AIR®
AMERICAN INSTITUTES FOR RESEARCH®

Project RISE Final Report

AUGUST 2018

Natalya Gnedko-Berry | Scott Houghton | So Jung Park |
Liu Feng | Trisha Borman

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Project RISE Final Report

AUGUST 2018

Natalya Gnedko-Berry | Scott Houghton | So Jung Park |
Liu Feng | Trisha Borman

1000 Thomas Jefferson Street NW
Washington, DC 20007-3835
202.403.5000

www.air.org

5444_08/18

Contents

Tables

Page

Figures

Page

Introduction

Schools in the Internationals Network for Public Schools (hereafter, Internationals) are public middle schools, high schools, and academies (i.e., schools within schools) that serve secondary English language learners (ELLs)—in particular, ELLs who have been enrolled in U.S. schools for fewer than 4 years and who are at the lowest levels of English proficiency. Currently, 29 schools and academies are in the network, with 9,000 students from more than 100 countries.

With funding from a 5-year Investing in Innovation (i3) development grant from the U.S. Department of Education's Office of Innovation and Improvement, Internationals received an opportunity to implement key components of its model in comprehensive high schools that serve both ELLs and non-ELLs. Internationals designed an intervention—Project RISE (**R**ealizing **I**nstructional **S**upports for **E**nglish Language Learners)—that aimed to improve the educational outcomes of ELLs in Grades 9–12 in two schools by changing teachers' instructional practices and several key structures in which students and teachers are organized. The program aimed to establish the following: (a) collaborative, interdisciplinary teaching teams responsible for (b) shared heterogeneous cohorts of students in classrooms that (c) integrate instruction of language and content using (d) collaborative experiential projects across the curriculum. To establish these program components in participating schools, Project RISE provided instructional resources and school-level supports, such as coaching and professional development, to teachers and school leaders, including principals and assistant principals. The Project RISE logic model, presented in Appendix A, outlines all the intervention components.

The implementation of Project RISE occurred during four school years, from 2013–14 through 2016–17, in two comprehensive high schools identified by the districts as persistently low performing: School A is in a large urban West Coast school district, and School B is in a large urban East Coast school district. In the first year of implementation, 496 students were enrolled at School A, with approximately 45% identified as ELLs. By the final year of implementation, the proportion of ELLs increased by 11 points to approximately 56%, but student enrollment remained similar at 463 students. School B was larger than School A, with an enrollment of 2,313 students in Year 1, and 2,049 students in the final program year. The proportion of ELLs at School B increased from 15% in Year 1 to 18% in the final program year. At both schools, students were identified as ELLs based on a standardized state assessment. The principal in each school remained the same during all four implementation years. Teachers at School A volunteered for the intervention. At School B, school administrators assigned teachers to the intervention.

Evaluation Overview

The American Institutes for Research (AIR), an external evaluator of Project RISE, was charged with providing both a formative evaluation and an impact evaluation.

Formative Evaluation

A formative evaluation of Project RISE was conducted during the 4-year implementation period. The purpose of this evaluation was to document and examine the implementation of Project RISE, such as successes and challenges of implementing the instructional and structural components of the program. The current report includes a summary of the findings across the 4-year formative evaluation period.

Impact Evaluation

An impact evaluation of Project RISE was conducted in 2018, a year after the project's completion. The timing of the impact evaluation was determined by the availability of data required to examine impact, with all the data available in early 2018. The purpose of the impact evaluation was to examine the impact on Project RISE on student outcomes, including achievement, credit accumulation, high school graduation, and attendance. The current report includes a summary of findings across the final 2 years of Project RISE, during which impact was assessed.

Purpose and Organization

The purpose of this document is to summarize the findings of the formative and impact evaluations of Project RISE. First, we present the results of the formative evaluation accumulated during the 4-year evaluation period while the project was being implemented. Then we present the results of the impact evaluation conducted after Project RISE was completed.

Formative Evaluation Results

Introduction to the Formative Results Summary

AIR conducted a formative evaluation of Project RISE during the 4-year implementation period. The evaluation resulted in six interim formative reports. The reports synthesized findings from semistructured interviews with teachers, school administrators, and coaches; the analysis of school documents pertaining to scheduling and lesson planning; and program fidelity documentation.

The summary of the formative evaluation findings presented in this report draws on the six previous interim reports. In those reports, the AIR research team devoted a section to summarizing results that emerged across schools by highlighting the similarities and differences between Schools A and B. In each report, the AIR research team also presented main findings regarding the fidelity of implementation. For the final reporting of results, we identified themes that emerged across these sections of the six interim reports. In the discussion that follows, we briefly examine the methods of data collection and summarize the themes that arose in the interviews and document analysis, followed by the themes that emerged in the analysis of implementation fidelity.

Methods

Interviews and Document Review

Across the 4-year formative evaluation period, the AIR team conducted 115 interviews with 39 participants. Of the total 39 participants, 27 were Project RISE teachers[1]; nine were school leaders, including principals, assistant principals, and school-based coaches; and three were Internationals instructional leadership coaches (ILCs) funded by Project RISE to provide school-level support to participating teachers and administrators. In school years 2013–14 through 2015–16, we interviewed all participating teachers and school leaders. In 2016–17, we interviewed school leaders exclusively because the focus of interviews was on program

[1] Project RISE teachers provided instruction to ELLs grouped into heterogeneous cohorts per Project RISE design and participated in Project RISE professional development and instructional supports. Project RISE teachers had flexibility in the extent of participation in professional development and instructional supports, but they typically participated in the yearly summer learning institute and other professional development opportunities organized by Internationals and engaged in weekly meetings with other Project RISE teachers and the ILC.

sustainability, a topic that school leaders are best suited to address. Depending on the length of participation in the intervention, all participants were interviewed one to six times. Interviews were recorded and transcribed; the transcripts were then coded, first by using a priori codes and subsequently through application of emergent codes based on the themes that occurred most frequently. The document review examined the schools' planning templates and lesson plans to search for themes of Project RISE instructional elements and analyzed the school's teaching schedules to understand the extent to which Project RISE structure changes had been incorporated into the schools' programming.

Implementation Fidelity

To examine implementation fidelity, the AIR research team used two data sources in each of the 4 years of formative evaluation: the activity log completed by the ILCs for each school, and additional documentation collected by Internationals staff throughout the school year, such as meeting agendas and sign-up sheets. We analyzed these data to calculate implementation fidelity on indicators based on the logic model for the project (see Appendix A). Altogether, these reports provided formative feedback to Internationals regarding participants' perceptions of the instructional and structural elements of Project RISE and the level of implementation fidelity at each school.

Results

Across-Year Summary of Findings From Interviews and Document Analysis

The findings in this section are organized into subsections comprising Project RISE professional development and coaching; implementation of instructional and structural components of the project and their schoolwide influence; school-level facilitators and barriers to implementation; and sustainability. We provide background information where necessary to help the reader understand the context in which the findings have emerged.

Professional Development

Participants in both schools credited supports provided by Project RISE for improvements in ELL instruction.

Project RISE provided professional development to teachers and school leaders at both participating schools. Each school year began for teachers, before classes commenced, with an annual professional development institute that focused on the integration of language into content instruction and student collaborative learning. An ILC at each site, who was hired and trained by Internationals, delivered ongoing support to teachers and school leaders

throughout the school year through group professional development sessions and individual coaching. Both schools participated in intervisitations, in which school staff visited existing schools in the Internationals network to observe program implementation and engage in professional development with Internationals teachers. In addition, Internationals staff members consulted with the school principal and the ILC to monitor and adjust program implementation. Per the project design, the ILC's presence on-site was reduced from twice a week to one day each week beginning in the third year of the program, and one day every other week starting in the fourth year.

The AIR study results suggest that some challenges in Year 1 accompanied the rollout of professional development. School A participants reported interpersonal difficulties among the ILC, teachers, and school leaders. Further, teachers at School A expressed mixed perceptions about the usefulness of the team meetings and coaching. For example, some teachers suggested that one-on-one meetings could be improved through more focus on lesson planning and feedback that is more closely tailored to their needs. In School B, 10 teachers were assigned during the summer to participate in Project RISE, and attended the summer institute. However, eight of these 10 teachers were reassigned at the beginning of the school year to other teaching roles that prevented them from participating in Project RISE. As a result, only two of the six teachers ultimately involved in Project RISE at School B attended the summer institute. The other four teachers who participated in Project RISE were assigned to the program after the summer institute had already taken place. This shift could have introduced the initial challenges that teachers reported concerning the implementation of the Project RISE instructional components. Despite these Year 1 challenges, participants in both schools reported that the supports provided by Project RISE were useful.

In subsequent years, participants from both schools reported that Project RISE professional development—including individual teacher coaching, ongoing team professional development sessions, and intervisitations—was effective in improving their ELL instruction. Participants specifically cited the usefulness of learning strategies for teaching ELLs, the value of team meetings and collaboration among teachers, and the capacity-building support of the ILC, including facilitating teacher meetings and offering ideas for improving lessons. Participants from both schools consistently identified individual coaching and group professional development sessions and intervisitations, along with teacher collaboration, as key facilitators in implementing both instructional components of Project RISE: the integration of English and native language development into content instruction and the facilitation of student collaborative practices. As the ILCs in both schools reduced their presence on-site over the final years of implementation, teachers in both schools who were more experienced in the Project

RISE model assumed more responsibility for facilitating team meetings and for providing professional development to those colleagues who were participating in Project RISE, as well as those who were not. For example, beginning in the 2015–16 school year, Project RISE teachers from School B reported leading monthly, schoolwide workshops on Project RISE elements, which included general education teachers.

Instructional Components

Participants from both schools reported incorporating student collaborative work and combining English and native language development and content instruction in their classrooms, with the latter element taking approximately 2 years to master. Participants credited these instructional components with facilitating students' academic and English development and with providing social-emotional benefits.

Language and Content Integration

One of the core instructional components of Project RISE is the integration of English and native language development into content instruction. The AIR study results suggest that proficiency in this practice took approximately 2 years for teachers to develop. In Years 1 and 2, participants in both schools reported an increase in language development integration practices, such as introducing academic vocabulary to students before starting a lesson as well as using various kinds of visual aids. In Year 3 of implementation, teachers reported that they were confident about integrating English development in content classrooms. In Year 3, School A participants described the use of more purposeful approaches to incorporating English and native language development into content instruction, such as reducing the number of language development goals and focusing only on the few that could be explored effectively in an instructional unit. School B participants reported that teachers began to focus more on sentence structure and were more seamlessly integrating English and native language development and content. Because of this integration, these participants explained that students were acquiring academic English while "doing something purposeful." In addition, teachers from both schools began to implement formal language development objectives in content lessons—for example, focusing on grammatical forms or language functions that were aligned to the scope and sequence of the coursework. These language objectives were included in planning documents in both schools.

Participants across Schools A and B perceived that incorporating English and native language development into content class instruction accelerated students' learning of both English and subject-matter content, compared to the traditional English as a second language (ESL) approach that separates language and content development. Because of students' better

access to academic content, participants thought that students were improving in three specific areas: They were no longer lagging in subject areas as much as they had previously, they were better prepared for standardized testing, and they were on a better pace to graduate from high school and pursue college education. Participants further reported that this approach to language and content integration made learning English more meaningful, because it was better contextualized and connected to more academic concepts. Teachers from both schools credited the support from other teachers, as well as coaching and professional development from Project RISE, as facilitators of implementing this practice.

Student Collaboration

Student collaboration is another core instructional component of Project RISE, the purpose of which is to develop a classroom environment in which students can support and learn from one another. Participants from both schools reported that in the first year of implementation, student collaboration accounted for the largest shift in teaching practice. In the first year and subsequent years of implementation, a commonly reported approach to grouping was to group students of different English ability who share a common home language to enable them to use their native language to help one another learn. In both schools, participants demonstrated an understanding of the difference between truly collaborative student learning, in which students engage with one another and contribute meaningfully to tasks, and group work in which contributions are not equal or engagement is low. Feedback across schools suggests that teachers' use of student collaborative practices increased over time, and that they grouped students more purposefully as the project progressed, considering their academic, English, and native language abilities. For example, collaboration strategies were evident in one teacher's lesson plan, which included group directions with roles (e.g., facilitator, organizer, resource manager), and asked group members to discuss the meaning of a text, write their own analysis, and read it aloud. An ELL unit plan checklist further showed increasing intentionality around student collaboration relative to lesson plans before Project RISE. For example, the unit plan checklist called for "student opportunities to talk and discuss to build academic language" and "activities that promote collaboration."

Participants across both schools credited student collaboration with improvements in students' ability to communicate in English, and with providing a safe environment for language practice. Participants also thought that student collaboration facilitated the development of students' social skills by providing opportunities to interact with a diverse group of students and increased a sense of community and confidence among students. Participants credited Project

RISE supports and teacher collaboration for allowing teachers to learn how to implement and use student collaboration techniques.

The AIR study findings suggest that substantial academic and language heterogeneity in Project RISE classes sometimes created a barrier to implementing student collaborative structures effectively. Teachers explained that it could be a challenge to ask two students to collaborate and meet both of their needs when, for example, one student is very academically proficient and the other has had only limited, formal schooling.

Influence on Schoolwide Instruction

The primary goal of Project RISE was to improve the educational outcomes of ELLs through the program's core instructional and structural components. In the final year of implementation, the AIR research team also was interested in the perceived influence of Project RISE beyond its intended goals. As such, we asked school leaders in both schools to discuss whether and how Project RISE has influenced instruction in general education classes. The site leaders at School A reported that Project RISE practices regarding the integration of content and language development had spread to general education classes; indeed, schoolwide lesson and unit plans emphasize the use of language objectives, language functions, and introduction of new vocabulary during content classes. The influence of Project RISE on schoolwide language development practices in School B was more modest, with some evidence in the schoolwide unit planning template that general education teachers were asked to consider the language needs of ELLs—for example, by modifying instructional strategies.

School leaders at School A reported a considerable spread of the Project RISE approach to student collaboration to all classrooms, including general education classrooms. This was reflected in a schoolwide template that includes a collaborative task section in which teachers are asked to describe the key collaborative skills they intend to teach, and to choose from a list of interaction strategies. School leaders at School B thought that Project RISE promoted the use of student collaboration primarily in the ELL Academy—a smaller learning community within the school that is focused solely on ELLs—consistent with the project design. School leaders in both schools reported that the means through which Project RISE instructional practices spread to general education classrooms included teacher collaboration and schoolwide professional development.

Structural Components

Participants from both schools reported that heterogeneous student cohorts fostered differentiated instruction, opportunities for students to learn from their peers, and social-emotional benefits. Collaborative interdisciplinary teaching teams reportedly allowed teachers to share best practices and establish consistency across classes, improving their ability to support the needs of students.

Heterogeneous Student Cohorts

Student cohorts that are heterogeneous in English level and academic proficiency are a key structural component of Project RISE, necessitating detracking of students from traditional ESL course sections that are tracked on English level. This component also includes the scheduling of students to allow for travel between classes with the same group of peers. Both schools implemented the structural component of heterogeneous cohorts, although the timing and extent of implementation varied across the schools. In the first year of implementation (2013–14), School A established an interdisciplinary team of five teachers sharing a heterogeneous cohort of ELLs in grades 9 and 10. Teachers at School B who were part of a previously created ELL Academy within the larger school participated in Project RISE professional development but did not share student cohorts. School B maintained two levels of English classes in Year 1 (reduced from five levels before Project RISE) and expressed more initial concerns about teaching students of a wide range of English proficiency in the same classroom.

In the second year of implementation, School A expanded the cohort structure to ELLs in grades 11 and 12 but allowed for greater course flexibility outside the core disciplines for these students. This led to students being cohorted for only some of their classes. School B introduced a cohort structure in grade 9. ELLs in grade 10 experienced a hybrid approach in which they took the core classes with the same classmates and had the same Project RISE teachers, but they did not remain with the same classmates in all classes. In grades 11 and 12, ELLs at School B were not grouped in cohorts, although they received instruction from some Project RISE teachers. In subsequent years, the cohort structure remained largely the same as in Year 2.

Across schools, participants identified three primary benefits of heterogeneous student cohorts: (a) increasing the potential for differentiated instruction; (b) providing a space for students at the emergent levels of English proficiency to learn from their peers whose English is more developed; and (c) providing opportunities for students whose English is more developed to reinforce their learning and practice leadership. Teachers from both schools also repeatedly mentioned the challenges of addressing the needs of all students in a heterogeneous class;

specifically, they cited the difficulties in planning for different access points and finding the level of instruction that is appropriate for all students. They reported that students with more developed English proficiency might benefit from or prefer more challenging or fast-paced instruction than they receive in classes that also address the needs of students at the emergent levels of English proficiency. Although teachers from School B primarily highlighted the benefits for students at the emergent levels, some expressed concern about the pace of the class for students in the very beginning stages of learning English.

Participants across both schools perceived the social-emotional benefits of the cohort model, such as supportive relationships among students and teachers, a sense of belonging, and a place for students to grow in confidence. In conjunction with teacher collaboration, participants across schools explained that student cohorts increased teachers' ability to address the needs of individual students more effectively by allowing them to discuss strategies and share best practices with colleagues in other disciplines who were teaching the same students.

Collaborative Interdisciplinary Teaching Teams

The second structural element of Project RISE is a formal structure that allows teachers to collaborate on interdisciplinary teams. Both schools implemented this structure. Participants from both schools attended team meetings that included professional development twice per week beginning in the first year of implementation. School A also implemented an extra preparation period in the second year of implementation to give teachers time to collaborate and adapt their curricula to Project RISE. This schedule of collaborative planning time every day, along with grade-level team meetings twice per week, continued over the remaining years of the project. Ninth-grade teachers at School B had block scheduling, shared a common planning time, and collaborated twice a week. Not all tenth-grade teachers could meet at the same time because they did not have block scheduling; instead, they met during one of two periods, depending on their availability. Tenth-grade teacher meetings in School B did not exclusively focus on Project RISE instructional strategies.

Participants from both schools reported that the collaborative interdisciplinary teaching teams established as part of Project RISE provided a platform for teachers to share best practices, better support the needs of students, and establish consistent instructional practices across classrooms. Participants noted that the opportunity to share best practices among teachers helped them first to implement and then improve both student collaboration and language development strategies. The main reported advantage of the Project RISE approach was interdisciplinary teacher collaboration, facilitated by scheduled meeting times dedicated to professional development.

Influence on School Structures

The AIR study team was interested in the perceived influence of Project RISE on broader school structures, which is another potential unintended benefit of implementation, rather than a stated project goal. According to school leaders at School A, Project RISE directly influenced schoolwide teacher collaboration structures, namely interdisciplinary teaching teams for all teachers. Beginning in 2015–16, ELL and general education teachers at School A were grouped into interdisciplinary teaching teams that collaborated weekly. The AIR team's analysis of both Project RISE and schoolwide schedules confirms the implementation of these interdisciplinary structures, showing common planning time across grade-level teachers who teach ELLs or general education classes. School A also adopted the Project RISE student cohort model schoolwide, which allowed students to take classes with the same group of peers. School leaders at School B reported that Project RISE invigorated the previously intended but underdeveloped collaboration among ESL and content teachers of ELLs.

School-Level Facilitators and Barriers

The AIR study findings suggest that a small school could provide conditions that are more favorable to implementation than a large school. These results also suggest that a supportive relationship with key personnel, including the ILC and administrators, is an important facilitator of implementation.

Participants from both schools discussed the size of their school as a perceived factor that influenced the schools' ability to implement Project RISE. Participants from School A thought that the small size of their school fostered a sense of community among teachers that was conducive to project implementation. Participants from School B thought that the large size of their school presented challenges to implementing the structural components of Project RISE. For example, the ILC noted difficulties in implementing the structural components, such as teacher teams sharing students, because of the complexity associated with assigning schedules in a large school. Teachers further explained that large class sizes complicated the use of strategies that they learned in Project RISE.

Participants at both schools frequently discussed key school personnel and relationships as facilitators of and barriers to implementation. Participants from School A described personnel challenges in the first year of implementation, including a difficult relationship among Project RISE teachers and the ILC and the absence of a school-based ELL instructional coach. Teachers also described a perceived low engagement of the school principal in Project RISE as a barrier. The study findings suggest that in Year 2 and subsequent years, these barriers at School A were alleviated when the new ILC joined the school along with a school-based coach and as the

principal became more engaged in the program. School B participants indicated from the beginning that the school leaders engaged in program activities and clearly expressed expectations for teacher participation in the program, and the ILC established strong relationships with school staff, which increased program support and buy-in. Across both schools, support services provided by Project RISE—such as ILC coaching, professional development, and instructional tools, resources, and strategies—often were highly credited as facilitators of teachers' ability to implement the new approach to instructing ELLs.

Sustainability

Participants from both schools expressed a desire to maintain the structural elements of Project RISE or the entire model past the grant-funded period. Both schools reported taking steps to ensure sustainability, such as transitioning professional development responsibilities from the ILCs to teachers in later years of implementation.

In the third year of the study, the AIR research team asked participants about the sustainability of the Internationals model beyond the grant-funded period. The most commonly expressed theme was the participants' desire to maintain either the Project RISE structural elements that they had implemented during the 4-year period or the entire Project RISE model. Both schools reported taking steps to ensure sustainability. For example, beginning in the third year, teachers began to take on leadership roles in organizing collaborative meetings or delivering professional development aimed at ELL instruction—the role the ILCs had fulfilled in earlier years. School leaders from both schools were confident that once the project ends, the schools could, with the leadership of teachers and coaches, maintain professional development practices introduced by Project RISE. As evidence of their commitment to the project, school leaders at School B created professional development and planning documents that specifically address the needs of ELLs as additional supports for ELL teachers. These documents include a matrix of language objectives, a curriculum map template, and a variety of instructional strategies for ELLs. In addition, school leaders from both schools expressed concerns about maintaining the integrity of the program over time because of high teacher turnover at School A and the loss of the ILC at School B, who had served as the primary advocate for the program. Participants from both schools mentioned strong program advocates who could help ensure the sustainability of Project RISE.

Across-Year Summary of Findings for Implementation Fidelity

The AIR analysis suggests adequate implementation fidelity of key components of Project RISE, with two exceptions: ILC support in Year 1, and Internationals support of teacher and school leaders in Years 3 and 4.

The AIR study team's analysis of implementation fidelity focused on the implementation of key components of Project RISE: supports delivered by the ILCs, by Internationals, and by the school. Each key component consisted of implementation indicators aligned with the logic model for the project. At the end of each implementation year, the study team calculated the fidelity of implementation for each indicator, then aggregated indicator results to the key component level. Finally, we calculated sample-level fidelity for each key component across schools. We rated the sample level fidelity as *adequate* if both schools had achieved a medium score or higher across indicators, or *not adequate* if one or both schools failed to meet this threshold. The logic model for Project RISE is included in Appendix A. The fidelity matrix that specifies key components, their indicators, thresholds of implementation fidelity, and scores for each year of analysis is provided in Appendix B.

The sample-level fidelity was rated *adequate* for most components across all years, with a few exceptions. ILC support was rated *not adequate* in the first year of implementation because of a lower rating in both schools on the measure of individual coaching. Although teacher observations were sufficiently frequent in School A, they did not always include a one-on-one meeting to debrief or provide feedback. In School B, the number of observations and feedback sessions did not reach the frequency required for the high-fidelity threshold. Support of teachers and school leaders by Internationals was rated as *not adequate* in Years 3 and 4 due exclusively to lower than intended levels of teacher participation in the two annual visits to existing Internationals schools.

Conclusions

Throughout the 4-year period of Project RISE implementation, teachers and administrators described evolving implementation of all four Project RISE components, as well as improvements in ELL instruction and student learning. Participants credited Project RISE professional development with improving teachers' instruction of ELLs and facilitating the implementation of student collaboration and language development practices. They found that the implementation of these core instructional components enhanced students' academic and English language development and provided social-emotional benefits.

The implementation of the structural elements of Project RISE enhanced the implementation of the instructional components, and further supported instruction and learning for ELLs. Heterogenous student cohorts reportedly provided opportunities for students of varying English and academic proficiency to learn from one another and facilitated teachers' use of differentiated instruction. Participants also reported that collaborative interdisciplinary teaching teams supported teachers in sharing best practices with one another and establishing instructional consistency across classes, which helped teachers implement the project's instructional components and support the needs of ELLs. Because of the perceived program's success, participants from both schools aimed to maintain the structural elements of Project RISE or the entire program model and began taking steps to ensure this sustainability in the final years of the grant-funded period.

Impact Evaluation Results

Introduction to the Impact Results Summary

AIR completed the impact evaluation of Project RISE in 2018, which is the year immediately following the completion of project implementation. The summary of the impact evaluation of Project RISE is presented in this report for the first time. Therefore, we describe in detail the impact evaluation's methods and results. The level of detail is greater in this impact section compared with the formative section because details of the formative evaluation were documented in formative reports disseminated earlier.

The impact evaluation examined the impact of Project RISE on the academic and nonacademic outcomes of ELLs. The evaluation addressed four research questions:

RQ1: What is the impact of Project RISE on the achievement of ELLs in mathematics and English language arts (ELA)?

RQ2: What is the impact of Project RISE on credit accumulation of ELLs?

RQ3: What is the impact of Project RISE on the high school graduation of ELLs?

RQ4: What is the impact of Project RISE on the attendance of ELLs?

In reviewing the results of the impact evaluation, it is important to understand how the treatment was operationalized. Student-level treatment was operationalized as being taught by at least four different Project RISE teachers (i.e., teachers teaching students in heterogeneous cohorts and receiving Project RISE professional development and instructional support) in a single school year in any semester. For example, we considered a student treated if the student was taught by four Project RISE teachers in the fall and spring semesters of one school year or if the student was taught by two Project RISE teachers in the fall semester and two different Project RISE teachers in the spring semester of one school year. Full treatment consisted of 2 or more years of being taught by four or more Project RISE teachers. The years of full treatment did not need to be consecutive.[2] Partial treatment consisted of 1 year of being taught by four or more Project RISE teachers. For all research questions, we examined the impact of Project RISE on ELLs who received full or partial treatment.

[2] The proportion of ELLs in the final sample who received full treatment in nonconsecutive years was small—7%.

We examined outcomes in Years 3 and 4 of the implementation, which coincided with academic years 2015–16 and 2016–17. We combined outcomes for both years. That is, we do not report a separate set of outcomes for 2015–16 and 2016–17; instead we report, for example, an outcome of achievement that could have been recorded in 2015–16 by students who were tested in that year or 2016–17 by students who were tested then. We further describe how we operationalized and computed outcomes in the methods section. The final 2 years of implementation were selected as the outcome years; given the complexity of the intervention and the levels of instructional and structural supports, the program developers did not expect outcomes to emerge earlier.

The definition of treatment, the treatment dosage, and the selection of outcome years were determined a priori during the intervention and the evaluation planning phase. The definition of treatment and the selection of outcome years were affected by the limitations of the i3 grant that funded Project RISE. That is, the Internationals' school development and support model was designed to provide 4 or more consecutive years of students being taught by teachers who have expertise in the Internationals' approach and are teaching in a context in which students are grouped into heterogeneous cohorts, plus teachers working in interdisciplinary teams. However, for Project RISE, the duration of treatment was necessarily constricted given the 4-year school-level implementation period allowed by the funding source.

Methods

In this section, we describe the data and samples used in the impact evaluation, our matching procedure, the baseline equivalence of analytic samples, the computation of the outcome measures, and the analytic approach for estimating impact.

Data

We used three data sources. Student rosters from the two participating schools, which were provided by school administrators, comprise one data source. We used this data source to identify students who received treatment (i.e., students who were taught by four or more different Project RISE teachers within a school year) and treatment dosage (i.e., 1 year of treatment or 2 or more years of treatment). We also used district administrative records as a source of data on students' demographic and background characteristics and outcome information. Finally, we used data reported on districts' public websites as a source of information on school characteristics.

Sample

The initial sample for this study included all ELLs from 11 schools: two schools that participated in Project RISE and nine comparison schools. Participating schools (also called treatment schools) were selected by the program planners. We selected comparison schools by using a matching procedure, which is described in the next section. Districts identified ELLs based on standardized tests (discussed in the section on student matching). Our initial sample included any student identified as an ELL at one of the participating or comparison schools in any implementation year. Our initial sample included 9,482 ELLs: 1,406 in treatment schools and 8,076 in comparison schools. Of the 1,406 ELLs in the treatment schools, 1,016 were enrolled in grades where treatment was available.[3] Among these ELLs, we located 634 ELLs who were served: 252 ELLs received full treatment, and 382 ELLs received partial treatment. The remaining 382 ELLs were not considered treated because they were not taught by four or more Project RISE teachers in a single school year (193 ELLs were never taught by any Project RISE teachers).

Table 1 presents the background characteristics of ELLs who received full or partial treatment. Our findings suggest that the two groups were similar on most characteristics. However, Hispanic students were overrepresented in the full treatment group, and White students were overrepresented in the partial treatment group. Neither our data for the impact study nor data for the implementation study offer clues for why these racial groups were disproportionately represented in the different treatment groups. All our analyses, however, controlled for racial group composition.

Table 1. Characteristics of ELLs Who Received Full or Partial Treatment

Characteristic	Full treatment (*n* = 252)	Partial treatment (*n* = 382)	Chi-squared test *p*-value[a]
Female	35%	38%	0.586
White	13%	29%	0.000
Hispanic	49%	40%	0.032
Black	2%	3%	0.990
Asian	35%	28%	0.073
Race_Other	0	1%	1.000
Eligible for free or reduced-price lunch	69%	69%	1.000

[3] The selection of ELLs into Project RISE was carried out at the schools. Recall that treatment was not available in all grades by design or because of implementation differences. For example, treatment was not available to students in Grades 11 and 12 in any year at School B, which was a difference in implementation.

Characteristic	Full treatment (*n* = 252)	Partial treatment (*n* = 382)	Chi-squared test *p*-value[a]
Receiving special education services	1%	0	0.896

[a]The purpose of the chi-square test is to determine whether a characteristic is disproportionately represented among the compared groups. A desirable outcome for this analysis is a nonsignificant finding of $p > 0.05$, which indicates that the groups are similar.

To create the analytic sample, we identified treated ELLs who had valid outcome and baseline measures (i.e., complete cases). We then matched these students to similar students from the comparison schools in a procedure that we describe in the following section. The availability of data varied by outcome and treatment dosage. Therefore, the analytic sample for each outcome also varied. Table 2 shows the total number of students in each analytic sample by outcome and treatment dosage.

Table 2. Number of Students in Each Analytic Sample, by Outcome and Treatment Dosage

Outcomes	Full Treatment		Partial Treatment	
	Treatment	Comparison	Treatment	Comparison
Mathematics	74	186	104	330
ELA	44	96	60	186
Credit accumulation	149	361	142	515
Attendance	144	379	150	486
Graduation	56	124	25	66

Matching Procedure

We used a two-step matching procedure. We first selected schools with similar demographic composition and performance levels as the treatment schools and then selected students within those schools who were most similar to students in the treatment schools.

Step One: School Matching

We selected comparison schools using school-level data from 2012–13, which was the year prior to the implementation of Project RISE. We selected comparison schools using a stratification approach. We began matching with School B in District B because District B is a large district that offers a large pool of possible comparison schools. We stratified all public high schools based on five characteristics available in the public school data: (a) school size, (b) the percentage of ELLs,

(c) progress report scores,[4] (d) the percentage of students eligible for free or reduced-priced lunch (FRPL), and (e) the percentage of minority students. Our selection of these characteristics was driven by the belief that, collectively, they accounted for a large proportion of variation in school environment that could affect outcomes. Strata were created by placing schools into three groups based on the ordered data value for each of the five school characteristics (high/large stratum for the highest 25%, medium stratum for the middle 50%, and low/small stratum for the lowest 25%). With three categories for each characteristic, we had 243 (i.e., 3^5) strata. The treatment school was in the stratum of large school size, high percentage of ELLs, medium progress score, low percentage of FRPL eligibility, and low percentage of minority students. Five schools of the 393 high schools not implementing Project RISE in District B were in the same stratum as the treatment school. These five schools became our comparison schools.

For District A, we followed the same stratification procedure. Because District A had only 16 public high schools, we needed to reduce the number of strata to find an acceptable number of matched comparison schools. Therefore, we spilt schools into high and low groups (rather than high/large, medium, and low/small groups) based on the median value of each school characteristic. With five covariates and two categories, we had 32 (i.e., 2^5) strata. The treatment school was in the stratum of large school size, high percentage of ELLs, high progress report score, high percentage of FRPL eligibility, and low percentage of minority students. Only one school not participating in Project RISE was found in the same stratum.

To find additional comparison schools, we first prioritized the five school characteristics by considering correlations among them and their relevance to the study. Then we dropped the variables one at a time in prioritized order to reduce the number of strata and find additional comparison schools. However, as the number of criteria went down to two (school size and the percentage of ELLs), we were not locating any additional comparison schools. We then lowered our selection criteria to either school size or the percentage of ELLs. We found three schools similar in size to the treatment school, and two schools with similar percentages of ELLs as the treatment school.[5] We decided to combine the two lists. The lists had one school in common, leaving us with four schools. We added one school previously identified using all criteria for a total of five

[4] The progress report score measures the ability of a school to help students progress toward the eventual goal of graduation. The measures included in this score focus on the capacities students develop while at school, not the capacities they bring with them on the first day. The measure is based on six metrics focused on credit accumulation and six metrics focused on high school examinations.

[5] We initially had five schools for each criterion with one school in common. Of the eight schools that were different, we excluded five schools because of one or more of the following reasons: a low percentage of ELLs (less than 2.5%), designation as an alternative school, and participating in another intervention.

comparison schools. During the outcome analysis, we learned that the district closed one of the comparison schools, which left us with a sample of four comparison schools for District A.

Our sample of comparison schools across the districts thus was four for District A and five for District B, yielding nine comparison schools.

Step Two: Student Matching

To select comparison students, we used all ELLs from the nine comparison schools. We used Mahalanobis matching within propensity calipers and propensity score matching nearest neighbor with calipers.[6] We implemented a 1:4 matching within calipers (the number of standard deviations [*SD*] of the distance measure within which to draw a comparison unit) set to 0.20 *SD* or 0.25 *SD*, as recommended by Rosenbaum and Rubin (1985). Comparison students outside this distance were dropped, allowing us to ensure a closer match than would have been possible without calipers. The 1:4 matching also allowed us to select a comparison group larger than 1:1 matching, thus improving impact estimates. We used the MatchIt package (Ho, Imai, King, & Stuart, 2011) in R.

We matched students within each cohort on data prior to the outcome year. A cohort coincides with each implementation year for a total of four cohorts. The cohorts contain different grades. Recall that 2015–16 and 2016–17 are the two outcome years. Therefore, Cohort 1 (students who could participate in Project RISE starting in 2013–14) contains Grades 9 and 10 because students in those grades could have contributed outcomes in the two outcome years. However, Cohort 1 does not contain Grades 11 and 12 because students in these grades would have graduated before 2015–16 and, therefore, could not have contributed outcome data. The inclusion of grades also was affected by implementation differences, which we described in the implementation section of this report (see Structural Components section on pages 9–11). That is, School B administrators included only Grades 9 and 10 in Project RISE. Appendix C shows all cohorts and grades contributing to outcomes in each school.

We used the following variables in the matching procedure: (a) student English language proficiency test scores (ESL in District A, and ESL1 or ESL2 in District B),[7] (b) special education status, (c) FRPL status, (d) gender, (e) racial or ethnic group, and (f) grade levels. We used the measure of English language proficiency as a measure of prior achievement. We selected this measure during the

[6] We tested Mahalanobis matching within propensity score calipers, which works well when using relatively few covariates and a small sample size (Rubin, 1979) and propensity score matching nearest neighbor with calipers and selected the matching approach that gave the best balance results.

[7] We have named the tests as ESL, ESL1, and ESL2 to preserve districts' anonymity.

planning stages for the current study because research and testing materials available at that time suggested a strong correlation between the ESL tests conducted in Districts A and B and achievement outcomes (e.g., .59 correlation between ESL2 and high school test scores in District B). In addition, once we had access to data for the current study, we found a substantially greater number of students with ESL scores than other test scores that could have been used as a pretest measure of achievement, such as prior mathematics or ELA tests, leading to our decision to continue using ESL tests to preserve a greater number of students in the sample.

Note that students in District A whose home language is not English take the ESL test within 30 days after first enrolling in public schools. If a student is classified as an ELL, that student will take the ESL test each following year until reclassification. In District B, students whose home language is not English take ESL1 once within 10 days of initial enrollment (i.e., these students are new to the district). If classified as an ELL, a student will take ESL2 in the spring of each academic year until reclassification. Most students in our sample were not new to the district and therefore took ESL2. For new students, we fitted matching models using ESL1 scores, which allowed us to retain as many treated students as possible in the analytic sample. All other covariates included in the matching models were the same for both districts. All matching covariates were from the year prior to participation, which varied by cohort. For instance, Cohort 1 students who participated in Project RISE starting in 2013–14 were matched using measures from 2012–13, whereas Cohort 2 students who participated in Project RISE starting in 2014–15 were matched using measures from 2013–14. The cohort table in Appendix C shows the year of baseline data we used to match all cohorts.

Baseline Equivalence

We assessed the equivalence of the matched comparison groups of students by evaluating balance on key baseline measures (student's preintervention achievement score, FRPL status, special education status, gender, and race/ethnicity).[8] Consistent with What Works Clearinghouse standards (IES, 2017), we considered treatment and matched comparison groups to be balanced if the standardized mean difference (SMD) in baseline measures between the two groups of students in the sample was less than or equal to 0.25 *SD.*

We evaluated baseline equivalence separately for each analytic sample (i.e., for each outcome and treatment dosage). Tables of statistics for each sample are presented in Appendix D. The results indicate that the treatment and comparison groups of students were balanced on

[8] Binary dummy variables were created to convert the race/ethnicity categorical variable into binary variables (e.g., White, Black) during the propensity score matching process.

preintervention characteristics for all but two samples: (a) In the sample for mathematics outcome, partial treatment was not balanced on the proportion of Black students (Table D2 in Appendix D), and (b) in the sample for cumulative credit outcome, partial treatment was not balanced on the proportion of special education students (Table D6 in Appendix D). We did not anticipate that the baseline differences on these characteristics would affect the study's findings because they were not considered for selection into treatment. In addition, we controlled for the possible influence of these characteristics in the impact models.

Outcome Measures

We examined the impact of Project RISE on academic achievement, cumulative credit, high school graduation, and attendance. In the following subsections, we describe how each outcome was operationalized and computed. Note that first we computed outcomes in each district and then combined outcomes across districts and cohorts for use in the impact models.

Academic Achievement

We examined two achievement outcomes: mathematics and ELA. In District A, the state's standardized achievement test is given in Grade 11. In District B, students can test in January, June, or August of any grade level. Of all test scores in 2015–16 or 2016–17, we used the highest score as an outcome for students in District B. Therefore, the grade level for test taking in District B could be any grade level. We pooled scores across grades, cohorts, and districts. Prior to pooling data, we standardized test scores in both districts within a grade to capture differences in achievement independent of the grade of test taking. Because we pooled data across grades, cohorts, and outcome years, our impact models for achievement and all other outcomes controlled for these variables, as we discuss in the analysis section.

Cumulative Credit

Cumulative credit is another achievement outcome in the current study. It indicates the number of credits students successfully accumulate throughout high school, with the number typically increasing in higher grades. This outcome does not differentiate between core credits and other credits, such as electives or advanced courses. The information on the type of credits was not available for the impact study. We used the most recent credit accumulation number as an outcome (i.e., 2016–17 if available, otherwise 2015–16). For District A, our data included cumulative credit computed by the district. District B does not report cumulative credit but attempted credits and earned credits by year. To calculate credit accumulation, we summed the earned credits from 2011–12 through 2016–17.

High School Graduation

Graduation in the current study reflects a 4-year or 5-year high school graduation rate; some students in our sample had 5 years to graduate within the duration of Project RISE (e.g., 10th graders who were treated in 2013–14). Graduation is a binary outcome variable. Students who earned a high school diploma or a high school equivalency diploma during 2015–16 or 2016–17 were coded as 1; all others were coded as 0. Available data did not allow us to differentiate the academic standing of students who did not graduate (e.g., whether they dropped out, transferred, or remained enrolled).

Attendance

We computed the attendance rate by dividing the number of days present by the number of days enrolled. We used the most recent attendance rate as an outcome (i.e., 2016–17 if available, otherwise 2015–16). Prior to deciding to use the most recent attendance rate, we compared attendance between the two outcome years and found it similar. The outcome of attendance was within the range of 0 to 1 and highly skewed. Therefore, we transformed it using the logit transformation.

Analytic Approach

We estimated impact separately for all five outcomes and for the full and partial treatment samples. The analysis used the same basic two-level hierarchical linear model. The first level of this model uses students' preintervention achievement test scores and student demographics to estimate the average performance of students within a school. The equation is as follows:

Level 1: Student Level: $Y_{ij} = \pi_{0j} + \pi_{1j}X_{ij} + \varepsilon_{ij}$

where Y_{ij} is the outcome measure for student i of school j; X_{ij} is a set of student-level characteristics, including the preintervention measure of achievement (ESL, ESL1, or ESL2), grade level, the propensity score of being in the treatment group or Mahalanobis distance, cohort (i.e., 1 to 4), outcome year (i.e., 2015–16 or 2016–17), FRPL eligibility, gender, race/ethnicity, and special education status; π_{0j} is the intercept, or the average achievement of students for school j; and ε_{ij} is the student-level random error term, assumed to be independently and identically distributed (i.e., within school residual).

The Level 2 model predicts the average achievement for each school (intercept from the Level 1 model) using school characteristics, including the treatment indicator (*T*). The equation for this model is as follows:

Level 2: School Level: $\pi_{0j} = \beta_{00} + \beta_{01}T_j + \beta_{02}V_j + \beta_{03}D_j + r_{0j}$

where π_{0j} is the average achievement of students for school j; β_{00} is the mean outcome of comparison students; T_j is the treatment indicator and is 1 for Project RISE schools and 0 if otherwise; V_j is a vector of school-level covariates, including the percentage of FRPL eligible students, the percentage of students with a disability, the percentage of students in different racial groups, and the percentage of female students; D_j is a district indicator and is 1 if a school is in District A, and 0 if a school is in District B; and r_{0j} is the school-level random error term. β_{01} is the coefficient of interest or the treatment effect on student outcomes for those students served by Project RISE.[9] Our selection of Level 1 and Level 2 covariates was driven by similar prior research, the contribution of covariates to model fit, and the correlation between them. The results tables in the following section specify the covariates that we used in the final models.

To estimate models for graduation and attendance, we used the following link function to allow the transformed predicted value to take on any real value:

$$Y_{ij} = \log\left(\frac{\varphi_{ij}}{1-\varphi_{ij}}\right)$$

To facilitate the interpretation of regression coefficients, we calculated effect sizes for all outcomes. As the effect size for academic achievement and cumulative credit, we used Hedges' *g* to show group differences in *SD*s (Hedges, 1981). As the effect size for graduation and attendance, we converted log odds ratio to the SMD using *d* = log odds ratio * $\sqrt{3}/\pi$ (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Results

We begin the presentation of the study's findings with a summary of the descriptive statistics for the treatment and comparison groups of students. Next, we present the results for each research question. For each outcome, we present findings for students who received full treatment (2 or more years) and partial treatment (1 year).

[9] Our analysis did not include district × treatment interaction or separate district analyses for two reasons: (a) Each district had one treatment school, and a treatment effect in that district could not be attributed to the program but rather to conditions at that school (*n* = 1 confound), and (b) sample sizes were too small to conduct statistical significance testing for most outcomes for individual districts.

Descriptive Findings

In Table 3, we summarize the descriptive statistics for the study's samples, including unadjusted means and *SD*s by treatment dosage. A review of means suggests varying differences between the groups. For example, although the mathematics mean for the full treatment group favors comparison ELLs, the mathematics mean for the partial treatment group favors treated ELLs. A similar inconsistency is observed for other outcomes apart from ELA: for ELA, the comparison group is favored regardless of the treatment dosage (i.e., the mean for the comparison group is greater than the mean for the treatment group for full and partial treatment groups).

Table 3. Descriptive Statistics of Student Outcomes, by Treatment Definition and Treatment Status

Treatment dosage	Outcome[a]	Treatment			Comparison		
		Mean	SD	N	*Mean*	SD	N
Full treatment	Mathematics	0.14	0.87	74	0.26	0.97	186
	ELA	-0.47	0.80	44	-0.04	0.85	96
	Cumulative credit	150.02	100.65	149	138.00	111.93	361
	Graduation	0.86	0.35	56	0.85	0.36	124
	Attendance	0.91	0.15	144	0.88	0.22	379

Treatment dosage	Outcome[a]	Treatment			Comparison		
		Mean	SD	N	*Mean*	SD	N
Partial treatment	Mathematics	0.47	0.78	104	0.42	0.94	330
	ELA	0.05	1.17	60	0.25	1.00	186
	Cumulative credit	132.19	67.62	27	150.55	65.93	120
	Graduation	0.44	0.51	25	0.62	0.49	66
	Attendance	0.85	0.21	150	0.85	0.23	486

[a]For the outcome of achievement, mean values reflect a difference between groups and the district average for ELLs; for cumulative credit, mean values reflect the total earned credit count of any credit type; for graduation, mean values reflect the proportion of students who graduated in 4 or 5 years; for attendance, mean values reflect the proportion of days attended relative to days enrolled.

What Is the Impact of Project RISE on the Achievement of ELLs in Mathematics and English Language Arts?

Our results, summarized in Table 4, suggest no statistically significant difference between students who received full or partial treatment and comparison students in mathematics or ELA. Because we used standardized test scores in the model, the estimates presented in Table 4 show the difference between groups in *SD*s. Positive values in this and all following analyses indicate that treatment ELLs had a better outcome than comparison ELLs. The effect sizes indicate the difference between treatment and comparison students in *SD*s. Because we have already used standardized scores as outcomes, effect sizes closely resemble regression estimates. However, the effect sizes compare groups to one another directly versus the

regression estimates that compare groups based on how they compare with the district average for ELLs. Effect sizes for the full treatment group are small, whereas effect sizes for the partial group are approaching a medium level.[10]

Table 4. Impact Estimates of Participation in Project RISE on the Achievement of ELLs in Mathematics and English Language Arts, by Treatment Dosage

Outcome	Full treatment (math, *n* = 260)/(ELA, *n* = 140)		Partial treatment (math, *n* = 434)/(ELA, *n* = 246)	
	Estimate (standard error)	Effect size	Estimate (standard error)	Effect size
Mathematics	-0.16 (0.46)	-0.17	0.40 (0.70)	0.44
ELA	0.29 (0.50)	0.35	-0.43 (0.67)	-0.41

Note. Student-level covariates: cohort, grade, pretest, outcome year, race, FRPL eligibility, and special education status. School-level covariates: proportion female, proportion FRPL eligible, proportion special education, proportion in different racial categories, and district.

What Is the Impact of Project RISE on Credit Accumulation of ELLs?

Our results suggest no statistically significant difference between ELLs who received full or partial treatment and comparison ELLs on credit accumulation. We summarize these results in Table 5. The regression estimates in Table 5 show the average difference in credit accumulation for treated students relative to comparison students. Effect sizes for this analysis suggest small differences between the groups.

Table 5. Impact Estimates of Participation in Project RISE on Credit Accumulation, by Treatment Dosage

Outcome	Full treatment (*n* = 510)		Partial treatment (*n* = 659)	
	Estimate (standard error)	Effect size	Estimate (standard error)	Effect size
Cumulative credits	5.65 (26.91)	0.05	-15.76 (13.38)	-0.26

Note. Student-level covariates: cohort, grade, pretest, outcome year, race, FRPL eligibility, and special education status. School-level covariates: proportion female, proportion FRPL eligible, proportion special education, proportion in different racial categories, and district.

[10] We considered effect sizes small if they were less than or equal to 0.2, medium if they were greater than 0.2 but less than or equal to 0.5, and large if they were greater than 0.5.

What Is the Impact of Project RISE on the High School Graduation of ELLs?

Our results, summarized in Table 6, suggest no statistically significant differences between ELLs who received full or partial treatment and comparison ELLs in graduation rates. The regression estimates in Table 6 show the average difference in graduation rates for treated students relative to comparison students. The effect sizes suggest a small difference between ELLs who received full treatment and comparison ELLs and a negligible difference between ELLs who received partial treatment and comparison ELLs.

Table 6. Impact Estimates of Participation in Project RISE on High School Graduation of ELLs, by Treatment Dosage

Outcome	Full Treatment (*n* = 180)		Partial Treatment (*n* = 91)	
	Estimate (standard error)	Effect size	Estimate (standard error)	Effect size
Graduation	-0.144 (3.390)	-0.01	0.594 (0.975)	0.22

Note. Student-level covariates: pretest, racial minority, FRPL eligibility, gender, and distance. School-level covariates: proportion FRPL eligible, proportion special education, proportion of White students, and district.

What Is the Impact of Project RISE on the Attendance of ELLs?

Our results suggest no statistically significant differences between ELLs who received full or partial treatment and comparison ELLs in attendance rates. The regression estimates in Table 7 show the average difference in attendance rates for treated students relative to comparison students. The effect sizes suggest a medium difference between ELLs who received partial treatment and comparison ELLs and a small difference between ELLs who received full treatment and comparison ELLs.

Table 7. Impact Estimates of Participation in Project RISE on the Attendance of ELLs, by Treatment Dosage

Outcome	Full treatment (*n* = 523)		Partial treatment (*n* = 636)	
	Estimate (standard error)	Effect size	Estimate (standard error)	Effect size
Attendance	-0.058 (0.656)	-0.03	-0.953 (0.690)	-0.49

Note. Student-level covariates: cohort, outcome year, pretest, race/ethnicity, eligibility for FRPL, special education status, and distance. School-level covariates: proportion female, proportion FRPL eligible, proportion special education, proportion in different racial categories, and district.

Conclusions

We examined the impact of Project RISE on the achievement of ELLs in mathematics and ELA, credit accumulation, high school graduation, and school attendance. We examined the impact for ELLs who received full treatment (i.e., taught by four or more Project RISE teachers for 2 or more years) and partial treatment (i.e., taught by four or more Project RISE teachers for 1 year). We found no statistically significant impact of Project RISE on the outcomes we examined. Recall that Project RISE was supported by an i3 development grant, which enables the development of innovative practices. Therefore, the implementation of Project RISE had a degree of flexibility to allow the program planners to develop an intervention that could best meet the needs of schools, teachers, and students. For example, in the formative evaluation section of this report, we noted that School B deviated from the program design in how the school administration scheduled students into heterogeneous cohorts. This and other deviations from the original design, although they presented themselves as necessary for implementation, could have affected the program's ability to produce a statistically significant impact. In fact, not producing statistically significant impact is common among interventions under development. A recent analysis of i3 interventions found that only 8% of development grants produced a significant impact on student outcomes versus 50% of scale-up grants and 40% of validation grants that implemented previously tested interventions (Boulay et al., 2018).

Another characteristic of interventions under development is that they are implemented on a smaller scale than tested interventions. This was the case with Project RISE, in which two schools participated. The small sample size of two treated schools and nine matched comparison schools affected our ability to detect statistically significant findings. In addition, the small sample size affected the precision of our statistical models by producing inflated standard errors. We observed inflated standard errors for all estimates of impact. Large standard errors suggest that the analytic models had low predictive ability, possibly affecting both the precision and the direction of the estimates.

Finally, the study used a quasi-experimental design. Although it is the strongest design for causal inference in the absence of random assignment, its main limitation is the use of observable characteristics to construct a comparison group, which overlooks unobservable characteristics that can impact findings (Shadish, Cook, & Campbell, 2002). Examining Project RISE using random assignment, coupled with a sufficient sample and faithful implementation, will help make confident conclusions about impact. Given the design limitations of our impact evaluation combined with additional limitations outlined in the preceding discussion, we encourage a guarded interpretation of the impact findings.

References

Boulay, B., Goodson, B., Olsen, R., McCormick, R., Darrow, C., Frye, M., . . . Sama, M. (2018). *The Investing in Innovation Fund: Summary of 67 evaluations: Final report.* Washington, DC: U.S. Department of Education. Retrieved from https://ies.ed.gov/ncee/pubs/20184013/pdf/20184013.pdf

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to meta-analysis*. New York, NY: Wiley.

Hedges, L. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128.

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2011*).* MatchIt: Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software, 42*(8), 1–28.

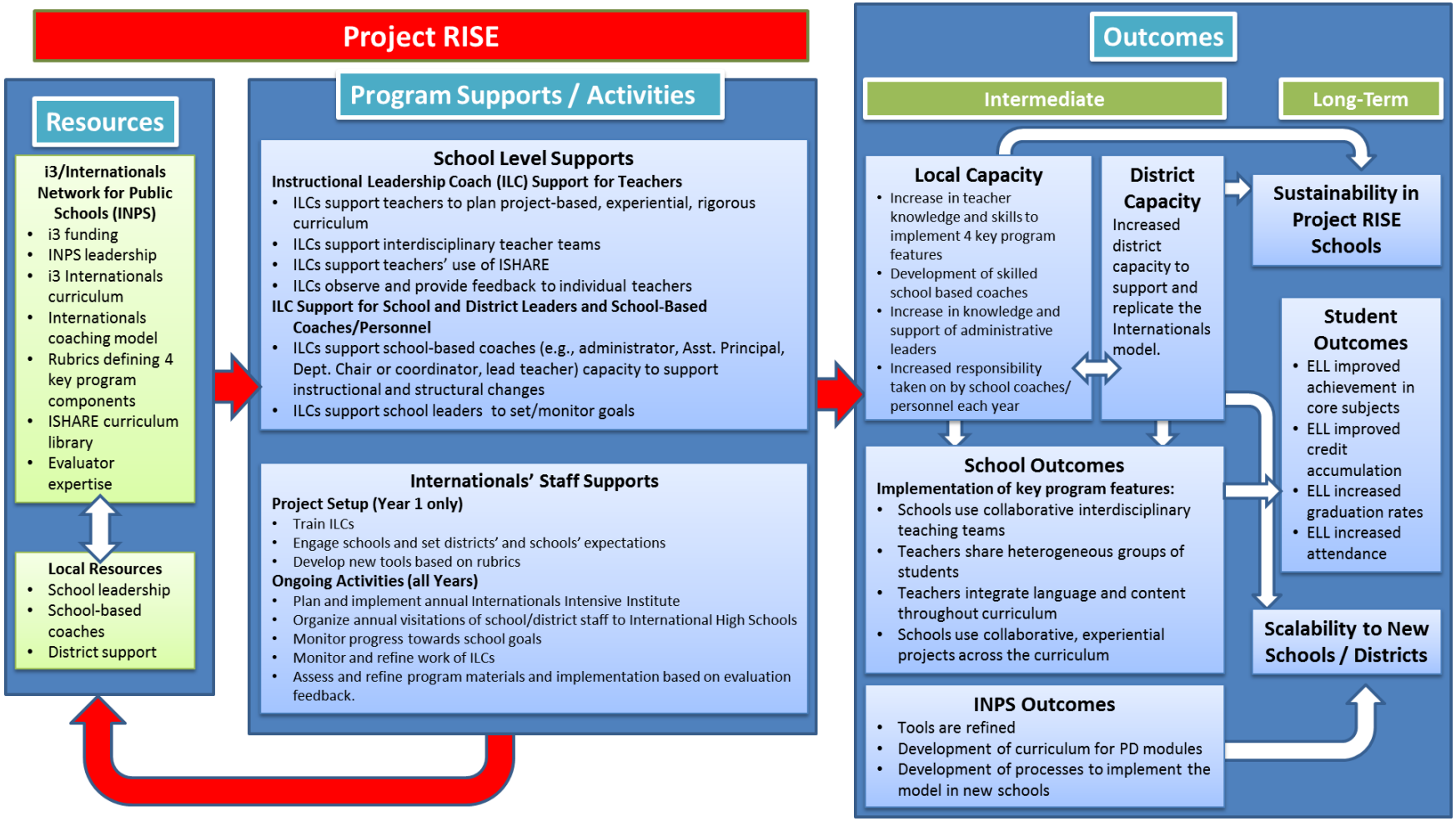
IES. (2017). *What Works Clearinghouse: Procedures Handbook Version 4.0.* Washington, DC: U.S. Department of Education, Institute of Education Sciences. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_procedures_handbook_v4.pdf

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

Rubin, D. B. (1979). Using multivariate matched sampling and regression adjustment to control bias in observational studies. *Journal of the American Statistical Association, 74*(366a), 318–328.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal infeence.* Boston, MA: Houghton Mifflin.

Appendix A. Project RISE Logic Model



Appendix B. Project RISE Fidelity of Implementation Matrix

No.	Indicator of Component	Levels of Fidelity	2013–14		2014–15		2015–16		2016–17	
			A	B	A	B	A	B	A	B
			Score	Score	Score	Score	Score	Score	Score	Score
Key Component 1: Instructional Leadership Coach (ILC) Supports										
ILC Support for Teachers										
1	Support teachers to plan instruction.	1 = ILCs are on-site less than 40% of expected days. [1] 2 = ILCs are on-site 40% to 60% of expected days. 3 = ILCs are on-site more than 60% of expected days.	3	3	3	3	3	3	3	3
2	Form interdisciplinary teacher teams.	1 = ILCs are on-site less than 40% of expected days. 2 = ILCs are on-site 40% to 60% of expected days. 3 = ILCs are on-site more than 60% of expected days.	3	3	3	3	3	3	3	3
3	Teachers use Internationals virtual repository.	1 = ILCs are on-site less than 40% of expected days. 2 = ILCs are on-site 40% to 60% of expected days. 3 = ILCs are on-site more than 60% of expected days.	3	3	3	3	3	3	3	3
4	Provide feedback to individual teachers, including meetings, observations, coplanning, written feedback, and discussions.	1 = Feedback occurs less than once per week. [2] 2 = Feedback occurs at least once per week.	1	1	2	2	2	2	2	2

No.	Indicator of Component	Levels of Fidelity	2013–14		2014–15		2015–16		2016–17	
			A	B	A	B	A	B	A	B
			Score	Score	Score	Score	Score	Score	Score	Score
ILC Support for School Leaders and School-Based Coaches and Personnel										
5	ILCs support the capacity of school-based coaches to support instructional and structural changes.	1 = ILC meets with school coach less than four times each year. 2 = ILC meets with school coach four to eight times. 3 = ILC meets with school coach more than eight times.	1	3	3	3	3	3	3	3
6	ILCs support school leaders to set and monitor goals.	1 = ILC meets with principal less than twice each year to discuss goals. 2 = ILC meets with principal at least twice each year.	2	2	2	2	2	2	2	2
		Sample-Level Fidelity for Indicators 1–6	*Not Adequate*		*Adequate*		*Adequate*		*Adequate*	
Key Component 1: Threshold for Implementation Fidelity										
Indicators		**School-Level Scores**	**Sample-Level Fidelity Threshold**							
1–6		LOW: Any indicator < 2 MEDIUM: All indicators ≥ 2 HIGH: All indicators ≥ 2 and at least two indicators > 2	*Adequate threshold* of implementation fidelity achieved if both schools are at MEDIUM score or higher.							
Key Component 2: Internationals Supports for Project Set-up										
7	Train ILCs.	1 = ILC not yet hired and trained at beginning of school year. 2 = ILC hired and trained at beginning of school year.	2	2	2	N/A	N/A	N/A	N/A	N/A

No.	Indicator of Component	Levels of Fidelity	2013–14		2014–15		2015–16		2016–17	
			A	B	A	B	A	B	A	B
			Score	Score	Score	Score	Score	Score	Score	Score
8	Engage schools and set the expectations of districts and schools.	1 = Memoranda of understanding (MOUs) not signed, or MOUs contain minimal language describing district and school roles and expectations. 2 = MOUs are signed, and they specify roles.	2	2	N/A	N/A	N/A	N/A	N/A	N/A
9	Develop and modify tools and professional development to be applicable to a traditional, comprehensive high school.	1 = Tools and professional development modules are not modified. 2 = Tools and professional development modules are modified.	2	2	N/A	N/A	N/A	N/A	N/A	N/A
		Sample-Level Fidelity for Indicators 7–9	*Adequate*		*Adequate*		*N/A*		*N/A*	
Key Component 2: Threshold for Implementation Fidelity										
Indicators (Year 1 Only or Additional Years If New Hires)		**School-Level Scores**	**Sample-Level Fidelity Threshold**							
7–9		LOW: Any indicator < 2 HIGH: All indicators = 2	*Adequate threshold* of implementation fidelity achieved if both schools are at HIGH score.							
Key Component 3: Internationals Support for Teachers and School Leaders										
10	Plan and implement annual training institute.	1 = 0% to 30% of teachers attend (each day if multiple days are offered). 2 = 31% to 60% of teachers attend (each day if multiple days are offered).	3	2	3	3	3	3	3	2

No.	Indicator of Component	Levels of Fidelity	2013–14		2014–15		2015–16		2016–17	
			A	B	A	B	A	B	A	B
			Score	Score	Score	Score	Score	Score	Score	Score
		3 = More than 61% of teachers attend (each day if multiple days are offered).								
11	Organize annual visitations of school and district staff to Internationals high schools.	1 = Fewer than 50% of invited participants attend one visit. 2 = Fewer than 50% of invited participants attend two visits. 3 = More than 50% of invited participants attend two visits.	3	3	3	2	1	1	1	1
12	Monitor progress toward school goals (meetings take place after a walk-through).	1 = One formal meeting held per year to monitor progress. 2 = Two formal meetings held per year; one per semester.	2	2	2	2	2	2	2	2
13	Monitor and refine work of ILCs.	1 = No formal evaluation conducted, no leadership team observations conducted, or no direction provided to ILCs by leadership team. 2 = Formal evaluation and observation of ILC conducted.	2	2	2	2	2	2	2	2
		Sample-Level Fidelity for Indicators 10–13	*Adequate*		*Adequate*		*Not Adequate*		*Not Adequate*	

No.	Indicator of Component	Levels of Fidelity	2013–14		2014–15		2015–16		2016–17	
			A	B	A	B	A	B	A	B
			Score	Score	Score	Score	Score	Score	Score	Score
Key Component 3: Threshold for Implementation Fidelity										
Indicators		**School-Level Scores**	**Sample-Level Fidelity Threshold**							
10–13		LOW: Any indicator < 2	*Adequate threshold* of implementation fidelity achieved if both schools are at MEDIUM score or higher.							
		MEDIUM: All indicators ≥ 2								
		HIGH: All indicators ≥ 2 and at least one indicator > 2								
Key Component 3: School-Level Supports [3]										
14	School leadership designates a school-based staff member to support instructional and structural changes.	1 = A school-based staff member not designated.	–	–	3	3	3	3	3	3
		2 = A school-based staff member designated for a portion of school year.								
		3 = A school-based staff member designated for the duration of school year.								
15	School leadership establishes a structure that supports sharing of students by Project RISE teachers.	1 = Project teachers are not sharing students.	–	–	3	2	3	2	3	2
		2 = Up to one-half of Project RISE teachers are sharing students.								
		3 = One-half to all Project RISE teachers are sharing students.								
16	School leadership designates clear time for Project RISE team meetings.	1 = Team meetings are not taking place.	–	–	2	2	2	2	2	2
		2 = Team meetings are taking place once per week.								
		Sample-Level Fidelity	–	–	*Adequate*		*Adequate*		*Adequate*	

No.	Indicator of Component	Levels of Fidelity	2013–14		2014–15		2015–16		2016–17	
			A	B	A	B	A	B	A	B
			Score	Score	Score	Score	Score	Score	Score	Score
Key Component 4: Threshold for Implementation Fidelity (Measured Starting in Year 2)										
Indicators		**School-Level Scores**	**Sample-Level Fidelity Threshold**							
14–16		LOW: Any indicator < 2	*Adequate threshold* of implementation fidelity achieved if both schools are at MEDIUM score or higher.							
		MEDIUM: All indicators ≥ 2								
		HIGH: All indicators ≥ 2 and at least one indicator > 2								

[1] The expected days were computed based on the academic calendars. For example, in 2016–17, that number was 18 days for School A and 17 days for School B (excluding vacation weeks and weeks with 2 or more holidays and 1 week for testing), translating to approximately 11 days for School A and 10 days for School B, for the highest implementation threshold of 60%.

[2] The expected days were computed based on the academic calendars. For example, in 2016–17, School A was in session 36 weeks (excluding vacation weeks and weeks with 2 or more holidays and 1 week for testing), and School B was in session 34 weeks, with ILCs expected to be on-site every other week. This translates into 17–18 observations/feedback sessions needed for the highest threshold of implementation fidelity. For School A, in 2016–17, the AIR study team scored 17 out of 18 as a "2," because the frequency shows that the ILC essentially provided feedback once per week, thereby reaching the highest threshold of implementation fidelity.

[3] Key component 3 was introduced to the fidelity matrix in 2014–15.

Appendix C. Student Cohorts

Figures C1 and C2 show the composition of student cohorts in each school. The study included four cohorts who could have received full (2 or more years) or partial (1 year) treatment depending on the year that each cohort could have participated in Project RISE. Because the study had two outcome years (2015–16 and 2016–17), some cohorts at School A included different grades. Grades 9 and 10 in Cohort 1 and Grades 9–11 in Cohort 2 at School A were included in the study. Students in these grades would still be enrolled in high school in 2015–16 or 2016–17, whereas students in other grades would have likely graduated. Because of how Project RISE was implemented at School B, each cohort included only Grades 9 and 10. The baseline year varies for the cohorts, which reflects the different time that each cohort could have started participation in Project RISE.

Figure C1. Study Cohorts at School A: Grade Composition, Possible Treatment Dosage, and Baseline Years

	Cohort 1		Cohort 2			Cohort 3				Cohort 4			
Year 0 **2012–13**	8th grade Baseline	9th grade Baseline											
	↓	↓											
Program Year 1 **2013–14**	**9th Grade** 1 Yr of Trt	**10th Grade** 1 Yr of Trt	8th grade Baseline	9th grade Baseline	10th grade Baseline								
	↓	↓	↓	↓	↓								
Program Year 2 **2014–15**	**10th Grade** 2 Yrs of Trt	**11th Grade** 2 Yr of Trt	**9th Grade** 1 Yr of Trt	**10th Grade** 1 Yr of Trt	**11th Grade** 1 Yr of Trt	8th grade Baseline	9th grade Baseline	10th grade Baseline	11th grade Baseline				
	↓	↓	↓	↓	↓	↓	↓	↓	↓				
Program Year 3 **2015–16**	**11th Grade** 3 Yrs of Trt	**12th Grade** 3 Yrs of Trt	**10th Grade** 2 Yrs of Trt	**11th Grade** 2 Yr of Trt	**12th Grade** 2 Yr of Trt	**9th Grade** 1 Yr of Trt	**10th Grade** 1 Yr of Trt	**11th Grade** 1 Yr of Trt	**12th Grade** 1 Yr of Trt	8th grade Baseline	9th grade Baseline	10th grade Baseline	11th grade Baseline
	↓	↓	↓	↓		↓	↓	↓		↓	↓	↓	↓
Program Year 4 **2016–17**	**12th Grade** 4 Yrs of Trt		**11th Grade** 3 Yrs of Trt	**12th Grade** 3 Yrs of Trt		**10th Grade** 2 Yrs of Trt	**11th Grade** 2 Yr of Trt	**12th Grade** 2 Yr of Trt		**9th Grade** 1 Yr of Trt	**10th Grade** 1 Yr of Trt	**11th Grade** 1 Yr of Trt	**12th Grade** 1 Yr of Trt

Figure C2. Study Cohorts at School B: Grade Composition, Possible Treatment Dosage, and Baseline Years

	Cohort 1		Cohort 2		Cohort 3		Cohort 4	
Year 0 **2012–13**	8th grade Baseline	9th grade Baseline						
	↓	↓						
Program Year 1 **2013–14**	**9th Grade** 1 Yr of Trt	**10th Grade** 1 Yr of Trt	8th grade Baseline	9th grade Baseline				
	↓	↓	↓	↓				
Program Year 2 **2014–15**	**10th Grade** 2 Yrs of Trt	**11th Grade** 2 Yr of Trt	**9th Grade** 1 Yr of Trt	**10th Grade** 1 Yr of Trt	8th grade Baseline	9th grade Baseline		
	↓	↓	↓	↓	↓	↓		
Program Year 3 **2015–16**	**11th Grade** 3 Yrs of Trt	**12th Grade** 3 Yrs of Trt	**10th Grade** 2 Yrs of Trt	**11th Grade** 2 Yr of Trt	**9th Grade** 1 Yr of Trt	**10th Grade** 1 Yr of Trt	8th grade Baseline	9th grade Baseline
	↓	↓	↓	↓	↓	↓	↓	↓
Program Year 4 **2016–17**	**12th Grade** 4 Yrs of Trt		**11th Grade** 3 Yrs of Trt	**12th Grade** 3 Yrs of Trt	**10th Grade** 2 Yrs of Trt	**11th Grade** 2 Yr of Trt	**9th Grade** 1 Yr of Trt	**10th Grade** 1 Yr of Trt

Appendix D. Baseline Equivalence of the Analytic Samples

The tables in this appendix show baseline characteristics for all analytic samples we used in the study. The last column in each table shows the SMD between treatment and comparison groups of students.

Academic Achievement: Mathematics

Table D1. Treatment and Comparison Group Characteristics at Baseline for Mathematics: Full Treatment

Baseline covariates	Treatment (*n* = 74)	Matched comparison (*n* = 186)	SMD
Preintervention achievement	-0.87	-0.85	0.01
Female	31%	29%	0.05
FRPL eligibility	84%	83%	0.02
Special education status	3%	2%	0.06
Asian	50%	61%	0.22
White	23%	20%	0.08
Hispanic	22%	16%	0.15
Black	5%	3%	0.15

Table D2. Treatment and Comparison Group Characteristics at Baseline for Mathematics: Partial Treatment

Baseline covariates	Treatment (*n* = 104)	Matched comparison (*n* = 330)	SMD
Preintervention achievement	-0.53	-0.54	0.01
Female	31%	30%	0.01
FRPL eligibility	80%	81%	0.04
Special education status	0%	0%	<0.01
Asian	30%	29%	0.01
White	45%	47%	0.03
Hispanic	20%	24%	0.09
Black	5%	0%	0.27

Academic Achievement: ELA

Table D3. Treatment and Comparison Group Characteristics at Baseline for ELA: Full Treatment

Baseline covariates	Treatment (*n* = 44)	Matched comparison (*n* = 96)	SMD
Preintervention achievement	-0.92	-0.84	0.10
Female	34%	28%	0.12
FRPL eligibility	86%	89%	0.09
Special education status	5%	3%	0.10
Asian	66%	68%	0.05
White	11%	11%	0.03
Hispanic	23%	18%	0.12
Black	0%	1%	0.14

Table D4. Treatment and Comparison Group Characteristics at Baseline for ELA: Partial Treatment

Baseline covariates	Treatment (*n* = 60)	Matched comparison (*n* = 186)	SMD
Preintervention achievement	-0.39	-0.33	0.06
Female	23%	25%	0.03
FRPL eligibility	82%	80%	0.05
Special education status	0%	0%	<0.01
Asian	30%	34%	0.10
White	55%	51%	0.08
Hispanic	12%	14%	0.07
Black	3%	0%	0.21

Credit Accumulation

Table D5. Treatment and Comparison Group Characteristics for Credit Accumulation: Full Treatment

Baseline covariates	Treatment (*n* = 149)	Matched comparison (*n* = 361)	SMD
Preintervention achievement	-0.66	-0.66	0.01
Female	38%	35%	0.05
FRPL eligibility	86%	87%	0.04
Special education status	1%	1%	0.02
Asian	54%	59%	0.11
White	16%	18%	0.05
Hispanic	26%	21%	0.12
Black	3%	0.00	0.22

Table D6. Treatment and Comparison Group Characteristics for Credit Accumulation: Partial Treatment

Baseline covariates	Treatment (*n* = 149)	Matched Comparison (*n* = 361)	SMD
Preintervention achievement	-0.32	-0.40	0.14
Female	54%	56%	0.06
FRPL eligibility	58%	69%	0.21
Special education status	29%	17%	0.31
Asian	50%	56%	0.15
White	0%	3%	0.24
Hispanic	50%	41%	0.23
Black	0%	0%	<0.01

Graduation

Table D7. Treatment and Comparison Group Characteristics for Graduation: Full Treatment

Baseline covariates	Treatment (*n* = 56)	Matched comparison (*n* = 124)	SMD
Preintervention achievement	-0.2	-0.16	0.05
Female	40%	45%	0.09
FRPL eligibility	92%	93%	0.04
Special education status	0%	0%	<0.01
Asian	74%	70%	0.10
White	10%	7%	0.10
Hispanic	13%	21%	0.21
Black	0%	0%	<0.01

Table D8. Treatment and Comparison Group Characteristics for Graduation: Partial Treatment

Baseline covariates	Treatment (*n* = 25)	Matched comparison (*n* = 66)	SMD
Preintervention achievement	-0.69	-0.76	0.10
Female	35%	36%	0.02
FRPL eligibility	80%	80%	0.00
Special education status	0%	0%	<0.01
Asian	32%	32%	0.00
White	32%	32%	0.01
Hispanic	35%	36%	0.01
Black	0%	0%	<0.01

Attendance

Table D9. Treatment and Comparison Group Characteristics for Attendance: Full Treatment

Baseline covariates	Treatment (*n* = 144)	Matched comparison (*n* = 379)	SMD
Preintervention achievement	-0.65	-0.64	0.00
Female	39%	36%	0.05
FRPL eligibility	86%	85%	0.04
Special education status	1%	1%	0.01
Asian	53%	54%	0.03
White	17%	19%	0.06
Hispanic	24%	25%	0.01
Black	3%	0%	0.23

Table D10. Treatment and Comparison Group Characteristics for Attendance: Partial Treatment

Baseline covariates	Treatment (*n* = 150)	Matched comparison (*n* = 486)	SMD
Preintervention achievement	-0.6	-0.55	0.04
Female	37%	36%	0.01
FRPL eligibility	76%	79%	0.06
Special education status	5%	2%	0.12
Asian	31%	31%	0.01
White	39%	39%	0.02
Hispanic	27%	29%	0.04
Black	3%	1%	0.19

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

Established in 1946, the American Institutes for Research (AIR) is an independent, nonpartisan, not-for-profit organization that conducts behavioral and social science research on important social issues and delivers technical assistance, both domestically and internationally, in the areas of education, health, and workforce productivity.

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH
1000 Thomas Jefferson Street NW
Washington, DC 20007-3835 | 202.403.5000
www.air.org

LOCATIONS

Domestic: Washington, DC (HQ) | Monterey, Sacramento, and San Mateo, CA | Atlanta, GA | Honolulu, HI | Chicago and Naperville, IL Indianapolis, IN | Metairie, LA | Waltham, MA | Frederick and Rockville, MD | Chapel Hill, NC | New York, NY | Columbus, OH | Cayce, SC Austin, TX | Reston, VA

International: El Salvador | Ethiopia | Haiti | Honduras | Zambia

5444_08/18