

γ_{03}: Regression coefficient for school-level TELPAS rating of beginning

$grand\_school\_beginning_j$: Dummy variable indicating that school j received TELPAS rating of beginning (grand-mean centered)

γ_{04}: Regression coefficient for school-level TELPAS rating of advanced

$grand\_school\_advanced_j$: Dummy variable indicating that school j received TELPAS rating of advanced (grand-mean centered)

γ_{0k}: Vector of regression coefficients for the k district dummy variables

$\sum grand\_district\_dummy_j$: Vector of district dummy variables for school j (grand-mean centered)

u_{0j}: Random school effect for school j

r_{ij}: Residual for student i in school j

The independent variables, except for the treatment indicator, were grand-mean centered to facilitate interpretation of the intercept (Enders & Tofighi, 2007). The model above was also adapted to estimate program impacts on teachers' use of research-based ESL strategies and focus on CALP, as measured by TBOP and TOR, with teachers nested within schools.

Similar hierarchical linear models—without the pretest and school covariates—were used to estimate baseline equivalence for all analytic samples. Baseline equivalence was satisfied ($\leq$ 0.25 standard deviations) for all student and teacher outcomes, after applying propensity score weighting in some cases (WWC, 2017).

Propensity score weighting. Baseline equivalence was not satisfied for the teacher analytic samples ($>$ 0.25 standard deviations) because the pretests were administered to teachers after treatment had already begun. Baseline equivalence was also violated for a few student outcomes due to unacceptably large differences in EL students' baseline achievement. To account for these baseline differences, propensity score weighting was incorporated into the hierarchical linear

model outlined above—both in models estimating intervention impacts and in models estimating baseline differences between treatment and control groups. Propensity score weighting was designed to make the "weighted intervention and comparison groups more similar" (WWC, 2017, p. 31).

We used an R package, Twang, to obtain the propensity score weights across the three treatment conditions (T1, T2, and Business-as-Usual) and calculate the average treatment effect (ATE) for each treatment and by grade (Ridgeway, McCaffrey, Morral, Burgette, & Griffin, 2014). The propensity score models included a subset of the pretests, and were estimated separately for each grade level.[7] To achieve baseline equivalence, we created propensity score weights at both the individual and school levels and incorporated both weights into the hierarchical linear model. We created propensity score weights at the school level by aggregating individual ratings or scores to the school level and re-running the Twang package at the school level. To incorporate propensity score weights at both the individual and school levels in the hierarchical linear model, we used Stata with the [pweight=student/teacher weight] option in the level-1 model and the pweight (school weight) in the level-2 model. We also used Stata's svy command to calculate the means and standard deviations of the pretest and posttest observation scores; for these descriptive statistics, only the weights from the level-1 model were applied.

Implementation study. To determine whether ELLA-V was implemented with fidelity, we analyzed the proportion of teachers and schools who participated at high levels of fidelity in each of the key program components—virtual teacher professional development (VPD), virtual mentoring and coaching (VMC), and distribution of ELLA-V materials.

The fidelity of implementation was analyzed for each program component for each school year from the 2013-14 through 2016-17 school years. Fidelity of VPD, VMC, and curricular materials were measured at the school level. VPD was considered to have been implemented with fidelity in a school if all treatment teachers in the school participated in all but two professional development sessions. VMC was considered to have implemented with fidelity in a school if all treatment teachers in the school participated in at least one coaching session. The distribution of curricular materials was considered to be implemented with fidelity if the school received the curriculum materials. The component level threshold for fidelity of implementation at the sample level was 90%. That is, 90% of schools had to have achieved high fidelity for the program component to be implemented with fidelity at the sample level.

Teachers were excluded from the fidelity sample if (a) they did not attend any of the VPD training sessions; (b) they (or their schools) withdrew from the study, or (c) they left their schools. If all treatment teachers in a specific grade level at a single school site were removed from the fidelity analyses, then the school site was excluded from the fidelity sample for the particular grade level.

[7] For teachers, propensity scores were also estimated separately for each outcome measure to achieve baseline equivalence. For students, propensity scores were estimated only once per grade.

Qualitative data sources—treatment and control teacher surveys, treatment teacher focus groups, treatment teacher ePortfolios, treatment and control principal surveys, and treatment principal interviews—were analyzed using multi-level triangulation to ensure inter-rater reliability and code consistency. First, each data source was coded by treatment and grade according to themes using Miles, Huberman, and Saldaña's (2014) qualitative analysis methods. One reviewer initially created a code, and these first-cycle codes were then verified by a second coder. Second, coded data were reviewed by analysts who developed second-cycle pattern codes by treatment by grade. Finally, coders and analysts discussed each pattern code by data source by treatment by grade for consistency, after which they developed themes for each treatment condition and grade level, as well as across treatment conditions and grade levels.

Findings

Impact on Student Outcomes

ELLA-V provided ESL-relevant professional development, coaching, and curricula to increase teacher capacity to meet the academic needs of EL students and ultimately improve ELs' academic performance and English language proficiency. ELLA-V materials featured state-mandated literacy and science content, while incorporating best practices for ELs to acquire English as a second language. The ELLA-V curricula targeted different content areas across treatments and grades. ELLA-V positively impacted student achievement in a few content areas when the ELLA-V curricula targeted those content areas.

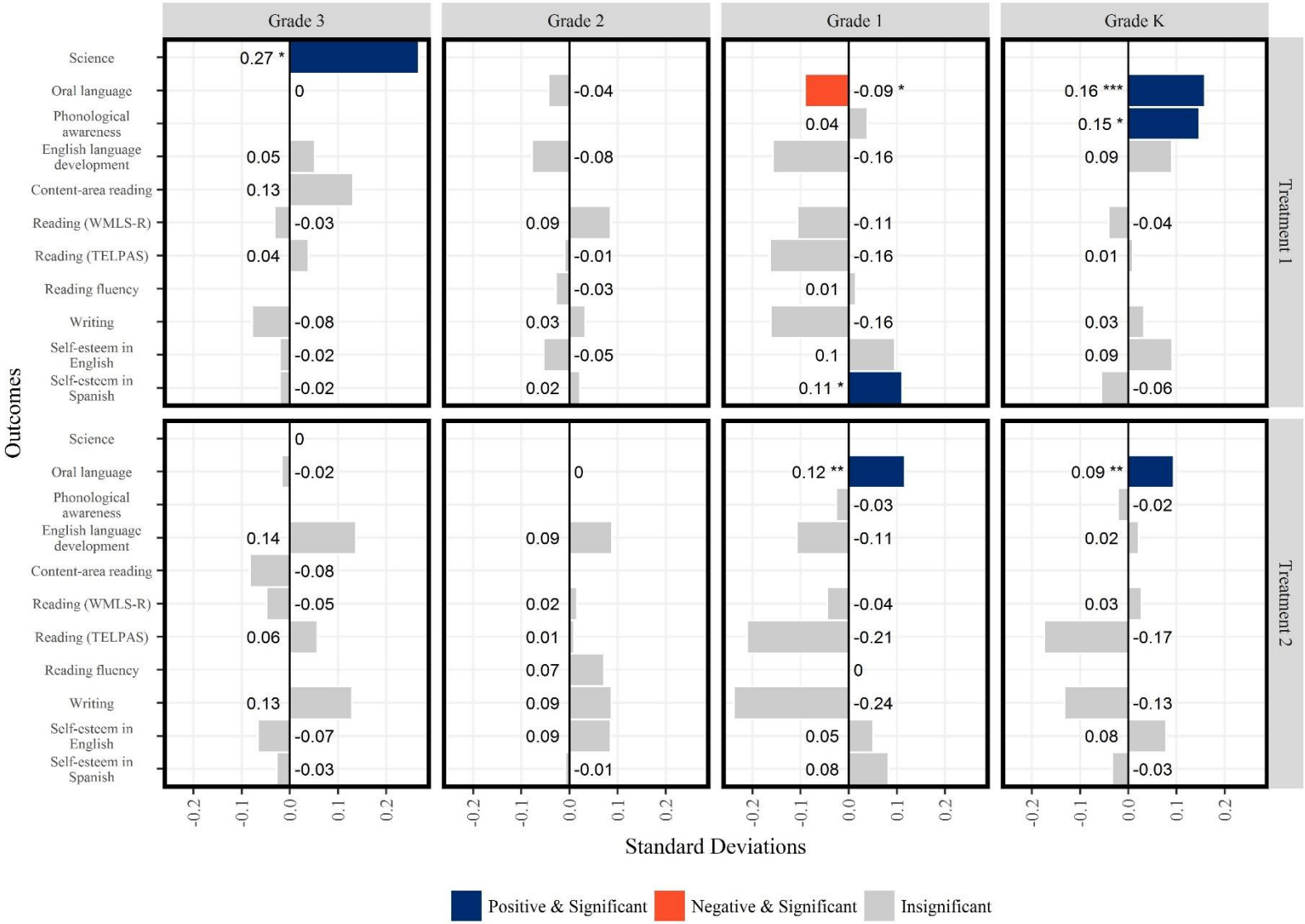
ELLA-V resulted in average improvements in science achievement for third-grade students who were exposed to intensive science-infused literacy ELLA-V curriculum (Treatment 1) compared with business-as-usual students. Yet third-grade students who were exposed to a science-infused oral language curriculum (Treatment 2) had similar gains in science as their business-as-usual peers.

ELLA-V also positively impacted oral language development in younger grades where the ELLA-V curricula had the strongest emphasis on oral language (grade K in Treatment 1 and grades K–1 in Treatment 2). Similarly, kindergarten students who were exposed to ELLA-V curriculum that emphasized phonemic awareness (Treatment 1) outperformed business-as-usual students in phonemic awareness. Conversely, ELLA-V produced negative average effects on EL students' oral language for first-grade students in Treatment 1. Findings showed no difference between treatment and business-as-usual students in oral language or phonemic awareness in other treatment-grade combinations.

There were no observed impacts of ELLA-V on EL students' English language development or reading (measured in multiple ways) for any treatment or grade. Similarly, student writing was mostly unaffected by the intervention, though student writing was not a substantial focus of ELLA-V.

Finally, student survey results showed no differences in treatment and business-as-usual students' self-esteem in their English and Spanish classes, with the exception of first-grade students in Treatment 1 exhibiting greater self-esteem in using the Spanish language. However, the majority of treatment teachers reported via qualitative data that ELLA-V had resulted in increased student confidence and self-esteem in speaking English. Figure 2 provides an overview of program impacts on student outcomes.

Figure 2. Summary of effects of ELLA-V on student outcomes.



NOTE—All outcomes measures were not administered to students in all grades.

Science. ELLA-V positively impacted science achievement for third-grade students in Treatment 1 who were exposed to ELLA-V curricula that focused on reading in the content area of science. Specifically, third-grade students in Treatment 1 outperformed business-as-usual students on the ITBS science test by 5.6 points, or 0.27 standard deviations, on average. Third-grade students in Treatment 2 were exposed to a science-infused oral language curriculum, yet Treatment 2 and business-as-usual students did not differ in science achievement. Students in grades K–2 were not tested in science. Table 5 shows program impacts relative to the business-as-usual condition and outlines the unadjusted mean for the business-as-usual students, impact estimate, standard error of the estimate (SE), p-value of the impact estimate, and standardized effect size. The standardized effect size provides the effect of the ELLA-V program on students' science achievement in terms of standard deviations.

Table 5. *Estimated impact of ELLA-V on ITBS science.*

			Treatment 1				Treatment 2			
Outcome	Grade	Unadjusted Control Mean	Impact Estimate	SE	P-value	Std. Effect Size	Impact Estimate	SE	P-value	Std. Effect Size
ITBS Science	3	185.94	5.63	2.83	.047	**0.27**	-0.07	2.29	.975	0.00

Science content was infused throughout all curricula to varying degrees, but the positive impact of ELLA-V on science achievement for third-grade students in Treatment 1 can be explained by curricula differences across treatments and grade levels. Third-grade students in Treatment 1 were exposed to an intensive science-infused literacy curriculum, while science was a lesser focus for third-grade students in Treatment 2. Treatment 2 instead emphasized oral language and engaging in academic conversations, while incorporating science vocabulary.

Oral language. The average ELLA-V kindergarten student in both Treatments 1 and 2 significantly outperformed the average business-as-usual student in oral language development as assessed by the WMLS oral language subtest. Treatment 1 kindergarten students improved their oral language by 0.16 standard deviations (or 4.2 points), and Treatment 2 kindergarten students by 0.09 standard deviations (or 2.4 points), more so than business-as-usual peers and on average. First-grade students in Treatment 2 also showed average gains in oral language development that were 0.12 standard deviations (or 2.4 points) higher than business-as-usual students. Conversely, first-grade students in Treatment 1 had significantly lower average gains in oral language development by 0.09 standard deviations (or 1.9 points lower), compared with business-as-usual peers. Table 6 provides impact estimates of ELLA-V treatments on EL students' oral language, relative to the business-as-usual condition.

Table 6. *Estimated impact of ELLA-V on WMLS-R oral language.*

			Treatment 1				Treatment 2			
Outcome	Grade	Unadjusted Control Mean	Impact Estimate	SE	P-value	Std. Effect Size	Impact Estimate	SE	P-value	Std. Effect Size
WMLS-R Oral	3	82.11	-0.01	0.69	.983	0.00	-0.28	0.72	.697	-0.02
	2	81.62	-0.71	0.59	.225	-0.04	-0.01	0.53	.983	0.00
	1	77.06	-1.92	0.76	.011	**-0.09**	2.44	0.79	.002	**0.12**
	K	66.99	4.15	0.92	.000	**0.16**	2.43	0.90	.007	**0.09**

These findings can be explained at least partially by curricula differences across treatments and grades. ELLA-V curricula focused primarily on developing EL students' oral language skills for kindergarten students in both treatments and for first-grade students in Treatment 2. For first-grade students in Treatment 1, curricula focused on oral language development during the first semester of the school year, and then emphasized learning to read during the second semester. Thus, the oral language focus of ELLA-V was most pronounced for kindergarten students in both treatments and for first-grade students in Treatment 2, and was consistent with the statistically significant positive effects. The negative result for Treatment 1 first-grade students was unexpected given that ELLA-V was designed to support EL students' language acquisition throughout all treatments and grades, and oral language is one component of language acquisition. There was no difference in oral language development for business-as-usual and treatment second- or third-grade students, but oral language was not the primary focus in these grades.

Phonological awareness. In addition to exhibiting gains in oral language, kindergarten students in Treatment 1 had significantly higher average gains in phonological awareness, compared with business-as-usual students, by 0.15 standard deviations, or 0.40 points on TOPA 2nd Edition Plus. The curriculum for kindergarten students in Treatment 1 specifically targeted phonological awareness in addition to oral language development and vocabulary building. There was no difference in phonological awareness for first-grade students in either treatment or kindergarten students in Treatment 2, relative to business-as-usual peers; however, phonological awareness was not emphasized for these treatment-grade combinations. Table 7 outlines impact estimates of ELLA-V treatments on EL students' phonological awareness, relative to the business-as-usual condition.

Table 7. *Estimated impact of ELLA-V on TOPA 2nd Edition Plus phonological awareness.*

			Treatment 1				Treatment 2			
Outcome	Grade	Unadjusted Control Mean	Impact Estimate	SE	P-value	Std. Effect Size	Impact Estimate	SE	P-value	Std. Effect Size
TOPA	1	6.37	0.11	0.36	.767	0.04	-0.08	0.33	.818	-0.03
	K	7.59	0.40	0.16	.010	**0.15**	-0.06	0.16	.722	-0.02

English language development. There was no difference in EL students' English language development, as measured by the listening and speaking subscales of the TELPAS test[8],

[8] Students' TELPAS scores on the two subscales (listening and speaking) were averaged to construct this measure.

for treatment and business-as-usual students in any grade level. Moreover, effects of ELLA-V on EL students' English language development according to TELPAS were directionally both positive and negative for treatment students, as well as not statistically significant. Table 8 shows impact estimates of ELLA-V treatments on EL students' English language development, relative to the business-as-usual condition.

Table 8. *Estimated impact of ELLA-V on TELPAS English language development (ELD).*

		Treatment 1					Treatment 2				
Outcome	Gr.	C Mean	Impact Est.	SE	P	Std. Eff.	C Mean	Impact Est.	SE	P	Std. Eff.
TELPAS ELD	3	3.28	0.04	0.07	.560	0.05	3.28	0.10	0.07	.141	0.14
	2	3.00[a]	-0.06	0.09	.471	-0.08[a]	3.09	0.07	0.07	.290	0.09
	1	2.50	-0.13	0.11	.202	-0.16	2.39[a]	-0.09	0.09	.338	-0.11[a]
	K	1.71	0.08	0.11	.497	0.09	1.71	0.02	0.11	.861	0.02

NOTE—[a]The model also incorporated propensity score weighting to establish baseline equivalence.

This finding appears to contradict the earlier one that younger treatment students outperformed business-as-usual peers in oral language. One explanation of these seemingly contradicting findings is the difference in instruments. The WMLS-R oral language subtest was scaled on monolingual English speakers, whereas the TELPAS is administered to and therefore normed from non-native English speakers. While EL students' scores on the WMLS-R oral language subtest and TELPAS were correlated (ρ = .55), the two instruments measured different constructs. Another potential explanation is that it is generally more difficult to identify program impacts on state or district tests as opposed to low-stakes assessments (Irby et al., 2010; Tong et al., 2008).[9] The WMLS-R oral language subtest is a low-stakes assessment, whereas the TELPAS is a high-stakes state assessment. Hence, differences in instruments may help to explain these seemingly contradictory findings.

Reading. Another component of EL students' English language acquisition that was targeted by ELLA-V was reading. Treatment 1 in grade 2 primarily focused on reading. Effects of ELLA-V were estimated on several reading outcomes, including reading achievement (STAAR reading), English language development in reading (WMLS-R reading and TELPAS reading subtests)[10], and reading fluency (DIBELS). There was no difference in reading performance for ELLA-V and business-as-usual students for any reading outcome, treatment, or grade. Table 9 provides impact estimates of ELLA-V treatments on EL students' reading performance.

[9] Additionally, using study data from the What Works Clearinghouse language arts and mathematics protocols as of January 2018, the average effect size of educational programs was 0.29 when using low-stakes assessments and 0.13 when using state or district assessments.

[10] Note that adjustments for multiple comparisons were not applied because there was only one outcome measure per domain for confirmatory contrasts.

Table 9. *Estimated impact of ELLA-V on reading outcomes.*

		Treatment 1					Treatment 2				
Outcome	Gr.	C Mean	Impact Est.	SE	P	Std. Eff.	C Mean	Impact Est.	SE	P	Std. Eff.
STAAR Reading	3	1369.1	16.74	12.77	.190	0.13	1369.1	-10.26	13.08	.433	-0.08
WMLS-R Reading	3	99.54	-0.54	0.97	.579	-0.03	99.54	-0.85	1.00	.393	-0.05
	2	99.53[a]	1.39	0.76	.067	0.09[a]	101.41	0.25	0.53	.642	0.02
	1	106.42	-1.90	1.16	.101	-0.11	106.42	-0.78	0.95	.409	-0.04
	K	92.68	-0.91	1.86	.626	-0.04	92.68	0.58	1.51	.703	0.03
DIBELS Reading Fluency	2	87.73	-0.94	1.48	.524	-0.03	87.73	2.43	1.56	.120	0.07
	1	46.53	0.40	1.66	.811	0.01	46.53	-0.14	1.43	.920	0.00
TELPAS Reading	3	2.68	0.04	0.07	.586	0.04	2.68	0.06	0.06	.380	0.06
	2	2.57[a]	-0.01	0.09	.922	-0.01[a]	2.64	0.01	0.10	.928	0.01
	1	2.13[a]	-0.15	0.11	.170	-0.16[a]	2.13[a]	-0.19	0.10	.059	-0.21[a]
	K	1.40	0.01	0.12	.952	0.01	1.40	-0.13	0.11	.254	-0.17

NOTE—[a]The model also incorporated propensity score weighting to establish baseline equivalence.

Writing. ELLA-V did not target EL students' writing, though the curricula for second- and third-grade students in Treatment 2 contained a small writing component. Impacts of ELLA-V were estimated on EL students' progress in writing in English using the TELPAS writing subtest. The estimated impacts of ELLA-V on writing for second- and third-grade students in Treatment 2 compared with business-as-usual students were directionally positive, but they were not statistically significant. There were no statistically significant differences in students' writing performance for treatment and business-as-usual students in other grades, as anticipated, given that ELLA-V did not target EL students' writing. Table 10 outlines impact estimates of ELLA-V treatments on EL students' English proficiency in writing, relative to the business-as-usual condition.

Table 10. *Estimated impact of ELLA-V on TELPAS writing.*

		Treatment 1					Treatment 2				
Outcome	Gr.	C Mean	Impact Est.	SE	P	Std. Eff.	C Mean	Impact Est.	SE	P	Std. Eff.
TELPAS Writing	3	2.78	-0.07	0.08	.385	-0.08	2.78	0.12	0.08	.111	0.13
	2	2.53[a]	0.03	0.10	.760	0.03[a]	2.64	0.08	0.09	.390	0.09
	1	2.02[a]	-0.14	0.12	.231	-0.16[a]	2.02[a]	-0.19	0.10	.059	-0.24[a]
	K	1.35	0.02	0.10	.823	0.03	1.35	-0.09	0.10	.353	-0.13

NOTE—[a]The model also incorporated propensity score weighting to establish baseline equivalence.

Self-esteem. With one exception, ELLA-V did not impact EL students' self-esteem in their English and Spanish classes. The exception was that first-grade students in Treatment 1 reported higher average self-esteem in their Spanish class than business-as-usual students. The survey instrument used to gauge EL students' self-esteem contained a 4-point survey scale and may not have been adequately precise to detect program impacts, however. Table 11 shows impact

estimates of ELLA-V treatments on EL students' self-esteem in their English and Spanish classes, relative to the business-as-usual condition.

Table 11. *Estimated impact of ELLA-V on self-esteem.*

			Treatment 1				Treatment 2			
Outcome	Grade	Unadjusted Control Mean	Impact Estimate	SE	P-value	Std. Effect Size	Impact Estimate	SE	P-value	Std. Effect Size
Self-esteem in English	3	1.67	-0.01	0.02	.761	-0.02	-0.02	0.02	.301	-0.07
	2	1.65	-0.02	0.02	.388	-0.05	0.02	0.02	.147	0.09
	1	1.52	0.03	0.02	.144	0.10	0.02	0.02	.373	0.05
	K	1.40	0.04	0.03	.119	0.09	0.03	0.03	.182	0.08
Self-esteem in Spanish	3	1.40	-0.01	0.04	.744	-0.02	-0.01	0.03	.637	-0.03
	2	1.51	0.01	0.03	.723	0.02	0.00	0.03	.883	-0.01
	1	1.54	0.05	0.02	.034	**0.11**	0.03	0.02	.108	0.08
	K	1.47	-0.03	0.03	.391	-0.06	-0.02	0.03	.622	-0.03

The majority of treatment teachers, through interviews, focus groups, and surveys, reported that ELLA-V fostered higher self-esteem and confidence for EL students across all grades. Teachers believed that ELLA-V created a risk-free, supportive environment in which students could experiment using the English language. ELLA-V provided standardized routine and structure for each lesson that taught students what to expect, resulting in increased student confidence. As one teacher stated, "*When they know what to expect, then their self-esteem increases because they know how to act and what to say.*" As a result of ELLA-V, teachers reported increased student engagement in terms of volunteering to answer questions or sharing responses with the class. One teacher remarked, "*Now it's like all students want to answer and have their opinion or ideas heard.*"

Teachers also reported that the structure, along with more and varied student groupings, provided more opportunities for students to practice speaking English with their classmates, thereby leading students to take more risks with the English language. One teacher commented, "*As students got more comfortable with the routines and activities of the project, they were taking more risks and attempting to communicate outside of their normal vocabulary*." Students also became less wary of making mistakes while practicing their English. One teacher remarked, "*Students know the routines and feel comfortable to make mistakes, as well as to celebrate gained knowledge*." Another teacher stated, for example, "*Students know they are allowed to make mistakes and no one will make fun or feel threatened if they cannot say the words correctly. They take chances in answering the best they can, knowing that someone will be there to help them to be successful.*" Many teachers reflected that establishing a safe classroom space where students could take risks with English ultimately increased student self-esteem and confidence.

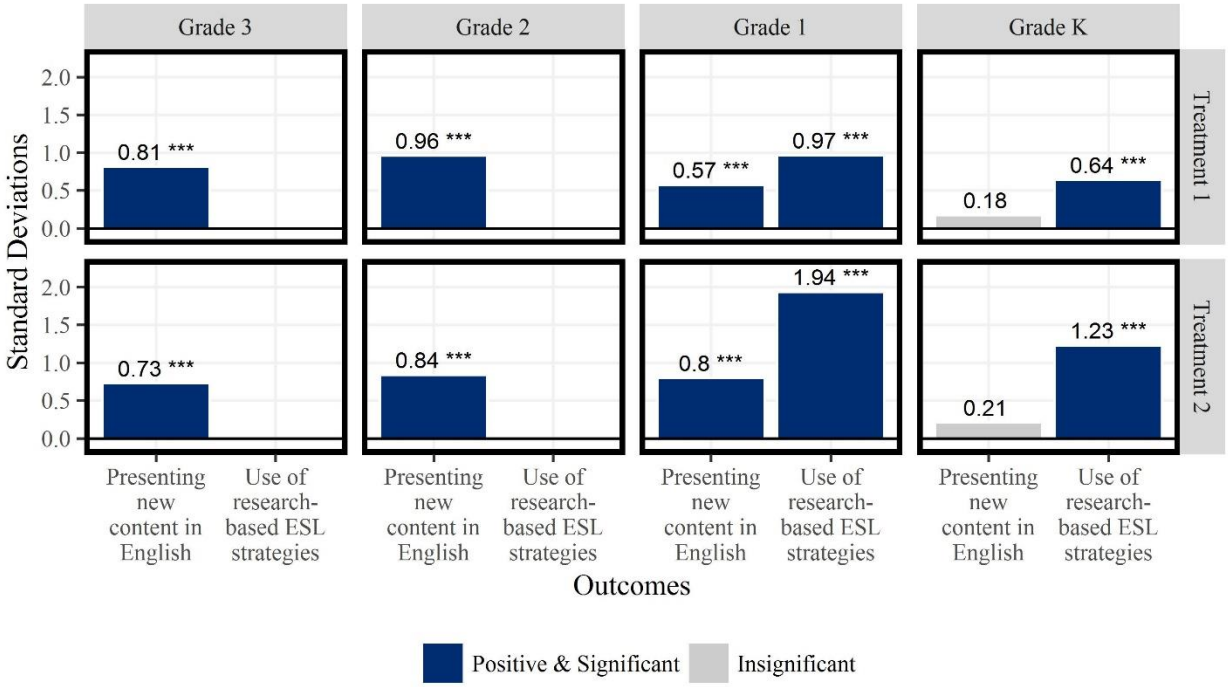
Teachers also attributed improved student confidence in speaking English to more opportunities to practice English outside of their ESL class. One teacher said, "*Students' confidence level is so high that they are now speaking English all through the day*." Other teachers remarked that ELLA-V had a positive impact on student confidence because students were more

comfortable interacting with other English speakers, including other students in the school. One teacher commented, "*I am proud of them and they know it. I can see most feel secure. The other day, a student approached me and told me her mom was proud of her, and I could see the pride on her face.*"

Impact on Teacher Outcomes

The ELLA-V professional development and coaching covered pedagogical strategies for ELs, such as structured opportunities for students to converse, less talking by the teacher, instruction clarifications, student engagement questioning strategies, structured student activities, and use of multiple forms of communication. The intervention also focused on increasing the amount of instructional time dedicated to developing ELs' cognitive-academic language proficiency (CALP). Findings showed that ELLA-V teachers implemented research-based ESL strategies to a greater extent and spent more instructional time presenting new academic content in English, compared with business-as-usual teachers. Figure 3 provides a summary of program impacts on teacher outcomes, and the following sections further detail the findings.

Figure 3. Summary of effects of ELLA-V on teacher outcomes.



NOTE—Teachers in the second and third grades were not observed using the TOR instrument.

Use of ESL-relevant strategies. ELLA-V increased treatment teachers' use of ESL-relevant instructional strategies, as measured by observers using the TOR instrument. Teachers in both Treatment 1 and Treatment 2 in grades K–1 exhibited increased implementation of research-based ESL strategies by 0.43–1.23 points, on average, on a 4-point survey scale, compared with business-as-usual teachers. These differences translated to an effect sizes ranging from 0.64–1.94

standard deviations across the treatments and grades. The effect sizes were very large, which is often the case when the outcome variable is a survey scale. Table 12 provides impact estimates of ELLA-V treatments on teachers' use of ESL strategies, relative to the business-as-usual condition. Note that teachers in grades 2–3 were not observed using the TOR instrument.

Table 12. *Estimated impact of ELLA-V on teachers' use of ESL strategies (TOR).*

			Treatment 1				Treatment 2			
Outcome	Grade	Unadjusted Control Mean	Impact Estimate	SE	P-value	Std. Effect Size	Impact Estimate	SE	P-value	Std. Effect Size
Use of ESL strategies	1	-0.69	0.70	0.20	.000	**0.97**	1.23	0.16	.000	**1.94**
	K	-0.31	0.43	0.14	.002	**0.64**	0.87	0.18	.000	**1.23**

NOTE—All models also incorporated propensity score weighting to establish baseline equivalence. Treatment teachers were exposed to the intervention prior to the baseline measure.

Qualitative data supported these findings. Regardless of grade level or treatment group, ELLA-V teachers consistently referred to four ESL strategies they started using in their classes due to the ELLA-V intervention: grouping strategies, differentiated instruction, sentence stems, and vocabulary building with visuals.

As a result of ELLA-V, treatment teachers reported relying more frequently on grouping of students, including heterogeneous (mixed level) and homogenous (same level) grouping, as well as peer-to-peer tutoring. Group activities included Think-Pair-Share, teamwork, partner talk, opportunities for students to help each other, conversation, and "ask a friend" to support both high- and low-proficiency students in the same class. Of the grouping strategies mentioned, teachers most frequently used the Think-Pair-Share strategy and found it to be the most influential in improving student confidence and oral language development. One teacher stated that by "*Allowing time for the students to stop and think, then share their thoughts and engage in academic conversation, the students have strengthened their language abilities by leaps and bounds.*"

Treatment teachers also reported using more differentiated instruction than they had done in the past. One example of differentiated instruction was kindergarten teachers asking students who were more proficient in English to write answers to questions before stating them or to answer "*what if*" questions requiring detailed answers, while asking students who were less proficient to answer questions using teacher-modeled sentence stems. Another example was the increased use of graphic organizers and visuals for students with lower English proficiency, while increasing extended discussion or text connections for students with higher English proficiency.

Differentiated instruction also included teachers' use of scaffolding and modeling based on students' language proficiency levels. Teachers noted they frequently aligned scaffolds to their students' levels, such as providing scaffolds in a student's native language or utilizing visuals to build vocabulary. One first-grade teacher noted positive improvements in EL students' academic language as a result of scaffolding: "*I think my students are using more academic language as we have scaffolded the language to the point to where they may apply it*." Teachers also commented

on the benefits of modeling: "*This project has really given me a lot of insight on how modeling and hands-on activities will help English Language Learners learn and retain information.*"

Treatment teachers also reported greater use of visuals and visual cues to build EL students' academic vocabulary. Visuals and visual cues included gestures, vocabulary cards, total physical response, word walls, graphic organizers, and fold-ables (e.g., three dimensional graphic organizers). While visuals were primarily used to support EL students with limited English proficiency, a few teachers indicated that they had used more complex visuals and graphic organizers for students with greater English proficiency.

Finally, treatment teachers reported using sentence stems to a greater extent than they had prior to ELLA-V. Teachers attributed improved spoken English by their EL students, particularly those with limited English proficiency, to the use of sentence stems. One teacher stated, "*The students that used to give answers only in Spanish are taking the risk now and responding to questions in English using the sentence stem.*" Another teacher remarked that as a result of using sentence stems, "*Most of my ELL students can now answer in complete sentences, and most of them can write in complete sentences.*"

Increased time spent presenting new academic content in English. ELLA-V prepared teachers to teach new academic content in English, while supporting the academic language needs of their EL students. The average treatment teacher in each grade except for kindergarten was observed spending more time teaching new academic content in English than the average teacher in business-as-usual schools. Specifically, treatment teachers in grades 1–3 averaged approximately 13–52% more time presenting new academic content while speaking in English than business-as-usual teachers. These differences translated into effect sizes ranging from 0.57–0.96 standard deviations, depending on the treatment and grade. Surprisingly, there were no observed differences in instructional time spent on presenting new content in English for kindergarten treatment and business-us-usual teachers, though the effect sizes were directionally positive for kindergarten treatment teachers. Table 13 outlines impact estimates of ELLA-V treatments on time spent presenting new content in English, relative to the business-as-usual condition.

Table 13. *Estimated Impact of ELLA-V on Time Spent Presenting New Content in English (TBOP).*

			Treatment 1				Treatment 2			
Outcome	Grade	Unadjusted Control Mean	Impact Estimate	SE	P-value	Std. Effect Size	Impact Estimate	SE	P-value	Std. Effect Size
Time Spent	3	0.55	0.21	0.06	.000	**0.81**	0.21	0.04	.000	**0.73**
	2	0.46	0.24	0.06	.000	**0.96**	0.22	0.06	.000	**0.84**
	1	0.68	0.09	0.03	.004	**0.57**	0.11	0.03	.000	**0.80**
	K	0.71	0.04	0.06	.513	0.18	0.05	0.05	.338	0.21

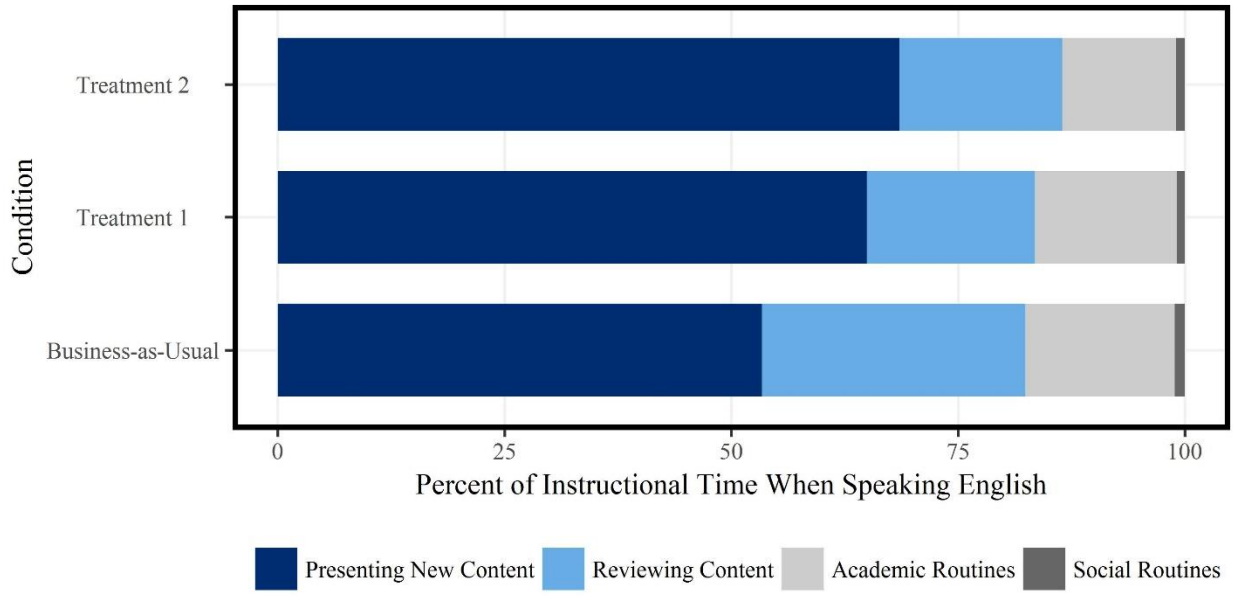
NOTE—All models also incorporated propensity score weighting to establish baseline equivalence. Treatment teachers were exposed to the intervention prior to the baseline measure.

Additionally, descriptive statistics revealed that teachers in all treatment and business-as-usual conditions spoke English about 90% of the time when observed at the end of the school year. Teachers of younger students spoke English to a greater extent than teachers of older students, as shown in Table 14. Moreover, as shown in Figure 4, treatment and business-as-usual teachers spent similar amounts of time on social and academic routines, yet treatment teachers spent more time presenting new academic content in English, while business-as-usual teachers spent more time reviewing academic content in English. These descriptive findings provide further evidence that ELLA-V teachers in grades 1–3 were targeting ELs' CALP more so than business-as-usual teachers.

Table 14. *Percentage of instructional time spent speaking English.*

	Business-as-Usual	Treatment 1	Treatment 2
Grade	%	%	%
3	84	82	82
2	90	87	89
1	91	93	94
K	94	97	97

Figure 4. Proportion of instructional time spent on various activities when speaking in English for all grade levels combined.



Fidelity of Program Implementation

ELLA-V included three major program components: virtual professional development (VPD), virtual mentoring and coaching (VMC), and curricular materials. Treatment teachers received approximately 90 minutes of VPD every two weeks from September to May, on average totaling three hours per month. VPD fidelity was measured in this study by teacher attendance rates for the professional development training sessions. Teachers also received at least one and up to three rounds of VMC, depending on teacher need, which occurred between January and May. VMC fidelity was measured by coach observation feedback rubrics, which recorded participation. Finally, schools received curricular materials at the start of each school year, and fidelity was measured by delivery receipts of curricular materials.

Fidelity of VPD, VMC, and curricular materials were measured at the school level. VPD was considered to have been implemented with fidelity in a school if all treatment teachers in the school participated in all but two professional development sessions. VMC was considered to have implemented with fidelity in a school if all treatment teachers in the school participated in at least one coaching session. The distribution of curricular materials was considered to be implemented with fidelity if the school received the curriculum materials. The component level threshold for fidelity of implementation at the sample level was 90%. That is, 90% of schools had to have achieved high fidelity for the program component to be implemented with fidelity at the sample level.

The fidelity of implementation for each program component was analyzed from the 2013-14 through 2016-17 school years. Across years of implementation, all key program components

were implemented with fidelity except for VPD for grades 3 and K, which were 57 and 12 percentage points below the fidelity threshold of 90%, respectively. Therefore, ELLA-V was mostly implemented with fidelity across the treatment years. Table 15 summarizes the fidelity for program component and year of implementation, with additional details on fidelity reported in Appendix C.

Table 15. *Fidelity of implementation scores for ELLA-V key components.*

Implementation Year & Grade	Key Component	Sample Size	Fidelity Score	Implemented with Fidelity?
2013–14	VPD	40 schools	43%	N
Grade 3	VMC	40 schools	100%	Y
	Materials	40 schools	100%	Y
2014–15	VPD	45 schools	98%	Y
Grade 2	VMC	45 schools	100%	Y
	Materials	45 schools	100%	Y
2015–16	VPD	39 schools	100%	Y
Grade 1	VMC	39 schools	100%	Y
	Materials	39 schools	100%	Y
2016–17	VPD	42 schools	88%	N
Grade K	VMC	42 schools	100%	Y
	Materials	42 schools	100%	Y

NOTE—Fidelity of implementation was calculated across both Treatments 1 and 2.

Perceived Effectiveness of Program

The vast majority of teachers and principals reported that the ELLA-V professional development, coaching, and curricula were effective in supporting them to meet the needs of their EL students. Participants were asked to reflect on the usefulness of these three key program components, as well as provide feedback for program improvement. The following sections summarize teacher and principal responses.

Professional development. The vast majority of teachers (around 85%) across treatments and grades perceived that the ELLA-V online professional development was effective, and participants used the words, "*helpful,"* "*supportive*," "*engaging*," and "*effective*" in describing the training. Moreover, even seasoned teachers found the professional development to be helpful. One teacher said, "*After 18 years of teaching English as a second language, I can honestly say that I learned effective new ways to teach spelling, reading, and writing*." Another teacher added, "*Although I have taught for 10 years, I have enjoyed using some of the program's strategies when teaching other subjects, such as Social Studies and Spanish Reading.*" Additionally, teachers found the cross-content application of ELLA-V strategies to be particularly useful. One kindergarten teacher stated, "*Excellent ESL strategies and I use them on other subjects such as science and social studies, and even during Spanish Reading*." Another teacher noted, "*I've caught myself using several strategies in other subjects.*"

Teachers also found the online professional development empowering. They felt encouraged to ask questions, and they commented that the training was valuable because it provided opportunities to discuss doubts and clarify concerns. Several teachers specifically cited lesson previewing as a helpful means of addressing instructional issues. They also liked working in a group and believed that doing so helped them comprehend the material. Teachers cited inter-teacher collaboration as one benefit of the professional development.

Overall, teachers believed that the professional development was worthwhile and necessary to implement ELLA-V strategies with fidelity. One teacher remarked, "*Without these trainings, I think we were lost. I needed to see the modeling, it showed me how to present my teaching to my students.*" This sentiment was similarly expressed by another teacher, "*The online training was awesome because we were able to see what, when, where, why and how, before we started our lessons. It helped us be prepared.*" Finally, one teacher expressed that the training provided, "*excellent support, was responsive to struggles of teachers, and filled with lifelong learning.*" Similarly, principals reported that the virtual online training was both helpful and effective. One principal noted it was a "*good training tool*." Another added*, "ELLA-V provides the opportunity to gain greater knowledge of EL strategies teachers can implement in the classroom*."

Teachers also provided a few recommendations for program improvement across different grades and treatments, and they consistently shared these suggestions:

- *Length of training.* Many teachers expressed that the professional development sessions were too long, and some requested that the duration of each session be changed from one and a half hours to one hour. In lieu of decreasing the duration of each session, others suggested being able to access the online training from home and on demand.
- *Repetitive content.* Multiple teachers and principals commented that the content presented in the trainings felt repetitive at times. Teachers also suggested ways to make the trainings feel less repetitive. They recommended having teachers read the Power Point presentations for themselves, as opposed to the trainers reading them aloud word-for-word. Other suggestions were to reduce the amount of time spent on personal introductions or introducing material at the beginning of each session.
- *Technical challenges.* Several teachers experienced technical problems with the platform. Specifically, several mentioned problems using the microphones and poor video quality as being the main technical issues.
- *Inadequate coverage of certain topics.* While principals believed that the professional development was beneficial to teachers, many expressed that additional professional development was needed in the areas of sheltered instruction, problem solving, writing strategies, monitoring teachers' fidelity of program implementation, and transitioning students to upper-grade levels.

Mentoring and coaching. Most teachers across grade levels and treatments agreed that the ELLA-V online coaching was effective, consistently commenting that the support was "*great*," "*helpful*," "*excellent*," "*lively*," and "*informative*." Teachers particularly appreciated the feedback

they received from the coaches. One teacher noted that it was "*good to have constructive criticism*," and that the coaching "*helped identify skills and strategies needed*." Teachers generally liked having detailed directions and constructive criticism. One teacher remarked, "*There are things that you do not notice that you do and the coaching helps you with it. Someone is there to help guide you in the right direction.*"

Teachers also appreciated that the coaching helped them understand what to expect in the lessons and guided them in how to prepare for the lessons, including help with lesson pacing and enrichment. One teacher noted that the coaching was especially effective when teachers were unfamiliar with a particular strategy because coaches provided step-by-step instructions that clarified doubts and answered questions. Finally, coaches nudged teachers to reflect on their practice and grow professionally. One teacher concluded that the coaching was an "*excellent and a fundamental part of the program.*"

Teachers also provided recommendations for how to improve the coaching. Several teachers found the earpiece hard to use, and a few described the process of instantaneous feedback as "*nerve-wracking*" because it was difficult to listen to both the coach and respond to students at the same time. One teacher suggested that instead of real-time feedback, they would rather record a teaching session and get feedback at a later time. However, some teachers who did not receive real-time feedback wished that they had received feedback more frequently and sooner after their observation.

Curricula. Nearly all teachers in grades K–1 liked the curricula materials, whereas about half of teachers in grades 2–3 felt similarly. Positive teachers thought the curricula (a) were simple to understand and easily implemented, (b) provided helpful structure and routines to improve learning environments, (c) offered useful tools, resources, and instructional practices, (d) incorporated student-to-student collaborative opportunities into lessons, and (e) adequately focused on the content area.

Teachers particularly liked the lesson plans, which they believed were sufficiently "*teacher friendly*," "*detailed*," "*well-paced*," "*thorough*," "*well-structured*," "*well-planned*," and "*simple and to the point.*" The reading books were also well received. Teachers liked the interactive aspect of the books, including the associated songs and engaging visuals. One teacher remarked, "*Students looked forward to the new book each week; in fact, they were sad when we finished studying the last book.*" Another teacher similarly expressed that the "*students want to read every day*." Teachers also commented that the books were age-appropriate and helped students develop their vocabulary. Teachers of students in grades 2–3 added that the books were relatable to students, and that students enjoyed the book topics, including both the fiction and non-fiction ones. Around 40% of teachers believed that the books helped students improve their reading skills, such as identifying the main idea, spelling correctly, summarizing the story, making story predictions, and skimming the story. One teacher commented, "*The books provided age-appropriate content with visuals and challenging vocabulary that helped my students improve their second language acquisition.*"

Some teachers voiced concerns about the curricula, however. About 10% of first- and second-grade teachers in both treatment groups noted that the ELLA-V lessons required more time than the 45 minutes scheduled for the ESL block. Teachers across all grade levels had similar suggestions for how to improve the curricula, including that (a) some books were too difficult for low-proficiency students, (b) more readings should be available online, (c) there should be more opportunities for writing practice, and (d) some graphic organizers were too detailed for students. There were also mixed responses regarding whether ELLA-V aligned well with district and state curricula. These critiques and suggestions provide formative feedback for future iterations of ELLA-V.

Conclusion

Consistent with the earlier ELLA program, ELLA-V improved EL teachers' quality of instruction, which led to improvements in oral language and phonological awareness for younger students and in science for third-grade students who were exposed to a literacy-infused science curriculum. Higher quality of instruction for treatment teachers was evident in increased use of ESL strategies (e.g., grouping activities, differentiated instruction, visuals for learning new vocabulary, and sentence stems) and a greater emphasis on cognitive-academic language proficiency compared with business-as-usual teachers.

With one exception, ELLA-V did not impact EL students' English language development, reading, writing, or self-esteem. Texas A&M researchers have found that ELs learn academic language incrementally, starting with oral language, and then pre-reading skills, and finally reading and writing (Tong et al., 2014). Given the backwards research design where students in each grade were exposed to the intervention for only one school year, EL students in older grades may not have reached their maximum potential under this intervention because they did not benefit from the cumulative effect of this intervention.

Another limitation is that treatment teachers were exposed to the intervention for only one school year, which may not have been adequate time for teachers to fully implement or students to fully benefit from the program. The professional development started in September, leaving ELLA-V teachers essentially 6–7 months to improve their instruction before EL student academic performance was re-assessed. Research has shown that practitioners may experience an "implementation dip," which is a short-term decrease in performance and confidence while new reforms are initiated (Fullan, 2004). Teachers in the treatment groups were asked to implement new instructional techniques, whereas teachers in the business-as-usual group could work to improve what they were already doing.

Different assessments may also help to explain some of the seemingly contradictory findings of program impacts on student outcomes. It is generally more difficult to identify program impacts on state or district tests as opposed to low-stakes assessments (Irby et al., 2010; Tong et al., 2008). In this study, there were positive impacts of ELLA-V on EL students' oral language using a low-stakes assessment, but no observed effects on EL students' English language

development using a high-stakes assessment. Moreover, some instruments were normed for monolingual English speakers, whereas other instruments were designed specifically for ELs. Therefore, tests normed for different student populations may measure different constructs even within the same domain (Bedore & Peña, 2008). Finally, the instrument used to measure EL students' self-esteem in this study may not have been adequately precise, given that study teachers overwhelming attributed ELs' improved confidence in speaking English to the intervention.

Qualitative findings showed that the vast majority of treatment teachers and principals believed that the ELLA-V professional development, coaching, and curricula were effective in supporting them to meet the needs of their EL students. Teachers benefitted from the professional development, and even veteran teachers reported that they had learned something new. Teachers also appreciated the constructive criticism they received from the coaches. Teacher feedback about the curricula was more mixed, with teachers in grades K–1 overwhelming liking the curricula, while only about half of teachers in grades 2–3 liked the curricula.

This report concludes that the ELLA-V was mostly implemented with fidelity and yielded improved outcomes for EL students in some content areas. More research is needed to identify the cumulative effects across multiple grade levels of the ELLA-V approach (oral language to pre-reading to reading and writing) on EL students' academic performance and English language proficiency. The report also highlights the ongoing need for a system of supports for teachers of ELs. Professional development and coaching together positively impacted teacher quality, yet student outcomes were impacted only when curricula also targeted the content area.

References

Ballantyne, K.G., Sanderman, A.R., & Levy, J. (2008). *Educating English language learners: Building teacher capacity*. Washington, DC: National Clearinghouse for English Language Acquisition. Available at http://www.ncela.gwu.edu/practice/mainstream_teachers.htm

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11*(1), 1–29. https://doi.org/10.2167/beb392.0

Breunig, N. A. (1998). Measuring the instructional use of Spanish and English in elementary transitional bilingual classrooms. *Dissertation Abstracts International, 59*(04), 1046A.

Bruce, K. L., Lara-Alecio, R., Parker, R., Hasbrouck, J. E., Weaver, L., & Irby, B. (1997). Accurately describing the language. *Bilingual Research Journal, 21*(2&3), 123-145.

Bryan, L. A., & Atwater, M. M. (2002). Teacher beliefs and cultural models: A challenge for science teacher preparation programs. *Science Education, 86*(6), 821-839. https://doi.org/10.1002/sce.10043

Buxton, C., & Allexsaht-Snider, M. (2016). *Supporting K-12 English language learners in science: Putting research into teaching practice*. New York, NY: Routledge.

Casteel, C.J., & Ballantyne, K.G. (Eds.). (2010). *Professional development in action: Improving teaching for English learners.* Washington, DC: National Clearinghouse for English Language Acquisition. Available at http://www.ncela.gwu.edu/files/uploads/3/PD_in_Action.pdf

Correll, P. K. (2016). *Teachers' preparation to teach English language learners (ELLs): An investigation of perceptions, preparation, and current practices* (Doctoral dissertation). Retrieved from https://uknowledge.uky.edu/edc_etds/19/

Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Trowbridge, England: Cromwell Press Ltd.

Darling-Hammond, L., & Richardson, N. (2009). Teacher learning: What matters? *Educational Leadership, 66*(5), 46-53. Retrieved from http://eds.b.ebscohost.com.proxy1.library.jhu.edu/ehost/pdfviewer/pdfviewer?vid=4&sid=3497d70b-1467-4196-a72c-08a6c258cedc%40sessionmgr103

Delaney, Y. A. (2012). Research on mentoring language teachers: Its role in language education. *Foreign Language Annals, 45*(1), 184-202. doi:10.1111/j.1944-9720.2011.01185.x

Dunbar, S., & Welch, C. (2015). *Iowa assessments: Research and development guide*. Orlando,

FL: Houghton Mifflin Harcourt. Retrieved from http://itp.education.uiowa.edu/ia/documents/Research-Guide-Form-E-F.pdf

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12*(2), 121-138. doi:10.1037/1082-989X.12.2.121

Fullan, M. (2004). *Leading in a culture of change*. San Francisco, CA: Wiley.

Good, R. H., & Kaminski, R. A. (2002). *Dynamic Indicators of Basic Early Literacy Skills* (6th ed.). Eugene, OR: Institute for the Development of Educational Achievement. Available at http://oregonreadingfirst.uoregon.edu/downloads/assessment/admin_and_scoring_6th_ed.pdf

Irby, B. J., Tong, F., Lara-Alecio, R., Mathes, G. P., Acosta, S., & Guerrero, C. (2010). Quality of instruction, language of instruction, and Spanish-speaking English learners' performance on a state high-stakes reading assessment. *TABE Journal, 12*(1), 1-42.

Irby, B., Tong, F., Lara-Alecio, R., Meyer, D., & Rodríguez, L. (2007). The critical nature of language of instruction compared to observed practices and high stakes tests in transitional bilingual classroom. *Research in the Schools, 14*(2), 27–36.

Irby, B. J., Tong, F., Nichter, M, Lara-Alecio, R., Hassey, F., & Guerrero, C. (2011). Hispanic English learners' self esteem related to instructional program type, language of instruction, and gender. *TABE Journal, 13*(1), 26-48.

Lara-Alecio, R., & Parker, R. I. (1994). A pedagogical model for transitional English bilingual classrooms. *Bilingual Research Journal, 18*(3-4), 119-133.

Lara-Alecio, R., Tong, F., Irby, B. J., & Mathes, G. P. (2009). Teachers' pedagogical differences among bilingual and structured English immersion kindergarten classrooms in a randomized trial study. *Bilingual Research Journal, 32*(1), 77-100. doi:10.1080/15235880902965938

Lee, O., & Buxton, C. A. (2013). Teacher professional development to improve science and literacy achievement of English language learners. *Theory Into Practice, 52*, 110-117. doi:10.1080/00405841.2013.770328

Lee, O., Hart, J. E., Cuevas, P., & Enders, C. (2004). Professional development in inquiry-based science for elementary teachers of diverse student groups. *Journal of Research in Science Teaching, 41*(10), 1021-1043. doi:10.1002/tea.20037

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). London, England: SAGE Publications, Inc.

Pruitt, S. L., & Wallace, C. S. (2012). The effect of a state department of education teacher

mentor initiative on science achievement. *Journal of Science Teacher Education, 23*, 367-385. doi:10.1007/s10972-012-9280-5

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed., Vol. 1). London, England: SAGE Publications, Inc.

Ridgeway, G., McCaffrey, D. F., Morral, A. R., Burgette, L. F., & Griffin, B. A. *Toolkit for weighting and analysis of nonequivalent groups: A tutorial for the R TWANG package*. Santa Monica, CA: RAND Corporation, 2014. Retrieved from https://www.rand.org/pubs/tools/TL136z1.html

Samson, J. F., & Collins, B. A. (2012). *Preparing all teachers to meet the needs of English language learners: Applying research to policy and practice for teacher effectiveness.* Washington, DC: Center for American Progress. Available from https://files.eric.ed.gov/fulltext/ED535608.pdf

Tarr, J. E., Reys, R. E., Reys, B. T., Chávez, Ó., Shih, J., & Osterlind, S. J. (2008). The impact of middle-grades mathematics curricula and the classroom learning environment on student achievement. *Journal for Research in Mathematics Education, 39*(3), 247-280. Retrieved from http://www.jstor.org.proxy1.library.jhu.edu/stable/pdf/30034970.pdf?refreqid=excelsior%3A38419830eaefc0fd7e535f384bff4591

Texas Education Agency. (2017a). *Enrollment in Texas public schools, 2016-17.* (Document No. GE17 601 12). Austin TX: Author. Available at https://tea.texas.gov/acctres/enroll_index.html

Texas Education Agency. (2017b). *Texas Academic Performance Report: 2016-17 State Performance.* Austin, TX: Author. Available at https://rptsvr1.tea.texas.gov/perfreport/tapr/2017/state.pdf

Tong, F., Irby, B. J., Lara-Alecio, & Koch, J. (2014). Integrating literacy and science for English language learners: From learning-to-read to reading-to-learn. *The Journal of Educational Research, 107*(5), 410-426. https://doi.org/10.1080/00220671.2013.833072

Tong, F., Irby, B. J., Lara-Alecio, R., & Mathes, G. P. (2008). English and Spanish acquisition by Hispanic second grades in developmental bilingual programs: A 3-year longitudinal randomized study. *Hispanic Journal of Behavioral Sciences, 30*(4), 500-529.

Tong, F., Irby, B. J., Lara-Alecio, R., Yoon, M., & Mathes, G. P. (2010). Hispanic English learners' response to a longitudinal English instructional intervention and the effect of gender: A multilevel analysis. *The Elementary School Journal, 110*(4), 542-566.

Tong, F., Irby, B., Tang, S., Lin, S., Guerrero, C., Lara-Alecio, R., & Lopez, T. (2017). *Indicators of inter-rater reliability for classroom observation instruments as fidelity of implementation in large-scale RCTs*. Paper presented at the annual meeting of American Educational Research Association, San Antonio, TX.

Tong, F., Lara-Alecio, R., Irby, B. J., Mathes, G. P., & Kwok, O. (2008). Accelerating early academic oral English development in transitional bilingual and structured English immersion programs. *American Educational Research Journal, 45*(4), 1011-1044.

Tong, F., Luo, W., Irby, B. J., Lara-Alecio, R., & Rivera, H. (2017). Investigating the impact of professional development on teachers' instructional time and English learners' language development: A multilevel cross-classified approach. *International Journal of Bilingual Education and Bilingualism*, *20*(3), 292-313. http://dx.doi.org/10.1080/13670050.2015.1051509

Torgesen, J. K., & Bryant, B. R. (2004). *Test of Phonological Awareness: Second Edition Plus (TOPA 2+)*. Austin, TX: Pro-Ed.

U. S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, What Works Clearinghouse. (2017). *What Works Clearinghouse: Procedures and standards handbook version 4.0.* Washington, DC: Author.

Valdés, G. (2004). Between support and marginalisation: The development of academic language in linguistic minority children. *International Journal of Bilingual Education and Bilingualism, 7*(2), 102-132.

Woodcock, R. W., Muñoz-Sandoval, A. F., Ruef, M. L., Alvarado, C. G. (2005). *Woodcock-Muñoz Language Survey-Revised*. Rolling Meadows, IL: Riverside Publishing. Available at http://www.seisd.net/common/pages/DisplayFile.aspx?itemId=2069212

Yoon, K. W., Duncan, T., Lee, S. W., Scarloss, B., & Shapley, K. (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. Retrieved from https://files.eric.ed.gov/fulltext/ED498548.pdf

Overview of Appendices

The technical appendices include the following sections: A) Program Description, B) Descriptive Statistics, C) i3 Tables, and D) Instruments. Appendix A provides an overview of all curriculum models as well as the implementation of each. Appendix B provides descriptive statistics for each outcome and pretest measure. Appendix C includes all required i3 tables including a master list of contrasts, program impact, cluster attrition, baseline equivalence, and fidelity of implementation. Finally, Appendix D contains the instruments used in the study. Throughout the appendices, the three treatment conditions will be referred to by the following abbreviations: T1 (Treatment 1), T2 (Treatment 2), and BAU (Business-as-Usual).

Appendix A: Program Description

The ELLA-V program utilized several curriculum models, which varied across grade levels and across T1 and T2. The curricula are described below, and Table A1 outlines which curriculum was used in each treatment and grade level.

- ***Santillana Intensive English (SEI)***. This curriculum provided a series of scripted lessons aligned with the English Language Proficiency Standards (ELPS) and addressed effective reading practices in phonemic awareness, phonics, vocabulary development, reading fluency, and reading comprehension. The curriculum featured a systematic approach to language instruction (engage, explore, teach, practice, apply, relate, and extend).
- ***Early Interventions in Reading (EIR-I and EIR-II)***. This curriculum was aligned with ELPS and reading standards. It was taught in whole-group instruction and addressed five central strands: phonemic awareness; letter-sound correspondences; word recognition; spelling and fluency strategies; and comprehension strategies.
- ***Content Reading Integrating Science for English Language and Literacy Acquisition (CRISELLA)***. This curriculum was aligned with state and national science standards for science-embedded English language development. It included pre-reading skills, vocabulary building activities, partner reading, graphic organizers, hands-on inquiry activities, cooperative grouping, scaffolded and leveled questions, vocabulary extensions, fluency practice, and direct teaching of reading skills.
- ***Story Re-Telling and Higher-Order Thinking for English Literacy and Language Acquisition (STELLA)***. This curriculum included authentic children's narrative and expository literature, and it featured one book per week along with scripts to support instruction. It was developed to increase oral language, implement Bloom's Taxonomy with leveled questions, and align vocabulary with ELPS and pre-selected EL strategies.
- ***Academic Oral Language in Science (AOLS)***: This curriculum was aligned with state ELPS and science standards. The curriculum was designed to facilitate development of EL students' oral language.

- ***Academic Oral and Written Language in Science (AOWLS)***. This curriculum was aligned with state ELPS and science standards. The curriculum was designed to facilitate development of EL students' oral and written language.

Table A1. *Curricula and dosage by grade and treatment.*

Grade	Treatment 1 Curriculum	Treatment 1 Dosage	Treatment 2 Curriculum	Treatment 2 Dosage
K	SEI	28 weeks 45 min. per day	STELLA	28 weeks 35 min. per day
			AOLS	28 weeks 10 min. per day
1	SEI	1-14 weeks 45 min. per day	STELLA	28 weeks 35 min. per day
	EIR-I	15-28 weeks 45 min. per day	AOLS	28 weeks 10 min. per day
2	EIR-II	28 weeks 45 min. per day	STELLA	28 weeks 35 min. per day
			AOWLS	28 weeks 10 min. per day
3	CRISELLA	28 weeks 45 min. per day	STELLA	28 weeks 35 min. per day
			AOWLS	28 weeks 10 min. per day

Appendix B: Descriptive Statistics

Appendix B contains descriptive statistics for the pretest and outcome measures by grade. These tables include the following statistics: sample size, mean, standard deviation, minimum, and maximum. Table B1 contains descriptive statistics for the student measures, and Table B2 contains descriptive statistics for the teacher measures. There were many different measures, and the following tables outline the range of possible values on the measures by grade and pre- or posttest. Note that the descriptive statistics reported here for student and teacher outcomes were not adjusted for propensity score weighting and therefore reflect unadjusted scores.

Student scores were provided to the CRRE by Texas A&M University. CRRE derived teachers' TBOP and TOR scores from data provided by Texas A&M University. The TBOP scores were the proportion of time the teacher spent presenting new academic content while speaking in English. The TOR scores were derived from observers' ratings using item response theory.

Table B1. *Descriptive statistics for student measures.*

Gr.	Domain	Outcome Measure	N	Mean	SD	Min	Max	Pretest Measure	N	Mean	SD	Min	Max
3	Science	ITBS Science	1931	187.3	20.32	134	265	ITBS Science	1931	170.03	14.23	122	240
3	Oral language	WMLS Oral	1790	81.37	15.98	1	132	WMLS Oral	1790	77.90	17.54	1	135
2	Oral language	WMLS Oral	1993	80.53	16.03	1	129	WMLS Oral	1993	74.35	20.24	1	129
1	Oral language	WMLS Oral	1728	76.78	21.40	6	121	WMLS Oral	1728	63.44	25.01	1	118
K	Oral language	WMLS Oral	1755	70.32	25.55	1	134	WMLS Oral	1755	55.24	32.16	1	128
1	Phonological awareness	TOPA	1711	6.31	2.77	1	15	TOPA	1711	5.71	2.11	1	15
K	Phonological awareness	TOPA	1726	7.73	2.70	1	13	TOPA	1726	7.65	2.55	1	15
3	English language development	TELPAS ELD	1836	3.31	0.73	1	4	TELPAS ELD	1836	2.79	0.86	0	4
2	English language development	TELPAS ELD	1764	3.01	0.80	1	4	TELPAS ELD	1764	2.27	0.80	1	4
1	English language development	TELPAS ELD	1602	2.37	0.82	1	4	TELPAS ELD	1602	1.62	0.72	1	4
K	English language development	TELPAS ELD	1833	1.79	0.84	0	4	TELPAS ELD	1833	88.52	18.79	50	145
3	Reading achievement	STAAR Reading	1641	1372.63	126.35	966	1982	WMLS Reading	1641	95.68	16.74	1	155
3	English language development in reading	WMLS Reading	1653	98.23	17.08	3	153	WMLS Reading	1653	95.34	16.94	1	155

Gr.	Domain	Outcome Measure	N	Mean	SD	Min	Max	Pretest Measure	N	Mean	SD	Min	Max
2	English language development in reading	WMLS Reading	1991	99.67	15.97	13	142	WMLS Reading	1991	97.98	17.61	1	145
1	English language development in reading	WMLS Reading	1728	105.09	17.92	21	145	WMLS Reading	1728	89.91	20.90	1	140
K	English language development in reading	WMLS Reading	1705	92.73	21.44	21	152	WMLS Reading	1705	83.95	22.25	8	144
3	English language development in reading	TELPAS Reading	1836	2.71	0.98	0	4	TELPAS Reading	1836	2.66	0.97	1	4
2	English language development in reading	TELPAS Reading	1763	2.48	0.93	1	4	TELPAS Reading	1763	1.98	0.86	1	4
1	English language development in reading	TELPAS Reading	1596	2.06	0.90	0	4	TELPAS Reading	1596	1.30	0.63	1	4
K	English language development in reading	TELPAS Reading	1833	1.39	0.72	0	4	TELPAS Reading	1833	88.52	18.79	50	145
2	Reading fluency	DIBELS	1995	84.12	34.15	0	229	DIBELS	1995	49.44	28.22	0	186
1	Reading fluency	DIBELS	1727	45.18	28.45	0	155	DIBELS	1727	16.94	18.65	0	104
3	Writing	TELPAS Writing	1836	2.78	0.90	0	4	TELPAS Writing	1836	2.41	0.88	0	4
2	Writing	TELPAS Writing	1763	2.53	0.89	0	4	TELPAS Writing	1763	1.84	0.78	1	4
1	Writing	TELPAS Writing	1602	1.89	0.84	1	4	TELPAS Writing	1602	1.26	0.58	1	4

Gr.	Domain	Outcome Measure	N	Mean	SD	Min	Max	Pretest Measure	N	Mean	SD	Min	Max
K	Writing	TELPAS Writing	1833	1.35	0.68	0	4	TELPAS Writing	1833	88.52	18.79	50	145
3	Self-esteem in English	SEI English	1915	1.66	0.32	0	2	SEI English	1915	1.58	0.36	0	2
2	Self-esteem in English	SEI English	1995	1.64	0.28	0	2	SEI English	1995	1.56	0.32	0	2
1	Self-esteem in English	SEI English	1726	1.54	0.35	0	2	SEI English	1726	1.43	0.41	0	2
K	Self-esteem in English	SEI English	1776	1.43	0.42	0	2	SEI English	1776	1.34	0.49	0	2
3	Self-esteem	SEI Spanish	1914	1.37	0.55	0	2	SEI Spanish	1914	1.39	0.55	0	2
2	Self-esteem	SEI Spanish	1995	1.48	0.43	0	2	SEI Spanish	1995	1.49	0.41	0	2
1	Self-esteem	SEI Spanish	1725	1.56	0.42	0	2	SEI Spanish	1725	1.54	0.42	0	2
K	Self-esteem	SEI Spanish	1776	1.45	0.45	0	2	SEI Spanish	1776	1.43	0.47	0	2

Table B2. *Descriptive statistics for teacher measures.*

			Outcome					Pretest				
Gr.	Domain	Measure	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
3	Presenting new content in English	TBOP	112	0.64	0.29	0.00	1.00	112	0.42	0.30	0.00	0.98
2	Presenting new content in English	TBOP	132	0.62	0.25	0.00	0.95	132	0.55	0.25	0.00	0.93
1	Presenting new content in English	TBOP	116	0.73	0.14	0.30	0.97	116	0.62	0.18	0.15	0.97
K	Presenting new content in English	TBOP	122	0.75	0.22	0.03	1.00	122	0.64	0.25	0.00	0.98
1	Use of research-based ESL strategies	TOR	114	-0.02	0.87	-2.14	1.35	114	-0.04	0.85	-2.73	1.23
K	Use of research-based ESL strategies	TOR	126	0.02	0.76	-2.38	1.35	126	0.04	0.75	-2.32	1.23

Appendix C: i3 Tables

Appendix C contains all tables required of evaluations funded by the Investing in Innovation (i3) Fund. This section includes:

- Master list of contrasts
- Impact tables
- Cluster attrition tables
- Baseline equivalence tables
- Fidelity of implementation tables

Contrast IDs found in each table identify the grade and treatment for which each outcome was analyzed.

Master list of contrasts. Tables C1-C2 provide a master list of student contrasts, and Table C3 provides a list of teacher contrasts. These tables also include the outcome and pretest measures as well as the timing of the administration of the measures. Finally, these tables include whether the contrast was confirmatory (C) or exploratory (E).

Table C1. *Master list of student contrasts (Treatment 1).*

Contrast ID	Treatment Group	Control Group	Domain	Outcome Measure		Pretest Measure		C/E
T1_Students_1_Gr3	Gr3 students in T1	Gr3 students in BAU schools	Science	ITBS Science	Spring 2014	ITBS Science	Fall 2013	C
T1_Students_2_Gr3	Gr3 students in T1	Gr3 students in BAU schools	Oral language	WMLS Oral	Spring 2014	WMLS Oral	Fall 2013	E
T1_Students_3_Gr2	Gr2 students in T1	Gr2 students in BAU schools	Oral language	WMLS Oral	Spring 2015	WMLS Oral	Fall 2014	E
T1_Students_4_Gr1	Gr1 students in T1	Gr1 students in BAU schools	Oral language	WMLS Oral	Spring 2016	WMLS Oral	Fall 2015	E
T1_Students_5_GrK	GrK students in T1	GrK students in BAU schools	Oral language	WMLS Oral	Spring 2017	WMLS Oral	Fall 2016	C
T1_Students_6_Gr1	Gr1 students in T1	Gr1 students in BAU schools	Phonological awareness	TOPA	Spring 2016	TOPA	Fall 2015	E
T1_Students_7_GrK	GrK students in T1	GrK students in BAU schools	Phonological awareness	TOPA	Spring 2017	TOPA	Fall 2016	C
T1_Students_8_Gr3	Gr3 students in T1	Gr3 students in BAU schools	English language development	TELPAS ELD	Spring 2014	TELPAS ELD	Spring 2013	E
T1_Students_9_Gr2	Gr2 students in T1	Gr2 students in BAU schools	English language development	TELPAS ELD	Spring 2015	TELPAS ELD	Spring 2014	E
T1_Students_10_Gr1	Gr1 students in T1	Gr1 students in BAU	English language development	TELPAS ELD	Spring 2016	TELPAS ELD	Spring 2015	E
T1_Students_11_GrK	GrK students in T1	GrK students in BAU schools	English language development	TELPAS ELD	Spring 2017	TVIP	Spring 2016	E
T1_Students_12_Gr3	Gr3 students in T1	Gr3 students in BAU	Reading achievement	STAAR Reading	Spring 2014	WMLS Reading	Fall 2013	C
T1_Students_13_Gr3	Gr3 students in T1	Gr3 students in BAU schools	English language development in reading	WMLS Reading	Spring 2014	WMLS Reading	Fall 2013	C
T1_Students_14_Gr2	Gr2 students in T1	Gr2 students in BAU schools	English language development in reading	WMLS Reading	Spring 2015	WMLS Reading	Fall 2014	E
T1_Students_15_Gr1	Gr1 students in T1	Gr1 students in BAU schools	English language development in reading	WMLS Reading	Spring 2016	WMLS Reading	Fall 2015	E
T1_Students_16_GrK	GrK students in T1	GrK students in BAU schools	English language development in reading	WMLS Reading	Spring 2017	WMLS Reading	Fall 2016	E
T1_Students_17_Gr3	Gr3 students in T1	Gr3 students in BAU schools	English language development in reading	TELPAS Reading	Spring 2014	TELPAS Reading	Spring 2013	E
T1_Students_18_Gr2	Gr2 students in T1	Gr2 students in BAU schools	English language development in reading	TELPAS Reading	Spring 2015	TELPAS Reading	Spring 2014	E

Contrast ID	Treatment Group	Control Group	Domain	Outcome Measure		Pretest Measure		C/E
T1_Students_19_Gr1	Gr1 students in T1	Gr1 students in BAU schools	English language development in reading	TELPAS Reading	Spring 2016	TELPAS Reading	Spring 2015	E
T1_Students_20_GrK	GrK students in T1	GrK students in BAU schools	English language development in reading	TELPAS Reading	Spring 2017	TVIP	Spring 2016	E
T1_Students_21_Gr2	Gr2 students in T1	Gr2 students in BAU schools	Reading fluency	DIBELS	Spring 2015	DIBELS	Fall 2014	E
T1_Students_22_Gr1	Gr1 students in T1	Gr1 students in BAU schools	Reading fluency	DIBELS	Spring 2016	DIBELS	Fall 2015	C
T1_Students_23_Gr3	Gr3 students in T1	Gr3 students in BAU schools	Writing	TELPAS Writing	Spring 2014	TELPAS Writing	Spring 2013	C
T1_Students_24_Gr2	Gr2 students in T1	Gr2 students in BAU schools	Writing	TELPAS Writing	Spring 2015	TELPAS Writing	Spring 2014	E
T1_Students_25_Gr1	Gr1 students in T1	Gr1 students in BAU schools	Writing	TELPAS Writing	Spring 2016	TELPAS Writing	Spring 2015	E
T1_Students_26_GrK	GrK students in T1	GrK students in BAU schools	Writing	TELPAS Writing	Spring 2017	TVIP	Spring 2016	E
T1_Students_27_Gr3	Gr3 students in T1	Gr3 students in BAU schools	Self-esteem in English class	SEI English	Spring 2014	SEI English	Fall 2013	E
T1_Students_28_Gr2	Gr2 students in T1	Gr2 students in BAU schools	Self-esteem in English class	SEI English	Spring 2015	SEI English	Fall 2014	E
T1_Students_29_Gr1	Gr1 students in T1	Gr1 students in BAU schools	Self-esteem in English class	SEI English	Spring 2016	SEI English	Fall 2015	E
T1_Students_30_GrK	GrK students in T1	GrK students in BAU schools	Self-esteem in English class	SEI English	Spring 2017	SEI English	Fall 2016	E
T1_Students_31_Gr3	Gr3 students in T1	Gr3 students in BAU schools	Self-esteem	SEI Spanish	Spring 2014	SEI Spanish	Fall 2013	E
T1_Students_32_Gr2	Gr2 students in T1	Gr2 students in BAU schools	Self-esteem	SEI Spanish	Spring 2015	SEI Spanish	Fall 2014	E
T1_Students_33_Gr1	Gr1 students in T1	Gr1 students in BAU schools	Self-esteem	SEI Spanish	Spring 2016	SEI Spanish	Fall 2015	C
T1_Students_34_GrK	GrK students in T1	GrK students in BAU schools	Self-esteem	SEI Spanish	Spring 2017	SEI Spanish	Fall 2016	E

NOTES—1. The research design for all domains was RCT with school assignment. 2. In all cases, exposure to the treatment was one school year. 3. The unit of observation for all domains was the student. 4. The student sample included all study participants who had non-missing pretest and posttest scores. 5. The scale for all measures was continuous; note that TELPAS is measured on a four-point scale. 6. ELD=English language development.

Table C2. *Master list of student contrasts (Treatment 2).*

Contrast ID	Treatment Group	Control Group	Domain	Outcome Measure		Pretest Measure		C/E
T2_Students_1_Gr3	Gr3 students in T2	Gr3 students in BAU schools	Science	ITBS Science	Spring 2014	ITBS Science	Fall 2013	C
T2_Students_2_Gr3	Gr3 students in T2	Gr3 students in BAU schools	Oral language	WMLS Oral	Spring 2014	WMLS Oral	Fall 2013	E
T2_Students_3_Gr2	Gr2 students in T2	Gr2 students in BAU schools	Oral language	WMLS Oral	Spring 2015	WMLS Oral	Fall 2014	E
T2_Students_4_Gr1	Gr1 students in T2	Gr1 students in BAU schools	Oral language	WMLS Oral	Spring 2016	WMLS Oral	Fall 2015	E
T2_Students_5_GrK	GrK students in T2	GrK students in BAU schools	Oral language	WMLS Oral	Spring 2017	WMLS Oral	Fall 2016	C
T2_Students_6_Gr1	Gr1 students in T2	Gr1 students in BAU schools	Phonological awareness	TOPA	Spring 2016	TOPA	Fall 2015	E
T2_Students_7_GrK	GrK students in T2	GrK students in BAU schools	Phonological awareness	TOPA	Spring 2017	TOPA	Fall 2016	C
T2_Students_8_Gr3	Gr3 students in T2	Gr3 students in BAU schools	English language development	TELPAS ELD	Spring 2014	TELPAS ELD	Spring 2013	E
T2_Students_9_Gr2	Gr2 students in T2	Gr2 students in BAU schools	English language development	TELPAS ELD	Spring 2015	TELPAS ELD	Spring 2014	E
T2_Students_10_Gr1	Gr1 students in T2	Gr1 students in BAU schools	English language development	TELPAS ELD	Spring 2016	TELPAS ELD	Spring 2015	E
T2_Students_11_GrK	GrK students in T2	GrK students in BAU schools	English language development	TELPAS ELD	Spring 2017	TVIP	Spring 2016	E
T2_Students_12_Gr3	Gr3 students in T2	Gr3 students in BAU schools	Reading achievement	STAAR Reading	Spring 2014	WMLS Reading	Fall 2013	C
T2_Students_13_Gr3	Gr3 students in T2	Gr3 students in BAU schools	English language development in reading	WMLS Reading	Spring 2014	WMLS Reading	Fall 2013	C
T2_Students_14_Gr2	Gr2 students in T2	Gr2 students in BAU schools	English language development in reading	WMLS Reading	Spring 2015	WMLS Reading	Fall 2014	E
T2_Students_15_Gr1	Gr1 students in T2	Gr1 students in BAU schools	English language development in reading	WMLS Reading	Spring 2016	WMLS Reading	Fall 2015	E
T2_Students_16_GrK	GrK students in T2	GrK students in BAU schools	English language development in reading	WMLS Reading	Spring 2017	WMLS Reading	Fall 2016	E
T2_Students_17_Gr3	Gr3 students in T2	Gr3 students in BAU schools	English language development in reading	TELPAS Reading	Spring 2014	TELPAS Reading	Spring 2013	E
T2_Students_18_Gr2	Gr2 students in T2	Gr2 students in BAU schools	English language development in reading	TELPAS Reading	Spring 2015	TELPAS Reading	Spring 2014	E

Contrast ID	Treatment Group	Control Group	Domain	Outcome Measure		Pretest Measure		C/E
T2_Students_19_Gr1	Gr1 students in T2	Gr1 students in BAU schools	English language development in reading	TELPAS Reading	Spring 2016	TELPAS Reading	Spring 2015	E
T2_Students_20_GrK	GrK students in T2	GrK students in BAU schools	English language development in reading	TELPAS Reading	Spring 2017	TVIP	Spring 2016	E
T2_Students_21_Gr2	Gr2 students in T2	Gr2 students in BAU schools	Reading fluency	DIBELS	Spring 2015	DIBELS	Fall 2014	C
T2_Students_22_Gr1	Gr1 students in T2	Gr1 students in BAU schools	Reading fluency	DIBELS	Spring 2016	DIBELS	Fall 2015	E
T2_Students_23_Gr3	Gr3 students in T2	Gr3 students in BAU schools	Writing	TELPAS Writing	Spring 2014	TELPAS Writing	Spring 2013	C
T2_Students_24_Gr2	Gr2 students in T2	Gr2 students in BAU schools	Writing	TELPAS Writing	Spring 2015	TELPAS Writing	Spring 2014	E
T2_Students_25_Gr1	Gr1 students in T2	Gr1 students in BAU schools	Writing	TELPAS Writing	Spring 2016	TELPAS Writing	Spring 2015	E
T2_Students_26_GrK	GrK students in T2	GrK students in BAU schools	Writing	TELPAS Writing	Spring 2017	TVIP	Spring 2016	E
T2_Students_27_Gr3	Gr3 students in T2	Gr3 students in BAU schools	Self-esteem in English class	SEI English	Spring 2014	SEI English	Fall 2013	E
T2_Students_28_Gr2	Gr2 students in T2	Gr2 students in BAU schools	Self-esteem in English class	SEI English	Spring 2015	SEI English	Fall 2014	E
T2_Students_29_Gr1	Gr1 students in T2	Gr1 students in BAU schools	Self-esteem in English class	SEI English	Spring 2016	SEI English	Fall 2015	E
T2_Students_30_GrK	GrK students in T2	GrK students in BAU schools	Self-esteem in English class	SEI English	Spring 2017	SEI English	Fall 2016	E
T2_Students_31_Gr3	Gr3 students in T2	Gr3 students in BAU schools	Self-esteem	SEI Spanish	Spring 2014	SEI Spanish	Fall 2013	E
T2_Students_32_Gr2	Gr2 students in T2	Gr2 students in BAU schools	Self-esteem	SEI Spanish	Spring 2015	SEI Spanish	Fall 2014	E
T2_Students_33_Gr1	Gr1 students in T2	Gr1 students in BAU schools	Self-esteem	SEI Spanish	Spring 2016	SEI Spanish	Fall 2015	C
T2_Students_34_GrK	GrK students in T2	GrK students in BAU schools	Self-esteem	SEI Spanish	Spring 2017	SEI Spanish	Fall 2016	E

NOTES—1. The research design for all domains was RCT with school assignment. 2. In all cases, exposure to the treatment was one school year. 3. The unit of observation for all domains was the student. 4. The student sample included all study participants who had non-missing pretest and posttest scores. 5. The scale for all measures was continuous; note that TELPAS is measured on a four-point scale. 6. ELD=English language development.

Table C3. *Master list of teacher contrasts.*

Contrast ID	Treatment Group	Control Group	Domain	Outcome Measure	Outcome Measure Timing	Pretest Measure	Pretest Measure Timing
T1 versus BAU							
T1_Teachers_1_Gr1	Gr1 teachers in T1	Gr1 teachers in BAU schools	Use of ESL strategies	TOR	Spring 2016	TOR	Fall 2015
T1_Teachers_2_GrK	GrK teachers in T1	GrK teachers in BAU schools	Use of ESL strategies	TOR	Spring 2017	TOR	Fall 2016
T1_Teachers_3_Gr3	Gr3 teachers in T1	Gr3 teachers in BAU schools	Presenting new content in English	TBOP	Spring 2014	TBOP	Fall 2013
T1_Teachers_4_Gr2	Gr2 teachers in T1	Gr2 teachers in BAU schools	Presenting new content in English	TBOP	Spring 2015	TBOP	Fall 2014
T1_Teachers_5_Gr1	Gr1 teachers in T1	Gr1 teachers in BAU schools	Presenting new content in English	TBOP	Spring 2016	TBOP	Fall 2015
T1_Teachers_6_GrK	GrK teachers in T1	GrK teachers in BAU schools	Presenting new content in English	TBOP	Spring 2017	TBOP	Fall 2016
T2 versus BAU							
T2_Teachers_1_Gr1	Gr1 teachers in T2	Gr1 teachers in BAU schools	Use of ESL strategies	TOR	Spring 2016	TOR	Fall 2015
T2_Teachers_2_GrK	GrK teachers in T2	GrK teachers in BAU schools	Use of ESL strategies	TOR	Spring 2017	TOR	Fall 2016
T2_Teachers_3_Gr3	Gr3 teachers in T2	Gr3 teachers in BAU schools	Presenting new content in English	TBOP	Spring 2014	TBOP	Fall 2013
T2_Teachers_4_Gr2	Gr2 teachers in T2	Gr2 teachers in BAU schools	Presenting new content in English	TBOP	Spring 2015	TBOP	Fall 2014
T2_Teachers_5_Gr1	Gr1 teachers in T2	Gr1 teachers in BAU schools	Presenting new content in English	TBOP	Spring 2016	TBOP	Fall 2015
T2_Teachers_6_GrK	GrK teachers in T2	GrK teachers in BAU schools	Presenting new content in English	TBOP	Spring 2017	TBOP	Fall 2016

NOTES—1. The research design for all domains was RCT with school assignment. 2. In all cases, exposure to the treatment was one school year. 3. The unit of observation for all domains was the teacher. 4. The teacher sample included all study participants who had non-missing pretest and posttest scores. 5. The scale for all measures was continuous; note that TBOP is a proportion. 6. All teacher analyses were exploratory.

Impact tables. Table C4 provides the impact estimates of ELLA-V on student outcomes when T1 was compared with the BAU condition. Table C5 provides the impact estimates of ELLA-V on student outcomes when T2 was compared with the BAU condition. Table C6 provides the impact estimates for teacher outcomes (both T1 v. BAU and T2 v. BAU). Tables C7–C9 list the statistical models that were used to estimate program impacts. All impact estimates were calculated by grade and separately for T1 and T2.

For each outcome measure, one grade level for each treatment (T1 or T2) was selected as the confirmatory contrast; the remaining contrasts were analyzed for exploratory purposes. The confirmatory Contrast IDs are highlighted in purple. Statistically significant and positive effects are highlighted in blue, and negative effects are highlighted in red.

Table C4. *Impact estimates for student outcomes for T1 versus BAU.*

Contrast ID	Outcome Measure	T Sch. N	C Sch. N	T Stu. N	C Stu. N	Unadj. T SD	Unadj. C SD	Pooled SD	Impact Est.	Impact SE	Std. Effect Size	P
T1_Students_1_Gr3	ITBS Science	21	21	745	572	20.99	21.15	21.06	5.63	2.83	**0.27**	0.047
T1_Students_2_Gr3	WMLS Oral	21	21	711	506	15.46	16.50	15.90	-0.01	0.69	0.00	0.983
T1_Students_3_Gr2	WMLS Oral	23	24	684	690	17.05	15.58	16.33	-0.71	0.59	-0.04	0.225
T1_Students_4_Gr1	WMLS Oral	20	21	605	561	21.33	20.97	21.16	-1.92	0.76	**-0.09**	0.011
T1_Students_5_GrK	WMLS Oral	21	24	563	583	24.97	27.02	26.03	4.15	0.92	**0.16**	0.000
T1_Students_6_Gr1	TOPA	20	21	594	560	2.55	2.89	2.72	0.11	0.36	0.04	0.767
T1_Students_7_GrK	TOPA	21	24	541	582	2.73	2.68	2.71	0.40	0.16	**0.15**	0.010
T1_Students_8_Gr3	TELPAS ELD	21	21	706	553	0.73	0.74	0.74	0.04	0.07	0.05	0.560
T1_Students_9_Gr2	TELPAS ELD[a]	23	24	612	599	0.80[a]	0.82[a]	0.81[a]	-0.06[a]	0.09[a]	-0.08[a]	0.471
T1_Students_10_Gr1	TELPAS ELD	20	21	555	532	0.82	0.90	0.86	-0.13	0.11	-0.16	0.202
T1_Students_11_GrK	TELPAS ELD	21	24	584	608	0.80	0.86	0.83	0.08	0.11	0.09	0.497
T1_Students_12_Gr3	STAAR Reading	21	21	639	472	130.37	124.99	128.12	16.74	12.77	0.13	0.190
T1_Students_13_Gr3	WMLS Reading	21	21	650	470	16.28	17.08	16.62	-0.54	0.97	-0.03	0.579
T1_Students_14_Gr2	WMLS Reading[a]	23	24	684	688	15.86	16.38[a]	16.12[a]	1.36[a]	0.76[a]	0.09[a]	0.067
T1_Students_15_Gr1	WMLS Reading	20	21	605	561	18.52	16.89	17.76	-1.90	1.16	-0.11	0.101
T1_Students_16_GrK	WMLS Reading	21	24	534	573	22.81	20.99	21.88	-0.91	1.86	-0.04	0.626
T1_Students_17_Gr3	TELPAS Reading	21	21	706	553	0.97	1.00	0.99	0.04	0.07	0.04	0.999
T1_Students_18_Gr2	TELPAS Reading[a]	23	24	611	599	0.93[a]	0.92[a]	0.92[a]	-0.01[a]	0.09[a]	-0.01[a]	0.922

Contrast ID	Outcome Measure	T Sch. N	C Sch. N	T Stu. N	C Stu. N	Unadj. T SD	Unadj. C SD	Pooled SD	Impact Est.	Impact SE	Std. Effect Size	P
T1_Students_19_Gr1	TELPAS Reading[a]	20	21	550	532	0.91[a]	0.95[a]	0.93[a]	-0.15[a]	0.11[a]	-0.16[a]	0.170
T1_Students_20_GrK	TELPAS Reading	21	24	584	608	0.74	0.75	0.74	0.01	0.12	0.01	0.999
T1_Students_21_Gr2	DIBELS	23	24	686	690	34.38	32.97	33.68	-0.94	1.48	-0.03	0.524
T1_Students_22_Gr1	DIBELS	20	21	605	560	27.16	28.74	27.93	0.40	1.66	0.01	0.811
T1_Students_23_Gr3	TELPAS Writing	21	21	706	553	0.84	0.92	0.88	-0.07	0.08	-0.08	0.385
T1_Students_24_Gr2	TELPAS Writing[a]	23	24	612	598	0.89[a]	0.90[a]	0.90[a]	0.03[a]	0.10[a]	0.03[a]	0.760
T1_Students_25_Gr1	TELPAS Writing[a]	20	21	555	532	0.84[a]	0.88[a]	0.86[a]	-0.14[a]	0.12[a]	-0.16[a]	0.231
T1_Students_26_GrK	TELPAS Writing	21	24	584	608	0.67	0.70	0.68	0.02	0.10	0.03	0.823
T1_Students_27_Gr3	SEI English	21	21	740	566	0.30	0.33	0.32	-0.01	0.02	-0.02	0.761
T1_Students_28_Gr2	SEI English	23	24	686	690	0.28	0.28	0.28	-0.02	0.02	-0.05	0.388
T1_Students_29_Gr1	SEI English	20	21	604	560	0.34	0.36	0.35	0.03	0.02	0.10	0.144
T1_Students_30_GrK	SEI English	21	24	566	594	0.43	0.43	0.43	0.04	0.03	0.09	0.119
T1_Students_31_Gr3	SEI Spanish	21	21	739	566	0.57	0.53	0.55	-0.01	0.04	-0.02	0.744
T1_Students_32_Gr2	SEI Spanish	23	24	686	690	0.42	0.43	0.43	0.01	0.03	0.02	0.723
T1_Students_33_Gr1	SEI Spanish	20	21	604	560	0.40	0.43	0.41	0.05	0.02	**0.11**	0.034
T1_Students_34_GrK	SEI Spanish	21	24	566	594	0.44	0.44	0.44	-0.03	0.03	-0.06	0.391

NOTES—1. [a] indicates that the baseline mean difference between the treatment and comparison groups was >0.25 before using propensity score weighting. 2. The degrees of freedom for all models were infinity.

Table C5. *Impact estimates for student outcomes for T2 versus BAU.*

Contrast ID	Outcome Measure	T Sch. N	C Sch. N	T Stu. N	C Stu. N	Unadj. T SD	Unadj. C SD	Pooled SD	Impact Est.	Impact SE	Std. Effect Size	P
T2_Students_1_Gr3	ITBS Science	19	21	614	572	18.33	21.15	19.74	-0.07	2.29	0.00	0.975
T2_Students_2_Gr3	WMLS Oral	19	21	573	506	16.14	16.50	16.31	-0.28	0.72	-0.02	0.697
T2_Students_3_Gr2	WMLS Oral	21	24	619	690	14.99	15.58	15.30	-0.01	0.53	0.00	0.983
T2_Students_4_Gr1	WMLS Oral	20	21	562	561	21.04	20.97	21.01	2.44	0.79	**0.12**	0.002
T2_Students_5_GrK	WMLS Oral	22	24	609	583	24.30	27.02	25.67	2.43	0.90	**0.09**	0.007
T2_Students_6_Gr1	TOPA	20	21	557	560	2.86	2.89	2.88	-0.08	0.33	-0.03	0.818
T2_Students_7_GrK	TOPA	22	24	603	582	2.68	2.68	2.68	-0.06	0.16	-0.02	0.722
T2_Students_8_Gr3	TELPAS ELD	19	21	577	553	0.70	0.74	0.72	0.10	0.07	0.14	0.141
T2_Students_9_Gr2	TELPAS ELD	21	24	553	599	0.74	0.81	0.78	0.07	0.07	0.09	0.290
T2_Students_10_Gr1	TELPAS ELD[a]	20	21	515	532	0.71[a]	0.87[a]	0.80[a]	-0.09[a]	0.09[a]	-0.11[a]	0.338
T2_Students_11_GrK	TELPAS ELD	22	24	641	608	0.85	0.86	0.86	0.02	0.11	0.02	0.861
T2_Students_12_Gr3	STAAR Reading	19	21	530	472	121.66	124.99	123.24	-10.26	13.08	-0.08	0.433
T2_Students_13_Gr3	WMLS Reading	19	21	533	470	17.95	17.08	17.55	-0.85	1.00	-0.05	0.393
T2_Students_14_Gr2	WMLS Reading	21	24	619	688	15.32	15.46	15.40	0.25	0.53	0.02	0.642
T2_Students_15_Gr1	WMLS Reading	20	21	562	561	18.05	16.89	17.48	-0.78	0.95	-0.04	0.409
T2_Students_16_GrK	WMLS Reading	22	24	598	573	20.55	20.99	20.77	0.58	1.51	0.03	0.703
T2_Students_17_Gr3	TELPAS Reading	19	21	577	553	0.96	1.00	0.98	0.06	0.06	0.06	0.759
T2_Students_18_Gr2	TELPAS Reading	21	24	553	599	0.92	0.92	0.92	0.01	0.10	0.01	0.999

Contrast ID	Outcome Measure	T Sch. N	C Sch. N	T Stu. N	C Stu. N	Unadj. T SD	Unadj. C SD	Pooled SD	Impact Est.	Impact SE	Std. Effect Size	P
T2_Students_19_Gr1	TELPAS Reading[a]	20	21	514	532	0.79[a]	0.95[a]	0.88[a]	-0.19[a]	0.10[a]	-0.21[a]	0.059
T2_Students_20_GrK	TELPAS Reading	22	24	641	608	0.69	0.75	0.72	-0.13	0.11	-0.17	0.508
T2_Students_21_Gr2	DIBELS	21	24	619	690	34.58	32.97	33.74	2.43	1.56	0.07	0.120
T2_Students_22_Gr1	DIBELS	20	21	562	560	29.41	28.74	29.07	-0.14	1.43	0.00	0.920
T2_Students_23_Gr3	TELPAS Writing	19	21	577	553	0.93	0.92	0.93	0.12	0.08	0.13	0.111
T2_Students_24_Gr2	TELPAS Writing	21	24	553	598	0.85	0.90	0.88	0.08	0.09	0.09	0.390
T2_Students_25_Gr1	TELPAS Writing[a]	20	21	515	532	0.73[a]	0.88[a]	0.81[a]	-0.19[a]	0.10[a]	-0.24[a]	0.059
T2_Students_26_GrK	TELPAS Writing	22	24	641	608	0.66	0.70	0.68	-0.09	0.10	-0.13	0.353
T2_Students_27_Gr3	SEI English	19	21	609	566	0.34	0.33	0.34	-0.02	0.02	-0.07	0.301
T2_Students_28_Gr2	SEI English	21	24	619	690	0.26	0.28	0.27	0.02	0.02	0.09	0.147
T2_Students_29_Gr1	SEI English	20	21	562	560	0.34	0.36	0.35	0.02	0.02	0.05	0.373
T2_Students_30_GrK	SEI English	22	24	616	594	0.42	0.43	0.42	0.03	0.03	0.08	0.182
T2_Students_31_Gr3	SEI Spanish	19	21	609	566	0.55	0.53	0.54	-0.01	0.03	-0.03	0.637
T2_Students_32_Gr2	SEI Spanish	21	24	619	690	0.42	0.43	0.42	0.00	0.03	-0.01	0.883
T2_Students_33_Gr1	SEI Spanish	20	21	561	560	0.42	0.43	0.42	0.03	0.02	0.08	0.108
T2_Students_34_GrK	SEI Spanish	22	24	616	594	0.46	0.44	0.45	-0.02	0.03	-0.03	0.622

NOTES—1. [a] indicates that the baseline mean difference between the treatment and comparison groups was >0.25 before using propensity score weighting. 2. The degrees of freedom for all models were infinity.

Table C6. *Impact estimates for teacher outcomes.*

Contrast ID	Outcome Measure	T Sch. N	C Sch. N	T Tch. N	C Tch. N	Unadj. T SD	Unadj. C SD	Pooled SD	Impact Est.	Impact SE	Std. Effect Size	P
T1 versus BAU												
T1_Teachers_1_Gr1	TOR	20	21	39	39	0.77	0.68	0.73	0.70	0.20	**0.97**	0.000
T1_Teachers_2_GrK	TOR	21	24	41	44	0.65	0.68	0.67	0.43	0.14	**0.64**	0.002
T1_Teachers_3_Gr3	TBOP	20	21	37	39	0.17	0.32	0.26	0.21	0.06	**0.81**	0.000
T1_Teachers_4_Gr2	TBOP	24	24	45	46	0.13	0.33	0.25	0.24	0.06	**0.96**	0.000
T1_Teachers_5_Gr1	TBOP	20	21	39	39	0.14	0.16	0.15	0.09	0.03	**0.57**	0.004
T1_Teachers_6_GrK	TBOP	21	24	41	41	0.19	0.24	0.22	0.04	0.06	0.18	0.513
T2 versus BAU												
T2_Teachers_1_Gr1	TOR	20	21	38	39	0.58	0.68	0.63	1.23	0.16	**1.94**	0.000
T2_Teachers_2_GrK	TOR	22	24	41	44	0.72	0.68	0.70	0.87	0.18	**1.23**	0.000
T2_Teachers_3_Gr3	TBOP	19	21	36	39	0.24	0.32	0.28	0.21	0.04	**0.73**	0.000
T2_Teachers_4_Gr2	TBOP	22	24	41	46	0.17	0.33	0.27	0.22	0.06	**0.84**	0.000
T2_Teachers_5_Gr1	TBOP	20	21	38	39	0.10	0.16	0.14	0.11	0.03	**0.80**	0.000
T2_Teachers_6_GrK	TBOP	21	24	40	41	0.20	0.24	0.22	0.05	0.05	0.21	0.338

NOTES—1. All measures failed baseline equivalence and were adjusted using propensity score weighting. 2. The degrees of freedom for all models were infinity.

Table C7. *Statistical models used to estimate program impacts on student outcomes for T1 versus BAU.*

Contrast ID	Outcome Measure	Model
T1_Students_1_Gr3	ITBS Science	mixed itbs1 t1 grand_* if t2!=1 & grade==3 || schid: ;
T1_Students_2_Gr3	WMLS Oral	mixed wmls_oral1 t1 grand_* if t2!=1 & grade==3 || schid: ;
T1_Students_3_Gr2	WMLS Oral	mixed wmls_oral1 t1 grand_* if t2!=1 & grade==2 || schid: ;
T1_Students_4_Gr1	WMLS Oral	mixed wmls_oral1 t1 grand_* if t2!=1 & grade==1 || schid: ;
T1_Students_5_GrK	WMLS Oral	mixed wmls_oral1 t1 grand_* if t2!=1 & grade==0 || schid: ;
T1_Students_6_Gr1	TOPA	mixed topa_pa1 t1 grand_* if t2!=1 & grade==1 || schid: ;
T1_Students_7_GrK	TOPA	mixed topa_pa1 t1 grand_* if t2!=1 & grade==0 || schid: ;
T1_Students_8_Gr3	TELPAS ELD	mixed telpas_eld1 t1 grand_* if t2!=1 & grade==3 || schid: ;
T1_Students_9_Gr2	TELPAS ELD	mixed telpas_eld1 t1 grand_* if t2!=1 & grade==2 [pweight= stuwgt] || schid: ; pweight(schwgt)
T1_Students_10_Gr1	TELPAS ELD	mixed telpas_eld1 t1 grand_* if t2!=1 & grade==1 || schid: ;
T1_Students_11_GrK	TELPAS ELD	mixed telpas_eld1 t1 grand_* if t2!=1 & grade==0 || schid: ;
T1_Students_12_Gr3	STAAR Read	mixed staar_read t1 grand_* if t2!=1 & grade==3 || schid: ;
T1_Students_13_Gr3	WMLS Read	mixed wmls_read1 t1 grand_* if t2!=1 & grade==3 || schid: ;
T1_Students_14_Gr2	WMLS Read	mixed wmls_read1 t1 grand_* if t2!=1 & grade==2 [pweight= stuwgt] || schid: ; pweight(schwgt)
T1_Students_15_Gr1	WMLS Read	mixed wmls_read1 t1 grand_* if t2!=1 & grade==1 || schid: ;
T1_Students_16_GrK	WMLS Read	mixed wmls_read1 t1 grand_* if t2!=1 & grade==0 || schid: ;
T1_Students_17_Gr3	TELPAS Read	mixed telpas_read1 t1 grand_* if t2!=1 & grade==3 || schid: ;
T1_Students_18_Gr2	TELPAS Read	mixed telpas_read1 t1 grand_* if t2!=1 & grade==2 [pweight= stuwgt] || schid: ; pweight(schwgt)
T1_Students_19_Gr1	TELPAS Read	mixed telpas_read1 t1 grand_* if t2!=1 & grade==1 [pweight= stuwgt] || schid: ; pweight(schwgt)
T1_Students_20_GrK	TELPAS Read	mixed telpas_read1 t1 grand_* if t2!=1 & grade==0 || schid: ;
T1_Students_21_Gr2	DIBELS	mixed dibels_tot1 t1 grand_* if t2!=1 & grade==2 || schid: ;
T1_Students_22_Gr1	DIBELS	mixed dibels_tot1 t1 grand_* if t2!=1 & grade==1 || schid: ;
T1_Students_23_Gr3	TELPAS Writing	mixed telpas_write1 t1 grand_* if t2!=1 & grade==3 || schid:
T1_Students_24_Gr2	TELPAS Writing	mixed telpas_write1 t1 grand_* if t2!=1 & grade==2 [pweight= stuwgt] || schid: ; pweight(schwgt)
T1_Students_25_Gr1	TELPAS Writing	mixed telpas_write1 t1 grand_* if t2!=1 & grade==1 [pweight= stuwgt] || schid: ; pweight(schwgt)
T1_Students_26_GrK	TELPAS Writing	mixed telpas_write1 t1 grand_* if t2!=1 & grade==0 || schid: ;
T1_Students_27_Gr3	SEI English	mixed se1_english t1 grand_* if t2!=1 & grade==3 || schid: ;
T1_Students_28_Gr2	SEI English	mixed se1_english t1 grand_* if t2!=1 & grade==2 || schid: ;
T1_Students_29_Gr1	SEI English	mixed se1_english t1 grand_* if t2!=1 & grade==1 || schid: ;
T1_Students_30_GrK	SEI English	mixed se1_english t1 grand_* if t2!=1 & grade==0 || schid: ;
T1_Students_31_Gr3	SEI Spanish	mixed se1_spanish t1 grand_* if t2!=1 & grade==3 || schid: ;
T1_Students_32_Gr2	SEI Spanish	mixed se1_spanish t1 grand_* if t2!=1 & grade==2 || schid: ;
T1_Students_33_Gr1	SEI Spanish	mixed se1_spanish t1 grand_* if t2!=1 & grade==1 || schid: ;
T1_Students_34_GrK	SEI Spanish	mixed se1_spanish t1 grand_* if t2!=1 & grade==0 || schid: ;

NOTES—1. Stata version 15.0 was used to estimate all models. 2. Grand_* indicates that all covariates (e.g., the pretest, school-level TELPAS rating of beginning, school-level TELPAS rating of advanced, district dummies, and school-level percentage EL) were included in the model, and all were grand-mean centered.

Table C8. *Statistical models used to estimate program impacts on student outcomes for T2 versus BAU.*

Contrast ID	Outcome Measure	Model
T2_Students_1_Gr3	ITBS Science	mixed itbs1 t2 grand_* if t1!=1 & grade==3 || schid: ;
T2_Students_2_Gr3	WMLS Oral	mixed wmls_oral1 t2 grand_* if t1!=1 & grade==3 || schid: ;
T2_Students_3_Gr2	WMLS Oral	mixed wmls_oral1 t2 grand_* if t1!=1 & grade==2 || schid: ;
T2_Students_4_Gr1	WMLS Oral	mixed wmls_oral1 t2 grand_* if t1!=1 & grade==1 || schid: ;
T2_Students_5_GrK	WMLS Oral	mixed wmls_oral1 t2 grand_* if t1!=1 & grade==0 || schid: ;
T2_Students_6_Gr1	TOPA	mixed topa_pa1 t2 grand_* if t1!=1 & grade==1 || schid: ;
T2_Students_7_GrK	TOPA	mixed topa_pa1 t2 grand_* if t1!=1 & grade==0 || schid: ;
T2_Students_8_Gr3	TELPAS ELD	mixed telpas_eld1 t2 grand_* if t1!=1 & grade==3 || schid: ;
T2_Students_9_Gr2	TELPAS ELD	mixed telpas_eld1 t2 grand_* if t1!=1 & grade==2 || schid: ;
T2_Students_10_Gr1	TELPAS ELD	mixed telpas_eld1 t2 grand_* if t1!=1 & grade==1 [pweight= stuwgt] || schid: ; pweight(schwgt)
T2_Students_11_GrK	TELPAS ELD	mixed telpas_eld1 t2 grand_* if t1!=1 & grade==0 || schid: ;
T2_Students_12_Gr3	STAAR Read	mixed staar_read t2 grand_* if t1!=1 & grade==3 || schid: ;
T2_Students_13_Gr3	WMLS Read	mixed wmls_read1 t2 grand_* if t1!=1 & grade==3 || schid: ;
T2_Students_14_Gr2	WMLS Read	mixed wmls_read1 t2 grand_* if t1!=1 & grade==2 || schid: ;
T2_Students_15_Gr1	WMLS Read	mixed wmls_read1 t2 grand_* if t1!=1 & grade==1 || schid: ;
T2_Students_16_GrK	WMLS Read	mixed wmls_read1 t2 grand_* if t1!=1 & grade==0 || schid: ;
T2_Students_17_Gr3	TELPAS Read	mixed telpas_read1 t2 grand_* if t1!=1 & grade==3 || schid: ;
T2_Students_18_Gr2	TELPAS Read	mixed telpas_read1 t2 grand_* if t1!=1 & grade==2 || schid: ;
T2_Students_19_Gr1	TELPAS Read	mixed telpas_read1 t2 grand_* if t1!=1 & grade==1 [pweight= stuwgt] || schid: ; pweight(schwgt)
T2_Students_20_GrK	TELPAS Read	mixed telpas_read1 t2 grand_* if t1!=1 & grade==0 || schid: ;
T2_Students_21_Gr2	DIBELS	mixed dibels_tot1 t2 grand_* if t1!=1 & grade==2 || schid: ;
T2_Students_22_Gr1	DIBELS	mixed dibels_tot1 t2 grand_* if t1!=1 & grade==1 || schid: ;
T2_Students_23_Gr3	TELPAS Writing	mixed telpas_write1 t2 grand_* if t1!=1 & grade==3 || schid:
T2_Students_24_Gr2	TELPAS Writing	mixed telpas_write1 t2 grand_* if t1!=1 & grade==2 || schid:
T2_Students_25_Gr1	TELPAS Writing	mixed telpas_write1 t2 grand_* if t1!=1 & grade==1 [pweight= stuwgt] || schid: ; pweight(schwgt)
T2_Students_26_GrK	TELPAS Writing	mixed telpas_write1 t2 grand_* if t1!=1 & grade==0 || schid: ;
T2_Students_27_Gr3	SEI English	mixed se1_english t2 grand_* if t1!=1 & grade==3 || schid: ;
T2_Students_28_Gr2	SEI English	mixed se1_english t2 grand_* if t1!=1 & grade==2 || schid: ;
T2_Students_29_Gr1	SEI English	mixed se1_english t2 grand_* if t1!=1 & grade==1 || schid: ;
T2_Students_30_GrK	SEI English	mixed se1_english t2 grand_* if t1!=1 & grade==0 || schid: ;
T2_Students_31_Gr3	SEI Spanish	mixed se1_spanish t2 grand_* if t1!=1 & grade==3 || schid: ;
T2_Students_32_Gr2	SEI Spanish	mixed se1_spanish t2 grand_* if t1!=1 & grade==2 || schid: ;
T2_Students_33_Gr1	SEI Spanish	mixed se1_spanish t2 grand_* if t1!=1 & grade==1 || schid: ;
T2_Students_34_GrK	SEI Spanish	mixed se1_spanish t2 grand_* if t1!=1 & grade==0 || schid: ;

NOTES—1. Stata version 15.0 was used to estimate all models. 2. Grand_* indicates that all covariates (e.g., the pretest, school-level TELPAS rating of beginning, school-level TELPAS rating of advanced, district dummies, and school-level percentage EL) were included in the model, and all were grand-mean centered.

Table C9. *Statistical models to estimate program impacts on teacher outcomes.*

Contrast ID	Outcome Measure	Model
T1 versus BAU		
T1_Teachers_1_Gr1	TOR	mixed tor_irt1 t1 grand* if t2!=1 & grade==1 [pweight= tchwgt] || schid: ; pweight(schwgt)
T1_Teachers_2_GrK	TOR	mixed tor_irt1 t1 grand* if t2!=1 & grade==0 [pweight= tchwgt] || schid: ; pweight(schwgt)
T1_Teachers_3_Gr3	TBOP	mixed dcifl2_prop3 t1 grand* if t2!=1 & grade==3 [pweight=tchwgt] || schid: ; pweight(schwgt)
T1_Teachers_4_Gr2	TBOP	mixed dcifl2_prop3 t1 grand* if t2!=1 & grade==2 [pweight= tchwgt] || schid: ; pweight(schwgt)
T1_Teachers_5_Gr1	TBOP	mixed dcifl2_prop3 t1 grand* if t2!=1 & grade==1 [pweight= tchwgt] || schid: ; pweight(schwgt)
T1_Teachers_6_GrK	TBOP	mixed dcifl2_prop3 t1 grand* if t2!=1 & grade==0 [pweight= tchwgt] || schid: ; pweight(schwgt)
T2 versus BAU		
T2_Teachers_1_Gr1	TOR	mixed tor_irt1 t2 grand* if t1!=1 & grade==1 [pweight= tchwgt] || schid: ; pweight(schwgt)
T2_Teachers_2_GrK	TOR	mixed tor_irt1 t2 grand* if t1!=1 & grade==0 [pweight= tchwgt] || schid: ; pweight(schwgt)
T2_Teachers_3_Gr3	TBOP	mixed dcifl2_prop3 t2 grand* if t1!=1 & grade==3 [pweight= tchwgt] || schid: ; pweight(schwgt)
T2_Teachers_4_Gr2	TBOP	mixed dcifl2_prop3 t2 grand* if t1!=1 & grade==2 [pweight= tchwgt] || schid: ; pweight(schwgt)
T2_Teachers_5_Gr1	TBOP	mixed dcifl2_prop3 t2 grand* if t1!=1 & grade==1 [pweight= tchwgt] || schid: ; pweight(schwgt)
T2_Teachers_6_GrK	TBOP	mixed dcifl2_prop3 t2 grand* if t1!=1 & grade==0 [pweight= tchwgt] || schid: ; pweight(schwgt)

NOTES—1. Stata version 15.0 was used to estimate all models. 2. Grand_* indicates that all covariates (e.g., the pretest, school-level TELPAS rating of beginning, school-level TELPAS rating of advanced, district dummies, and school-level percentage EL) were included in the model, and all were grand-mean centered. 3. The propensity score weights were different for the TBOP and TOR outcomes.

Cluster attrition tables. The following tables provide the cluster (school) attrition rates. Table C10 provides the cluster attrition for the student analyses for T1 versus BAU, and Table C11 provides the cluster attrition for the student analyses for T2 versus BAU. Table C12 provides the cluster attrition for the teacher analyses for T1 versus BAU and for T2 versus BAU. The cluster attrition rates (overall and differential) for all outcomes were acceptable according to the WWC (2017) standards.

Several schools attrited from the study. One T2 school that was randomly assigned did not participate in the study in any year. One T2 school that was randomly assigned prior to implementation in grade 3 did not begin implementation until the following year, in grade 2. All other attrited schools either declined to participate in the data collection or outcomes were not collected for these schools because one of the three schools in the original matched cluster attrited from the study.

Table C10. *Cluster attrition for student outcomes for T1 versus BAU.*

Contrast ID	Outcome Measure	C Sch. N	T1 Sch. N	N Sch. Randomized to C	N Sch. Randomized to T1	Attrited C Sch.	Attrited T1 Sch.	Overall Sch. Attrition Rate (%)	Diff. Sch. Attrition Rate (%)
T1_Students_1_Gr3	ITBS Science	21	21	21	21	0	0	0.00	0.00
T1_Students_2_Gr3	WMLS Oral	21	21	21	21	0	0	0.00	0.00
T1_Students_3_Gr2	WMLS Oral	24	23	25	24	1	1	4.08	0.17
T1_Students_4_Gr1	WMLS Oral	21	20	25	24	4	4	16.33	0.67
T1_Students_5_GrK	WMLS Oral	24	21	27	26	3	5	15.09	8.12
T1_Students_6_Gr1	TOPA	21	20	25	24	4	4	16.33	0.67
T1_Students_7_GrK	TOPA	24	21	27	26	3	5	15.09	8.12
T1_Students_8_Gr3	TELPAS ELD	21	21	21	21	0	0	0.00	0.00
T1_Students_9_Gr2	TELPAS ELD	24	23	25	24	1	1	4.08	0.17
T1_Students_10_Gr1	TELPAS ELD	21	20	25	24	4	4	16.33	0.67
T1_Students_11_GrK	TELPAS ELD	24	21	27	26	3	5	15.09	8.12
T1_Students_12_Gr3	STAAR Reading	21	21	21	21	0	0	0.00	0.00
T1_Students_13_Gr3	WMLS Reading	21	21	21	21	0	0	0.00	0.00
T1_Students_14_Gr2	WMLS Reading	24	23	25	24	1	1	4.08	0.17
T1_Students_15_Gr1	WMLS Reading	21	20	25	24	4	4	16.33	0.67
T1_Students_16_GrK	WMLS Reading	24	21	27	26	3	5	15.09	8.12
T1_Students_17_Gr3	TELPAS Reading	21	21	21	21	0	0	0.00	0.00
T1_Students_18_Gr2	TELPAS Reading	24	23	25	24	1	1	4.08	0.17

Contrast ID	Outcome Measure	C Sch. N	T1 Sch. N	N Sch. Randomized to C	N Sch. Randomized to T1	Attrited C Sch.	Attrited T1 Sch.	Overall Sch. Attrition Rate (%)	Diff. Sch. Attrition Rate (%)
T1_Students_19_Gr1	TELPAS Reading	21	20	25	24	4	4	16.33	0.67
T1_Students_20_GrK	TELPAS Reading	24	21	27	26	3	5	15.09	8.12
T1_Students_21_Gr2	DIBELS	24	23	25	24	1	1	4.08	0.17
T1_Students_22_Gr1	DIBELS	21	20	25	24	4	4	16.33	0.67
T1_Students_23_Gr3	TELPAS Writing	21	21	21	21	0	0	0.00	0.00
T1_Students_24_Gr2	TELPAS Writing	24	23	25	24	1	1	4.08	0.17
T1_Students_25_Gr1	TELPAS Writing	21	20	25	24	4	4	16.33	0.67
T1_Students_26_GrK	TELPAS Writing	24	21	27	26	3	5	15.09	8.12
T1_Students_27_Gr3	SEI English	21	21	21	21	0	0	0.00	0.00
T1_Students_28_Gr2	SEI English	24	23	25	24	1	1	4.08	0.17
T1_Students_29_Gr1	SEI English	21	20	25	24	4	4	16.33	0.67
T1_Students_30_GrK	SEI English	24	21	27	26	3	5	15.09	8.12
T1_Students_31_Gr3	SEI Spanish	21	21	21	21	0	0	0.00	0.00
T1_Students_32_Gr2	SEI Spanish	24	23	25	24	1	1	4.08	0.17
T1_Students_33_Gr1	SEI Spanish	21	20	25	24	4	4	16.33	0.67
T1_Students_34_GrK	SEI Spanish	24	21	27	26	3	5	15.09	8.12

Table C11. *Cluster attrition for student outcomes for T2 versus BAU.*

Contrast ID	Outcome Measure	C Sch. N	T2 Sch. N	N Sch. Randomized to C	N Sch. Randomized to T2	Attrited C Sch.	Attrited T2 Sch.	Overall Sch. Attrition Rate (%)	Diff. Sch. Attrition Rate (%)
T2_Students_1_Gr3	ITBS Science	21	19	21	21	0	2	4.76	9.52
T2_Students_2_Gr3	WMLS Oral	21	19	21	21	0	2	4.76	9.52
T2_Students_3_Gr2	WMLS Oral	24	21	25	24	1	3	8.16	8.50
T2_Students_4_Gr1	WMLS Oral	21	20	25	24	4	4	16.33	0.67
T2_Students_5_GrK	WMLS Oral	24	22	27	26	3	4	13.21	4.27
T2_Students_6_Gr1	TOPA	21	20	25	24	4	4	16.33	0.67
T2_Students_7_GrK	TOPA	24	22	27	26	3	4	13.21	4.27
T2_Students_8_Gr3	TELPAS ELD	21	19	21	21	0	2	4.76	9.52
T2_Students_9_Gr2	TELPAS ELD	24	21	25	24	1	3	8.16	8.50
T2_Students_10_Gr1	TELPAS ELD	21	20	25	24	4	4	16.33	0.67
T2_Students_11_GrK	TELPAS ELD	24	22	27	26	3	4	13.21	4.27
T2_Students_12_Gr3	STAAR Reading	21	19	21	21	0	2	4.76	9.52
T2_Students_13_Gr3	WMLS Reading	21	19	21	21	0	2	4.76	9.52
T2_Students_14_Gr2	WMLS Reading	24	21	25	24	1	3	8.16	8.50
T2_Students_15_Gr1	WMLS Reading	21	20	25	24	4	4	16.33	0.67
T2_Students_16_GrK	WMLS Reading	24	22	27	26	3	4	13.21	4.27
T2_Students_17_Gr3	TELPAS Reading	21	19	21	21	0	2	4.76	9.52
T2_Students_18_Gr2	TELPAS Reading	24	21	25	24	1	3	8.16	8.50

Contrast ID	Outcome Measure	C Sch. N	T2 Sch. N	N Sch. Randomized to C	N Sch. Randomized to T2	Attrited C Sch.	Attrited T2 Sch.	Overall Sch. Attrition Rate (%)	Diff. Sch. Attrition Rate (%)
T2_Students_19_Gr1	TELPAS Reading	21	20	25	24	4	4	16.33	0.67
T2_Students_20_GrK	TELPAS Reading	24	22	27	26	3	4	13.21	4.27
T2_Students_21_Gr2	DIBELS	24	21	25	24	1	3	8.16	8.50
T2_Students_22_Gr1	DIBELS	21	20	25	24	4	4	16.33	0.67
T2_Students_23_Gr3	TELPAS Writing	21	19	21	21	0	2	4.76	9.52
T2_Students_24_Gr2	TELPAS Writing	24	21	25	24	1	3	8.16	8.50
T2_Students_25_Gr1	TELPAS Writing	21	20	25	24	4	4	16.33	0.67
T2_Students_26_GrK	TELPAS Writing	24	22	27	26	3	4	13.21	4.27
T2_Students_27_Gr3	SEI English	21	19	21	21	0	2	4.76	9.52
T2_Students_28_Gr2	SEI English	24	21	25	24	1	3	8.16	8.50
T2_Students_29_Gr1	SEI English	21	20	25	24	4	4	16.33	0.67
T2_Students_30_GrK	SEI English	24	22	27	26	3	4	13.21	4.27
T2_Students_31_Gr3	SEI Spanish	21	19	21	21	0	2	4.76	9.52
T2_Students_32_Gr2	SEI Spanish	24	21	25	24	1	3	8.16	8.50
T2_Students_33_Gr1	SEI Spanish	21	20	25	24	4	4	16.33	0.67
T2_Students_34_GrK	SEI Spanish	24	22	27	26	3	4	13.21	4.27

Table C12. *Cluster attrition for teacher outcomes.*

Contrast ID	Outcome Measure	C Sch. N	T Sch. N	N Sch. Randomized to C	N Sch. Randomized to T	Attrited C Sch.	Attrited T Sch.	Overall Sch. Attrition Rate (%)	Diff. Sch. Attrition Rate (%)
T1 versus BAU									
T1_Teachers_1_Gr1	TOR	21	20	25	24	4	4	16.33	0.67
T1_Teachers_2_GrK	TOR	24	21	27	26	3	5	15.09	8.12
T1_Teachers_3_Gr3	TBOP	21	20	21	21	0	1	2.38	4.76
T1_Teachers_4_Gr2	TBOP	24	24	25	24	1	0	2.04	4.00
T1_Teachers_5_Gr1	TBOP	21	20	25	24	4	4	16.33	0.67
T1_Teachers_6_GrK	TBOP	24	21	27	26	3	5	15.09	8.12
T2 versus BAU									
T2_Teachers_1_Gr1	TOR	21	20	25	24	4	4	16.33	0.67
T2_Teachers_2_GrK	TOR	24	22	27	26	3	4	13.21	4.27
T2_Teachers_3_Gr3	TBOP	21	19	21	21	0	2	4.76	9.52
T2_Teachers_4_Gr2	TBOP	24	22	25	24	1	2	6.12	4.33
T2_Teachers_5_Gr1	TBOP	21	20	25	24	4	4	16.33	0.67
T2_Teachers_6_GrK	TBOP	24	21	27	26	3	5	15.09	8.12

Baseline equivalence tables. For all analytic samples, baseline equivalence on pretests was assessed using the same analytic model to estimate program impacts, except without the covariates. In other words, the baseline mean difference was estimated using an HLM model with the pretest as the dependent variable and the treatment indicator as the independent variable. Table C13 shows the baseline equivalence for the student outcomes for T1 versus BAU, and Table C14 shows the baseline equivalence for the student outcomes for T2 versus BAU. Table C15 shows the baseline equivalence for the teacher outcomes.

Baseline equivalence was initially not established in a few cases for student outcomes and not established in all cases for teacher outcomes. In these cases, propensity score weighting was applied to the models used to estimate the baseline mean difference (as well as the models used to estimate impacts); consequently, all baseline differences between treatment and comparison groups were <0.25 standard deviations. Note that all statistical models estimating program effects included the pretest as a covariate.

Table C13. *Baseline equivalence for student outcomes for T1 versus BAU.*

Contrast ID	Pretest Measure	T1 Stu. N	C Stu. N	Unadj T SD at Pretest	Unadj C SD at Pretest	Pooled SD for T and C	C Mean at Pretest	T/C Diff. at Pretest	Std. T/C Diff. at Pretest
T1_Students_1_Gr3	ITBS Science	745	572	13.28	15.08	14.09	171.30	-2.00	-0.14
T1_Students_2_Gr3	WMLS Oral	711	506	17.26	17.91	17.54	78.57	-1.05	-0.06
T1_Students_3_Gr2	WMLS Oral	684	690	20.86	20.60	20.73	75.34	-4.16	-0.20
T1_Students_4_Gr1	WMLS Oral	605	561	25.10	24.90	25.00	64.07	-2.90	0.16
T1_Students_5_GrK	WMLS Oral	563	583	32.54	32.26	32.40	53.13	5.08	0.16
T1_Students_6_Gr1	TOPA	594	560	2.02	2.21	2.11	5.96	-0.43	-0.20
T1_Students_7_GrK	TOPA	541	582	2.51	2.53	2.52	7.58	0.13	0.05
T1_Students_8_Gr3	TELPAS ELD	706	553	0.86	0.91	0.88	2.78	-0.04	-0.05
T1_Students_9_Gr2	TELPAS ELD	612	599	0.80	0.79	0.79	2.26	0.15	0.19
T1_Students_10_Gr1	TELPAS ELD[a]	555	532	0.74[a]	0.73[a]	0.74[a]	1.71[a]	-0.13[a]	-0.18[a]
T1_Students_12_Gr3	WMLS Reading[b]	639	472	16.55	17.18	16.82	96.24	-0.57	-0.03
T1_Students_13_Gr3	WMLS Reading	650	470	16.99	17.35	17.14	95.87	-0.60	-0.03
T1_Students_14_Gr2	WMLS Reading[a]	684	688	17.71[a]	17.36[a]	17.54[a]	98.22[a]	2.10[a]	0.12[a]
T1_Students_15_Gr1	WMLS Reading	605	561	21.23	21.18	21.20	90.91	-3.54	-0.17
T1_Students_16_GrK	WMLS Reading	534	573	22.52	22.00	22.26	83.15	1.46	0.07
T1_Students_17_Gr3	TELPAS Reading	706	553	0.98	0.95	0.97	2.65	0.02	0.02
T1_Students_18_Gr2	TELPAS Reading[a]	611	599	0.87[a]	0.87[a]	0.87[a]	1.99[a]	0.11[a]	0.12[a]
T1_Students_19_Gr1	TELPAS Reading[a]	550	532	0.58[a]	0.63[a]	0.60[a]	1.29[a]	0.05[a]	0.09[a]
T1_Students_21_Gr2	DIBELS	686	690	28.62	27.40	28.02	53.25	-6.37	-0.23
T1_Students_22_Gr1	DIBELS	605	560	17.25	19.25	18.24	18.63	-3.45	-0.19
T1_Students_23_Gr3	TELPAS Writing	706	553	0.86	0.92	0.89	2.43	-0.04	-0.05

Contrast ID	Pretest Measure	T1 Stu. N	C Stu. N	Unadj T SD at Pretest	Unadj C SD at Pretest	Pooled SD for T and C	C Mean at Pretest	T/C Diff. at Pretest	Std. T/C Diff. at Pretest
T1_Students_24_Gr2	TELPAS Writing[a]	612	598	0.81[a]	0.79[a]	0.80[a]	1.84[a]	0.19[a]	0.24[a]
T1_Students_25_Gr1	TELPAS Writing[a]	555	532	0.53[a]	0.57[a]	0.55[a]	1.26[a]	0.05[a]	0.10[a]
T1_Students_11_GrK T1_Students_20_GrK T1_Students_26_GrK	TVIP[c]	584	608	18.55	19.59	19.08	87.93	0.90	0.05
T1_Students_27_Gr3	SEI English	740	608	0.35	0.38	0.36	1.58	0.00	0.01
T1_Students_28_Gr2	SEI English	686	608	0.35	0.29	0.32	1.59	-0.05	-0.14
T1_Students_29_Gr1	SEI English	604	566	0.42	0.41	0.41	1.44	-0.01	-0.03
T1_Students_30_GrK	SEI English	566	690	0.48	0.50	0.49	1.31	0.05	0.10
T1_Students_31_Gr3	SEI Spanish	739	560	0.55	0.54	0.55	1.42	-0.01	-0.01
T1_Students_32_Gr2	SEI Spanish	686	594	0.44	0.40	0.42	1.52	-0.03	-0.06
T1_Students_33_Gr1	SEI Spanish	604	566	0.40	0.43	0.41	1.55	0.00	0.00
T1_Students_34_GrK	SEI Spanish	566	690	0.46	0.47	0.47	1.45	-0.03	-0.07

NOTES—1. [a] indicates that the measure initially failed baseline equivalence and was adjusted using propensity score weighting. 2. The source for the standard deviations was the sample. 3. The outcome measure was the same as pretest measure for all domains except in two cases. The pretest for STAAR Reading was WMLS Reading[b], and the pretest for all TELPAS outcomes for Kindergarten students only was TVIP[c].

Table C14. *Baseline equivalence for student outcomes for T2 versus BAU.*

Contrast ID	Pretest Measure	T2 Stu. N	C Stu. N	Unadj T SD at Pretest	Unadj C SD at Pretest	Pooled SD for T and C	C Mean at Pretest	T/C Diff. at Pretest	Std. T/C Diff. at Pretest
T2_Students_1_Gr3	ITBS Science	614	560	14.46	15.08	14.76	171.30	-1.56	-0.11
T2_Students_2_Gr3	WMLS Oral	573	506	17.55	17.91	17.72	78.57	-0.62	-0.03
T2_Students_3_Gr2	WMLS Oral	619	690	18.84	20.60	19.79	75.34	0.08	0.00
T2_Students_4_Gr1	WMLS Oral	562	561	24.76	24.90	24.83	64.07	2.56	0.10
T2_Students_5_GrK	WMLS Oral	609	583	31.67	32.26	31.96	53.13	2.65	0.08
T2_Students_6_Gr1	TOPA	557	561	2.08	2.21	2.14	5.96	-0.31	-0.14
T2_Students_7_GrK	TOPA	603	571	2.61	2.53	2.57	7.58	0.02	0.01
T2_Students_8_Gr3	TELPAS ELD	577	560	0.81	0.91	0.86	2.78	0.05	0.06
T2_Students_9_Gr2	TELPAS ELD	553	582	0.81	0.80	0.80	2.37	-0.12	-0.16
T2_Students_10_Gr1	TELPAS ELD[a]	515	532	0.69[a]	0.69[a]	0.69[a]	1.61[a]	0.02[a]	0.03[a]
T2_Students_12_Gr3	WMLS Reading[b]	530	561	16.59	17.18	16.87	96.24	-0.65	-0.04
T2_Students_13_Gr3	WMLS Reading	533	599	16.51	17.35	16.91	95.87	-0.40	-0.02
T2_Students_14_Gr2	WMLS Reading[a]	619	532	16.06	16.78	16.44	100.56	-1.15	-0.07
T2_Students_15_Gr1	WMLS Reading	562	470	20.02	21.18	20.60	90.91	1.09	0.05
T2_Students_16_GrK	WMLS Reading	598	688	22.27	22.00	22.14	83.15	0.82	0.04
T2_Students_17_Gr3	TELPAS Reading	577	690	0.98	0.95	0.97	2.65	-0.01	-0.02
T2_Students_18_Gr2	TELPAS Reading[a]	553	560	0.83	0.89	0.86	2.12	-0.15	-0.18
T2_Students_19_Gr1	TELPAS Reading[a]	514	532	0.59[a]	0.63[a]	0.60[a]	1.29[a]	0.05[a]	0.09[a]
T2_Students_21_Gr2	DIBELS	619	573	28.21	27.40	27.79	53.25	-3.60	-0.13
T2_Students_22_Gr1	DIBELS	562	472	19.34	19.25	19.29	18.63	-0.73	-0.04
T2_Students_23_Gr3	TELPAS Writing	577	599	0.86	0.92	0.89	2.43	0.01	0.01

Contrast ID	Pretest Measure	T2 Stu. N	C Stu. N	Unadj T SD at Pretest	Unadj C SD at Pretest	Pooled SD for T and C	C Mean at Pretest	T/C Diff. at Pretest	Std. T/C Diff. at Pretest
T2_Students_24_Gr2	TELPAS Writing[a]	553	532	0.74	0.82	0.78	1.98	-0.15	-0.19
T2_Students_25_Gr1	TELPAS Writing[a]	515	532	0.56[a]	0.57[a]	0.56[a]	1.26[a]	0.06[a]	0.11[a]
T2_Students_11_GrK T2_Students_20_GrK T2_Students_26_GrK	TVIP[c]	641	598	18.25	19.59	18.91	87.93	0.43	0.02
T2_Students_27_Gr3	SEI English	609	532	0.35	0.38	0.37	1.58	0.00	0.01
T2_Students_28_Gr2	SEI English	619	608	0.31	0.29	0.30	1.59	-0.05	-0.17
T2_Students_29_Gr1	SEI English	562	608	0.39	0.41	0.40	1.44	0.01	0.02
T2_Students_30_GrK	SEI English	616	608	0.49	0.50	0.50	1.31	0.05	0.11
T2_Students_31_Gr3	SEI Spanish	609	566	0.55	0.54	0.54	1.42	-0.02	-0.03
T2_Students_32_Gr2	SEI Spanish	619	690	0.40	0.40	0.40	1.52	-0.03	-0.08
T2_Students_33_Gr1	SEI Spanish	561	560	0.42	0.43	0.43	1.55	0.00	0.00
T2_Students_34_GrK	SEI Spanish	616	594	0.48	0.47	0.47	1.45	-0.01	-0.02

NOTES—1. [a] indicates that the measure initially failed baseline equivalence and was adjusted using propensity score weighting. 2. The source for the standard deviations was the sample. 3. The outcome measure was the same as pretest measure for all domains except in two cases. The pretest for STAAR Reading was WMLS Reading[b], and the pretest for all TELPAS outcomes for Kindergarten students only was TVIP[c].

Table C15. *Baseline equivalence for teacher outcomes.*

Contrast ID	Pretest Measure	T Tch. N	C Tch. N	Unadj T SD at Pretest	Unadj C SD at Pretest	Pooled SD for T and C	C Mean at Pretest	T/C Diff. at Pretest	Std. T/C Diff. at Pretest
T1 versus BAU									
T1_Teachers_1_Gr1	TOR	39	39	0.85	0.75	0.80	-0.07	0.05	0.06
T1_Teachers_2_GrK	TOR	41	44	0.80	0.70	0.75	0.04	0.07	0.10
T1_Teachers_3_Gr3	TBOP	37	39	0.30	0.32	0.31	0.42	-0.06	-0.20
T1_Teachers_4_Gr2	TBOP	45	46	0.14	0.27	0.21	0.55	0.03	0.16
T1_Teachers_5_Gr1	TBOP	39	39	0.17	0.19	0.18	0.61	0.00	0.00
T1_Teachers_6_GrK	TBOP	41	41	0.24	0.24	0.24	0.60	0.00	0.00
T2 versus BAU									
T2_Teachers_1_Gr1	TOR	38	39	0.92	0.75	0.84	-0.07	0.10	0.12
T2_Teachers_2_GrK	TOR	41	44	0.89	0.70	0.78	0.04	0.05	0.07
T2_Teachers_3_Gr3	TBOP	36	39	0.31	0.32	0.31	0.42	-0.04	-0.13
T2_Teachers_4_Gr2	TBOP	41	46	0.16	0.27	0.21	0.55	0.02	0.08
T2_Teachers_5_Gr1	TBOP	38	39	0.17	0.19	0.18	0.61	-0.01	-0.04
T2_Teachers_6_GrK	TBOP	40	41	0.25	0.24	0.25	0.60	0.03	0.14

NOTES—1. The source for the standard deviations was the sample. 2. The outcome measure was the same as pretest measure. 3. All measures initially failed baseline equivalence and were adjusted using propensity score weighting.

Fidelity of implementation. The following tables show that the key components of ELLA-V were mostly implemented with fidelity. Table C16 lists the three key program components and indicators for each component. The fidelity of each program component was measured using one unique indicator. Table C17 demonstrates whether each key program component was implemented with fidelity in each year of implementation (i.e., 2013–14 through 2016–17).

Fidelity was calculated for treatment teachers who had not attrited from the study and who participated, at least minimally, in the intervention. Teachers were excluded from the fidelity sample if (a) they did not attend any of the VPD training sessions; (b) they (or their schools) withdrew from the study, or (c) they left their schools. Note that if all treatment teachers in a specific grade level at a single school site were excluded from fidelity analyses, then the school site was excluded from the fidelity sample for the particular grade level.

Table C16. *List of key program components.*

Key Program Component	Indicator for Each Component	Data Source
Virtual Professional Development (VPD)	100% of treatment teachers in school missed no more than two PD trainings, and fidelity threshold was met in at least 90% of schools.	Teacher training attendance record
Virtual Mentoring and Coaching (VMC)	100% of treatment teachers in the school attended at least one coaching session, and fidelity threshold was met in at least 90% of schools.	Coach observation feedback rubric
Materials	90% of schools received curriculum materials.	Delivery receipts

Table C17. *Fidelity of implementation of each key program component by school year.*

Intervention Component	Implementation Year & Grade	Sample Size	Component Level Threshold for Fidelity of Implementation at the School Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity?
VPD	2013-2014 (Gr. 3)	40 schools	100% of teachers in school missed no more than two PD trainings	90% of schools met threshold	42.5%	N
VMC	2013-2014 (Gr. 3)	40 schools	100% of teachers in school attended at least one coaching session	90% of schools met threshold	100.0%	Y
Materials	2013-2014 (Gr. 3)	40 schools	School received curriculum materials	90% of schools met threshold	100.0%	Y
VPD	2014-2015 (Gr. 2)	45 schools	100% of teachers in school missed no more than two PD trainings	90% of schools met threshold	97.8%	Y
VMC	2014-2015 (Gr. 2)	45 schools	100% of teachers in school attended at least one coaching session	90% of schools met threshold	100.0%	Y
Materials	2014-2015 (Gr. 2)	45 schools	School received curriculum materials	90% of schools met threshold	100.0%	Y
VPD	2015-2016 (Gr. 1)	39 schools	100% of teachers in school missed no more than two PD trainings	90% of schools met threshold	100.0%	Y
VMC	2015-2016 (Gr. 1)	39 schools	100% of teachers in school attended at least one coaching session	90% of schools met threshold	100.0%	Y
Materials	2015-2016 (Gr. 1)	39 schools	School received curriculum materials	90% of schools met threshold	100.0%	Y
VPD	2016-2017 (Gr. K)	42 schools	100% of teachers in school missed no more than two PD trainings	90% of schools met threshold	88.1%	N
VMC	2016-2017 (Gr. K)	42 schools	100% of teachers in school attended at least one coaching session	90% of schools met threshold	100.0%	Y

Materials	2016-2017 (Gr. K)	42 schools	School received curriculum materials	90% of schools met threshold	100.0%	Y

NOTES—1. During their respective treatment year, four teachers left the Gr. K sample, five teachers left the Gr. 1 sample, five teachers left the Gr. 2 sample, and four teachers left the Gr. 3 sample.

Appendix D: Instruments

Figure D1. Transitional Bilingual Observation Protocol (TBOP) instrument.

TRANSITIONAL BILINGUAL OBSERVATION PROTOCOL

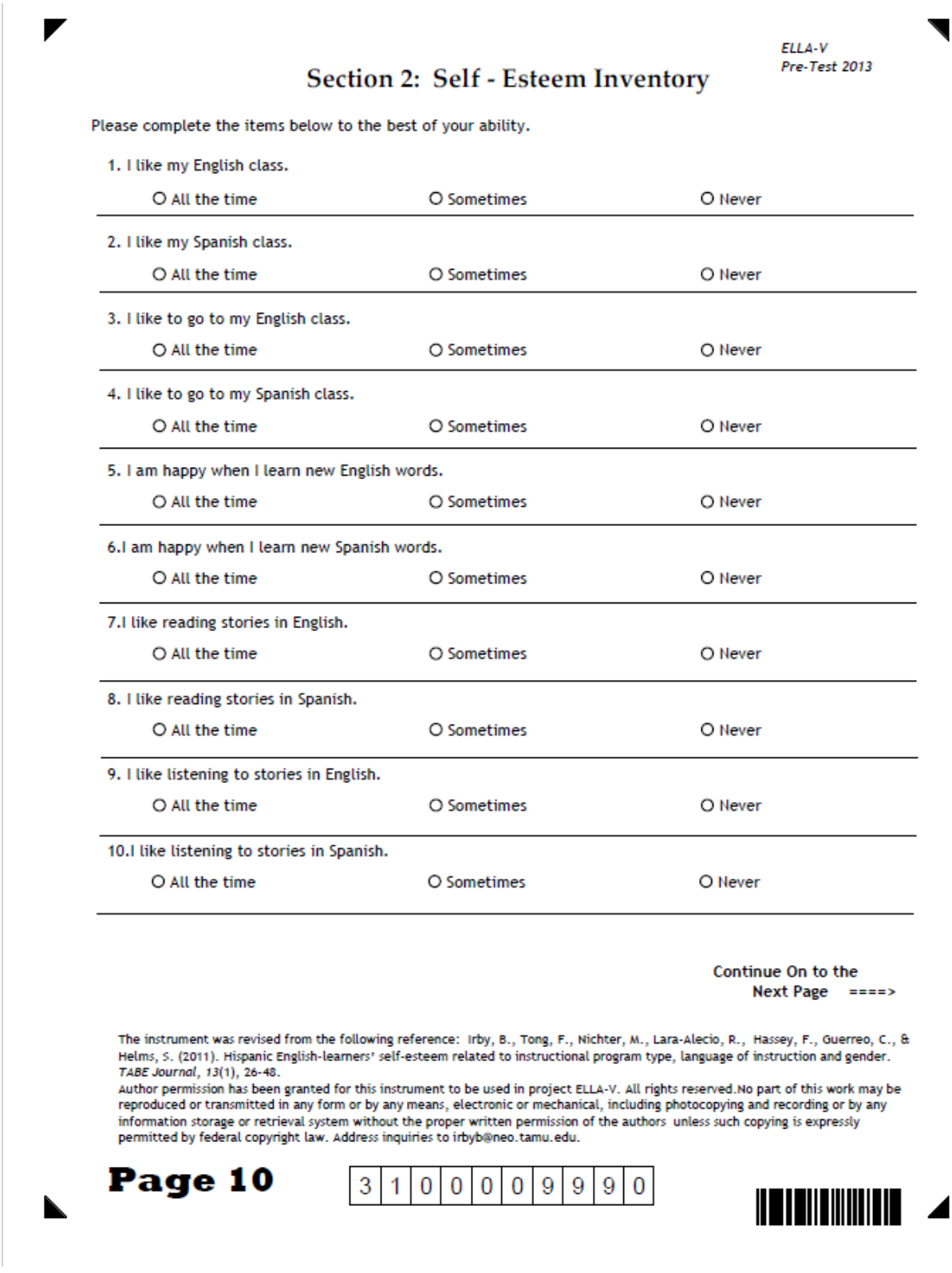
Observer: _____ Date: _____

Start Time: _____ End Time: _____

School: Teacher: Grade:

Time	Strategy	Curriculum Area	Physical Group	Activity Structure	Mode	Language Content	Lang of Instruction Teacher/Student	
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
CODING:	50 NA 51 QS 52 ALS 53 VS 54 MR 55 AO 56 CG 57 CC 58 LC 59 IT	1 read/lit 2 math 3 spell 4 hand 5 science 6 soc sci 7 health 8 PE 9 Music 10 Art 11 Lang 12 Compos 13 Non-ac 14 ESL	1 TC 2 LG 3 SG 4 Pairs 5 Single	1 lec/lis 2 lec/per 3 dir/lis 4 dir/per 5 dem/lis 6 led/per 7 ask/per 8 ask/ans 9 ans/ask 10 ev/per 11 obs/per 12 ev/dis 13 ev/cop 14 obs/dis 15 obs/cop 16 NA/feed 17 NA/free 18 NA/tran 19 NA/int 20 NA/out 21 interact	1 writing 2 reading 3 aural 4 verbal 5 wr-re 6 wr-au 7 wr-ver 8 re-wr 9 re-au 10 re-ver 11 au-wr 12 au-re 13 au-ver 14 ver-wr 15 ver-re 16 ver-au 17 au-re-ver 18 NA	1 social 2 academic 3 light cog 4 dns cog	1 L1 2 L2 3 L1-2 4 L2 -1 5 NA	1 L1 2 L2 3 L1-2 4 L2-1 5 NA

Figure D2. Student self-esteem inventory (SEI) instrument.



ELLA-V
Pre-Test 2013

Section 2: Self - Esteem Inventory

Please complete the items below to the best of your ability.

1. I like my English class.
   O All the time O Sometimes O Never

2. I like my Spanish class.
   O All the time O Sometimes O Never

3. I like to go to my English class.
   O All the time O Sometimes O Never

4. I like to go to my Spanish class.
   O All the time O Sometimes O Never

5. I am happy when I learn new English words.
   O All the time O Sometimes O Never

6.I am happy when I learn new Spanish words.
   O All the time O Sometimes O Never

7.I like reading stories in English.
   O All the time O Sometimes O Never

8. I like reading stories in Spanish.
   O All the time O Sometimes O Never

9. I like listening to stories in English.
   O All the time O Sometimes O Never

10.I like listening to stories in Spanish.
   O All the time O Sometimes O Never

Continue On to the Next Page ====>

The instrument was revised from the following reference: Irby, B., Tong, F., Nichter, M., Lara-Alecio, R., Hassey, F., Guerreo, C., & Helms, S. (2011). Hispanic English-learners' self-esteem related to instructional program type, language of instruction and gender. *TABE Journal*, *13*(1), 26-48.

Page 10

3	1	0	0	0	0	9	9	9	0

ELLA-V
Pre-Test 2013

Self - Esteem Inventory (Continued)

11.I understand my teacher when she speaks in English.
O All the time O Sometimes O Never

12. I understand my teacher when she speaks in Spanish.
O All the time O Sometimes O Never

13 I like to talk in English.
O All the time O Sometimes O Never

14. I like to talk in Spanish.
O All the time O Sometimes O Never

15. I am proud of my school work in English.
O All the time O Sometimes O Never

16. I am proud of my school work in Spanish.
O All the time O Sometimes O Never

17 I can speak to people in English.
O All the time O Sometimes O Never

18. I can speak to people in Spanish.
O All the time O Sometimes O Never

19. I can read well in English.
O All the time O Sometimes O Never

20. I can read well in Spanish.
O All the time O Sometimes O Never

21. I can write well in English.
O All the time O Sometimes O Never

22. I can write well in Spanish.
O All the time O Sometimes O Never

23. I am happy that I can answer questions in English.
O All the time O Sometimes O Never

24. I am happy that I can answer questions in Spanish.
O All the time O Sometimes O Never

==The End==

Page 11

3	1	0	0	0	0	9	9	9	0

Figure D3. Teacher observation record (TOR) instrument.

Teacher Observation Record (sample portion from Gr.3)

Teacher Observation Report

ELLA
07-08

Instructor's Last Name:

Observer's Last Name:

Campus Name:

Campus Number:

Teacher ID:

Content Reading Integrating Science for English Language and Literacy and Acquisition (CRISELLA)

Date: / /

Start Time: :

End Time: :

		1	2	3	4
A. Student Involvement 1. 90% or more = 4 2. 80-89% = 3 3. 70-79% = 2 4. 69% = 1	Does the teacher use questioning to promote discussion and writing and involve more of the ESL students in the lesson or activity?	○	○	○	○
B. Material Usage and Teacher Preparation 1. All material is prepared in orderly manner = 4 2. One item is lacking preparation = 3 3. Two items are lacking preparation = 2 4. Three or more items are lacking preparation = 1	Is the material ready and used according to the prescribed instructions? Does the teacher demonstrate preparation?	○	○	○	○
C. Leveled Questioning: 1. All levels = 4 2. 2 out of 3 levels = 3 3. 1 out of 3 levels = 2 4. No adjustments to any level = 1	Does the teacher address all ability levels (low, medium, and high)?	○	○	○	○
D. Teacher Talking, Reading, and Writing Time vs. Student Talking, Reading, and Writing Time: 1. All students are given enough time = 4 2. Most students are given enough time = 3 3. Few students are given enough time = 2 4. No time or very little time given to student involvement = 1	Does the teacher allow enough time for students to produce and practice the newly gained oral and written ESL skills?	○	○	○	○
E. Script: 1. Knows all content = 4 2. Knows most content = 3 3. Knows some content = 2 4. Not familiar with content = 1	Does the teacher know the lesson content and present it with confidence?	○	○	○	○

For Office use only. Please DO NOT fill in anything below this line

Batch No _____

Date Scanned _____