Evaluation of an Expansion Strategy for the Assessment-to-Instruction Professional Support System

Pei Zhu
Emma Alterman
Nicholas Commins
Peyton Nash

June 2023

mdrc

BUILDING KNOWLEDGE
TO IMPROVE SOCIAL POLICY

Evaluation of an Expansion Strategy for the Assessment-to-Instruction Professional Support System

Pei Zhu, Emma Alterman, Nicholas Commins, and Peyton Nash

mdrc

BUILDING KNOWLEDGE
TO IMPROVE SOCIAL POLICY

JUNE 2023

FUNDERS

For information about MDRC and copies of our publications, see our website: www.mdrc.org.

OVERVIEW

During the early elementary grades, many students do not achieve literacy proficiency because they do not receive effective personalized literacy instruction that addresses their needs. With strong evidence of effectiveness, the Assessment-to-Instruction (A2i) Professional Support System aims to meet these needs by combining data-driven technology with professional development activities to help teachers use differentiated small-group instruction to improve literacy achievement among students in Kindergarten through Grade 3. In 2017, the United2Read team was awarded a five-year expansion grant under the U.S. Department of Education's Educational Innovation and Research (EIR) program to study how to efficiently bring A2i and its evidence-driven literacy instructional practices to elementary school classrooms across the nation.

This report from the independent evaluation of this EIR expansion project examines the relative effectiveness of two support models—a mixed-mode professional development delivery model that aims to reduce costs through a combination of in-person and virtual teacher coaching, and the tested, resource-intensive, face-to-face delivery model—in creating robust implementations of the A2i Professional Support System and improvements in student literacy achievement. Fifty-nine schools in 20 school districts were randomly assigned either to a group of 30 schools that received the mixed-mode model or to a group of 29 schools that received the face-to-face model for three school years (2018–2019 to 2020–2021). Comparing student and teacher experiences in the two sets of schools measures the relative effectiveness of the two models. If both models are implemented adequately, a finding of no difference in the effects of the two models would indicate that they are equally effective approaches to implementing A2i. This report describes the evaluation and presents its findings, key among which are the following:

- The two professional development (PD) models had been in place for less than two school years when the COVID-19 pandemic struck in March 2020. During the pre-pandemic period, teachers in mixed-mode and face-to-face schools received training and support that differed in amounts and delivery modes as intended. However, in both groups of schools, teachers did not fully utilize the A2i technology and had difficulty differentiating instruction based on A2i-generated recommendations.

- In the second half of the implementation period, in-person coaching was not possible due to COVID-19-related school disruptions, which in turn muted the contrast in the amount and types of PD support received by teachers in mixed-mode and face-to-face schools. Teachers in both groups of schools also reduced their use of A2i components and differentiated small-group instruction.

- Across all three study years, student achievement in reading was similar in mixed-mode and face-to-face schools, and about 50 percent of the third-graders in the study were reading at or above proficiency level as defined by state or district standards at the end

of the evaluation. However, because A2i was not fully implemented over the study period, these results do not reflect the actual relative effectiveness of the two PD models.

- The cost differential between the two PD models was small because a large proportion of the cost was made up of fixed expenses that were equally distributed across students in both groups, and the COVID-19-related changes in PD delivery muted the difference in variable costs in later implementation years.

- The program reached schools beyond the study sample in the first two years, but adjustments had to be made to the expansion strategy during the pandemic to address the expansion goals.

CONTENTS

LIST OF EXHIBITS

TABLE

LIST OF EXHIBITS *(CONTINUED)*

FIGURE

ACKNOWLEDGMENTS

INTRODUCTION

How to effectively and efficiently improve student reading skills remains a pressing issue in education. Reading is a critical skill closely linked to academic and life success. Students who do not attain proficient literacy skills face serious challenges both academically and personally. They are more likely to be referred to special education, to repeat a grade, to drop out of high school, or to be arrested.[1] In recent decades, improving reading instruction and student reading achievement has been a major focus of federal, state, and local education programs. With progress in understanding the characteristics of effective reading instruction, student reading achievement had improved slowly over the 20 years prior to the onset of the COVID-19 pandemic in the spring of 2020. The National Assessment of Educational Progress reported that the percentage of fourth-graders reading at the "basic" level increased from 60 percent in 1998 to 66 percent in 2019.[2]

This trend of slow growth, however, has been reversed by the pandemic and its ensuing disruption to K-12 schooling across the country. According to the latest long-term trend analyses, average reading scores for 9-year-old students in 2022 declined by 5 points compared to 2020 scores. This represents the largest average score decline in reading since 1990.[3] As a result, in 2022 these 9-year-old students were reading at a similar level to two decades ago.[4] Such an unforeseen reversal heightens the need for evidence-based effective interventions that can meet students where they are, help them recover unfinished learning, and get them back on track in the post-COVID-19 era.

One of the reasons many students fail to attain proficient reading skills is that they do not receive the amounts and types of instruction needed for them to reach their potential.[5] Differentiated literacy instruction tailored to meet individual students' needs is one important way to help students learn effectively and achieve reading proficiency.[6] There is growing evidence that differentiated instruction informed by assessments of students' reading abilities improves learning more effectively than one-size-fits-all instruction

---

1. National Institute of Child Health and Human Development (2000); Reynolds and Ou (2004).

2. Such rates are lower for high needs students. For example, only 51 percent of students eligible for the National School Lunch Program scored above the basic level with 21 percent of them achieving the proficiency level. See National Assessment of Educational Progress (NAEP, 2019). NAEP is a congressionally authorized project of the National Center for Education Statistics within the Institute of Education Sciences in the U.S. Department of Education. It is known as "The Nation's Report Card." NAEP tests have been conducted periodically in a number of subject areas since 1969.

3. Carr (2022).

4. National Assessment of Educational Progress (2022).

5. Connor et al. (2009a).

6. Effective differentiated instruction often entails proactive differentiation in content, materials, and pacing, as well as the flexible use of small teaching-learning groups in the classroom. It addresses students' varying readiness levels, interests, and learning profiles, and has been found to be beneficial to student learning (Tomlinson et al., 2003).

observed in many classrooms.[7] However, even though teachers receive a variety of assessment and benchmarking data about their students, they are not adequately supported in using such data to determine which instructional practices are effective for which students, and they do not know how to integrate multiple practices to meet the diverse needs of students in busy classrooms and adjust instruction according to classroom and school contexts, such as changing levels of support or reading curricula.[8] As a result, it is challenging for teachers to implement differentiated instruction in the classroom.

The Assessment-to-Instruction (A2i) Professional Support System, developed by Carol Connor and Learning Ovations, Inc., provides a useful tool to bridge this assessment-to-instruction gap. The A2i system combines data-driven technology and professional development (PD) to support teachers' use of differentiated small-group instruction to improve literacy achievement for students in Kindergarten through Grade 3. Specifically, it uses information collected through a set of online adaptive reading assessments to compute recommended amounts and types of literacy instruction for each student, suggests ways to group students with similar needs, gives aligned lessons drawn from the curricula the teachers are using, and provides such information to teachers through a web-based platform. The system also offers embedded lesson planning tools to make it easier to follow through on these recommendations. To support teachers' use of the technology and help them implement differentiated small-group instruction in the classroom, Learning Ovations employed trained coaches to provide training throughout the implementation period.

The A2i system has compelling evidence of efficacy based on randomized controlled trials and quasi-experiments conducted in 28 schools in multiple districts since 2005. These studies repeatedly demonstrated that schools using A2i could accelerate gains in literacy during the crucial early elementary grades for all students, including many children from high poverty families.[9] In these studies of A2i, the system was delivered to schools and teachers through intensive in-person training and ongoing PD tailored to teachers' understanding, knowledge, and usage of the system. This high level of personal attention requires a substantial investment in training and support. Unfortunately, that characteristic makes it difficult and costly to implement A2i on a large scale.

In 2017, the United2Read team, comprised of Digital Promise Global, Learning Ovations, University of California, Irvine, and MDRC, was awarded a five-year expansion grant under the U.S. Department of Education's Educational Innovation and Research (EIR) grant to study how to bring A2i and its evidence-driven literacy instructional practices to elementary

---

7. Connor et al. (2011a); Connor et al. (2011b).

8. Brighton et al. (2005); Tomlinson (1995).

9. See Connor et al. (2007) and Conner et al. (2011a). Note that the sample of schools and students included in these prior studies differ from those included in the current study in that the study sample in this report includes a higher proportion of students of color and students from disadvantaged backgrounds.

school classrooms across the nation. Under this grant, Learning Ovations designed an alternative, mixed-mode PD delivery model that aims to lower costs by harnessing the power of technology. Unlike the resource-intensive, face-to-face PD approach similar to that used in the original studies of A2i, this mixed-mode model supports schools' and teachers' implementation of A2i through a combination *of in-person and virtual* teacher coaching, with a substantial proportion of the PD activities taking place remotely through video conferencing and other technologies. In addition, this mixed-mode model includes fewer PD events than the face-to-face model. This approach aims to sustain A2i's emphasis on personalized professional development while improving coaches' productivity by reducing travel time and allowing each coach to work with more schools and teachers.

As part of the EIR grant, MDRC conducted an independent evaluation to study the relative effectiveness of these two support models—the mixed-mode PD delivery model and the face-to-face PD delivery model—in producing robust implementation of the A2i Professional Support System and improvements in student literacy achievement. While this study also touches on some other grant activities, it focuses on addressing the following questions:

1. How well were the two PD delivery models implemented? Did PD implementation differ between the two models in delivery mode and frequency as expected? How much did teachers participate in the offered PD activities?

2. Did the two models affect teachers' use of the A2i technology differently, and if so, how?

3. Did these two PD delivery models produce differences in teachers' views of A2i and their instructional practices, and if so, how?

4. Did these two PD delivery models lead to differences in the cumulative effect of A2i on students' reading achievement three years after the program started?

In 2018, MDRC began a multiyear evaluation to assess these research questions. The study recruited 59 elementary schools from 20 school districts in 4 states to participate in this evaluation and, within each district or district group, randomly assigned about half of them to receive PD through the face-to-face model (the face-to-face schools) and the other half through the mixed-mode model (the mixed-mode schools).[10] In the 2018–2019 school year (Year 1), the two models were implemented in Kindergarten and Grade 1; in 2019–2020 (Year 2), the program was expanded to include Grade 2; and in 2020–2021 (Year 3), the program included Grade 3. The study assesses the differential effects of these two models by comparing the average outcomes between the two groups of schools. If both models were implemented adequately, a finding of no difference in the effects of the two models

---

10. The team recruited 60 schools. However, in one district, one school serves only Grades K-1, and another school serves only Grades 2-3. In random assignment, these two schools are treated as one unit. Therefore, there are 59 school units in the random assignment process.

would indicate that they are equally effective approaches to implementing A2i. Given that implementation cost is considered a significant barrier for program expansion, the study also analyzes the cost of each model to see if the mixed-mode model is less expensive to implement than the face-to-face model as intended. Finally, as part of the expansion grant, the United2Read team worked to bring the program to a broader set of schools (referred to as the "expansion schools") and students beyond the study sample.[11] This report describes this effort and the scope of the expansion.[12]

Evaluation Challenges

The onset of the global pandemic in March 2020 resulted in the closures of many U.S. schools to in-person learning. Most schools transitioned to remote education models for the remainder of the school year. The effect of COVID-19 lasted well into the 2020–2021 school year, as districts oscillated between virtual, hybrid, and in-person learning to balance the need to keep students and staff members safe with the need to provide an effective learning environment. Figure 1 illustrates how the onset of COVID-19 intersects with the study timeline. This drastic change in the learning environment in the middle of the study challenged the study in multiple ways.

Figure 1. COVID-19 and the Study Timeline
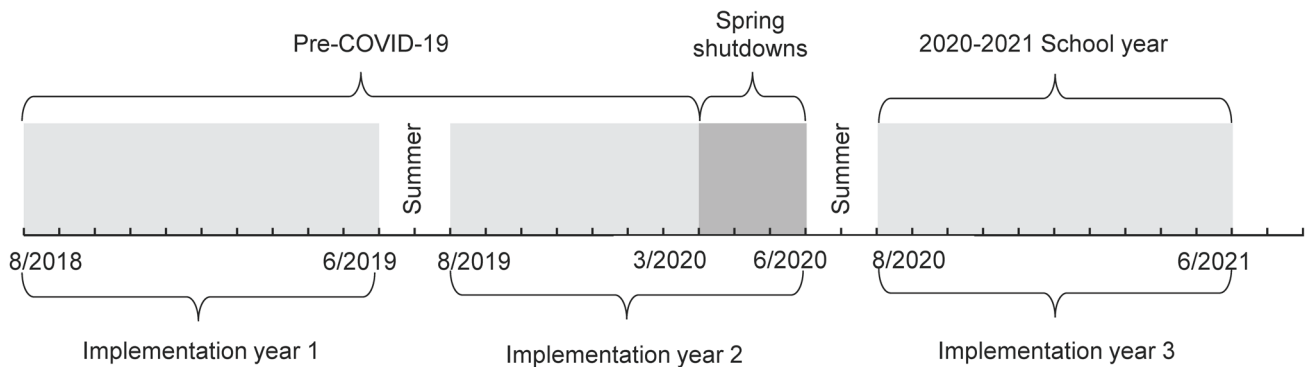
First, COVID-19-related school disruptions hampered the delivery of both PD models and the implementation of the A2i system in the second half of the evaluation period. Teachers missed many of the intended support sessions in Year 2 that were scheduled for March 2020 through June 2020 as all PD events were canceled due to school closures. In Year 3, due to restrictions on school access, PD activities were delivered only virtually to all

---

11. The expansion schools all paid a fee to receive the mixed-mode version of the program.

12. This study focuses on the relative performance of both PD delivery models as outlined by the research questions. It does not evaluate the whole EIR grant, which included other components such as the expansion efforts.

study schools. These changes significantly reduced the intended contrasts between the two models' PD delivery modes and PD amounts. Shifting learning arrangements also affected the implementation of other components of the A2i technology, such as the use of assessments and the A2i platform, as both teachers and students struggled with technical and resource issues associated with virtual learning. As a result, schools' experiences with A2i differed for the periods before and after the onset of COVID-19. The study reports information for these two periods separately, and these findings should be interpreted with the broader context in mind.

Second, COVID-19-related changes in state standardized test arrangements affected important planned evaluation analyses. The study initially intended to assess the effectiveness of the PD models by comparing the outcome trends of the mixed-mode and face-to-face schools with that of similar schools serving primary grades in the business-as-usual ways through a quasi-experimental analysis. The cancellation of the 2019–2020 state tests and the changes that occurred to the 2020–2021 state tests made it impossible to establish the outcome trend, rendering this planned exploratory analysis infeasible. Therefore, this report cannot speak for whether either model produced better student outcomes than the business-as-usual condition.

Lastly, the onset of COVID-19 hindered efforts to expand the program to more districts and schools as schools shifted their focus and resources to COVID-19-related priorities and put on hold additional initiatives to ease teachers' burdens. As a response to the shift to online and at-home learning, Learning Ovations pivoted from planned expansion strategies and developed new tools that allowed other adults to support reading learning during remote schooling as an alternative way to reach students.

In sum, these challenges changed the nature and course of the program's implementation and planned expansion. They had implications for the evaluation analysis and the interpretation of its findings. The rest of the report details these findings within the context of such changes. It starts with descriptions of the A2i Professional Support System, the two PD models, and background information about the evaluation. The report then provides separate evaluation findings for the pre-COVID-19 period and the 2020–2021 school year. That is followed by discussions of the costs of the two PD models and Learning Ovations' efforts to expand the program's reach to a broader population of schools and students. The final section offers reflections on what one can learn from this evaluation.

THE A2i PROFESSIONAL SUPPORT SYSTEM AND THE TWO PROFESSIONAL DEVELOPMENT DELIVERY MODELS UNDER EVALUATION

The A2i Professional Support System is the result of more than two decades of research by Carol Connor and her colleagues. At the core of this research is a set of reading instruction practices that are linked to students' reading growth. Guided by earlier

research on the components of reading skills and the elements of effective instruction, A2i was developed using extensive classroom observation data to identify these instructions. Researchers focused on two factors—the content of instruction (meaning-focused or code-focused) and who manages students' learning (the child or the teacher)—and developed a classification scheme that places reading instruction into one of four quadrants defined by these two factors.[13] The table embedded in Figure 2 illustrates each of the four types of instruction. Classroom observation data show that much of students' reading growth in early grades could be predicted by how much of each type of instruction they received during a school year.[14]

This research also demonstrated that no single type of effective reading instruction is appropriate for all students and the same instruction can produce very different outcomes for different students. Teachers, therefore, need to know how much of which type of instruction works for which students. Connor and colleagues derived a research-based algorithm to generate such information for teachers. The algorithm uses information on an individual student's initial vocabulary level and their ability to link spoken and written words—two malleable skill sets that were identified as most useful in predicting students' progress in early reading—to calculate the amount of each type of instruction the child needs.[15] Based on these results, it also provides recommendations for which students should be grouped together for similar instructional activities.

From this knowledge, the A2i Professional Support System was developed to help teachers deliver research-based reading instruction matched to the unique needs of individual students to support reading growth for all.

Elements of the A2i Professional Support System

The A2i Professional Support System is a suite of technology tools that aim to facilitate differentiated small-group instruction. Specified PD activities aid with implementation of the system.

A2i technology has three main components. First, it offers online adaptive assessments to measure students' skill levels on several dimensions of reading. These two assessments—Letter2Meaning and Words Match Game—measure students' decoding, vocabulary, and reading comprehension skills. The tests are designed to be administered frequently on A2i's web-based platform to monitor student progress. Second, the A2i software platform provides information and resources to support teachers' instructional planning. The platform uses students' assessment data to make recommendations for each student with the amount and type of each kind of instruction (teacher-managed and code-focused, teacher-managed and meaning-focused, child-managed and code-focused, and child-managed

---

13. Connor, Morrison, and Katch (2004); Connor, Morrison, and Petrella (2004).

14. Connor, Morrison, and Katch (2004).

15. Connor, Morrison, and Underwood (2007).

Figure 2. The Assessment-to-Instruction (A2i) Professional Support System

Professional Development Support

Professional development is provided by a Literacy Outcomes Specialist. To support use of the suite of A2i technology-based tools, there are three types of professional development:

1) A2i Professional Learning Communities or Literacy Huddles: Meetings with grade-level teams to help teachers implement the recommendations available on the A2i platform.

2) In-Class Coaching or Individualized Classroom Coaching: On-campus visits to classrooms or 1:1 video calls that may include a combination of modeling, practice and feedback, or observation.

3) On-demand support and Flex Days: Optional days for a school to request support in implementing A2i.

A2i Online Assessments

A2i adaptive assessments are administered to students periodically throughout the year to monitor student progress.

A2i Software Platform

Based on data from the A2i assessments, the A2i platform provides recommendations for grouping students and the number of minutes each student should get for each of four types of instruction. The types of instruction, with examples, are:

	Code-focused instruction	Meaning-focused instruction
Teacher-managed	The teacher is working with a small group of children providing explicit phonics instruction.	The teacher is reading to the students and leading a discussion of the text's meaning.
Child-managed	Children are working together to practice blending words from phenomes.	A student is reading a book of his choice in the library center.

Which can be supported by...

Lesson Planning Tools

The A2i system includes lesson planning tools for finding and tracking lessons aligned with the recommended time and types of instruction needed for each small group recommended by the platform. The system also features additional professional development tools.

Support differentiated small-group instruction in class

SOURCE: Authors' summary of key A2i components.

NOTES: Professional Learning Communities were renamed to Literacy Huddles in Year 2.
In-Class Coaching was renamed to Individualized Classroom Coaching in Year 2.

and meaning-focused). Based on the number of minutes and type of instruction, the A2i algorithm also suggests groupings of students within the class so that teachers may differentiate instruction according to students' needs. In addition, to make it easier to follow through on those recommendations, the system provides embedded lesson planning tools such as lesson menus with suggested activities linked to a district's particular reading program, as well as data visualization tools and other online resources.

Teachers receive PD that supports the use of these technology tools and helps them implement differentiated small-group instruction in the classroom. PD is provided by a trained coach, known as a Literacy Outcomes Specialist (LOS) throughout the implementation period. These coaches are employed centrally by Learning Ovations and are assigned to multiple schools. Depending on the size of a district, the coaches may work across districts or in just one district. There are three types of PD provided by the coaches to teachers. First, there are Professional Learning Communities or Literacy Huddles (LH).[16] These are 30–60-minute meetings with grade-level teams to help teachers use the A2i technology and implement recommendations available on the A2i platform. The LOS usually focuses on a specific topic at each meeting. Second, LOSs provide In-Class Coaching or Individualized Classroom Coaching (ICC),[17] which is either a one-on-one in-person meeting or video call of about 30–45 minutes with a teacher that may include a combination of modeling, practice and feedback, or observation, based on the teacher's needs, with a focus on helping teachers implement differentiated small-group instruction in the classroom. Finally, on-demand support and Flex Days (days on which the LOS is asked to come to the school to support specific activities, such as administering assessments or modeling lessons) allow additional access for teachers to call on the LOS for support as needed.

In addition to the support provided by the LOS, teachers receive local support through their administrator and a Literacy Champion. The administrator, generally the principal, receives professional development from the LOS to understand the program and its implementation within the building. Each school also selects a Literacy Champion to be a point of contact for teachers implementing the program and to provide support to any new teachers that may have joined the school after implementation was already underway. The roles the Literacy Champions play can include being the liaison between teachers and the LOS, coaching teachers on small-group instruction, supporting teachers with administering A2i assessments, understanding and using A2i data, and lesson planning. Literacy Champions may be classroom teachers or interventionists who provide extra support and interact directly with specific students. Together, this local support provides avenues to sustainability beyond the life of the grant.

---

16. Learning Ovations renamed Professional Learning Communities to Literacy Huddles in early 2020.

17. Learning Ovations renamed In-Class Coaching to Individualized Classroom Coaching in early 2020.

How the A2i System Might Affect Student Reading Achievement

The A2i Professional Support System aims to support teachers' use of differentiated small-group instruction. The result is improved reading achievement for students. Figure 3 illustrates how the A2i system is expected to change students' reading performance.

Figure 3. How the A2i Professional Support System May Affect Students' Reading Skills

Program Inputs

Professional development

Professional Learning Communities or A2i Literacy Huddles (30-minute grade-level meetings with Literacy Outcomes Specialist)

In-Class Coaching or Individualized Classroom Coaching (45-minute one-on-one coaching sessions)

On-demand support and Flex Days

A2i technology

Online student assessment

Software platform
Instruction recommendation for individual students
Lesson planning resources

Change in Teachers

Knowledge and beliefs

Knowledge of A2i recommended instructional strategies

Belief that they can implement the program components

Behavior

Use of A2i system-generated instruction recommendations and other resources

Use of differentiated small-group instruction

Change in Students

Short-term

Improved reading performance after one year of exposure

Long-term

Reading at or above grade level by end of Grade 3

The program provides teachers with PD activities, either virtually or in person, depending on their school's program condition. The PD supports their understanding and implementation of the A2i suite of technology, which includes online student assessments, instructional recommendations for individual students, and lesson planning tools. Through this PD, teachers gain knowledge of the instructional strategies and the understanding that they can implement the program in their own classrooms. This leads to changes in behavior, including increased use of the A2i system and the differentiated small-group instruction

that it suggests.[18] As a result of being in a classroom that uses the A2i recommendations and receiving differentiated small-group instruction, students are expected to improve their reading performance after one year of exposure to the program and to be able to read at or above grade level by the end of third grade.

The Two Professional Development Delivery Models and the Program Roll Out

This study looks at two PD delivery models: face-to-face and mixed-mode. The face-to-face model is similar to the original PD delivery model implemented and evaluated in prior studies of A2i. Schools randomly assigned to the face-to-face model receive all their PD in person from their LOS at their school. Schools in the mixed-mode group receive both in-person and virtual PD: initial activities and classroom coaching are delivered in person, but the Literacy Huddles are delivered virtually via Zoom. Another critical difference between these two PD models is that the face-to-face model has six planned Individualized Classroom Coaching sessions during a school year while the mixed-mode model has two. These differences are designed to test whether A2i can be effectively implemented with a reduced number of lower-cost PD supports.

PD activities planned for this study were intentionally not as numerous as those offered in prior studies of A2i, especially for mixed-mode models. Prior studies, for example, offered teachers a half-day initial workshop and monthly meetings. This study planned for two to seven Literacy Huddles for teachers. Furthermore, in previous studies, teachers received biweekly classroom-based coaching during the literacy block (from a research assistant), in contrast to this study, which was designed to include two to six face-to-face Individualized Classroom Coaching sessions over the school year.[19] This reduction in PD was intended to reduce the program's costs and to see if A2i could be implemented effectively with a lower level of support.

The A2i system and PD activities were rolled out to early elementary grade levels throughout the life of the project. In Year 1, kindergarten and first grade teachers were the only teachers to receive access to A2i and PD. In Year 2, second grade teachers were the focus of the PD effort, with kindergarten and first grade teachers still receiving some PD but scaled back from Year 1. In Year 3, third grade teachers were the focus, with teachers from kindergarten through second grade receiving less support. The grade levels that were new to the program in a given school year are referred to as the "focal grade" of that year. Table 1 presents the number and type of PD events planned for each model and year by grade. While the targeted grades rolled out as intended, the numbers and delivery modes of Literacy Huddles and Individualized Classroom Coaching sessions presented for Year 3

---

18. Such changes in teacher behavior can be complex and demanding. They would require teachers to plan their lessons and master teaching materials differently to allow differentiation. Teachers would also need to change classroom management practices to allow for a more decentralized classroom environment.

19. Connor et al. (2013).

Table 1. Planned Number and Type of Professional Development (PD) Events, by Grade, Year, and Program Condition

Grade	Type of PD	Year 1		Year 2		Year 3	
		Mixed-Mode	Face-to-Face	Mixed-Mode	Face-to-Face	Mixed-Mode	Face-to-Face
Grades K-1	Virtual Literacy Huddles	**7**	**0**	3 to 6	0	3 to 6	3 to 6
	In-Person Literacy Huddles	**0**	**7**	0	3 to 6	NA	NA
	Individualized Classroom Coaching	**2**	**6**	0 to 2	0 to 6	0 to 2	0 to 3
Grade 2	Virtual Literacy Huddles			**7**	**0**	3 to 6	3 to 6
	In-Person Literacy Huddles			**0**	**7**	NA	NA
	Individualized Classroom Coaching			**2**	**6**	0 to 2	0 to 3
Grade 3	Virtual Literacy Huddles					**7**	**7**
	In-Person Literacy Huddles					**NA**	**NA**
	Individualized Classroom Coaching					**2**	**6**

SOURCE: PD planning documents provided by Learning Ovations.

NOTES: Numbers reported for Years 1 and 2 reflect the original PD plan; numbers reported for Year 3 reflect adjustments made to the original plan. There were no in-person Literacy Huddles or Individualized Classroom Coaching planned for Year 3 due to ongoing challenges related to the COVID-19 pandemic.
  NA = Not Applicable.
  Bolded numbers are the planned numbers of PD events for the focal grade in a given implementation year.

reflect adjustments made to the original plan at the start of the 2020–2021 school year due to the COVID-19 pandemic.

STUDY DESIGN, DATA, AND STUDY SAMPLES

Study Design[20]

The study's primary goal is to assess whether a mixed-mode PD model that combines virtual and in-person delivery of PD activities can serve as a viable alternative to the original face-to-face delivery approach in supporting teachers' use of the A2i system, facilitating their implementation of differentiated small-group instruction, and improving students' reading skills in a large, diverse set of schools. To achieve this goal, the study recruited 59 elementary schools from 20 districts around the country and randomly assigned roughly

---

20. The study design was pre-registered with the Registry of Efficacy and Effectiveness Studies (REES, registry ID: 1790).

equal numbers of schools to either the mixed-mode or the face-to-face PD model within districts or district groups.[21] Early grade teachers in the mixed-mode schools received a combination of virtual and in-person training and support for the A2i Professional Support System, while those in the face-to-face schools received such training and support exclusively in person. This design determines that any differences in teacher experiences or student outcomes between these two groups of otherwise similar schools can be attributed to the different PD delivery models to which they were exposed. Therefore, if the two PD models were fully implemented, the difference in outcomes between these two groups of schools can be considered the differential impact between these two PD models. The study collected quantitative data on schools, teachers, and students to assess whether and to what degree the two PD models affected the teachers and students differently.

Data Sources

The study collected multiple sets of reading test scores to assess the differential impact of the two PD models on students' short- and long-term reading performances. In the spring of Year 1, the study administered the Gates-MacGinitie reading test (GMRT) to first-graders in the study sample to measure their reading level after several months of exposure to the program.[22] Over the course of the study, as part of the A2i system, teachers administered the A2i assessments—including the Letters2Meaning and Word Match Game assessments—multiple times during each school year to monitor students' reading levels and progress in vocabulary, decoding, and comprehension. The study obtained these assessment scores from the A2i system. Lastly, the study collected students' scores for the third grade state or district standardized reading tests as a measure for their general reading skill after three years in the program.[23] It used this measure to compare the cumulative impact of A2i on student reading skills between the mixed-mode and face-to-face schools after three years of exposure to the program.

To assess whether the two PD models were delivered to teachers as intended, the study collected Learning Ovations' PD planning documents and PD records to calculate the amounts and delivery modes of the Literacy Huddles and Individualized Classroom Coaching sessions for the mixed-mode and face-to-face schools separately. The PD records include the number of PD events that took place in each school by type and

---

21. Schools in 12 of the 20 study districts were randomly assigned within their district. Schools from the remaining eight districts, including seven single-school districts, were grouped into three blocks based on their geographic proximity, and randomly assigned within each block. In total, there are 15 random assignment blocks in the study.

22. The study used Edition Four, Forms S and T of the Gates-MacGinitie reading test.

23. The study planned to use the third grade state standardized English Language Arts (ELA) test scores as the long-term reading outcome to assess whether students with A2i exposure can read at or above proficiency level by Grade 3. However, due to COVID-19-related state test policy changes, not all study districts participated in standardized testing in the 2020–2021 school year. For districts that opted out of the state test, the study collected the spring 2021 test scores from district benchmark tests.

delivery mode for all three implementation years. For Years 2 and 3, the records also include PD delivery time and teacher attendance documented by LOSs at each PD event.

The study also obtained measures of teachers' usage of the A2i technology from the A2i system. The *frequency of A2i assessment* is the number of times students took the A2i assessments in a school year. It measures how well students' progress was monitored and used to generate instruction recommendations by the A2i system. In a normal school year, students are expected to take the assessments at least three times. *Teachers' use of the A2i platform* is captured by the number of minutes teachers spent on the A2i platform as recorded by the A2i system.

Lastly, the study administered two rounds of online *teacher surveys* in the spring of Year 1 and spring of Year 3. The survey data yielded information about teachers' backgrounds, their knowledge, experiences, and perceptions of the program's components, and their use of differentiated small-group instruction. The teacher surveys were intended to provide valuable information about whether the two PD models affected teachers' perceptions of the system and their instructional behaviors differently.

Table 2 summarizes these data sources, the data obtained from them, their collection times, and the unit of measure for each source by the research questions they address.

In addition to these data sources, whenever possible, the report references the qualitative research that the Digital Promise Global team conducted during the study period that captures teachers' experiences with the program. This qualitative research collected information through multiple rounds of interviews and focus group discussions with teachers, school staff members, school and district leaders, and LOSs from an initial set of 21 schools, including 16 mixed-mode schools and 5 face-to-face schools.[24] Among the 16 mixed-mode schools, 12 are in the study sample and 4 are expansion schools that are not part of the study sample.

Recruitment and Study Schools

The recruitment for the evaluation was led by Learning Ovations as part of the general outreach to schools, districts, and states for the Education, Innovation, and Research (EIR) expansion grant. The EIR grant presented the opportunity to offer considerable financial benefits to schools interested in using the A2i system. If a district committed to as many expansion schools (which paid a fee for the mixed-mode program) as study schools, then the study schools in the district could receive the program gratis for the duration of the study. Each study school had to serve students in Grades K through 3. They also had to be willing to participate in a random assignment experiment and to provide student records

---

24. Due to COVID-19, the number of schools and teachers the Digital Promise Global team was able to reach dropped in the second half of Year 2 (18 schools, including 2 new expansion schools) and Year 3 (14 schools). The team also had to adapt the in-person data collection to virtual in many cases during the pandemic (Kasad, Vang, and Young, 2021).

Table 2. Measures, Data Sources, and Data Collection Activities

Research Question	Measures	Data Sources	Collected	For
1. How well were the two professional development (PD) delivery models implemented and did their implementation differ?	Number of planned PD events	Learning Ovations documents	Throughout the study	All study schools
	Number of PD events delivered	Learning Ovations PD records	Throughout the study	All study schools
	PD received by teachers (# of minutes)	Learning Ovations PD records	Throughout the study	Grade K-2 teachers in Year 2 Grade K-3 teachers in Year 3
2. Did teachers' use of the A2i technology differ, and how?	Usage of A2i platform (# of minutes)	A2i platform	Throughout the study	Grade K-1 teachers in Year 1 Grade K-2 teachers in Year 2 Grade K-3 teachers in Year 3
	Frequency of A2i assessment (counts)	A2i platform	Throughout the study	Grade K-1 students in Year 1 Grade K-2 students in Year 2 Grade K-3 students in Year 3
3. Did teachers' views of A2i and their instructional practices differ, and how?	Use of small-group instruction (Y/N, frequency, and % of reading block time)	Teacher surveys	Spring of Year 1 Spring of Year 3	Grade K-1 teachers in Year 1 Grade K-3 teachers in Year 3
	Use of system recommendations for grouping and instructional time (%)	Teacher surveys	Spring of Year 1 Spring of Year 3	Grade K-1 teachers in Year 1 Grade K-3 teachers in Year 3
	Use of lesson planning tools (# of days)	Teacher surveys	Spring of Year 1 Spring of Year 3	Grade K-1 teachers in Year 1 Grade K-3 teachers in Year 3
	Knowledge of A2i instructions (score of 1 to 6)	Teacher surveys	Spring of Year 1 Spring of Year 3	Grade K-1 teachers in Year 1 Grade K-3 teachers in Year 3
4. What is the difference in the cumulative impact of A2i on students' reading achievement?	State/district reading test scores	District records	After Year 3	Grade 3 students in Year 3
	Gates-MacGinitie Reading Test scores	Study administered test	Spring of Year 1	Grade 1 students in Year 1
	A2i assessment scores	A2i platform	Throughout the study	Grade K-1 students in Year 1 Grade K-2 students in Year 2 Grade K-3 students in Year 3

SOURCE: Authors' summary of data collection activities.

NOTE: The 2018–2019 school year was Year 1, the 2019–2020 school year was Year 2, and the 2020–2021 school year was Year 3.

data for evaluation purposes. Given these requirements, it was not the intention of the study to recruit a sample of schools that was representative of the national population.

At the end of the recruitment phase, 20 school districts in 4 states agreed to participate in the study. The number of study schools in each district ranged from 1 to 10, producing a total sample of 59 schools. Table 3 shows the average characteristics of these schools at the time of random assignment and how these study schools compared with a national sample of elementary schools that serve students in lower elementary school grades.[25]

Even though recruitment did not target a specific school population, the study schools predominantly serve high needs students.[26] About three-quarters of the study schools are in urban fringe and large towns, with 80 percent eligible for Title I. Most students in the average study school are Hispanic (61 percent) and eligible for the free and reduced-price lunch program (72 percent). The average study school also has about 27 percent English learners and about 13 percent students with Special Education status. Finally, about 43 percent of third grade students in the average study school can read at or above proficiency level.

Overall, the study schools differ from schools in the national sample on most characteristics measured in Table 3. Compared with the national elementary school population, the study schools differ in their locality—three-quarters of them are in urban fringe or large towns instead of cities or rural areas. The study schools are also larger than the national sample on average in terms of students enrolled in Grades K-2. The study schools have more Hispanic students, English learners, and students eligible for free or reduced-price lunch than the schools in the national sample. Finally, the study schools have a lower proportion of students who scored at or above proficiency level in reading than the schools in the national sample. Therefore, findings from this study are most applicable in contexts where schools are similar to the study schools.

Random Assignment and Equivalence of Background Characteristics Among Schools, Students, and Teachers in the Randomized Groups

In the summer of 2018, the schools within each district or district group were randomly assigned to the mixed-mode model or the face-to-face model. The random assignment produced 30 mixed-mode schools and 29 face-to-face schools. The purpose of random assignment is to produce two groups of schools that are statistically equivalent on baseline characteristics at the start of the study. If the two groups are indeed equivalent

---

25. These comparisons use data from the 2017–2018 school year, before the current study began; at this point the schools' characteristics could not have been affected by the study itself and therefore represent true baseline values.

26. High needs students include students eligible for and receiving subsidies, students from low-income or economically disadvantaged backgrounds, English Language Learners, former English Language Learners, or students with disabilities.

Table 3. Comparison of Study Schools and Public Elementary Schools Nationally

Characteristics	Study Schools	Public Elementary Schools Nationally	Estimated Difference	P-Value for Estimated Difference
Urbanicity (% of schools)				
Large or middle-sized city[a]	16.9	29.1	-12.1*	0.041
Small city or suburban[b]	74.6	51.0	23.6*	0.000
Town or rural[c]	8.5	20.0	-11.5*	0.027
Title I status (% of schools)	79.7	54.5	25.1*	0.000
Male (average % of students)	51.9	51.4	0.6	0.243
Race/ethnicity (average % of students)				
White, non-Hispanic	26.5	38.0	-11.5*	0.007
Black, non-Hispanic	5.0	10.7	-5.7*	0.020
Hispanic	60.6	37.6	23.0*	0.000
Asian	4.4	8.4	-4.1*	0.024
Other	3.6	5.3	-1.7*	0.026
Proportion of students with				
IDEA plans (%)[d]	13.2	13.0	0.1	0.864
504 plans (%)[e]	1.1	1.5	-0.4	0.164
Limited English Proficiency (%)	26.7	16.6	10.1*	0.000
Free and reduced-price lunch status (%)	71.7	56.3	15.4*	0.000
Third-grade students reading at or above proficiency level (%)	43.1	60.0	-16.9	0.636
Mean school enrollment (n)	562.4	503.1	59.3	0.128
Kindergarten	97.0	79.9	17.1*	0.006
Grade 1	85.8	74.4	11.4*	0.039
Grade 2	87.2	74.4	12.8*	0.015
Grade 3	81.1	74.6	6.5	0.236
Mean number of teachers (n)	31.4	31.6	-0.2	0.944
Number of schools	59	10,560		

SOURCE: The Common Core of Data (CCD) for the 2017–2018 school year, Office for Civil Rights Data for the 2017–2018 school year, and state-reported school performance data for the 2017–2018 school year.

NOTES: Rounding may cause slight discrepancies in calculating means and differences. Sample size for each characteristic may vary due to missing values. *indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An omnibus test was used to determine whether there is a systematic difference between the study schools and the national sample, with respect to the characteristics included in this table. The p-value for this test is 0.944.

The national sample includes all public, regular elementary schools serving students in any of Grades K, 1, 2, or 3, and are not charter, magnet, or virtual schools.

[a]A city inside an urbanized area and inside a principal city with a population of more than 100,000.

[b]A city inside an urbanized area and inside a principal city with a population of less than 100,000 or a territory outside a principal city and inside an urbanized area.

[c]Territories outside urbanized areas.

[d]IDEA plans are written instructional plans for students with disabilities who are designated as special education students.

[e]504 plans provide students with identified disabilities accommodations to ensure academic success and access to the learning environment.

at the outset, and if any attrition from the sample over the course of the study is balanced across groups, one can be reasonably confident that any differences that are later found in outcomes between the two groups are due to the differences in the PD models they were exposed to.

Study Schools

As intended, random assignment created two groups of schools that were similar in baseline characteristics at the start of the study. Table 4 shows that all the estimated differences between mixed-mode and face-to-face schools are small enough to meet commonly used criteria for baseline equivalence, such as the one used by the What Works Clearinghouse.[27] In addition, there is no evidence of systematic differences between these two groups of schools.[28]

The number of schools that stayed in the study declined over the course of the study. Nine schools (five mixed-mode and four face-to-face schools) withdrew from the program after Year 1, with another ten schools (five mixed-mode and five face-to-face schools) withdrawing from the study after Year 2.[29] In the end, 40 schools from 13 districts stayed with the study for all 3 years.[30] By withdrawing from the study, these schools lost access to the A2i technology and did not receive any A2i support or participate in any subsequent data collection. Fortunately, all the withdrawn schools agreed to share their student records data with the study, which allowed the evaluation team to include these schools in the analyses based on school record information. Importantly, the attrition of schools from the sample did not affect the similarities between the mixed-mode and face-to-face schools that remained in the study, as demonstrated in Appendix Tables B.2–B.4.

Students

To answer the main research question about differential impacts the mixed-mode and face-to-face models might have on student reading performance, the study focused on a cohort

---

27. Even though the difference in the proportion of students with limited English proficiency status is statistically significant, the magnitude of the difference is about 0.11 standard deviations in effect size, below the 0.25 criterion noted in the What Works Clearinghouse Handbook (2022).

28. To examine whether there are any systematic differences between the two groups of schools, the study conducted an F-test in a regression model controlling for the background variables listed in Table 4. The p-value of the test is 0.939, indicating no such difference.

29. Such decisions were usually made at the district level. In addition to logistical challenges caused by the COVID-19 pandemic, other common threads identified in districts that chose to discontinue their active participation in the U2R grant include but are not limited to shifts in policy or personnel, differences in goals/instructional focus among schools within the districts, and differences in philosophies about how to change teacher behavior.

30. One additional school withdrew from the study in April of Year 3. Since this school complied with all data collection activities and stayed with the program for most of the duration of the study, it was still included in the analyses. Appendix Tables B.2 and B.3 compare the mixed-mode and face-to-face schools that remained in the study after Year 1 and Year 2, respectively.

Table 4. School Background Characteristics for Study Sample Schools, by Program Condition

Characteristics	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	P-Value for Estimated Difference
Urbanicity (% of schools)				
Large or middle-sized city[a]	16.7	16.7	0.0	1.000
Small city or suburban[b]	73.3	78.3	-5.0	0.160
Town or rural[c]	10.0	5.0	5.0	0.160
Title I status (% of schools)	76.7	82.2	-5.6	0.487
Male (average % of students)	52.4	51.3	1.1	0.080
Race/ethnicity (average % of students)				
White, non-Hispanic	27.0	26.2	0.7	0.810
Black, non-Hispanic	4.5	5.3	-0.8	0.690
Hispanic	59.5	61.6	-2.1	0.327
Asian	5.2	3.6	1.7	0.329
Other	3.7	3.3	0.4	0.551
Proportion of students with				
IDEA plans (%)[d]	12.5	13.3	-0.8	0.563
504 plans (%)[e]	1.2	1.0	0.2	0.242
Limited English Proficiency (%)	23.9	29.3	-5.5 *	0.024
Free and reduced-price lunch status (%)	69.7	73.3	-3.6	0.333
Third-grade students reading at or above proficiency level (%)	43.7	42.9	0.9	0.785
Mean school enrollment (n)	537.1	591.5	-54.4	0.209
Kindergarten	92.1	102.1	-9.9	0.152
Grade 1	81.3	90.3	-9.0	0.150
Grade 2	82.8	91.6	-8.7	0.138
Grade 3	79.4	83.2	-3.8	0.526
Mean number of school year teachers (n)	30.2	33.3	-3.1	0.316
Number of schools	30	29		

SOURCE: The Common Core of Data (CCD) for the 2017–2018 school year, Office for Civil Rights Data for the 2017–2018 school year, and state-reported school performance data for the 2017–2018 school year.

NOTES: The sample includes all 59 study schools that participated in the study. The estimated differences are regression-adjusted, controlling for the blocking of random assignment.

Rounding may cause slight discrepancies in calculating means and differences. The sample size for each characteristic may vary due to missing values. * indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

An F-test was used to determine whether there is a systematic difference between the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.939.

[a]A city inside an urbanized area and inside a principal city with a population of more than 100,000.

[b]A city inside an urbanized area and inside a principal city with a population of less than 100,000 or a territory outside a principal city and inside an urbanized area.

[c]Territories outside urbanized areas.

[d]IDEA plans are written instructional plans for students with disabilities who are designated as special education students.

[e]504 plans provide students with identified disabilities accommodations to ensure academic success and access to the learning environment.

of students as they moved through the grades, starting in first grade in Year 1 to third grade in Year 3.[31]

This primary student cohort consists of 4,733 first-graders enrolled in the study schools at the start of Year 1.[32] Using the demographic data received from students' district records, as well as their initial A2i assessment scores, Table 5 compares the background characteristics of these students in mixed-mode and face-to-face schools and finds no systematic differences between the two groups. On average, these students were 6 years old as of August 31, 2018; most of the students were Hispanic and from low-income backgrounds, as defined by their school districts.[33] About 11–12 percent were students with Special Education status and about 30 percent to 34 percent were English learners. More importantly, there was no difference between the mixed-mode and face-to-face school students in the timing and scores of their first A2i assessments in the fall of 2018—a proxy for their reading performance at the start of the study.

The study collected GMRT scores, A2i assessment scores, and third grade state or district standardized reading test scores for this cohort of students at various points in the study. Due to attrition, non-response, and mobility, the number of schools and students with valid data varies across the samples used for the analyses of these outcome measures. The study examined the response rates and student background characteristics for each sample used in these analyses and found no systematic differences between the two groups of schools.[34]

In addition to this primary student sample, the study also compared the impact of A2i between the two groups of schools using other available school and student samples to deepen the understanding of the relative effectiveness of the different levels of PD support. Such variations include restricting the sample to schools and first-graders who stayed with the study for all three years (the stable sample) and expanding the sample to include students from other grade levels and cohorts who were exposed to the program (the overall sample). Findings for these exploratory samples are included in Appendix D.

Teachers
The study fielded two rounds of online teacher surveys to collect information on teachers' use of differentiated small-group instruction, their knowledge of the literacy instruction used in the A2i system recommendations, and their use of the recommendations made by the system. The first round of teacher surveys was fielded in the spring of Year 1 to

---

31. Based on information from fall 2018, students in self-contained special education classes or dual-language immersion classes that did not use English for reading instruction, and students who enrolled for less than 30 days in Year 1, were excluded from the sample.

32. This cohort includes 2,372 students from the mixed-mode schools and 2,361 students from the face-to-face schools. See Appendix Table B.1 for details.

33. Defined by students' eligibility for free or reduced-price lunch or other available poverty indicators provided by study districts.

34. Appendix B provides findings for these checks.

Table 5. Student Background Characteristics for Study Schools, by Program Condition

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Difference		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Difference	Effect Size of Estimated Difference	P-Value for Estimated Difference
A2i assessments											
Letters2Meaning											
Scaled score	30	2,191	429.1	100.2	29	2,163	423.3	102.2	5.9	0.06	0.452
Grade equivalent			0.4				0.3				
At grade level (%)	30	2,191	48.7	50.0	29	2,163	46.1	49.9	2.6	0.05	0.475
Word Match Game											
Scaled score	30	2,221	466.2	12.9	29	2,161	465.6	13.0	0.5	0.04	0.555
Grade equivalent			-0.1				-0.1				
At grade level (%)	30	2,221	29.4	45.2	29	2,161	28.1	45.1	1.4	0.03	0.609
Time of test											
(Number of days)	30	2,216	25.3	22.1	29	2,151	24.7	17.9	0.6	0.03	0.721
Demographic characteristics (%)											
Age at start of Year 1	30	2,368	6.0	0.3	29	2,361	6.0	0.2	0.0	0.04	0.213
Female	30	2,363	49.0	50.0	29	2,354	49.5	50.0	-0.5	-0.01	0.753
Race/ethnicity											
Hispanic	30	2,329	58.5	48.4	29	2,328	58.8	49.4	-0.3	-0.01	0.915
Black, non-Hispanic	30	2,329	4.6	21.7	29	2,328	5.1	21.9	-0.5	-0.02	0.767
White, non-Hispanic	30	2,329	27.4	41.8	29	2,328	29.1	45.6	-1.7	-0.04	0.564
Asian	30	2,329	5.6	23.6	29	2,328	3.8	19.6	1.8	0.08	0.310
Native American or Alaskan Native	30	2,329	0.2	4.1	29	2,328	0.2	5.5	-0.1	-0.01	0.685
Other	30	2,329	3.8	19.6	29	2,328	3.0	18.7	0.9	0.05	0.330
Other background characteristics (%)											
Low-income background	30	2,053	69.3	45.7	29	2,115	73.7	45.4	-4.4	-0.10	0.134
English learner	30	2,060	30.2	46.6	29	2,118	34.4	47.7	-4.2	-0.09	0.180
Special education	30	2,053	11.6	31.5	29	2,117	10.5	30.8	1.2	0.04	0.493

(continued)

Table 5 (continued)

SOURCE: The Common Core of Data for the 2017–2018 school year, Office for Civil Rights Data for the 2017–2018 school year, and state-reported school performance data for the 2017–2018 school year. District administrative records for the 2017–2018 school year and A2i assessment data for the 2018–2019 school year were obtained for this study.

NOTES: This table is based on a total of 4,733 Grade 1 students (2,372 in mixed-mode schools and 2,361 in face-to-face schools) who were enrolled in the 59 study schools in the fall of 2018 (Year 1). ELA = English Language Arts. None of the differences between the program and non-program schools is statistically significant at the 0.05 level.

An F-test was used to determine whether there is a systematic difference between the students in the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.843.

all Kindergarten and Grade 1 regular classroom teachers in the study schools. Of these teachers, 321 from 58 study schools responded to the survey, leading to an overall response rate of 78 percent. The differential response rate was 1.2 percent between mixed-mode and face-to-face schools and was not statistically significant. Teachers from the two groups of schools who responded to the survey did not differ in their background characteristics.[35]

Due to the lasting effects of COVID-19, the second round of teacher surveys, fielded in the spring of Year 3, received a lower response rate. This survey targeted all Kindergarten to Grade 3 regular classroom teachers in the schools. Of these, 273 teachers from 37 schools, or 55 percent of the targeted sample, responded to the survey. There was no difference in response rate between the two groups, and the respondents from the mixed-mode and face-to-face schools were similar in their background characteristics.[36]

Overall, the evidence presented in this section demonstrates a high degree of similarity between the mixed-mode and face-to-face groups at school-, student-, and teacher levels at the start of the study. Moreover, subsequent attrition and changes in the sample did not disrupt this balance. Therefore, one can be confident that differences in outcomes between the two groups reflect the differential impacts of the two PD models rather than preexisting differences in the sample.

---

35. Appendix Table B.12 provides comparisons of teacher background characteristics for the respondents to the Year 1 teacher survey.

36. Appendix Table B.13 provides comparisons of teacher background characteristics for the respondents to the Year 3 teacher survey. These findings suggest that the high non-response rate for the Year 3 survey due to COVID-19 is unlikely to cause bias in the survey analysis. However, these findings need to be generalized with caution as the study found some small differences between students taught by teachers who responded to the Year 3 survey and those taught by teachers who did not respond (Appendix Table B.14).

IMPLEMENTATION AND DIFFERENTIAL IMPACT FINDINGS

The study intended to implement the A2i system and the two PD models in all study schools for three consecutive school years from the fall of 2018 through the spring of 2021. Data collected during this period would be used to assess whether the mixed-mode PD delivery model could be a viable, more cost-effective alternative to the face-to-face PD model in supporting the implementation of the A2i Professional Support System and improving lower elementary grade students' reading performance in the long run. For the mixed-mode model to be a viable alternative to the face-to-face model, one expects to see that (1) the delivered PD activities differ, as intended, in terms of quantity, delivery mode, and cost between the two models; (2) both models lead to similar levels of usage of the A2i technology and use of differentiated small-group instruction; and (3) both models lead to a similar level of student reading achievement at the end of the study. In other words, a "no differential impact" finding on student reading outcomes implies that the mixed-mode model might be a good substitute for the face-to-face approach.

However, this three-year implementation period was disrupted by COVID-19 in March 2020. Pandemic-related school closures and the ensuing shifts in learning modes fundamentally changed the teaching and learning environment for adults and children alike (see Figure 1). Such changes affect the context within which the study findings should be interpreted. Therefore, this section presents the study findings separately for the periods before and after the onset of COVID-19. These findings do not cover the period of school closures from March 2020 to June 2020 because no program-related PD or data collection took place during this period.

The Pre-COVID-19 Period

During the pre-COVID-19 period of the study from August 2018 to March 2020, the A2i program, supported by the two PD delivery models, was implemented in the study schools in a regular educational setting.

- **Teachers in mixed-mode and face-to-face schools received training and support that differed by amount and delivery mode as intended.**

As mentioned earlier, while the program provided all study schools with PD support to facilitate their use of A2i assessments and the A2i online platform throughout the pre-COVID-19 period, the planned amount and delivery mode of such support differed for the two groups of schools. The top panel in Figure 4 presents the number of PD events planned for and delivered in Year 1 and Year 2 (up until March 2020) by grade and by PD activity (Individualized Classroom Coaching and Literacy Huddles) and delivery mode separately for the two groups of schools. The gray bar or line on each graph represents the targeted number (for focal grades), or targeted range (for non-focal grades), of events for a given PD activity, while the dots represent the average number of PD events delivered to each group of schools.

Figure 4. Professional Development (PD) Delivered to and Received by Teachers, by Type, Year, Grade, and Program Condition, Pre-COVID-19

Number of PD events delivered
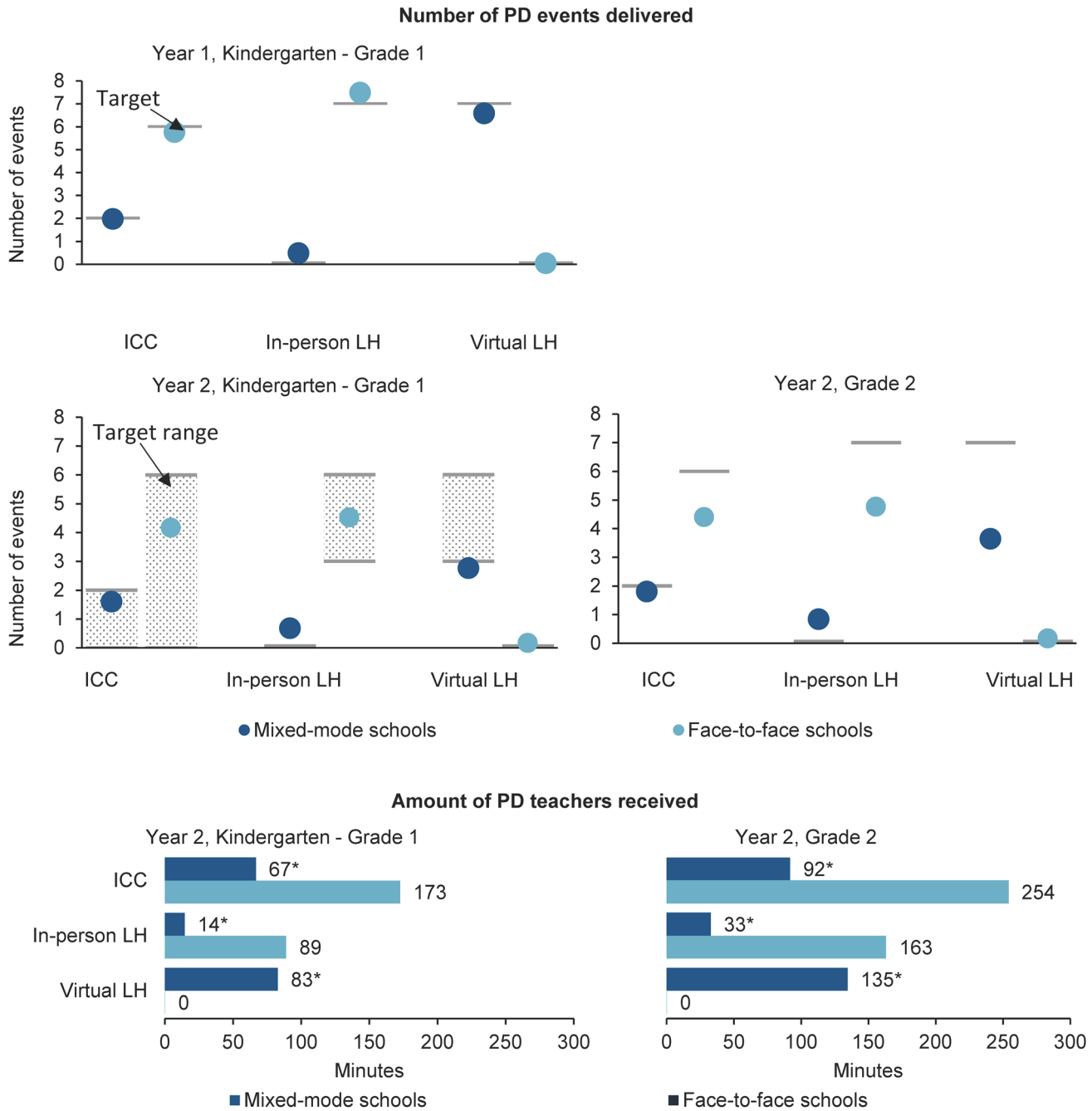
SOURCES: PD plan and delivery records and teacher-level and PD attendance records collected by Learning Ovations in Year 2 before school closures in March 2020.

NOTES: The targeted number or range of PD events reflects the Learning Ovations PD plan. The delivered number of PD events is the average number of events delivered to a given grade(s). ICC = Individualized Classroom Coaching, LH = Literacy Huddle. *indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

In Year 1, A2i training provided to teachers and school staff members largely occurred at a frequency and in the delivery modes consistent with Learning Ovations' implementation plan on average, creating the intended contrasts between the two PD delivery models. For example, the mixed-mode schools on average received 6.6 of the 7 planned virtual Literacy Huddles. In contrast, no virtual huddles were planned for schools in the face-to-face model, and none were delivered.[37]

The PD delivery was interrupted by the pandemic in the spring of Year 2. Due to school closures that started in March 2020, Learning Ovations had to cancel all PD activities for the last 3–4 months of that school year. As a result, the number of PD events that occurred during Year 2 was about a quarter to a third less than the planned number.[38] For example, on average, of the 7 planned A2i Literacy Huddles, 3.6 were delivered virtually to second grade teachers in the mixed-mode schools and 4.8 were delivered in-person to those in the face-to-face schools.

While the interrupted school year affected PD delivery in both mixed-mode and face-to-face schools alike, some contrasts remained in training delivered to these two groups of schools. For example, most A2i Literacy Huddles were conducted virtually in mixed-mode schools, while they were primarily in person in face-to-face schools. Moreover, as intended, more Individualized Classroom Coaching events took place in face-to-face schools than in mixed-mode schools. These contrasts reflect the planned variation between these two models that Learning Ovations managed to deliver before the school closures.

Teachers' PD attendance records from Year 2 also show evidence of intended contrasts in the amount and types of PD received by teachers from the two groups of schools (bottom panel of Figure 4). For example, second grade teachers (the focal grade in this year) in the face-to-face group on average had 254 minutes of Individualized Classroom Coaching per teacher, compared to 92 minutes received by the average teacher in the mixed-mode group. This difference is in line with the difference in planned Individualized Classroom Coaching sessions for these two groups of schools (6 sessions for face-to-face schools and 2 sessions for mixed-mode schools, with about 45 minutes per session). In addition, teachers from the mixed-mode schools primarily participated in the Literacy Huddles virtually, and those from the face-to-face schools only participated in person. Such contrasts in the PD

37. A small number of in-person Literacy Huddles took place in the mixed-mode schools. Those occurred when the LOS was already on campus for an in-person Individualized Classroom Coaching visit. Overall, most study schools received most of the planned PD: all but one mixed-mode school got at least half of the planned PD, 70 percent of them received at least three-quarters of the training, and 43 percent got the full dose. Similarly, all face-to-face schools received at least half of the PD, with 79 percent receiving at least three-quarters of the training, and 45 percent getting the full amount.

38. In Year 2, the LOS also offered on-demand Flex Days to teachers with unstructured, hands-on practice and collaboration. Records show that Flex Days on average occurred about 0.5 times for Kindergarten to Grade 1 teachers and 0.6 times for Grade 2 teachers in the mixed-mode schools, and 0.1 times for Kindergarten to Grade 1 teachers in the face-to-face schools.

amount and delivery mode between these two groups of schools were consistent across training types and grade levels in Year 2.[39]

- **In both groups of schools, most students took the A2i assessment frequently, but teachers did not fully utilize the A2i software platform.**

The PD activities are designed to support teachers' use of A2i technology, which contains two key components: the A2i assessments that help monitor student progress, and the A2i software platform, which provides system-generated instructional recommendations and lesson planning tools to help teachers more efficiently plan their classroom instruction. Past research has demonstrated that, for these elements to be effective, teachers need to use them with enough frequency or intensity.[40] The study used data obtained from the A2i platform and the Year 1 teacher survey to assess teachers' usage of these elements.

A2i Assessments

The A2i system includes two integrated online adaptive assessments—Letters2Meaning and Word Match Game—that capture students' vocabulary, decoding, and comprehension levels.[41] Students can take these assessments on laptops (such as Chromebooks) or desktop computers with a mouse and headphones. The assessments are designed with the expectation that young children should be able to complete them independently. The A2i algorithm uses the assessment data to generate instruction type, time, and grouping recommendations to help teachers plan and monitor students' language and literacy outcomes toward meeting grade-level expectations. Students are expected to take these assessments at least three times during a regular school year so that the A2i system can effectively monitor student progress and adjust recommendations. This expectation was reduced to twice for Year 2 due to school closures in the spring of that school year.

In Year 1 and Year 2, the majority of students in both groups met the expected testing frequency. Using the Letters2Meaning assessment as an example, on average, students in both groups took this assessment 4.3 times in Year 1. In Year 2, this number dropped to 3 times for mixed-mode schools and 2.7 times for face-to-face schools (top panel of Figure 5). Specifically, as shown in the appendix, over three-quarters of all students took the Leters2Meaning assessment three times or more in Year 1 in both groups of schools. In Year 2, over 80 percent of students took the assessment at least twice, and about 60

---

39. Taking into account the reduced amount of PD delivered in Year 2, 21 of the 25 mixed-mode schools and all of the 25 face-to-face schools that remained in the study in Year 2 had at least 60 percent of their teachers receiving the full amount of delivered PD, reaching the fidelity threshold for PD participation (see Appendix C for details about the calculation).

40. Connor et al. (2013).

41. A psychometric analysis of scores from these two assessments showed that they have a reliability greater than 0.7. See Connor et al. (2017).

Figure 5. Usage of A2i Technology, by Year and Program Condition, Pre-COVID-19

Year 1   Year 2

Number of A2i Letters2Meaning (L2M) assessments students took

Average number of L2M assessments: Year 1 — 4.3 (Mixed-mode), 4.3 (Face-to-face); Year 2 — 3.0 (Mixed-mode), 2.7 (Face-to-face)

Teacher usage of the A2i platform

Total minutes spent on platform: Year 1 — 109 (Mixed-mode), 142 (Face-to-face), Threshold, Estimated difference: -33.6* minutes; Year 2 — 169 (Mixed-mode), 193 (Face-to-face), -23.8 minutes

Percentage of teachers who spent more time than the threshold: Year 1 — 11 (Mixed-mode), 20 (Face-to-face), -9.4%*; Year 2 — 35 (Mixed-mode), 52 (Face-to-face), -16.7%*

■ Mixed-mode schools   ■ Face-to-face schools

SOURCES: A2i assessment data and A2i system log data from the A2i platform, collected throughout the pre-COVID-19 period in Year 1 and Year 2.

NOTE: *indicates the estimated difference is statistically significant at the 0.05 level with a two tailed t-test.

percent to 65 percent tested three or more times before school closures in March. The pattern for the Word Match Game assessment is similar.[42]

Responses to the Year 1 survey show that teachers in both groups felt confident in their students' ability to take the assessment independently and reported that they had adequate training for administering the assessment, but more assistance from school staff members might help them carry out this task more efficiently.[43]

---

42. See Appendix Figure C.1 for details. The proportion of students tested at or above the expected frequency varied by school in both years, ranging from 40 percent to close to 100 percent in Year 1 and from 60 percent to 100 percent in Year 2 except for a few outliers. See Appendix Figure C.2 for details.

43. More than half of the teachers who responded to the Year 1 teacher survey did not get any adult help administering the assessment. Only a little more than half of them felt they could do it efficiently, even though they claimed to have had enough training for the task (Appendix Table C.4).

A2i Software Platform

The A2i software platform contains student assessment data, personalized instruction recommendations generated by the A2i system based on assessment information, and materials that help teachers with their lesson planning. Active use of such information can support the instruction teachers provide to their students, as research has shown that the amount of time teachers spend on the A2i platform is positively associated with their use of differentiated in-class instruction and with their students' reading outcomes.[44] The expectation for adequate usage was that teachers should spend at least 240 minutes on the platform during a school year to benefit from the information provided by the system.[45] This benchmark was adjusted to 160 minutes in Year 2 to account for the school closures and disruptions resulting from the COVID-19 pandemic.

The A2i system log data recorded the number of minutes each teacher spent on the system's web pages during a school year.[46] Such data show that most teachers in mixed-mode and face-to-face schools did not spend the expected amount of time on the A2i software platform (bottom two panels of Figure 5).

In Year 1, about 20 percent of teachers in the face-to-face schools and 11 percent in the mixed-mode schools reached the benchmark level. The average amount of time teachers spent on the A2i platform was 109 minutes for the teachers in mixed-mode schools and 142 minutes for those in the face-to-face schools, far below the expected amount of 240 minutes. Several factors might have contributed to this lower-than-expected usage level. First, some of the time teachers spent on the platform might not have been attributed to them. One focus area of the PD sessions in Year 1 was to train teachers on navigating and using the A2i platform; therefore, a proportion of teachers' use of the platform took place under the supervision of the LOS. It is plausible that the time the LOS spent demonstrating the available information to the teachers or walking the teacher through their classroom information during PD sessions was recorded under the LOS's login and not the teachers' login information. Second, delays in rolling out some components of the platform during the first year might have hindered teachers' usage. For example, the dashboard/data visualization tools were released in the spring of 2019, and the curriculum/lesson plan improvement and resource pages were released late in the fall of 2019.

Year 2 saw an increase in the average A2i platform usage, even though the school year was cut short by school closures. On average, the mixed-mode group teachers spent

---

44. Connor et al. (2013).

45. This is the minimum viable amount of usage time suggested by past studies of the program (Connor et al., 2013).

46. The study grouped these web pages into four categories based on their content: (1) student information pages, including student information, test scores, graphs, progress, and classroom information; (2) resource pages, including resources and a video library; (3) A2i tool pages such as lesson plans, reports, and administrative information; and (4) A2i login pages, including the system portal and teachers' account pages, among other login pages. Teachers in both groups spent more time on pages related to student information than on other pages. Detailed findings can be found in Appendix Table C.5.

169 minutes on the platform, and the face-to-face group teachers spent 193 minutes, an increase of 56 percent and 36 percent from their Year 1 levels, respectively. This increase could reflect Learning Ovations' enhanced emphasis on the need for teachers to spend time on the A2i platform and to make use of the tools and resources for differentiation.[47] However, even with this increase, the overall usage level was still below expectations: only about 35 percent of the mixed-mode teachers and 52 percent of the face-to-face teachers met the reduced level of expected usage.

In both years, a higher proportion of teachers in the face-to-face schools met the expected usage level than teachers in the mixed-mode schools. One possible explanation is that more in-person interaction between teachers and LOSs might help build teacher buy-in to the program, which could lead to a higher degree of reliance on the platform. An alternative hypothesis is that teachers were more likely to log onto the A2i system during face-to-face training sessions than during virtual training sessions. However, a closer look at teacher usage timing does not support this explanation, as only about a quarter of teachers' usage time coincided with PD training days, and there was no noticeable difference between these two groups of teachers in the amount of time they spent on the platform during PD days.[48]

Lesson Planning Tools

The A2i platform also offers lesson planning tools. These tools provide resources such as daily lesson plan templates, group activity planners, and a library of instructional activities drawn from the core literacy curricula of many school districts and the Florida Center for Reading Research activities.[49] These tools are designed to help teachers prepare for their lessons, especially for planning differentiated instruction efficiently and effectively.

However, teachers in both groups of schools reported limited use of these tools in the Year 1 teacher survey. Respondents from both groups of schools said that, on average, they used the system-generated lesson suggestions for one day's lesson or less in a typical school week. Only 19 percent of the mixed-mode group teachers and 26 percent of the face-to-face group teachers reported that they used the system-generated lesson plan for some groups in their classes.[50] Qualitative research conducted by Digital Promise Global as part of this grant also indicates that, while the redesigned lesson menu is more user-friendly, it is still challenging to navigate. In addition, the activities suggested by the lesson menu tend to deviate from the teachers' pacing guide, making them hard for teachers to adopt, even though A2i was designed to support schools' existing curricula.[51]

---

47. Learning Ovations also updated and added tools and resources (for example, a resources page, a lesson menu, data tools, and a teacher dashboard) in Year 2 to the A2i platform to improve teachers' experiences and usage.

48. Appendix Table C.6 presents detailed findings of this analysis.

49. See www.fcrr.org.

50. Appendix Table C.8 presents teachers' responses to Year 1 survey questions related to their use of lesson planning tools.

51. Young, Vang, and Kasad (2020).

- **Teachers in both mixed-mode and face-to-face schools demonstrated a reasonable understanding of the recommended instructional practices. However, they expressed some uncertainty regarding the validity of the assessment, did not fully adopt A2i instructional recommendations, and found it challenging to implement differentiated instruction for small groups.**

Differentiating literacy instruction to meet individual needs is the driving premise of A2i. In A2i, this entails small-group instruction as the primary organizing structure in the classroom. Under this structure, each small group engages in reading practices that are appropriate to the group's learning level, so there is differentiation in instruction across small groups, even when different groups rotate through the same work centers (stations).[52] The whole suite of A2i technology centers around providing information (student assessment data), resources (grouping and instruction recommendations and lesson planning tools), and professional supports (PD activities provided by experts) to help teachers prepare and deliver high-quality, differentiated small group reading instruction to their students. Teachers' awareness and perceptions of these resources could affect their usage of the system and their adoption of differentiated small-group instruction in their classrooms. The study examines teachers' understanding of the system through teachers' responses to relevant survey questions collected in the spring of Year 1.

Knowledge of Instructional Recommendations

Teachers in both groups demonstrated a good understanding of the four types of instructional recommendations that A2i technology provides (see Figure 2 for descriptions). In the surveys, teachers were asked to classify six hypothetical classroom scenarios as one of the four types of system-recommended instructional activities. This exercise intends to test teachers' understanding of specific instructional strategies and the dimensions of instruction in the A2i system. In Year 1, teachers in both groups of schools on average answered about 75 percent of these questions correctly, with an average score of 4.8 and 4.5 (out of 6) for mixed-mode and face-to-face schools, respectively (top panel of Figure 6). The correct rate for each question ranges from 59 percent to 89 percent.[53]

A2i Assessments and Instructional Recommendations

Teachers need to believe in the program to make long-term changes to instructional practices. It is hard for teachers to buy in to the program if they have doubts about what the program wants them to do. The teacher survey asked teachers about their views on two key A2i components—the A2i assessments and the system-generated instructional recommendations—and found that teachers had reservations about both.

---

52. Note that after the onset of COVID-19 and pandemic-related restrictions, such as social-distancing, some teachers had tried to differentiate with student-level learning activities. This is different from the differentiated small-group instruction emphasized by A2i.

53. See Appendix Table C.9 for details.

Figure 6. Teachers' Reactions to A2i Components, by Program Condition, Pre-COVID-19

Knowlege of A2i-recommended instruction strategies

Total score
(0 to 6 points)
- 4.8
- 4.5

0   1   2   3   4   5   6

Beliefs about A2i assessments and recommendations

I believe that the data collected on A2i assessments is valid if taken at school.
- 45
- 54

I found the A2i assessment results useful for understanding my students' academic needs.
- 53
- 56

I found the A2i recommended amounts of each type of instruction useful.
- 57
- 60

0   10   20   30   40   50   60   70   80   90   100
Percentage agree

Agreement with A2i-generated recommendations

Grouping recommendations

Agreed with A2i-generated recommendations for majority of students
- 52
- 67

0   25   50   75   100

Made adjustments to A2i recommendations
- 89
- 87

0   25   50   75   100
Percentage

Time recommendations

Agreed with A2i-generated recommendations for majority of students
- 54
- 66

0   25   50   75   100

Made adjustments to A2i recommendations
- 84
- 77

0   25   50   75   100
Percentage

■ Mixed-mode schools   ■ Face-to-face schools

(continued)

Figure 6 (continued)

Use of differentiated small-group instruction

Do you use this practice in your reading block?
- 98
- 99

Percentage (0, 25, 50, 75, 100)

How many days do you use it in a typical week?
- 4.2
- 4.2

Number of days (0, 1, 2, 3, 4, 5)

More than half of reading block time is dedicated to differentiated small-group learning
- 74
- 69

Percentage (0, 25, 50, 75, 100)

■ Mixed-mode schools   ■ Face-to-face schools
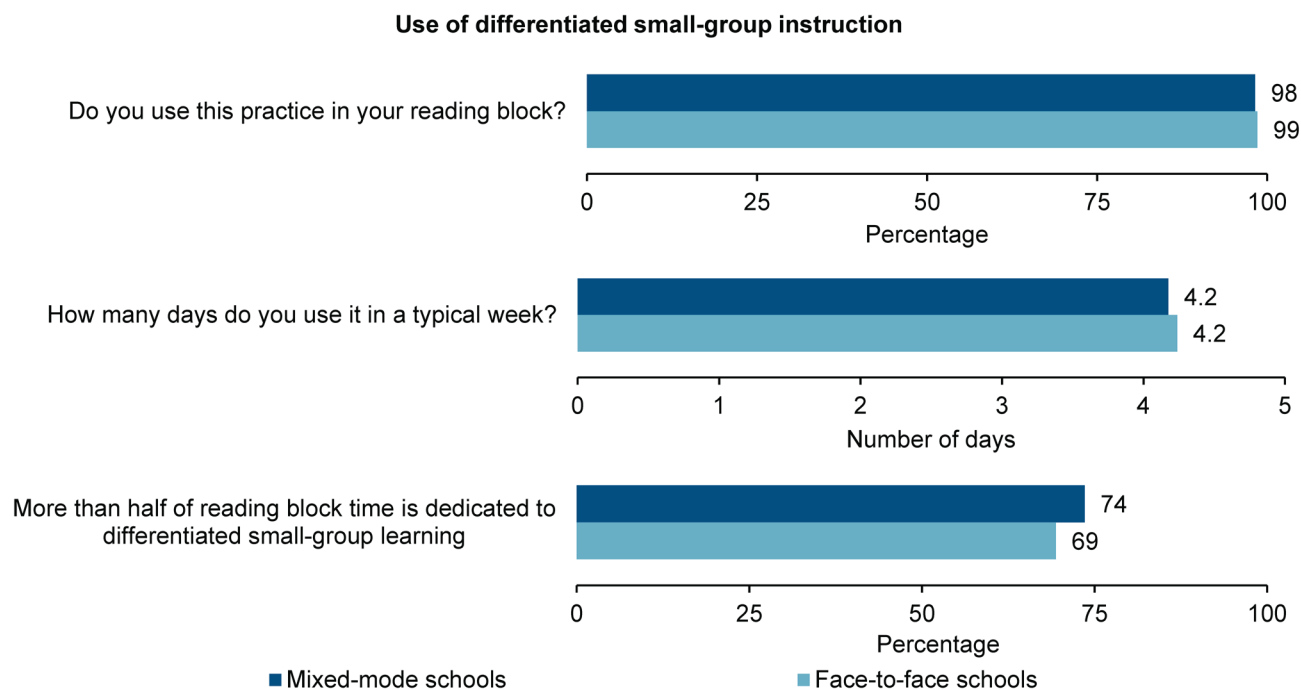
SOURCE: Teacher survey collected in spring 2019.

NOTE: None of the estimated differences between the mixed-mode and face-to-face schools is statistically significant at 0.05 level.

Teachers expressed some uncertainty about the A2i assessments' validity. Only 45 percent of teachers from mixed-mode schools and 54 percent of those from face-to-face schools agreed with the statement that the assessment "is valid," and less than 60 percent of the teachers found the A2i assessment helpful in understanding students' needs (second panel of Figure 6). Digital Promise Global's qualitative research also finds a noticeable portion of teachers struggled with the validity of the assessments, and most teachers expressed the need for more diagnostic information about students' reading skills and said that they used data from other assessments for such information.[54]

While close to 60 percent of teachers in both groups found the A2i-recommended amounts of each type of instruction useful (second panel of Figure 6), they also reported limited agreement with the system-generated instructional recommendations (third panel of Figure 6). Using individual students' assessment data, the A2i algorithm computes recommendations (in minutes per day) for each of the four types of literacy instruction and

---

54. Young, Vang, and Kasad (2020).

how students should be grouped for different activities.[55] When surveyed in the spring of Year 1, slightly more than 50 percent of teachers in mixed-mode schools and a little more than 65 percent of those in face-to-face schools said they agreed with the grouping and minutes recommendations for more than half of their students. More than 80 percent of teachers also reported making adjustments to the recommendations based on their professional judgment or looking at classwork, grades, or other diagnostic tests.

Teachers' general uncertainties about the two key components of the A2i technology might have hindered their use of the A2i system and might partially explain the less-than-expected amount of time they spent on the A2i platform.

Differentiated Small-Group Instruction
Most teachers in mixed-mode and face-to-face schools reported using small-group instruction for at least part of their literacy instructional time in the Year 1 teacher survey. More than 98 percent of all survey respondents said that they used small-group instruction in their reading blocks and that they used it about four days a week on average. About 30 percent of teachers in the mixed-mode group and 35 percent of teachers in the face-to-face group reported that they devoted more than half of their literacy instructional time to differentiated small-group instruction (last panel of Figure 6).

However, information collected from interviews of LOSs and teacher focus groups indicates that teachers faced challenges when differentiating small-group instruction.[56] In Year 1, while most teachers recognized the importance of small-group instruction and the need for differentiation, they struggled with managing differentiation while implementing the system-generated recommendations on type of instruction, minutes of instruction, and student groupings. Qualitative data collected in Year 2 showed some movement along the continuum toward differentiated small-group instruction in both types of schools. However, teachers continued to find it challenging to differentiate child-managed activities and still needed help managing different amounts of time for each type of instruction for individual students.

- **There were similar patterns of continuous growth in students' reading skills in both the mixed-mode and face-to-face schools during the pre-COVID-19 period.**

Past research on the effectiveness of A2i has demonstrated that the amounts and types of reading instruction required to achieve reading proficiency differ for students, individualized small-group instruction tailored to students' needs can lead to better reading outcomes, and the A2i system is an effective tool to help teachers and students achieve this goal.[57] This study builds on the existing evidence for A2i and focuses on whether students who were exposed to two versions of the PD delivery model for the A2i system reached the

---

55. Connor et al. (2007).

56. Young, Kasad, and Vang (2019); Young, Vang, and Kasad (2020).

57. Connor et al. (2013).

same reading proficiency level. To this end, the study found that, for a cohort of students who started the program as first-graders and moved along the grade levels with the program rollout, there were no measurable differences in the average reading performance between those in the mixed-mode and face-to-face schools at the end of Year 1 and in Year 2 before school closures.[58]

The study measured students' vocabulary, decoding, and comprehension skills by students' scores from the last round of recorded Letter2Meaning and Word Match Game assessments by the end of Year 1 and prior to school closures in Year 2.[59] A student is considered reading at grade level in a given assessment if their grade equivalent score from the last assessment in the school year exceeded the expected grade equivalent based on the test month of the assessment. For example, if a school year started in September and a second-grader took the last assessment in April (the 8th month of the school year), the expected grade equivalent for this student would be 2.8. If this student's actual grade equivalent score from the assessment is equal to or greater than 2.8, then they are considered to be reading at grade level, and vice versa.

The study found that both groups of schools had similar proportions of students reading at grade level at the end of both years (top panel of Figure 7). For example, at the start of the program (fall of Year 1), about 49 percent of primary cohort students in the mixed-mode schools and 46 percent in the face-to-face schools were reading at grade level based on their first score in the Letters2Meaning assessment. After about one year of exposure to the program, this proportion rose to about 55 percent–56 percent in spring Year 1, with no difference between the two groups of schools. This proportion continued to increase in both groups to about 68 percent–69 percent as measured by the last assessment taken before the school closures in Year 2. Students' performance on the Word Match Game assessments exhibited a similar growth pattern during the pre-COVID-19 period for both groups of schools, even though most growth appears to have occurred during Year 1. Similar findings are also observed when the sample is limited to the 50 schools that remained in the study for these two years and to students for whom all outcome measures were available.[60]

The study also fielded the Gates-MacGinitie reading test to first-graders in the spring of Year 1 to get an exploratory assessment of reading differences between these groups of schools after one year of implementation (bottom panel of Figure 7). About 84 percent of first-graders in both groups of schools took the test.[61] Students from the mixed-mode and

---

58. There was no difference in students' average reading performance between the mixed-mode and face-to-face schools using other sample configurations (see Appendix D for more details).

59. The timing of the last assessment varies by student. The average test date was 247 days from the start of school for Year 1, 166 days for Year 2, and 242 days for Year 3. There was no difference between mixed-mode and face-to-face schools.

60. See Appendix Table D.2 for detailed findings for this sample.

61. Appendix Table B.1 provides more details on the response rates of the Gates-MacGinitie test.

Figure 7. Estimated Differential Impacts on Student Reading Outcomes, by Outcome and Program Condition, Pre-COVID-19

Proportion of students reading at or above grade level as measured by A2i assessment

GMRT (Spring of Year 1)
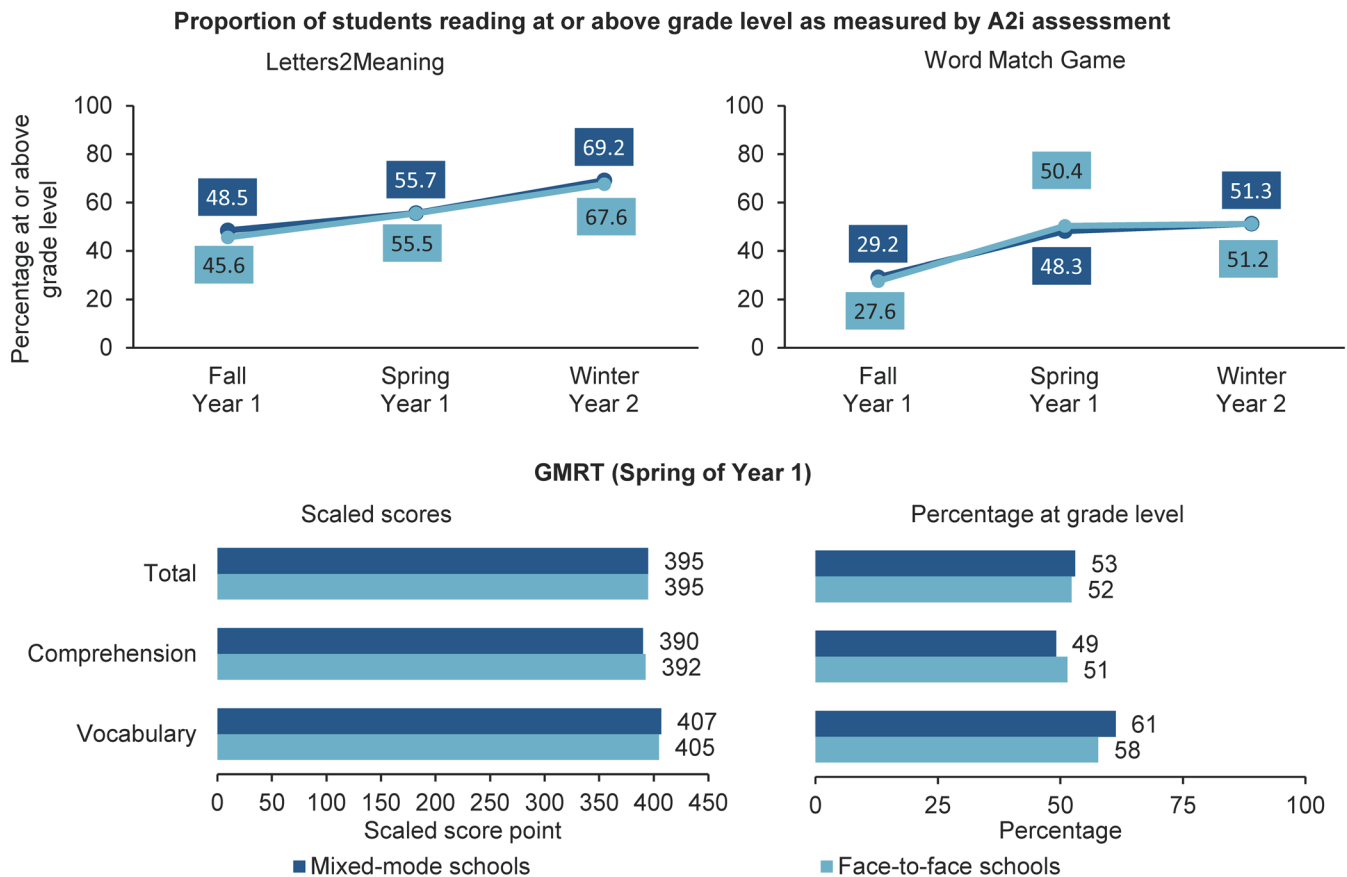
SOURCES: Gates-MacGinitie Reading Test (GMRT) data collected in the spring of 2019; A2i assessment data collected by the A2i platform from summer 2018 to early spring 2020, and district records data for the 2017-2018 school year.

NOTES: See Appendix Table B.1 for sample sizes in each estimation. None of the estimated differences is statistically significant at the 0.05 level.

face-to-face schools performed at about the same level for the test, with similar average total scores and scores on the vocabulary and comprehension subtests. The average total scaled scores for both groups of schools translate to a grade equivalent of 1.9, and about 53 percent of the mixed-mode group students and 52 percent of those in the face-to-face group scored at or above grade level as measured by the total Gates-MacGinitie score.[62]

It is worth noting that the observed increase in students' reading performance could reflect the natural growth in the reading level that young children would have experienced under normal circumstances. For example, the gain in the Letters2Meaning scaled score

---

62. The study fielded the Gates-MacGinitie test to all study schools between late March and early May of 2019.

from fall to spring in Year 1 was equivalent to about one standard deviation, and the spring to spring increase between Year 1 and Year 2 was equal to about 0.7 standard deviations in effect size.[63] These rates of growth are smaller than the developmental trajectory for nationally representative student samples, which reported the average estimated grade to grade reading gains for the transition from Kindergarten to Grade 1 to be 1.52 standard deviations, and from Grade 1 to Grade 2 to be 0.97 standard deviations, based on information for seven nationally normed standardized reading tests.[64] Such trajectories in early elementary grades were similar for students who were eligible for free or reduced-price lunch or whose reading test scores were low enough to place them at the 25th percentile of their district.[65] The growth rates observed in this study are also smaller than the grade to grade gains reported in a prior study of the A2i program with a different student sample. That study indicates a gain of 1.71 standard deviations between Kindergarten and Grade 1, and a 1.35 standard deviation gain from Grade 1 to Grade 2 for a sample of students exposed to the A2i program.[66] Without a comparison to a business-as-usual condition in similar schools not implementing A2i, the study cannot determine whether the observed growth in students' reading levels was the result of A2i or a reflection of expected natural growth.

Overall, before the onset of the COVID-19 pandemic that forced schools across the country to close in March 2020, teachers in the mixed-mode and face-to-face schools received training and support that differed in amount and delivery mode as intended. However, in both groups of schools, teachers' average usage of such technology was less than the predetermined thresholds, which might be linked to teachers' skepticism about the assessment and system recommendations. Teachers in both groups also found it challenging to implement differentiated small-group instruction in class. This period saw continued growth in students' reading performances, with higher proportions of students reading at grade level compared to where they were at the start of the program in both groups of schools. The rate of growth observed here, however, was smaller than the typical developmental trajectory and those reported in prior studies of A2i.

To address issues that arose from the implementation of A2i, the United2Read team developed and used an internal feedback circle to continuously adjust the program and its implementation. Each year, researchers from Digital Promise Global provided Learning Ovations with information about school and teacher experiences with the program and issues they discovered during their qualitative research. Learning Ovations took action to

---

63. The effect sizes were computed as the ratio between the difference in mean scaled scores and the pooled standard deviations for each pair of adjacent grades.

64. These tests include CAT/5, SAT 9, TerraNova, Gates-MacGinitie, MAT 8, and TerraNova CAT (Bloom, Hill, Black, and Lipsey, 2008).

65. Bloom, Hill, Black, and Lipsey (2008).

66. Conor et al. (2013). The student sample in this prior study differs from the study sample in that 81 percent of the participating students were White and on average 47 percent of students in participating schools received free or reduced-price lunch, as compared to 27 percent and 80 percent, respectively, in the current sample of schools.

address these issues in a timely manner. For example, Digital Promise Global identified opportunities to better clarify what the program is and is not for various stakeholders at the beginning of A2i implementation. In response, Learning Ovations created a number of new structures and materials, including a documented scope and sequence of professional development to share with teachers and school and district administrators, a revamped lesson planning tool for teachers to easily create lessons aligned with A2i recommendations, standardized materials for LOSs to use during professional development, and an accompanying rubric specifying developmental stages for LOSs to use in supporting teachers and schools. Notably, to facilitate teachers' implementation of A2i, Learning Ovations launched the Learning Ovations Framework in January 2020. The purpose of this framework was to align coaching better with teachers' needs based on their progress and to provide them with a clearer sense of what an A2i classroom can look like and what steps are needed to make differentiation part of their routine. By doing so, the team continuously made improvements to the A2i system and its implementation throughout the pre-COVID-19 period.

In March 2020, the COVID-19 pandemic struck the country, causing unprecedented disruptions to the school system. Most schools were closed for in-person instruction immediately and stayed closed for the rest of Year 2. By the time schools restarted in the fall of 2020, most students had been out of their normal learning environment for five to six months.

The 2020–2021 School Year

COVID-19-related disruptions to schools continued well into the 2020–2021 school year. Across the country, districts and schools toggled between in-person and online teaching as they balanced the need to keep students and staff members safe with the need to provide an effective learning environment. Students also reported experiencing multiple obstacles to learning, such as feeling depressed or anxious, as well as having health-related disruptions at home.[67] Teachers and students in this study also experienced significant changes in the learning environment. Teachers reported that COVID-19 negatively affected the length of their literacy instructional time and available time to plan for individualized instruction, likely due to changing learning arrangements and time spent on technical issues during remote learning and on complying with in-person restrictions such as cleaning and social distancing. Adult help, such as co-teachers, teachers' aides, and volunteers, was less available in their classes as schools refocused resources on students' well-being and other priorities, such as complying with distance and sanitary guidelines. They also faced students who were less engaged or had trouble following instruction during virtual learning, likely due to struggles with internet connections and Zoom fatigue.[68] While Learning Ovations adjusted PD activities and A2i technology to mitigate the disruptions in instruction caused by the pandemic and to facilitate implementation during this difficult time, the

---

67. YouthTruth (2021).

68. See Appendix Table C.10 for details of teachers' responses to survey questions related to these issues.

challenges brought on by COVID-19 still deeply affected PD delivery, teacher behavior, and student performance in this study.

- **Changes associated with COVID-19 in schools significantly affected the delivery of the planned professional development supports for both mixed-mode and face-to-face schools and reduced the contrast between them.**

Due to COVID-19-related restrictions on access to school premises, Learning Ovations switched all Literacy Huddles and Individualized Classroom Coaching to virtual mode in the 2020–2021 school year, meaning teachers in the face-to-face schools received all their Literacy Huddles virtually instead of in person, and teachers in both groups received the Individualized Classroom Coaching virtually as well. This shift diminished the crucial distinction in delivery mode between the two models in Year 3.

In addition, both groups received fewer PD events than expected. Scheduling such events was challenging because of frequent changes to learning environments in schools. Consequently, the number of PD events that took place was either below the target or at the low end of the targeted range (left panel of Figure 8). For example, only 4.8 of the 7 planned Literacy Huddles and 3.2 of the 6 planned Individualized Classroom Coaching sessions were delivered virtually to third grade teachers in the face-to-face schools. Similarly, third grade teachers in the mixed-mode schools received 4.6 of the 7 Literacy Huddles and 1.6 of the 2 planned Individualized Classroom Coaching sessions. Overall, 4 mixed-mode schools (or 20 percent) and 9 face-to-face schools (or 45 percent) were offered the full amount of planned PD.[69]

Relatedly, teachers received less PD in Year 3 than they did in Year 2, even though Year 2 was cut short by COVID-19 (bottom panel of Figure 4 and right panel of Figure 8). The decline was particularly noticeable in the Individualized Classroom Coaching sessions. In mixed-mode and face-to-face schools, teachers in the focal grade in Year 3 (grade 3) logged 31 minutes and 84 minutes of Individualized Classroom Coaching, compared to 92 minutes and 254 minutes logged by teachers in the focal grade in Year 2 (grade 2), respectively. It was not surprising to see such a decline given that schools and teachers were under a great deal of pressure. It was also difficult to schedule and deliver online coaching during the school year. Even though there remained some differences in the amount of PD received by teachers in these two groups of schools, the crucial distinction between the two PD models—virtual versus in-person delivery—was non-existent. Due to the diminished contrasts between these two models, any comparisons of implementation and impact between them do not accurately reflect their true differences if fully delivered as intended.

- **The remote administration of A2i assessment during Year 3 exacerbated teachers' concerns about the program's validity.**

---

69. Appendix Table C.2 provides the details of this calculation.

Figure 8. Professional Development (PD) Delivered to and Received by Teachers, by Type, Grade, and Program Condition, 2020-2021 School Year
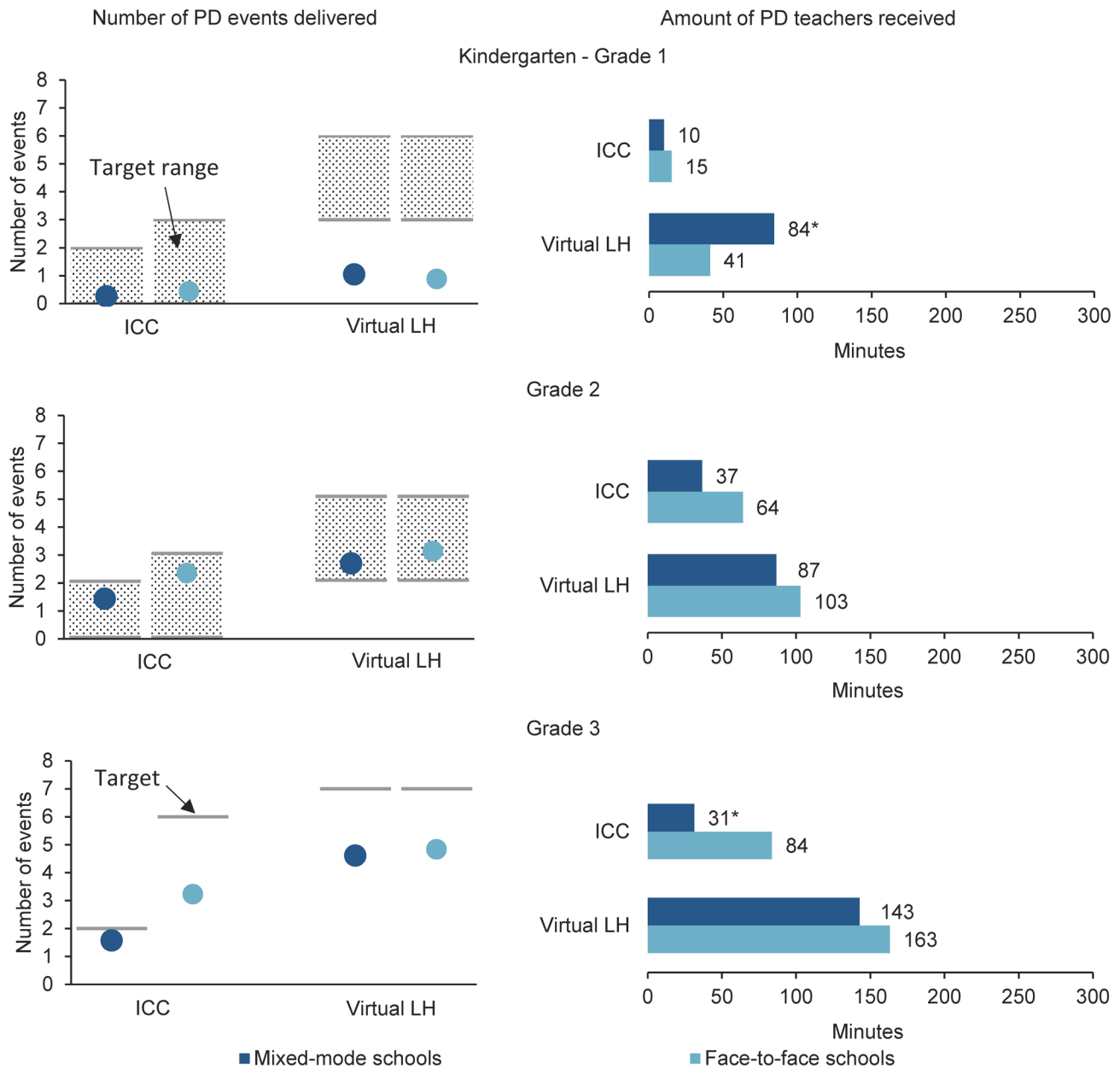
SOURCES: PD plan and delivery records and teacher-level PD attendance records collected by Learning Ovations in the 2020-2021 school year.

NOTES: The targeted number/range of PD events reflects Learning Ovations' PD plan. The number of PD events delivered is the average number of events delivered to a given grade(s) across the mixed-mode schools or face-to-face schools. ICC = Individualized Classroom Coaching, LH = Literacy Huddle. *indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

In Year 3, as schools oscillated between different learning arrangements, many students took at least some of the A2i assessments remotely.[70] Teachers did not think the assessment scores were accurate or of value if the students who took it remotely had help from family members at home: less than 20 percent of surveyed teachers believed in the validity of the assessments when taken at home (second panel of Figure 9).

Figure 9. Teachers' Knowledge and Beliefs About A2i, by Program Condition, 2020-2021 School Year
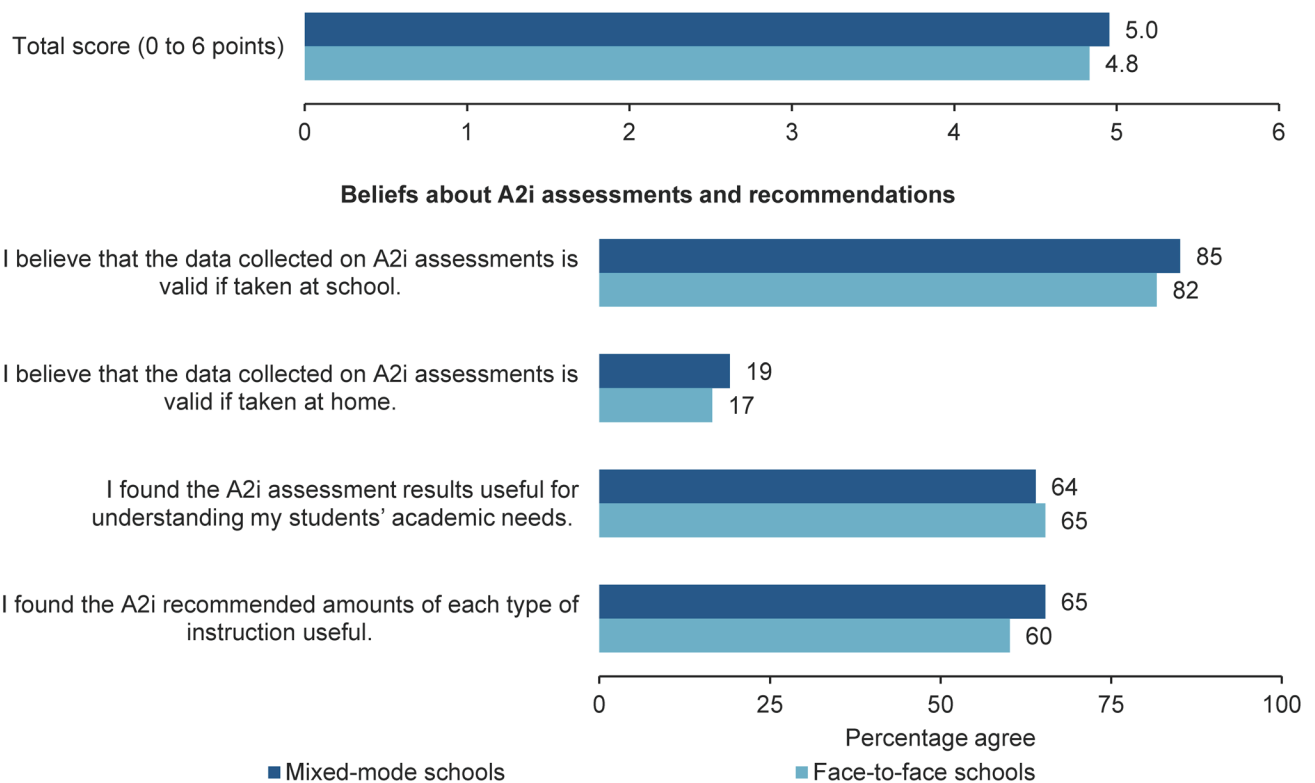
SOURCE: Teacher survey collected in spring 2021.

NOTE: None of the estimated differences between the mixed-mode and face-to-face schools is statistically significant at the 0.05 level.

---

70. The Year 3 teacher survey shows that about three-quarters of the students were learning remotely when they took their first two A2i assessments, and most of these students took the assessments virtually.

- **Such concerns, together with logistical challenges, might have affected teachers' use of the A2i technology.**

The frequency with which students took the A2i assessments decreased in Year 3, with the average number of assessments students took dropping from 4.3 in Year 1 (the last whole school year available for assessments) to 2.2 in Year 3 (top panel of Figure 10). The proportion of students in the mixed-mode and face-to-face schools who took the A2i assessments three times or more dropped from over 80 percent in Year 1 to below 65

Figure 10. Usage of A2i Technology, by Program Condition, 2020-2021 School Year

Number of A2i Letters2Meaning (L2M) assessments students took

Agreement with A2i-generated recommendations

Teacher usage of the A2i platform

■ Mixed-mode schools　　　　■ Face-to-face schools
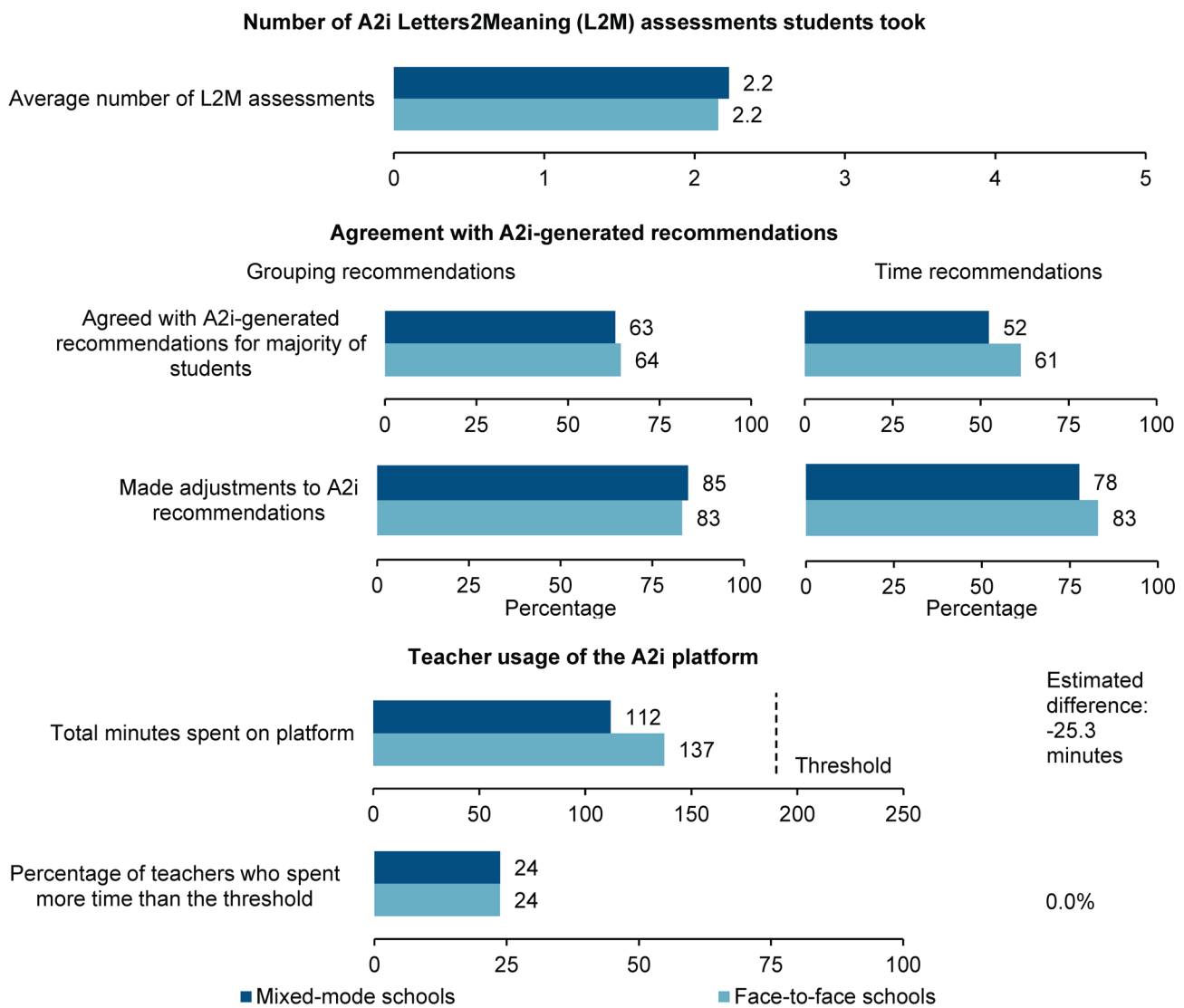
(continued)

Figure 10 (continued)

Use of differentiated small-group instruction

Do you use this practice in your reading block?
- 73
- 76

How many days do you use it in a typical week?
- 3.0
- 2.7

More than half of reading block time dedicated to differntiated small-group learning
- 19
- 19

■ Mixed-mode schools   ■ Face-to-face schools
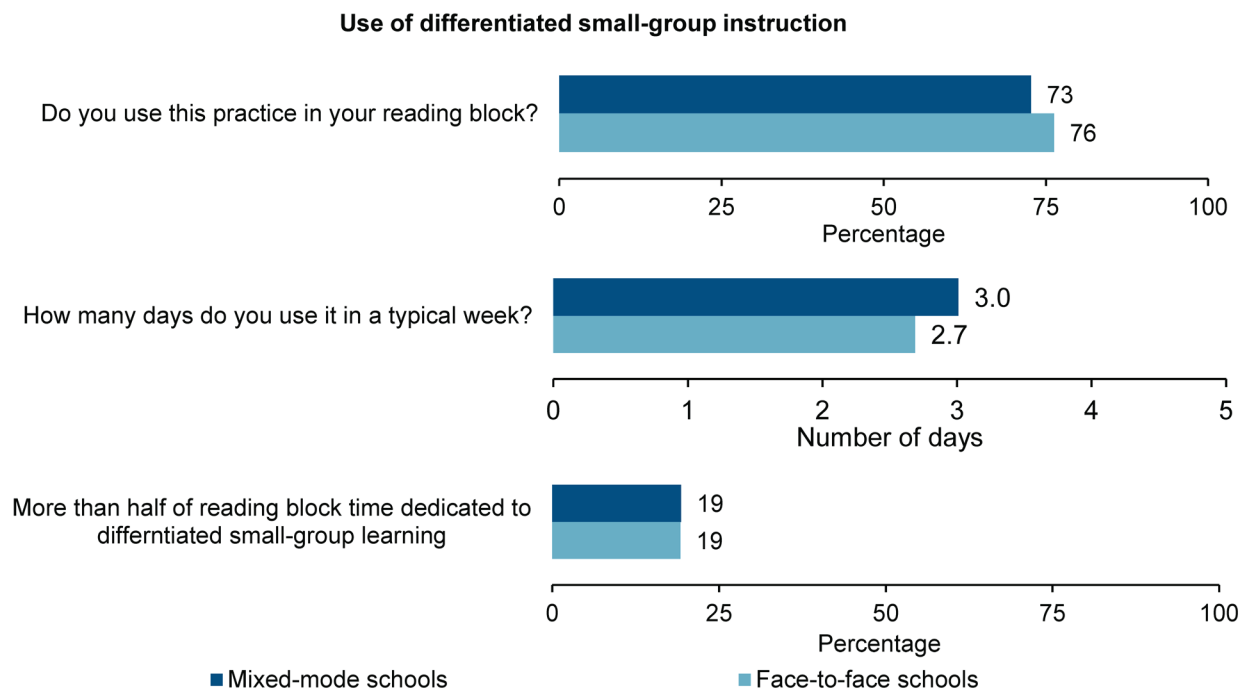
SOURCES: Teacher survey collected in spring 2021. A2i usage data obtained from the A2i platform for this study.

NOTES: None of the estimated differences between the mixed-mode and face-to-face schools is statistically significant at the 0.05 level.

percent in Year 3.[71] Such a drop could reflect teachers' concerns about the validity of the assessment. If teachers did not think the assessment captured the students' true reading level, they would be less motivated to administer the assessment. It could also reflect the logistical challenges teachers and schools were faced with during this school year: Teachers found it more challenging to administer the assessment remotely, as it required more time and effort to address technical issues and to monitor students as they took the assessment at home than in the typical school environment. As a result, many districts and schools reduced the expected number of A2i assessments or even made them optional to ease teachers' workloads.[72]

Teachers also did not fully adopt the A2i system-generated instructional recommendations in Year 3. Like Year 1, less than two-thirds of the teachers agreed with the grouping or instruction time recommendations for the majority of their students, and over three-quarters of them reported changing the recommendations for at least some of the students in their

---

71. See Appendix Figure C.1 for details. In addition, none of the study schools reached the expectation of having at least 95 percent of students tested at least three times in Year 3 (Appendix Table C.2).

72. Kasad, Vang, and Young (2021).

classes (second panel of Figure 10). Less than half of the survey respondents reported that they created small groups or gave instruction based on A2i time recommendations.[73] LOSs also reported that teachers in only 36 percent of the schools included in the qualitative research used A2i data to inform their instruction in fall 2020, and this proportion dropped to 21 percent in spring 2021.[74] This is not surprising given that teachers did not believe in the validity of the remotely administered A2i assessments and the recommendations were based on the assessment results.

As a result, teachers' use of the A2i platform was below expectations in both groups of schools (third panel of Figure 10). Even though the expected usage level was adjusted from 240 minutes in Year 1 to 190 minutes to account for the challenges teachers faced in Year 3, only about 24 percent of teachers in both groups reached this adjusted expectation. As in Year 1, teachers in neither group used the lesson planning tools much, reporting that they used lesson suggestions for less than one day's lessons in a typical week.[75]

- **Teachers in both groups of schools found it difficult to use differentiated small-group instruction in Year 3 due to the changing learning environment and COVID-19-related restrictions.**

While close to three-quarters of survey respondents in both groups still reported that they used small-group instruction and that they used it for around three days a week, these numbers are lower than what was reported by teachers in Year 1 when almost all teachers reported that they used such instruction and for over four days a week (last panel of Figure 10). Teachers also reported spending a smaller proportion of reading instruction time on small-group instruction: about 20 percent of the surveyed teachers reported spending half or more of reading block time on small-group instruction, a sharp drop from over 70 percent in Year 1. The qualitative research conducted by Digital Promise Global corroborates this decrease in differentiated small-group instruction. Teachers found that implementing small-group instruction became logistically challenging during remote learning. Even with in-person instruction, various COVID-19 and social distancing restrictions made it hard for teachers to pull small groups for differentiated instruction. LOSs also reported little consistency in differentiating literacy instruction at grade or school levels. In addition, the lack of LOS monitoring and support in non-focal lower grades might also have contributed to the decline in those grade levels.[76]

---

73. Specifically, among the Year 3 teacher survey respondents, 43 percent of teachers in mixed-mode schools and 54 percent of those in face-to-face schools reported that they created small groups in their classes based on A2i recommendations. Among respondents, 41 percent and 37 percent of teachers from the mixed-mode and face-to-face schools, respectively, said that they gave instruction based on A2i recommendations (Appendix Table C.9).

74. Kasad, Vang, and Young (2021).

75. See Appendix Table C.8 for teachers' responses to survey questions on the use of lesson planning tools.

76. Kasad, Vang, and Young (2021).

- **Against this background, both groups of schools reported a decrease in Year 3 from the pre-COVID-19 level in the proportion of students reading at or above grade level.**

The proportion of students reading at grade level dropped by around 10 percent or more between Year 2 and Year 3 as measured by the A2i assessments, reversing the upward trends from the first two implementation years (top panel of Figure 11). This shift in direction occurred in mixed-mode and face-to-face schools to a similar degree. By the end of Year 3, among third-graders who started the program in first grade in both groups of schools, about 50 percent tested at grade level in the Letter2Meaning assessment and about 41 percent tested at grade level for the Word Match Game assessment.

Figure 11. Estimated Differential Impacts on Student Reading Outcomes, by Outcome and Program Condition, 2020-2021 School Year

Proportion of students reading at or above grade level as measured by A2i assessment

State or district standardized third grade reading test (spring of year 3)
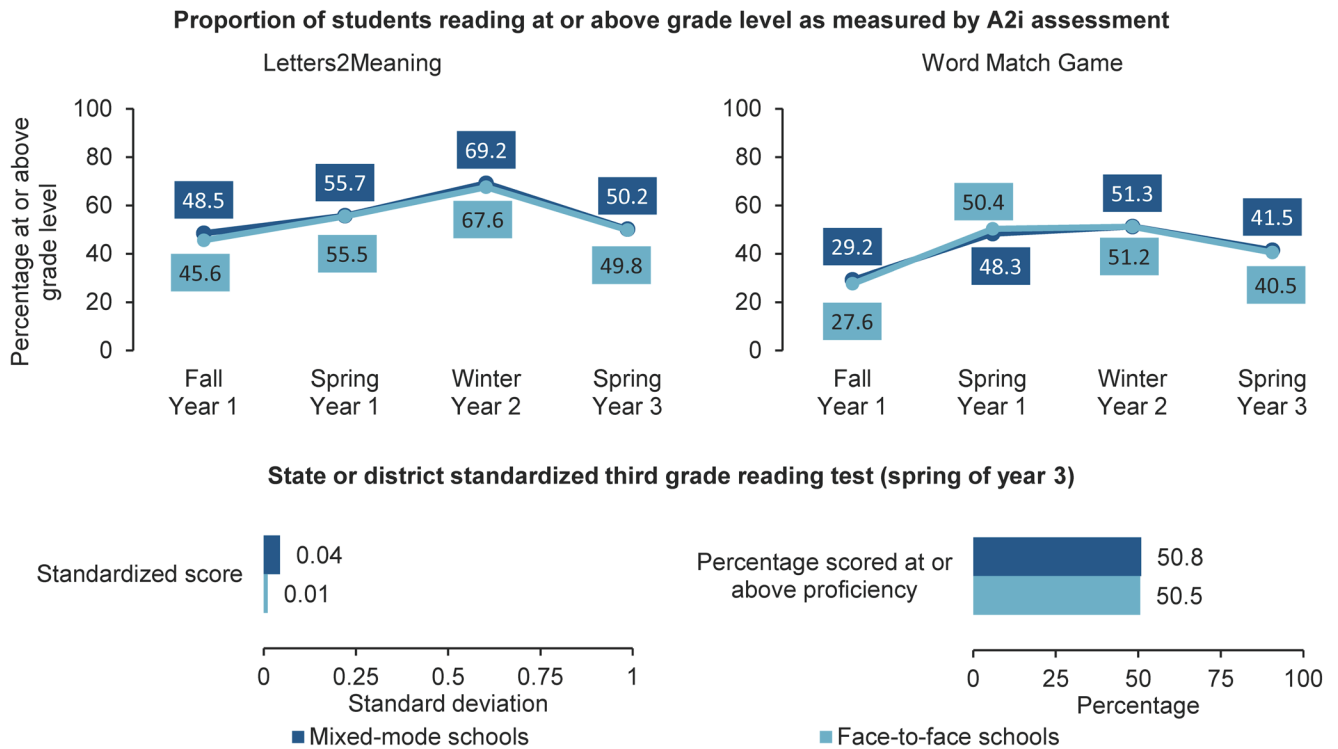
SOURCES: A2i assessment data collected by the A2i platform from summer 2018 to early spring 2020; district records data for the 2017-2018 and 2020-2021 school years.

NOTES: See Appendix Table B.1 for sample sizes in each estimation. None of the estimated differences is statistically significant at the 0.05 level.

In addition, the study found no difference in students' general reading performance between the two groups of schools as measured by the state or district standardized reading test scores for third-graders, which were available for 55 of the 59 study schools for the 2020–2021 school year. Based on these standardized tests, about 50 percent to 51 percent of the students in the primary cohort from both groups of schools were considered to be reading at or above proficiency level by their state or district definitions (bottom panel of Figure 11).

These patterns are in line with what has been reported on student learning during the pandemic nationally. Three nationwide studies examined the fall-to-spring growth for the 2020–2021 school year using interim/benchmark assessments. They found that, while students made learning gains during this school year, the gains were smaller relative to a typical school year. As a result, students' achievement levels at the end of the school year, measured as percentiles or percentage on grade level, was lower than the historical level.[77] Data from a special administration of the National Assessment of Education Progress long-term trend reading assessment for 9-year old students also showed a decline of 5 scaled score points compared to 2020, the largest decline since 1990.[78] In addition, reports from multiple states with available state reading test data show that the average percentage of students who were reading at or above proficient level dropped by 6 percentage points on their spring 2021 state assessment.[79]

Learning Ovations adjusted implementation strategies to address emerging issues and continued to improve the program during this difficult school year, both in response to qualitative data collected by Digital Promise Global during the second year of implementation and to the realities of classroom teaching during the pandemic. For example, in response to earlier findings about teachers not understanding the validity or use of A2i assessments, Learning Ovations conducted and published additional academic research communicating their validity and use. Earlier findings also indicated that teachers would benefit from more modeling of differentiated small-group instruction. Unable to provide this support in person, Learning Ovations pivoted and created a library of videos that walk teachers through how A2i can be understood and applied in their classroom with LOS narration describing specific examples or details. However, given the challenges and competing priorities that schools, teachers, and students were faced with in the 2020–2021 school year, the potential of such adjustments and improvements might not have been fully realized.

In sum, the 2020–2021 school year saw muted contrasts between the two PD models as implemented, reduced levels of usage for the key A2i components and differentiated small-group instruction, and a decline in the proportion of students reading at grade level compared to pre-COVID-19 levels that was consistent with the decline observed across the nation.

---

77. Curriculum Associates (2021); Lewis, Kuhfeld, Ruzek, and McEachin (2021); Renaissance (2021).

78. National Assessment of Educational Progress (2022).

79. This is the aggregated number based on reports from 16 states that had participation rates of at least 85 percent. See Kuhfeld et al. (2022).

DIFFERENTIAL COST OF THE TWO PD MODELS

A key barrier to bringing A2i to scale is the cost of delivering the A2i Professional Support System, and an important part of that cost comes from the PD required to support effective A2i implementation. The face-to-face PD approach tested in prior studies of A2i is effective but resource intensive. The mixed-mode PD delivery model is intended to lower the cost of on-site PD and ongoing coaching by applying technology to reduce the cost of human capital. Specifically, in this model, Literacy Huddles were mediated virtually through video conferencing, and the number of Individualized Classroom Coaching sessions were reduced. These adjustments were expected to reduce coaches' travel time and related labor costs and thus improve their productivity and allow each coach to work with more schools and teachers.

The study measured the total and variable costs per student of the two PD models for each implementation year to examine their relative costs. The total cost took into account both fixed costs, such as the cost of running the Learning Ovations organization and developing the A2i technology and materials, and variable costs, which include LOS salaries and travel costs.[80] Because each LOS worked with both face-to-face and mixed-mode schools, the variable costs take into account the proportion of their time preparing for and delivering PD for each model.

An analysis of the costs per student of these two models reveals that, as intended, the per student variable cost was lower for the mixed-mode model than for the face-to-face model across all three years (top panel of Figure 12). In Year 1, the variable cost per student for the mixed-mode model was about 29 percent lower than that for the face-to-face model. The differential costs for the subsequent two years were muted due to COVID-19-related disruptions in PD delivery. In Year 2, all PD activities were canceled after March; thus, the variable cost differential only reflects the activities that took place before then. In Year 3, all PD events took place virtually, diminishing the difference in the costs related to preparing for and delivering PD between the two models. The remaining difference in variable cost in Year 3 reflects the different amount of PD planned for and delivered to the two categories of schools. Therefore, the reported differential cost does not reflect what it would be in a normal school environment.

The differential variable costs per student led to differential total costs per student (bottom panel of Figure 12). Year 1 saw the largest difference in total costs when the mixed-mode model cost was approximately $17 (or 6 percent) per student less than the face-to-face model. This difference was reduced to about $9 and $7 per student in Year 2 and Year 3, respectively. Since most differences in total cost between these two models come from differences in variable costs by design, and since variable costs constituted a small proportion of the total costs, ranging from 17 percent to 36 percent across years, it is

---

80. Due to a lack of data, the team made a number of assumptions about the parameter values used in this calculation. See Appendix E for details.

Figure 12. Program Cost per Student, by Program Condition and Year

Variable cost per student

		Cost reduction:
Year 1	$42.04 / $59.27	29%
Year 2	$35.15 / $44.16	20%
Year 3	$28.17 / $35.41	20%

Total cost per student

Year 1	$249.03 / $266.26	6%
Year 2	$124.90 / $133.91	7%
Year 3	$91.11 / $98.35	7%

■ Mixed-mode schools   ■ Face-to-face schools
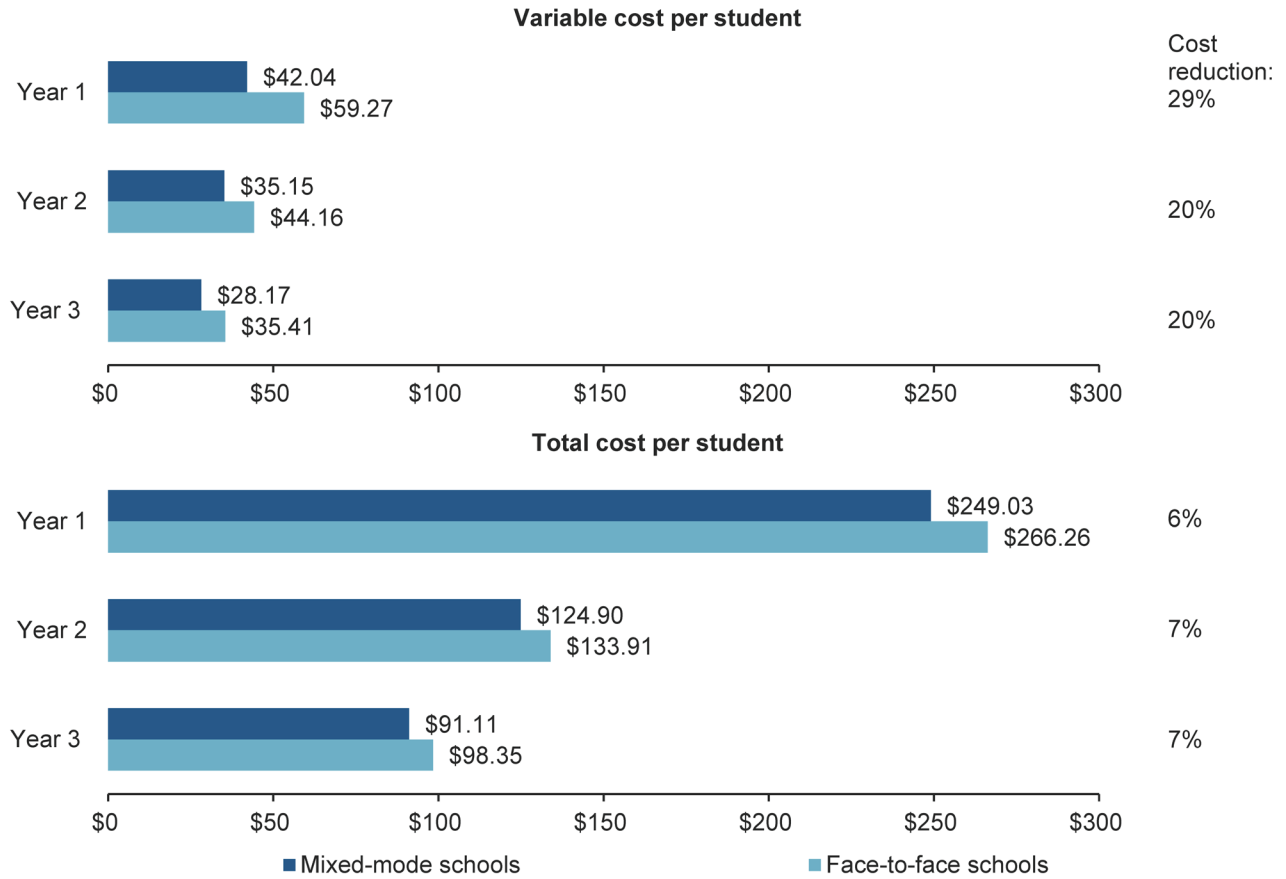
SOURCES: United2Read project annual progress reports and professional development planning and delivery records collected by Learning Ovations for 2018-2019, 2019-2020, and 2020-2021 school years.

not surprising that the percentage difference in total costs were much smaller than the difference in variable costs.

Figure 12 also shows that the program's total cost per student was highest in Year 1. This high initial cost is likely due to fixed startup costs that happened at the beginning of the grant, including the hiring and training of LOSs, onboarding schools, and setting up the technology infrastructure for these schools. The cost per student dropped precipitously between Years 1 and 2 and continued to drop in Year 3 as the A2i infrastructure was set up, as Learning Ovations found increased efficiencies, and as more students were added to the study when the program expanded to new grade levels in the study schools. Therefore, the total cost reported for Years 2 and 3 might reflect the program cost at "steady-state" from an organizational development perspective, meaning the actual regular program cost when all elements of the program had been put into place and the program had reached its intended scale, even with disruptions due to the pandemic.

Overall, this analysis suggests that the mixed-mode model's capacity for cost-reduction is limited to the reductions in variable costs because of the design of A2i. The fixed costs are the main driver of the overall cost of the program, especially at the startup stage. Future efforts to lower the cost of the program may want to focus on creating efficiencies within program operations to reduce fixed costs.

EXPANDING THE REACH OF A2i

In addition to testing the cost effectiveness of the mixed-mode and face-to-face models of PD delivery, this grant sought to increase the reach of A2i by expanding to more students, schools, and districts beyond those involved in the evaluation. Table 6 summarizes the results of such expansion efforts across the three implementation years. Overall, the COVID-19 pandemic hampered the team's ability to reach the expansion goal set forth at the start of the grant. In response, the team made adjustments to reach students through alternative routes.

Table 6. Schools and Students Served by A2i, by Year

| School Year | Implementation | Number of Districts | | Number of Schools | | Number of Students | | |
		Target	Actual	Target	Actual	Target	Actual	Detail
2018–2019	Year 1	9	20	110	103	17,600	23,838	
2019–2020	Year 2	10	17	150	91	32,800	27,864	26,786 A2i; 1,078 SLC
2020–2021	Year 3	12	18	215	91	55,200	35,365	26,873 A2i; 8,492 A2i CLSS

SOURCES: United2Read project annual progress reports for the 2018–2019, 2019–2020, and 2020–2021 school years.

NOTES: SLC refers to the Summer Literacy Coach platform; A2i CLSS refers to the A2i Community Literacy Support System.

The United2Read team primarily worked with schools and districts to reach schools beyond the study sample in the pre-COVID-19 period. To recruit schools with the best chances for sustainability, Learning Ovations identified system readiness as a key driver to find districts that would have success with A2i implementation. Learning Ovations built their recruitment and onboarding strategies on the premise that the most effective implementation models start with clear expectations and adult and system readiness for change, and focused recruitment efforts on these schools. To aid in this expansion, Learning Ovations also explored two additional approaches: (1) harnessing the technology of IBM Watson to index lessons for any reading curricula and (2) creating regional field offices of support staff members and coaches. After Year 1, it was determined that neither approach was the right fit for the program or the right investment of resources at the time. See Appendix F for more details. While exploring those approaches, Learning Ovations also continued their efforts to

recruit schools and districts to the program more generally and was on track to reach the expansion target. Each expansion school paid a fee to access A2i and relevant training and support provided by Learning Ovations. The mixed-mode model was used to provide such training and support to expansion schools.

The onset of the COVID-19 pandemic in the spring of Year 2, took place at a time when districts traditionally make decisions about adopting new programs for the next school year (2020–2021). For the most part, district recruiting was suspended for the rest of that school year and it was unrealistic to secure district commitments for the fall of 2020. As a result, no new districts or schools were added to the A2i program in the 2020–2021 school year.

Even though it was not possible to recruit new schools and districts during the pandemic, Learning Ovations took the opportunity to innovate and expanded the reach of their program beyond schools. As a response to the shift to online and at-home learning resulting from the pandemic, Learning Ovations developed new tools to be available to parties beyond schools (for example, parents and community centers). The Home Literacy Coach (HLC) platform was developed to support at-home literacy instruction and connect parents and communities to the Learning Ovations mission. The Summer Literacy Coach (SLC) platform supported summer programs in one large urban district in 2020 to help instructors (such as summer camp directors and program staff members who were also teachers) differentiate teaching and combat the student learning slide resulting from school closures and weak remote instruction in the spring. These platforms evolved into A2i Summer/A2i After-School and Reading CheckUp, all part of the A2i Community Learning Support System (CLSS) created to support students' continued learning with adults at home. These new options allowed other adults to support reading learning during remote schooling or other learning contexts such as after-school programs, summer programs, tutoring sessions, or in pre-K classrooms. By mid-2021, this work had reached approximately 8,500 children in multiple communities and states beyond the students in schools using A2i. These kinds of community-level efforts offer an additional potential entry point for district partnerships and could be considered a complementary scaling strategy to the district-focused expansion strategy.

LOOKING BACK, LOOKING FORWARD

This study intended to examine whether a more technology-based, mixed-mode delivery model for PD activities supporting the implementation of the A2i program can improve student reading performance at the same rate as the original face-to-face delivery model. In the summer of 2018, the study randomly assigned 59 elementary schools across the country to either receive the mixed-mode or face-to-face model. The study then implemented these two PD models and the A2i system for the next three years and compared the average student outcomes between the two groups.

However, the delivery of the PD models and the implementation of key A2i components among schools in the study started slow and were severely hampered by COVID-19.

As a result, teachers in both the mixed-mode and face-to-face schools did not receive adequate support to use the A2i system for their teaching at a level the models specified, and students in their classrooms did not receive the A2i-recommended amount and type of differentiated instruction for the full duration of the implementation period. Students' reading performance levels, as measured, therefore, did not reflect what they could have achieved had the PD models and the A2i system been fully implemented under normal circumstances, and the findings based on these measures do not provide a true picture of the two models' actual relative effect when adequately implemented.

In addition, due to disruptions in state and district standardized tests during the study period, it was not possible to compare the effectiveness of A2i (and the two PD delivery models) to business-as-usual instructional approaches, as the study originally planned. Therefore, the study cannot speak to whether the observed fluctuations in students' reading performance were due to the A2i program or natural growth in reading skills.

Nonetheless, a few themes emerged from schools' and teachers' experiences with A2i during this tumultuous period. These themes can provide valuable lessons for implementing the A2i Professional Support System in diverse settings.

First, it is important to make it clear to teachers what A2i can and cannot do and how it might differ from other tools they already use. For example, unlike many diagnostic reading tests that provide teachers with information that pinpoints specific skills students need to work on, A2i assessments are designed to feed into the A2i algorithm to generate instructional recommendations that suit students' individual needs. This difference led a noticeable proportion of teachers to question the validity and value of the A2i assessment, as they expected to get diagnostic information on their students' reading skills from these assessments. This misunderstanding of the purpose and value of key A2i components could have discouraged teachers' use of the A2i system. Therefore, being clear about the purpose of A2i assessments could help set the right expectations and build teachers' understanding of how A2i can make their jobs easier, and in turn make them more likely to implement A2i with fidelity.

Second, teachers need more support in adopting the A2i instructional recommendations and implementing differentiated small-group instruction in class. Tailoring instruction to suit the needs of individual students is arguably one of the hardest parts of teaching. Teachers need to translate what they know about the students, be it assessment scores, classroom observations, or other types of information, into actionable instruction that is appropriate for each student every day. One crucial contribution of the A2i technology is to translate this information for teachers. The A2i software is designed to provide teachers with (1) information on the type and amount of instruction that individual students need based on their assessment information, (2) suggestions for how to group students, and (3) resources teachers can use to plan their lessons accordingly. Such information intends to help teachers plan and prepare for differentiated instruction in their classes. However, this study found that teachers did not make adequate use of what the system had to offer, either because they had doubts about its value or because they found it hard to incorporate

the recommendations in their classrooms. This challenge was exacerbated by the reduced level of hands-on PD support, the constraints of remote and hybrid learning, and limited instructional time because of COVID-19. As a result, the system-recommended instruction was not fully implemented in class. More support, especially information on what an A2i classroom looks like, classroom management strategies for different learning environments, and leveled activities for different groups, would be essential to facilitate teachers' implementation of differentiation in small-group instruction.

Lastly, varying teacher needs, and diverse class or school contexts call for differentiated support for teachers. For example, teachers with little or no prior experience with small-group instruction might need different strategies than their peers who have already been using small groups in the classroom. Kindergarten classrooms might need more support to help set up routines and manage class flow than third grade classes, where students are more accustomed to the school environment. Taking time to learn about the classroom and school context and the challenges each teacher faces would help LOSs identify strategies that suit teachers' needs and better support their implementation of A2i.

Throughout the study period, Learning Ovations continuously adjusted its training and supports to address these and other emerging implementation issues. For example, to address teachers' concerns about the A2i assessments, materials explaining the nature and purpose of the assessments, including articles and description videos, were made available to teachers in Year 2. LOSs also incorporated more discussions on this topic in their training plans. Resources and materials were constantly added to the system to help teachers plan their lessons, including a comprehensive set of materials describing how to use A2i during virtual teaching, both for synchronous and asynchronous instructional time. However, the challenging nature of the learning environment during the second half of the study hampered the study's full potential to improve teachers' adoption of A2i and student reading performances could not be fairly tested.

In a period when the K-12 system nationwide was confronted with unprecedented challenges brought forth by the COVID-19 pandemic, the United2Read team was not able to meet the ambitious expansion goals it set for itself. Nonetheless, as a response to the shift to online and at-home learning resulting from the pandemic, Learning Ovations developed new tools for other adults, such as caregivers, tutors, and afterschool providers, to support reading learning during remote schooling. This effort has impacted approximately 8,500 children in multiple communities and states beyond the students in schools using A2i in the classroom in Year 3. By the end of the evaluation period, the A2i program reached a total of 35,365 students in 91 schools and other learning communities.

In addition to the planned routes of expansion, to ensure long-term sustainability and scale, Learning Ovations pursued partnerships and acquisition opportunities with companies that have missions similar to those of Learning Ovations and reached an agreement with Scholastic, the global children's publishing, education, and media company, to become part

of the Scholastic literacy platform.[81] Looking forward, drawing from implementation lessons learned through this study, and with support from Scholastic's resources and infrastructure, A2i will be able to reach a broader population of young children and to help them read successfully.

---

81. Seeking Alpha (2022).

The A2i Professional Support System

This appendix presents additional information about the A2i Professional Support System, described in the section titled "The A2i Professional Support System and the Two PD Delivery Models Under Evaluation" in the report, as well as the changes made to the program throughout the study in response to formative feedback.

I. THE A2I PROFESSIONAL SUPPORT SYSTEM

A2i Assessments

As part of the A2i system, the Word Match Game and Letters2Meaning assessments measure three constructs required by the A2i recommendation algorithms—vocabulary, decoding, and comprehension. The Word Match Game is a semantic (meaning) matching task where students are shown three or four words and are asked to click on two of the words that are semantically related. The Letters2Meaning assessment includes letter and letter-sound identification, word reading, spelling, and sentence construction (students make sentences from a pool of words). The online assessments are adaptive as subsequent items are automatically selected based on whether the student's previous responses were correct or incorrect and the difficulty level of the item. Students take these assessments independently on either laptop or desktop computers with a mouse and headphones. Each assessment provides scores as Age Equivalents, Grade Equivalents, and Developmental Scaled Scores. Prior research has shown that the assessments are psychometrically strong and are highly correlated with other standardized measures of language and reading.[1]

A2i Software Platform

The A2i software platform uses data from the assessments to make recommendations for student groupings, and the number of minutes each student should get for each of four types of instruction. It also provides lesson planning tools. Appendix Figure A.1 presents an example of the classroom view on the A2i platform. It shows the individual student instructional recommendations based on their A2i assessment scores, the resulting grouping recommendations, and the number of minutes and types of instruction for each group. Appendix Figure A.2 shows the accompanying lesson plans on the A2i platform, which provide suggestions for each student group for each type of instruction. The platform also allows teachers to check off which lessons they used.

Rollout of Professional Development and Support

Professional development support was provided to teachers by the Literacy Outcomes Specialists (LOSs). In Year 1, professional development was provided to kindergarten and first grade teachers. In Year 2, professional development was provided mainly to second grade teachers, with kindergarten and first grade teachers getting some ongoing supports.

---

1. Connor et al. (2017).

Appendix Figure A.1. Sample A2i Grouping and Instructional Recommendations
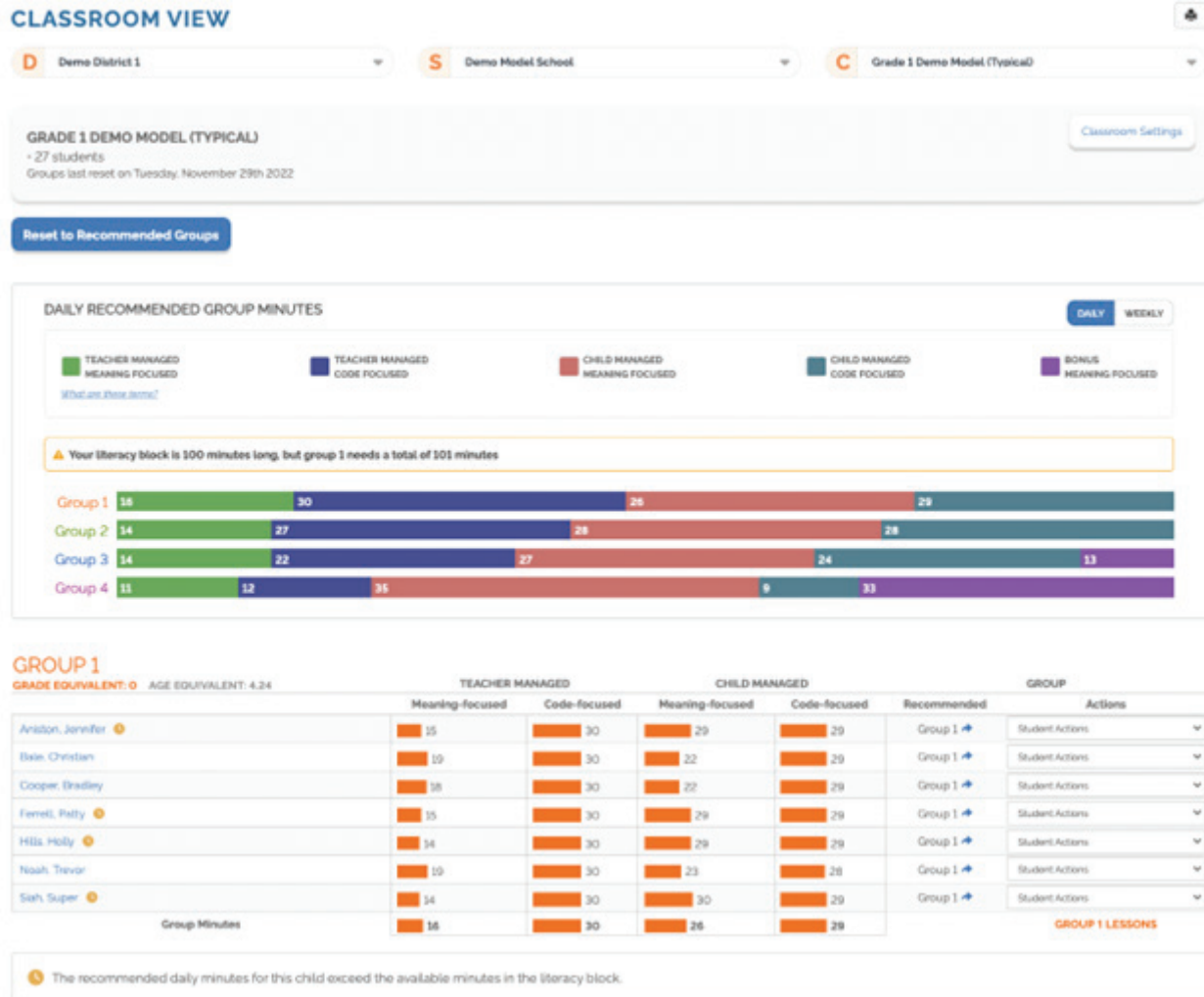
Source: Learning Ovations, Inc.

Finally, in Year 3, the program was rolled out to third grade teachers, with second grade teachers getting some ongoing supports and on-demand support available for kindergarten and first grade teachers. As a result, the way LOSs spent their time was designed to change throughout the life of the project. In Year 1, they spent half of their time with kindergarten teachers and half with first grade teachers. In Year 2, they spent half their time with second grade teachers, and one-fourth of their time with kindergarten and first grade teachers. In Year 3, they spent half their time with third grade teachers, one-fourth with second grade teachers, and one-fourth was made available for on-demand support. While these are approximate estimates, the design shows the changing focal grades during each year of implementation.

Appendix Figure A.2. Sample A2i Lesson Plan

Source: Learning Ovations, Inc.

The A2i Literacy Huddles were professional development events that were facilitated by the LOSs and took place with grade-level peers. Each A2i Literacy Huddle was scheduled to last 30 minutes. They followed a scope and sequence throughout the school year, provided by Learning Ovations, though they had the flexibility to be responsive to teacher needs. One topic that was covered was differentiation.

In-class coaching (later called "Individualized Classroom Coaching") was provided to teachers by their LOSs throughout the year, often around the same time as the A2i Literacy Huddles. Individualized Classroom Coaching may include the LOS modeling a technique and then debriefing the classroom teacher afterward, the LOS observing the teacher and

providing feedback, a one-on-one deep dive data conversation, or support in navigating the platform. The Individualized Classroom Coaching sessions were meant to be individualized to teachers' needs and abilities and were scheduled for 45 minutes each.

Finally, LOSs provided on-demand support through optional flex days. Should a school choose, they could schedule up to two flex days throughout the school year during which the LOS would come to campus and provide whatever support was needed. This time could be used to support and model best practices for conducting the assessments with students, developing and preparing small-group instruction materials, or observing the school to better understand the school's culture and needs.

II. CHANGES TO THE PROGRAM

The United2Read project included opportunities for Digital Promise Global to conduct annual interviews with teachers and other school staff members to understand their experiences with implementing the A2i Professional Support System and provide formative feedback to Learning Ovations based on what was gleaned from those conversations. Learning Ovations worked closely with Digital Promise Global to understand the implications from the interviews and address each issue, evolving the program in direct response to the teachers' voices. These implications and the resulting changes touched all levels of implementation—teachers, Literacy Champions, school administrators, district administrators, LOSs, and other Learning Ovations staff members. Appendix Table A.1 highlights selected examples of the changes and the timeframe in which they were implemented.

Appendix Table A.1. Implementation Adaptations Throughout the Study Period

Area of Change	Examples of Changes
A2i program services to schools	• Revamped the A2i Initial Training to better meet the needs of teachers and build rapport at the beginning of the process and added Flex Days and a scope and sequence for planning professional development to build rapport with schools (identified and addressed in Year 1)[a] • Provided flexibility for when teachers administer the first A2i assessment and begin small groups by working with districts to plan administration windows and providing support in-person and via the A2i portal (Year 1) • Created materials for Literacy Outcomes Specialists (LOSs) to use in professional development, including a scope and sequence, slide decks, and pacing guides (Years 1–2)[b] • Created a start-up package that provides teachers with sample classroom configurations, small group routines, and sets of leveled activities for different types of centers (Years 2–3) • Developed a library of standardized materials for LOSs to use in supporting Literacy Champions (LCs) (Years 2–3)[c] • Added new dashboards to support teachers in understanding and using A2i data reports and how to organize group instruction (Years 2–3) • Created a video library to visualize differentiated small-group instruction with A2i (Years 2–3)
Expectations for LOS coaching role	• Clarified and documented LOS and LC roles on the LOS homepage and in other materials (Year 1) • Provided support for LOSs to coach teachers on fundamental classroom management skills to facilitate shifts to differentiated small-group instruction (Year 1)
Learning Ovations organizational systems	• Created an onramp for districts to prepare for A2i implementation (Year 1) • Developed written materials and administrator dashboards on the A2i platform to facilitate administrator coaching and create buy-in (Years 2–3) • Expanded the role of the School Outcomes and Operations teams to act as liaisons with school and district leadership (Year 3)[d] • Created a new lesson planning dashboard with activity recommendations within the lesson plan, as well as improved/expanded indexing (Years 2–3)
Professional development supports for LOSs	• Provided LOSs with time to collaboratively solve specific problems of coaching practice and to be reflective practitioners (Years 1–2) • Created a rubric that specifies developmental stages of implementation to help guide LOSs on what professional development to provide to schools based on their existing literacy approaches and what aspects of A2i to emphasize based on teachers' readiness (Years 2–3)
Messaging to schools about A2i assessment	• Created materials to emphasize the purpose and the validity of A2i assessments, explain how adaptive testing produces valid recommendations, and the uses of A2i assessment data, including videos and academic papers (Years 1–3)

SOURCES: Authors' compilation of implementation documents from Learning Ovations and Digital Promise Global.

NOTES: [a]Flex Days are optional days for a school to request support in implementing A2i.
[b]Literacy Outcome Specialists (LOSs) are coaches who provide training throughout the implementation period. Materials such as a scope and sequence, slide decks, and pacing guides provide the content, presentation medium, and timeline for professional development sessions over the course of the school year.
[c]Each school selects a Literacy Champion (LC) from their staff to support implementation. The LC's responsibilities include liaising between teachers and the LOS, coaching teachers on differentiated small-group instruction, supporting teachers with administering A2i assessments, understanding and using A2i data, and lesson planning.
[d]The School Outcomes and Operations Teams work within Learning Ovations to liaise with and advise schools and districts around implementation and student performance.

Analytical Approaches and Samples

This appendix presents the statistical models the study used for estimating the differences between the mixed-mode and face-to-face schools. It also discusses several technical issues related to the estimation approach. The appendix begins by describing the statistical models for the analysis in this study, and it then discusses some key features in these models and specific issues related to the analyses. Lastly it describes various samples used in the analysis and presents findings from the baseline equivalence checks for each of them.

I. ESTIMATION MODEL

Since schools were randomly assigned to the mixed-mode and face-to-face models, differential impacts on student outcomes can be estimated by comparing the average outcomes of students in these two groups of schools. Similarly, the service contrasts can be estimated by comparing teacher or student experience in the mixed-mode and face-to-face schools. In practice, a multilevel regression model is used to estimate differences in outcomes between these two groups of schools. This makes it possible to account for the blocking of random assignment and for the fact that the data are clustered.

Specifically, a two-level hierarchical model with students or teachers nested within schools is used for impact estimations for student or teacher level outcomes reported in the report:

$$Y_{ik} = \sum_m \gamma_{0m} D_{km} + \sum_m \gamma_{1m} D_{km} T_k + \sum_l \alpha_{1kl} X_{ikl} + \delta_k + \varepsilon_{ijk} \qquad (1)$$

Where:

Y_{ik} = outcome of student/teacher i in school k.

X_{ikl} = A set of L student/teacher-level characteristics for student/teacher i in school k measured prior to their first exposure to the program, including students' first A2i assessment scores as their prior reading level measures. These covariates reduce within-school and between-school variations in the outcome measures, thereby increasing the precision of the impact estimates.

D_{km} = The mth Random assignment block indicator, which equals 1 if school k is in random assignment block m and 0 otherwise (m = 1, 2, …, M).

T_k = Indicator of program condition for school k, which equals 1 if school k is in the mixed-mode group and 0 otherwise.

δ_k = A random error term for school k, assumed to be independently and identically distributed across schools (that is, the between-school residual).

ε_{ik} = A random error term for student/teacher i in school k, assumed to be independently and identically distributed across students within classrooms (that is, the within-school residual).

Therefore,

$$\gamma_{1m} \quad = \quad \text{The difference between the school-level average of outcome Y in the mixed-mode schools and the face-to-face schools in block } m \text{ (that is, the impact of the program on outcome Y in block } m\text{).}$$

To obtain the overall estimated differential impact between the two program conditions, γ_1, estimated impacts for each block (γ_{1m}) were pooled by weighting each block by the number of mixed-mode schools in the block. This pooled overall estimate therefore represents the differential impact for the average mixed-mode school in the sample. Note that Equation (1) provides an "intent to treat" analysis of the differential impact, meaning the reported findings represent the estimated differential impacts between the two program conditions schools were assigned to rather than the conditions schools implemented. However, since the professional development models were provided and monitored closely by Learning Ovations, there were no "cross-over" schools (schools randomly assigned to one model but that somehow managed to receive the other).

Below are some key features of this estimation model:

- **Block fixed effects.** The statistical model presented above treats random assignment blocks as fixed effects based on the assumption that districts in the study are not representative of the national population of elementary schools due to the purposeful nature of site recruitment. This means that the findings presented by this study should be interpreted as the estimated impact of the program for the districts and schools in which the program was implemented. In other words, conclusions about the two models' relative effectiveness based on these models cannot be generalized to other schools or districts.

- **Covariates.** To improve the precision of the impact estimates, the student/teacher level analysis controls for random differences between students and teachers in the mixed-mode and face-to-face schools with respect to a range of background characteristics. The choice of these covariates was made prior to estimating impacts, based on their predictive power of the outcomes and data availability. For student level analysis, the covariates include students' grade, age, race and ethnicity, district-provided poverty indicator, English learner status, Individualized Education Plan status, and baseline A2i assessment scores. For teacher level analysis, the covariates include the amount of their reading instruction time and the number of days (in a week) they had a teaching assistant during reading instruction; teachers' teaching experience in general, in current school, and in current grade; as well as their highest degree earned and types of teaching certifications.

- **Missing values.** This study does not impute missing outcome values. Missing information on the covariates are imputed using a dummy variable approach. This approach consists of (1) imputing a value of "zero" for the missing values in each of the covariates, (2) creating a dichotomous indicator of missingness for each covariate,

and (3) including these indicators alongside the imputed covariates in the statistical model.[1] In theory it is not strictly necessary to control for these baseline characteristics because random assignment should ensure that students in the mixed-mode and face-to-face schools are similar in expectation. However, controlling for these background characteristics can improve the precision of the impact estimates.

II. PRIMARY COHORT OF STUDENTS

To answer the main research question about differential impacts the mixed-mode and face-to-face professional development models might have on student reading performance, the study focused on a cohort of students as they moved through the grades, starting in first grade in Year 1 to third grade in Year 3. The study collected the Gates-MacGinitie scores, A2i assessment scores, and third-grade state or district standardized reading test scores for this cohort of students at various points in the study. Due to attrition, non-response, and mobility, the number of schools and students with valid data varies across the samples used for the analyses of these outcome measures. Appendix Table B.1 presents the sample sizes and response rates for each outcome for the primary student cohort. Note that the student-level response rate is calculated based on students in the non-attrited schools. None of the estimated differences in response rates at both school and student levels is statistically significant across all outcome measures.

The study also checked the equivalence between the mixed-mode and face-to-face school samples on school and student background characteristics and found few differences between the schools and students in the two program conditions. Appendix Tables B.2–B.4 present such checks for school-level characteristics, and Appendix Tables B.5–B.9 present findings from the student-level equivalence checks for various analysis samples.

A common way to convey a study's statistical power is through the minimum detectable effect size (MDES). Formally, the MDES is the smallest true program impact, scaled as an effect size, that can be detected with a reasonable degree of power (in this case, 80 percent) for a given level of statistical significance (in this case, 5 percent for a two-tailed test). The number of students and schools in the sample are crucial factors that determine the degree to which the impacts on student and school outcomes can be estimated with enough precision to reject with confidence the hypothesis that the program had no effect. In general, larger sample sizes provide more precise impact estimates.

Appendix Table B.10 presents the MDES for the differential impact estimates for the primary cohort of students. The MDES in this table are based on the number of students and schools used in the actual impact estimation and the standard errors of the estimated impact of actual assignment to intervention. Hence, the values in the tables represent the actual precision of the analyses.

---

1. Puma et al. (2009).

Appendix Table B.1. Sample Sizes and Response Rates for the Primary Student Cohort, by Outcome and Program Condition

Primary Student Sample	Total		Mixed-Mode Schools		Face-to-Face Schools	
	School	Student	School	Student	School	Student
At start of Year 1	59	4733	30	2372	29	2361
With GMRT score						
Count (spring Year 1)	58	3893	29	1920	29	1973
Response rate (%)	98.3	84.2	96.7	84.8	100	83.6
With A2i assessment score						
Year 1						
Count	59	4367	30	2216	29	2151
Response rate (%)	100	92.3	100	93.4	100	91.1
Year 2						
Count	50	3243	25	1606	25	1637
Response rate (%)	84.7	81.5	83.3	82.6	86.2	80.4
Year 3						
Count	40	2278	20	1165	20	1113
Response rate (%)	67.8	71.8	66.7	74.2	69.0	69.4
With state/district test score (spring Year 3)						
Count	55	3130	28	1605	27	1525
Response rate (%)	93.2	72.1	93.3	73.4	93.1	70.7

SOURCE: Authors' calculations based on student level data collected and compiled by the study team.

NOTES: The sample includes all students who started the program as first-graders in the fall of 2018 in the 59 study schools. Student level response rates reported in the table are calculated as the ratio between the number of students in each analysis and the number of students present at the start of Year 1 in the non-attrited schools only.
   GMRT = Gates-MacGinitie Reading Test.
   None of the differences in the response rates between the mixed-mode schools and the face-to-face schools is statistically significant at the 0.05 level.

Appendix Table B.2. School Background Characteristics for Schools Remaining in the Study After Year 1, by Program Condition

Characteristics	Mixed-Mode Schools Mean	Face-to-Face Schools Mean	Estimated Difference	P-Value for Estimated Difference
Urbanicity (% of schools)				
Large or middle-sized city	20.0	20.0	0.0	1.000
Urban fringe and large town	68.0	74.0	-6.0	0.129
Small town and rural area	12.0	6.0	6.0	0.129
Title I status (% of schools)	72.0	84.0	-12.0	0.173
Female (school average % of students)	47.8	48.6	-0.8	0.241
Race/ethnicity (school average % of students)				
White non-Hispanic	28.6	27.4	1.2	0.720
Black non-Hispanic	4.5	5.4	-0.8	0.704
Hispanic	58.5	60.6	-2.2	0.334
Asian	4.4	3.1	1.3	0.335
Other	4.0	3.5	0.5	0.528
Proportion of students with				
IDEA plans (%)	12.1	13.4	-1.3	0.376
504 plans (%)	1.2	1.1	0.1	0.486
Limited English Proficiency (%)	24.3	30.6	-6.3 *	0.019
Free and reduced-price lunch status (%)	68.3	74.7	-6.5	0.115
Third grade students with at or above reading proficiency (%)	43.8	42.8	0.9	0.787
Mean school enrollment (n)	539.68	605.10	-65.42	0.170
Kindergarten	87.84	100.18	-12.34	0.103
Grade 1	78.64	88.40	-9.76	0.140
Grade 2	80.44	89.98	-9.54	0.132
Grade 3	79.84	83.54	-3.70	0.568
Mean number of school year teachers (n)	30.16	34.00	-3.84	0.250
Number of schools	25	25		

SOURCES: The Common Core of Data (CCD) for the 2017–2018 school year, Office for Civil Rights Data for the 2017–2018 school year, and state-reported school performance data from the 2017–2018 school year.

NOTES: The sample includes all 59 study schools that participated in the study. The estimated differences are regression-adjusted, controlling for the blocking of random assignment.

Rounding may cause slight discrepancies in calculating means and differences. Sample size for each characteristic may vary due to missing values. * indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test.

An F-test was used to determine whether there is a systematic difference between the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.919.

Appendix Table B.3. School Background Characteristics for Schools Remaining in Study After Year 2, by Program Condition

Characteristics	Mixed-Mode Schools Mean	Face-to-Face Schools Mean	Estimated Difference	P-Value for Estimated Difference
Urbanicity (% of schools)				
Large or middle-sized city	25.0	25.0	0.0	1.000
Urban fringe and large town	60.0	67.5	-7.5	0.137
Small town and rural area	15.0	7.5	7.5	0.137
Title I status (% of schools)	80.0	85.0	-5.0	0.596
Female (school average % of students)	47.9	48.6	-0.8	0.319
Race/ethnicity (school average % of students)				
White, non-Hispanic	26.9	25.4	1.5	0.696
Black, non-Hispanic	4.9	5.7	-0.8	0.783
Hispanic	61.2	63.0	-1.8	0.480
Asian	4.6	2.8	1.7	0.299
Other	2.4	3.1	-0.7	0.114
Proportion of students with				
IDEA plans (%)	12.2	14.0	-1.8	0.322
504 plans (%)	1.2	1.2	0.0	0.920
Limited English Proficiency (%)	26.7	29.2	-2.5	0.198
Free and reduced-price lunch status (%)	71.6	75.2	-3.6	0.408
Third grade students with at or above reading proficiency (%)	41.7	43.0	-1.4	0.676
Mean school enrollment (n)	550.75	600.13	-49.38	0.331
Kindergarten	41.44	42.83	-1.39	0.722
Grade 1	78.60	86.70	-8.10	0.176
Grade 2	80.00	87.28	-7.28	0.219
Grade 3	80.00	80.93	-0.93	0.883
Mean number of school year teachers (n)	31.65	35.30	-3.65	0.359
Number of schools	20	20		

SOURCES: The Common Core of Data (CCD) for the 2017-2018 school year, Office for Civil Rights Data for the 2017–2018 school year, state reported school performance data from the 2017-2018 school year.

NOTE: The sample includes the 40 schools that stayed in the study for all three years. The estimated differences are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating means and differences. The sample size for each characteristic may vary due to missing values. * indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.919.

Appendix Table B.4. Background Characteristics for Study Schools in the State and District Reading Assessment Analysis, by Program Condition

Characteristics	Mixed-Mode Schools Mean	Face-to-Face Schools Mean	Estimated Difference	P-Value for Estimated Difference
Urbanicity (% of schools)				
Large or middle-sized city	17.9	17.9	0.0	
Urban fringe and large town	71.4	76.8	-5.4	0.147
Small town and rural area	10.7	5.4	5.4	0.147
Title I status (% of schools)	75.0	84.5	-9.5	0.254
Female (school average % of students)	47.8	48.8	-1.1	0.103
Race/ethnicity (school average % of students)				
White, non-Hispanic	27.1	26.7	0.4	0.903
Black, non-Hispanic	4.6	5.4	-0.8	0.686
Hispanic	59.2	61.4	-2.2	0.324
Asian	5.3	3.3	2.1	0.251
Other	3.7	3.2	0.5	0.483
Proportion of students with				
IDEA plans (%)	12.2	13.3	-1.1	0.437
504 plans (%)	1.2	1.0	0.2	0.301
Limited English Proficiency (%)	24.3	30.4	-6.1 *	0.017
Free and reduced-price lunch status (%)	69.8	74.5	-4.7	0.220
Third grade students with at or above reading proficiency (%)	43.0	42.3	0.7	0.844
Mean school enrollment (n)	535.86	589.03	-53.17	0.236
Kindergarten	91.64	100.33	-8.68	0.225
Grade 1	80.79	89.45	-8.67	0.181
Grade 2	81.50	90.46	-8.96	0.142
Grade 3	79.67	82.06	-2.39	0.704
Mean number of school year teachers (n)	30.54	33.61	-3.07	0.332
Number of schools	28	27		

SOURCES: SOURCES: The Common Core of Data (CCD) for the 2017-2018 school year, Office for Civil Rights Data for 2017–2018 school year, state reported school performance data from the 2017-2018 school year.

NOTE: The sample includes the 55 study schools for which the study was able to obtain the third-grade state or district reading assessment scores in the spring of Year 3. The estimated differences are regression-adjusted, controlling for the blocking of random assignment. Rounding may cause slight discrepancies in calculating means and differences. The sample size for each characteristic may vary due to missing values. *Indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.939.

Appendix Table B.5. Background Characteristics for Primary Cohort Students in the State and District Reading Assessment Analysis Sample, by Program Condition

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Difference		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Difference	Estimated Difference in Effect Size	P-Value
Baseline reading test (scaled score)											
Letters2Meaning	28	1,527	431.93	99.12	27	1,463	422.98	101.45	8.94	0.09	0.304
Grade equivalent			0.39				0.31				
At grade level (%)	28	1,527	50.7	50.0	27	1,463	45.8	50.0	4.9	0.10	0.242
Word Match Game	28	1,545	466.51	12.94	27	1,463	465.45	12.91	1.06	0.08	0.283
Grade equivalent			0.00				-0.20				
At grade level (%)	28	1,545	30.1	45.5	27	1,463	27.5	45.0	2.6	0.06	0.321
Time of test (number of days)	28	1,543	25.46	22.21	27	1,454	23.72	14.85	1.75	0.09	0.355
Demographic characteristics (%)											
Female	28	1,605	48.9	50.0	27	1,525	50.5	50.0	-1.6	-0.03	0.401
Race/ethnicity											
Hispanic	28	1,576	58.3	47.9	27	1,511	59.9	48.5	-1.7	-0.03	0.643
Black, non-Hispanic	28	1,576	3.4	18.2	27	1,511	4.1	18.4	-0.7	-0.04	0.600
White, non-Hispanic	28	1,576	27.3	40.7	27	1,511	30.5	44.9	-3.2	-0.07	0.376
Asian	28	1,576	6.5	24.2	27	1,511	2.7	17.2	3.9	0.18	0.088
Native American or Alaskan Native	28	1,576	0.2	4.4	27	1,511	0.1	3.6	0.1	0.03	0.608
Other	28	1,576	4.4	21.2	27	1,511	2.8	17.7	1.6	0.08	0.178
Age at baseline (year)	28	1,607	6.0	0.3	27	1,525	6.0	0.2	0.0	-0.03	0.397
Other background characteristics (%)											
Low-income background status	28	1,442	69.2	45.4	27	1,400	74.7	44.2	-5.5	-0.12	0.099
English learner status	28	1,445	32.4	47.8	27	1,403	37.9	49.0	-5.5	-0.11	0.122
Special education status	28	1,442	10.7	30.4	27	1,401	9.6	29.2	1.2	0.04	0.429

(continued)

Appendix Table B.5 (continued)

SOURCES: District administrative records for the 2017–2018 school year and A2i assessment data for the 2018–2019 school year obtained for this study.

NOTES: This table is based on first grade students who were enrolled in the study schools in the fall of 2018 (Year 1) and whose third grade state or district reading assessment scores from Year 3 were made available to the study.

*indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the students in the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.965.

Appendix Table B.6. Background Characteristics for Primary Cohort Students in the Gates-MacGinitie Reading Test (GMRT) Analysis Sample, by Program Condition

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Difference		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Difference	Estimated Difference in Effect Size	P-Value
Baseline reading test (scaled score)											
Letters2Meaning	29	1,916	431.64	99.44	29	1,971	426.23	101.49	5.40	0.05	0.493
Grade equivalent			0.39				0.34				
At grade level (%)	29	1,916	50.7	50.0	29	1,971	47.7	50.0	3.0	0.06	0.414
Word Match Game	29	1,916	466.46	12.90	29	1,971	465.96	12.98	0.50	0.04	0.607
Grade equivalent			-0.10				-0.10				
At grade level (%)	29	1,916	30.3	45.7	29	1,971	28.3	45.3	2.1	0.05	0.491
Time of test (number of days)	29	1,915	24.75	21.86	29	1,971	24.37	17.66	0.38	0.02	0.838
Demographic characteristics (%)											
Female	29	1,918	50.7	50.0	29	1,971	50.8	50.0	-0.1	0.00	0.939
Race/ethnicity											
Hispanic	29	1,887	58.8	48.1	29	1,950	60.2	49.4	-1.4	-0.03	0.685
Black, non-Hispanic	29	1,887	3.8	19.3	29	1,950	4.4	21.4	-0.6	-0.03	0.732
White, non-Hispanic	29	1,887	28.4	42.3	29	1,950	29.0	45.9	-0.6	-0.01	0.868
Asian	29	1,887	4.7	21.4	29	1,950	3.1	17.7	1.5	0.08	0.267
Native American or Alaskan Native	29	1,887	0.1	3.3	29	1,950	0.2	5.1	-0.1	-0.02	0.570
Other	29	1,887	4.2	20.4	29	1,950	3.1	18.7	1.1	0.06	0.266
Age at baseline (year)	29	1,920	6.0	0.3	29	1,973	6.0	0.2	0.0	0.01	0.698
Other background characteristics (%)											
Low-income background status	29	1,701	69.0	45.8	29	1,789	72.7	45.7	-3.7	-0.08	0.238
English learner status	29	1,707	30.5	47.0	29	1,792	34.4	47.7	-3.9	-0.08	0.257
Special education status	29	1,701	9.6	29.3	29	1,789	8.7	28.2	0.9	0.03	0.403

(continued)

Appendix Table B.6 (continued)

SOURCES: District administrative records for the 2017–2018 school year and A2i assessment data for the 2018–2019 school year obtained for this study.

NOTES: This table is based on first grade students who were enrolled in the study schools in the fall of 2018 (Year 1) and had taken the study-administered GMRT test in the spring of 2019.

 *indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the students in the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.550.

Appendix Table B.7. Background Characteristics for Primary Cohort Students in the Year 1
A2i Assessment Analysis Sample, by Program Condition

| | Mixed-Mode Schools | | | | Face-to-Face Schools | | | | Estimated Difference | | |
Outcome	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Difference	Estimated Difference in Effect Size	P-Value
Baseline reading test (scaled score)											
Letters2Meaning	30	2,184	429.01	100.31	29	2,151	422.73	102.32	6.27	0.06	0.427
Grade equivalent			0.36				0.31				
At grade level (%)	30	2,184	48.5	49.9	29	2,151	45.6	49.9	2.9	0.06	0.436
Word Match Game	30	2,214	466.16	12.92	29	2,149	465.64	13.03	0.52	0.04	0.556
Grade equivalent			-0.10				-0.10				
At grade level (%)	30	2,214	29.2	45.0	29	2,149	27.6	44.9	1.6	0.04	0.548
Time of test (number of days)	30	2,216	25.33	22.07	29	2,151	24.70	17.94	0.63	0.03	0.721
Demographic characteristics (%)											
Female	30	2,207	49.5	50.0	29	2,144	50.2	50.0	-0.7	-0.01	0.650
Race/ethnicity											
Hispanic	30	2,174	58.7	48.1	29	2,118	58.7	49.4	0.0	0.00	0.989
Black, non-Hispanic	30	2,174	4.3	21.0	29	2,118	5.1	22.0	-0.8	-0.04	0.636
White, non-Hispanic	30	2,174	27.3	41.2	29	2,118	29.2	45.6	-2.0	-0.05	0.517
Asian	30	2,174	5.7	23.7	29	2,118	3.7	19.3	1.9	0.09	0.295
Native American or Alaskan Native	30	2,174	0.2	4.3	29	2,118	0.3	5.3	-0.1	-0.01	0.723
Other	30	2,174	3.9	19.6	29	2,118	3.0	18.7	0.9	0.04	0.337
Age at baseline (year)	30	2,216	6.0	0.3	29	2,151	6.0	0.2	0.0	0.04	0.248
Other background characteristics (%)											
Low-income background status	30	1,927	69.5	45.5	29	1,932	73.2	45.3	-3.6	-0.08	0.231
English learner status	30	1,934	30.0	46.7	29	1,935	34.3	47.8	-4.3	-0.09	0.188
Special education status	30	1,927	10.0	29.7	29	1,932	9.0	28.9	1.0	0.03	0.462

(continued)

Appendix Table B.7 (continued)

SOURCES: District administrative records for the 2017–2018 school year and A2i assessment data for the 2018–2019 school year obtained for this study.

NOTES: This table is based on first grade students who were enrolled in the study schools in the fall of 2018 (Year 1) and took the A2i assessments during Year 1.

   *indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the students in the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.012. This significant result is mostly driven by the fact that 30 students from one mixed-mode school were missing baseline Letters2Meaning test scores. If the missing indicator for this baseline assessment score is omitted from the F-test, the p-value becomes 0.134.

Appendix Table B.8. Background Characteristics for Primary Cohort Students in the Year 2 A2i Assessment Analysis Sample, by Program Condition

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Difference		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Difference	Estimated Difference in Effect Size	P-Value
Baseline reading test (scaled score)											
Letters2Meaning	25	1,565	435.35	98.52	25	1,613	425.08	100.97	10.28	0.10	0.246
Grade equivalent			0.42				0.33				
At grade level (%)	25	1,565	52.3	50.0	25	1,613	46.7	50.0	5.6	0.11	0.165
Word Match Game	25	1,565	466.72	12.84	25	1,613	465.78	12.95	0.93	0.07	0.363
Grade equivalent			0.00				-0.10				
At grade level (%)	25	1,565	30.9	45.9	25	1,613	28.4	45.3	2.5	0.06	0.410
Time of test											
(number of days)	25	1,562	25.31	22.62	25	1,601	24.20	17.03	1.10	0.06	0.577
Demographic characteristics (%)											
Female	25	1,563	49.7	50.0	25	1,613	49.5	50.0	0.3	0.01	0.891
Race/ethnicity											
Hispanic	25	1,533	57.4	48.0	25	1,593	58.7	49.4	-1.2	-0.02	0.747
Black, non-Hispanic	25	1,533	3.4	18.4	25	1,593	4.0	20.6	-0.6	-0.03	0.756
White, non-Hispanic	25	1,533	29.7	41.9	25	1,593	31.2	46.2	-1.5	-0.03	0.648
Asian	25	1,533	4.9	22.0	25	1,593	3.1	17.9	1.9	0.09	0.233
Native American or Alaskan Native	25	1,533	0.1	3.6	25	1,593	0.1	4.3	0.0	0.00	0.986
Other	25	1,533	4.4	21.0	25	1,593	3.0	18.7	1.4	0.07	0.203
Age at baseline (year)	25	1,565	6.0	0.3	25	1,613	6.0	0.2	0.0	0.01	0.791
Other background characteristics (%)											
Low-income background status	25	1,402	68.6	45.7	25	1,484	73.3	45.3	-4.7	-0.10	0.193
English learner status	25	1,407	32.0	47.9	25	1,487	36.6	48.4	-4.6	-0.10	0.218
Special education status	25225	1,402	9.7	29.3	25	1,485	9.5	29.3	0.2	0.01	0.908

(continued)

SOURCES: District administrative records for the 2017–2018 school year and A2i assessment data for the 2018–2019 school year obtained for this study.

NOTES: This table is based on first grade students who were enrolled in the study schools in the fall of 2018 (Year 1) and took the A2i assessments during Year 2.

   *indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the students in the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.393.

Appendix Table B.9. Background Characteristics for Primary Cohort Students in the Year 3 A2i Assessment Analysis Sample, by Program Condition

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Difference		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Difference	Estimated Difference in Effect Size	P-Value
Baseline reading test (scaled score)											
Letters2Meaning	20	1,137	433.73	96.41	20	1,079	425.12	98.97	8.61	0.09	0.347
Grade equivalent			0.41				0.33				
At grade level (%)	20	1,137	50.9	50.0	20	1,079	46.3	50.0	4.6	0.09	0.281
Word Match Game	20	1,137	466.33	12.91	20	1,079	465.84	13.07	0.49	0.04	0.633
Grade equivalent			-0.10				-0.10				
At grade level (%)	20	1,137	29.3	45.2	20	1,079	28.8	45.2	0.4	0.01	0.871
Time of test (number of days)	20	1,135	25.87	23.75	20	1,073	23.57	16.14	2.31	0.11	0.354
Demographic characteristics (%)											
Female	20	1,135	49.8	50.0	20	1,079	48.8	50.0	1.0	0.02	0.646
Race/ethnicity											
Hispanic	20	1,131	59.6	47.4	20	1,070	59.8	48.3	-0.2	0.00	0.961
Black, non-Hispanic	20	1,131	3.7	18.9	20	1,070	4.6	20.3	-0.9	-0.05	0.667
White, non-Hispanic	20	1,131	26.9	40.0	20	1,070	29.8	44.4	-2.8	-0.07	0.520
Asian	20	1,131	5.4	22.9	20	1,070	2.1	14.5	3.2	0.17	0.103
Native American or Alaskan Native	20	1,131	0.1	3.0	20	1,070	0.1	3.1	0.0	-0.02	0.779
Other	20	1,131	4.3	21.1	20	1,070	3.2	18.7	1.1	0.05	0.452
Age at baseline (year)	20	1,137	6.0	0.4	20	1,079	6.0	0.3	0.0	-0.01	0.799
Other background characteristics (%)											
Low-income background status	20	1,021	72.3	44.2	20	990	73.0	44.6	-0.7	-0.02	0.707
English learner status	20	1,024	35.9	49.0	20	993	36.1	48.6	-0.2	0.00	0.962
Special education status	20	1,021	9.7	29.2	20	990	9.1	28.3	0.7	0.02	0.680

(continued)

Appendix Table B.9 (continued)

SOURCES: District administrative records for the 2017–2018 school year and A2i assessment data for the 2018–2019 school year obtained for this study.

NOTES: This table is based on first grade students who were enrolled in the study schools in the fall of 2018 (Year 1) and took the A2i assessments during Year 3.

*indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the students in the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.547.

Appendix Table B.10. Realized Minimum Detectable Effect Size for the Primary Cohort

Outcome	Minimum Detectable Effect Size	Minimum Detectable Effect	Unit of Measure
State/district reading achievement test (Year 3)			
Standardized score	0.15	0.15	Scaled score point
Percentage scored at or above proficiency	0.15	7.66	Percentage
A2i Assessments			
Letters2Meaning			
Year 1 scaled score	0.17	15.84	Scaled score point
Year 1 percentage reading at or above grade level	0.15	7.62	Percentage
Year 2 scaled score	0.15	10.70	Scaled score point
Year 2 percentage reading at or above grade level	0.13	6.17	Percentage
Year 3 scaled score	0.15	11.97	Scaled score point
Year 3 percentage reading at or above grade level	0.15	7.59	Percentage
Word Match Game			
Year 1 scaled score	0.16	2.26	Scaled score point
Year 1 percentage reading at or above grade level	0.11	5.35	Percentage
Year 2 scaled score	0.16	1.90	Scaled score point
Year 2 percentage reading at or above grade level	0.13	6.63	Percentage
Year 3 scaled score	0.11	1.35	Scaled score point
Year 3 percentage reading at or above grade level	0.13	6.27	Percentage
GMRT			
Total			
Scaled score	0.14	6.25	Scaled score point
Percentage reading at or above grade level	0.12	5.88	Percentage
Comprehension			
Scaled score	0.17	7.34	Scaled score point
Percentage reading at or above grade level	0.15	7.50	Percentage
Vocabulary			
Scaled score	0.15	8.05	Scaled score point
Percentage reading at or above grade level	0.13	6.32	Percentage

SOURCES: Gates-MacGinitie Reading Test (GMRT) data collected in the spring of 2019. A2i assessment data collected by the A2i platform from summer 2018 to spring 2021. District records data for the 2017–2018 and 2020–2021 school years.

NOTE: The minimum detectable effect sizes in this table are calculated by multiplying the standard error of the estimated effects by 2.8 and dividing by the standard deviations of the student outcome measures in the face-to-face schools, assuming a statistical significance level of 0.05.

III. TEACHER SURVEY SAMPLES

The study fielded two rounds of online teacher surveys to collect information on teachers' use of differentiated small-group instruction, their knowledge of the literacy instruction used in the A2i system recommendations, and their use of the recommendations made by the system. The first round of the teacher survey was fielded in the spring of Year 1 to all kindergarten and first grade regular classroom teachers in the study schools. The second round was fielded in the spring of Year 3, targeting all kindergarten to third grade regular classroom teachers in the study schools that stayed in the study. Appendix Table B.11 shows the sample sizes and response rates for these two rounds of surveys, and Appendix Tables B.12–B.13 present the equivalence check for the respondent samples from these two rounds of surveys.

Table B.11. Sample Sizes and Response Rates for Teacher Surveys, by Round and Program Conditions

Outcome	Total		Mixed-Mode Schools		Face-to-Face Schools	
	School	Teacher	School	Teacher	School	Teacher
Year 1 teacher survey						
Total number	59	414	30	199	29	215
Respondents	58	321	29	153	29	168
Response rate (%)	98.3	77.5	96.7	76.9	100.0	78.1
Year 3 teacher survey						
Total number	40	499	20	225	20	274
Respondents	37	273	19	124	18	149
Response rate (%)	92.5	54.7	95.0	55.1	90.0	54.4

SOURCE: Authors' calculation based on teacher survey data collected in spring 2019 and spring 2021.

NOTE: None of the differences in the response rate between the mixed-mode schools and face-to-face schools are statistically significant at the 0.05 level. The test is based on a regression analysis that controls for the random assignment blocks.

Table B.12. Background Characteristics for Year 1 Teacher Survey Respondents, by Program Condition

Characteristics	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Estimated Difference in Effect Size	P-Value for Estimated Difference
Teacher experience (number of years)					
In any school	18.31	16.87	1.45	0.15	0.329
In current school	11.20	11.68	-0.47	-0.05	0.720
In current grade	9.85	9.20	0.65	0.08	0.550
Highest degree obtained (%)					
Bachelor's degree	31.7	30.2	1.5	0.03	0.815
Master's degree	58.3	61.9	-3.6	-0.07	0.583
Post-master's certificate	5.7	5.2	0.5	0.02	0.849
Doctorate degree	0.0	1.6	-1.6	-0.28	0.359
Other	4.3	1.7	2.6	0.15	0.231
Certification (%)					
General elementary grade teacher	94.9	96.9	-2.0	-0.10	0.456
Reading teacher	14.7	10.9	3.8	0.12	0.310
Special educator	8.6	7.1	1.5	0.05	0.709
English as a Second Language	13.4	18.1	-4.7	-0.13	0.287
Bilingual education	7.8	9.6	-1.8	-0.06	0.613
Other certification	12.0	9.2	2.7	0.08	0.469
No certification	1.4	0.0	1.4	0.18	0.131
Number of teachers	153	168			

SOURCE: Teacher survey data collected in the spring of 2019.

NOTES: Sample size for each characteristic may vary due to missing values.
   *indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the respondents in the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.445.

Table B.13. Background Characteristics for Year 3 Teacher Survey Respondents, by Program Condition

Characteristics	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Estimated Difference in Effect Size	P-Value for Estimated Difference
Teacher experience (number of years)					
In any school	18.49	18.03	0.47	0.05	0.710
In current school	12.33	13.46	-1.13	-0.13	0.348
In current grade	9.28	9.77	-0.49	-0.06	0.640
Highest degree obtained (%)					
Bachelor's degree	28.9	19.8	9.0	0.21	0.131
Master's degree	57.1	66.9	-9.8	-0.20	0.192
Post-master's certificate	14.1	10.8	3.2	0.11	0.457
Doctorate degree	0.0	0.8	-0.8	-0.09	0.515
None of the above	0.0	1.2	-1.2	-0.19	0.174
Certification (%)					
General elementary grade teacher	99.3	94.1	5.2	0.31	0.037
Reading teacher	16.1	19.2	-3.2	-0.08	0.548
Special educator	9.0	11.4	-2.4	-0.08	0.526
English as a Second Language	16.0	13.3	2.7	0.08	0.609
Bilingual education	9.1	11.2	-2.0	-0.07	0.634
Other certification	13.1	15.1	-1.9	-0.05	0.714
No certification	0.0	0.5	-0.5	-0.09	0.545
Number of teachers	124	149			

SOURCE: Teacher survey data collected in the spring of 2021.

NOTES: Sample size for each characteristic may vary due to missing values
  *indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the respondents in the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is 0.125.

Given that over 45 percent of the teachers did not respond to the Year 3 teacher survey, largely because of the effects of COVID-19, it is important to see whether the survey findings based on the respondent sample still apply to all teachers in the study. To do so, ideally one would compare the background characteristics of the respondents and non-respondents. However, such analysis is not feasible because the study does not have such information for the non-responding teachers in the sample. Instead, the study examined whether students taught by teachers who responded to the Year 3 survey differed from those who did not respond and found no systematic differences between these two groups, even though there are sporadic and small differences for some individual characteristics. Details of this check are presented in Appendix Table B.14. Findings presented in this section suggest that the validity of the survey analysis does not appear to be affected by teacher non-response.

Appendix Table B.14. Student Background Characteristics for all Students with Year 3 A2i Assessment Scores, by Year 3 Teacher Response Status to the Year 3 Teacher Survey

Outcome	Students with Responding Teachers				Students with Non-Responding Teachers				Estimated Difference			
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Difference		Estimated Difference in Effect Size	P-Value
Baseline reading test (scaled score)												
Letters2Meaning	37	5,060	398.57	106.18	40	3,768	407.18	107.95	-8.61	*	-0.08	0.001
Grade equivalent			0.12				0.18					
At grade level (%)	37	5,060	79.3	41.0	40	3,768	76.6	42.2	2.7	*	0.06	0.008
Word Match Game	37	5,062	463.26	13.62	40	3,769	464.14	14.21	-0.88	*	-0.06	0.009
Grade equivalent			-0.40				-0.30					
At grade level (%)	37	5,062	65.2	47.7	40	3,769	63.2	48.8	2.0		0.04	0.090
Time of test (number of days)	37	5,073	40.99	50.65	40	3,778	41.17	49.17	-0.19		0.00	0.878
Demographic characteristics (%)												
Age at baseline (year)	37	5,073	4.68	1.05	40	3,778	4.81	1.15	-0.13	*	-0.12	0.000
Female	37	4,735	48.3	50.0	40	3,568	50.1	50.0	-1.8		-0.04	0.150
Race/ethnicity												
Hispanic	37	4,719	59.7	48.2	40	3,543	60.1	48.9	-0.4		-0.01	0.616
Black, non-Hispanic	37	4,719	4.3	19.0	40	3,543	3.9	22.0	0.4		0.02	0.449
White, non-Hispanic	37	4,719	27.2	43.4	40	3,543	28.7	43.6	-1.4		-0.03	0.054
Asian	37	4,719	4.7	19.3	40	3,543	3.3	19.7	1.4	*	0.07	0.004
Native American or Alaskan Native	37	4,719	0.2	3.3	40	3,543	0.4	6.3	-0.2		-0.04	0.109
Other	37	4,719	3.9	19.3	40	3,543	3.7	20.6	0.2		0.01	0.695
Other background characteristics (%)												
Low-income background status	37	3,658	75.0	43.4	40	2,783	70.7	46.5	4.3	*	0.10	0.000
English learner status	37	3,606	33.8	48.7	40	2,640	32.1	47.1	1.7		0.04	0.251
Special education status	37	3,752	10.2	30.0	40	2,809	11.7	29.9	-1.6		-0.05	0.089

(continued)

Appendix Table B.14 (continued)

SOURCES: District administrative records from the 2017–2018 school year, A2i assessment data for the 2018–2019 school year obtained for this study, and Teacher Survey data collected in the spring of 2021.

NOTES: This table is based on students in grades K-3 who were enrolled in the study schools in the fall of 2020 and took at least one A2i assessment during Year 3.

   *indicates the estimated difference is statistically significant at the 0.05 level with a two-tailed t-test. An F-test was used to determine whether there is a systematic difference between the students in the mixed-mode schools and the face-to-face schools, with respect to the characteristics included in this table. The p-value for this test is less than 0.001.

C

Program Implementation

This appendix provides supplemental information for program implementation. It starts with information about the initial implementation fidelity measures proposed at the start of the study. The following parts present supplementary findings for implementing key A2i components, including the administration of the A2i assessment, teachers' use of the A2i platform, and teachers' perception of other A2i components. Lastly, it provides contextual information on teachers' perception of how COVID-19 affected their teaching environment in Year 3.

I. IMPLEMENTATION FIDELITY

As illustrated in the theory of action (Figure 3), the key input components of A2i include the professional development (PD) provided to teachers, A2i assessments, and teachers' use of the A2i platform. To assess whether schools implemented the A2i system with fidelity, one needs to examine whether schools had adequate exposure to or implementation of these key components. At the start of the evaluation, the study team proposed a set of implementation fidelity measures and related fidelity thresholds to assess program implementation in mixed-mode and face-to-face schools. Appendix Table C.1 details these proposed fidelity measures and the threshold values.

An individual is considered to be implementing with fidelity if this person (a teacher or a student) reaches the individual-level threshold for a given component. A school is considered to implement an A2i component with fidelity if there is a large enough proportion of teachers or students in the school reaching the individual threshold level specified in the table. An A2i component is implemented with fidelity if at least 60 percent of the schools implement it with fidelity. For example, in Year 1, a teacher is considered to be making adequate use of the A2i platform if he or she spent at least 240 minutes on the platform over the course of the school year. A school is considered to be using the A2i technology with fidelity if more than 60 percent of its Kindergarten and Grade 1 teachers (the grade levels covered by the study in Year 1) met the threshold of 240 minutes. The A2i technology is considered implemented with fidelity for the mixed-mode model if more than 60 percent of the mixed-mode schools implemented it with fidelity.

The threshold values for these fidelity measures vary by implementation year to account for COVID-19-related school disruptions. For example, the team reduced the threshold for the number of A2i assessments students needed to take in a school year from three times to two times to account for school closures in spring 2020. The threshold for the amount of PD delivered to and received by teachers and the minimum amount of time teachers spent on the A2i platform were also adjusted to reflect the disruptions.

The fidelity thresholds for PD vary by program condition, implementation year, and grade level. First, the thresholds differ by program condition because, by design, the mixed-mode schools receive less Individualized Classroom Coaching (ICC) than the face-to-face schools. Second, the team adjusted the minimum amount of PD required for fidelity to account for COVID-19-related disruptions. Specifically, for Year 2, the minimum necessary

Appendix Table C.1. Thresholds for Implementation Fidelity Calculation

Measure	Individual Level	School Level	Sample Level
		Fidelity Threshold	
Professional development (PD)			
Mixed-mode schools			
Year 1, Grades K-1[a]		Received 7 Literacy Huddles (LH) + 2 Individualized Classroom Coaching sessions (ICC)	
Year 2, Grades K-1	60 minutes		
Year 2, Grade 2	185 minutes		
Year 3, Grades K-1	0 minutes		60% of schools[a]
Year 3, Grade 2	60 minutes	60% of teachers per school[a]	
Year 3, Grade 3	300 minutes		
Face-to-face schools			
Year 1, Grades K-1[a]		Received 7 LHs + 2 ICCs	
Year 2, Grades K-1	60 minutes		
Year 2, Grade 2	275 minutes		
Year 3, Grades K-1	0 minutes		60% of schools[a]
Year 3, Grade 2	60 minutes	60% of teachers per school[a]	
Year 3, Grade 3	480 minutes		
A2i assessment frequency			
Year 1	3 times		
Year 2	2 times	95% of students per school	60% of schools
Year 3	3 times		
A2i platform usage			
Year 1	240 minutes		
Year 2	160 minutes	60% of teachers per school	60% of schools
Year 3	190 minutes		

SOURCE: Authors' summary.

NOTES: Fidelity thresholds for Year 2 and Year 3 were adjusted to account for COVID-19-related school disruptions.

[a]PD fidelity calculation for Year 1 is based on school-level PD delivery information because teacher PD participation data was not available for this year.

amount of Literacy Huddles was reduced by one-third, and that of ICC was reduced by half. Lastly, PD efforts differ across grade levels, with the focal grade in each year (Grades K-1 in Year 1, Grade 2 in Year 2, and Grade 3 in Year 3) receiving the "full" dosage and teachers in other grades (who would have been trained in prior implementation years) receiving scaled-back amounts. Therefore, the threshold values differ by grade level. In addition, because teacher-level PD attendance data were unavailable for Year 1, the study used the school-by-grade level measure of the number of PD events as an alternative fidelity measure for this year.

Appendix Tables C.2 and C.3 present implementation fidelity findings for the overall sample and the primary cohort separately. The overall sample includes all grade levels exposed to the A2i program, and the primary cohort includes the focal grades intended to receive the full dosage of PD each year. Findings presented in both tables show that only a small proportion of individuals and schools reached the specific thresholds for the three key measures each year. Therefore, except for PD attendance and use of assessment in Year 2, none of these key A2i components were implemented with adequate fidelity in the two groups of schools.

II. SUPPLEMENTARY FINDINGS FOR ADMINISTRATION OF THE A2I ASSESSMENT

Figures 5 and 10 in the report present the average number of times students took the A2i assessment each year. However, the number of assessments each student took varies. Appendix Figure C.1 shows such variation at the student level. Specifically, it shows the proportion of students who took A2i assessments at least three times. Across the three implementation years, a majority of students in the sample took a given A2i assessment at least three times, but this proportion decreased from year to year.

In addition, the proportion of students meeting the implementation fidelity threshold for minimum number of A2i assessments in each school varies, and the average number masks this variation across schools. Appendix Figure C.2 uses box-and-whisker charts to illustrate this variation by assessment, year, and program condition. Specifically, the colored box in the graph shows the inter-quartile (or the 25th to 75th percentile) range of the distribution, and the whiskers indicate the minimum and maximum values (excluding outliers) of the distribution. The horizontal line in the middle of the box represents the median value of the distribution, and the cross symbol indicates the mean level of the distribution. For example, in Year 1, the mean percentage of students in a mixed-mode school who took the Letters2Meaning assessment at least three times is 73 percent, as represented by the cross symbol, and the median level is 76 percent as indicated by the horizontal line in the dark blue box.

Teachers reported confidence in their capacity to administer the A2i assessment and in their students' ability to take the assessment. Appendix Table C.4 presents teachers' responses to questions related to administering the assessment in the Year 1 and Year 3 teacher surveys.

Appendix Table C.2. Implementation Fidelity for Overall Sample, by Year and Program Condition

Implementation Measure	Year 1 Program Condition		Year 2 Program Condition		Year 3 Program Condition	
	Mixed-Mode Schools	Face-to-Face Schools	Mixed-Mode Schools	Face-to-Face Schools	Mixed-Mode Schools	Face-to-Face Schools
Professional development delivery and attendance (minutes)						
Percentage of schools with fidelity[a]	43.3	44.8	84	100	20	45
Number of schools	30	29	25	25	20	20
Number of teachers	NA	NA	222	254	225	274
A2i assessment administration						
Percentage of schools with fidelity[b]	6.7	6.9	16	32	0	0
Number of schools	30	29	25	25	20	20
Number of students	5111	5119	5801	6111	5291	5758
A2i platform usage						
Percentage of schools with fidelity	0	6.9	20	44	5	10
Number of schools	30	29	25	25	20	20
Number of teachers	201	217	222	254	225	274

SOURCE: Professional development (PD) plan and delivery records provided by Learning Ovations, Inc. in Year 1; teacher-level PD attendance records collected by Learning Ovations in Year 2 and Year 3; and A2i assessment data and A2i system log data from the A2i platform collected throughout Years 1-3.

NOTES: The fidelity calculations are based on all students and teachers covered by the program in each year. The Year 1 sample includes Grades K through 1; the Year 2 sample includes Grades K through 2; and the Year 3 sample includes Grades K through 3.

A school is considered to be implementing PD or A2i platform usage with fidelity if at least 60 percent of teachers meet the fidelity threshold. A school is considered to be implementing A2i assessment with fidelity if at least 95 percent of students meet the fidelity threshold. A program condition (mixed-mode or face-to-face) is considered to be implementing a component with fidelity if at least 60 percent of the schools meet the fidelity threshold.

[a]The fidelity thresholds for Year 2 were adjusted to account for school closures from March to June 2020 when no program-related activities took place. Teacher-level PD attendance records were not available in Year 1. For this year, a school's fidelity status for implementing PD was based on whether the number of delivered PD sessions met the targeted number.

[b]The fidelity thresholds for Year 2 were adjusted to account for school closures from March to June 2020 when no program-related activities took place.

Appendix Table C.3. Implementation Fidelity for the Primary Cohort, by Year and Program Condition

Implementation Measure	Year 1 Program Condition		Year 2 Program Condition		Year 3 Program Condition	
	Mixed-Mode Schools	Face-to-Face Schools	Mixed-Mode Schools	Face-to-Face Schools	Mixed-Mode Schools	Face-to-Face Schools
Professional development (PD) delivery and attendance (minutes)						
Percentage of schools with fidelity[a]	46.7	58.6	80	80	10	0
Number of schools	30	29	25	25	20	20
Number of teachers	30	29	80	84	65	65
A2i assessment administration						
Percentage of schools with fidelity[a]	30	34.5	64	60	5	5
Number of schools	30	29	25	25	20	20
Number of students	2288	2223	1580	1653	1171	1173
A2i platform usage						
Percentage of schools with fidelity	6.7	20.7	40	56	35	25
Number of schools	30	29	25	25	20	20
Number of teachers	96	98	80	84	65	65

SOURCE: PD plan and delivery records provided by Learning Ovations Inc. in Year 1; teacher-level PD attendance records collected by Learning Ovations in Year 2 and Year 3; and A2i assessment data and A2i system log data from the A2i platform, collected throughout Years 1–3.

NOTES: The fidelity calculations are based on all students and teachers in the primary cohort of the sample. This cohort includes Grade 1 in Year 1, Grade 2 in Year 2, and Grade 3 in Year 3.

A school is considered to be implementing PD or A2i platform usage with fidelity if at least 60 percent of teachers meet the fidelity threshold. A school is considered to be implementing the A2i assessment with fidelity if at least 95 percent of students meet the fidelity threshold. A program condition (mixed-mode or face-to-face) is considered to be implementing a component with fidelity if at least 60 percent of the schools meet the fidelity threshold.

[a]The fidelity thresholds for Year 2 were adjusted to account for school closures from March to June 2020 when no program-related activities took place. Teacher-level PD attendance records were not available in Year 1. For this year, a school's fidelity status for implementing PD was based on whether the number of delivered PD sessions met the targeted number.

Appendix Figure C.1. Percentage of Students Who Took A2i Assessments at or Above the Minimum Required Frequency, by Year, Assessment and Program Condition

Letters2Meaning

Word Match Game

Year 1

Tested 3+ times
- 84.0 (Mixed-mode schools)
- 86.4 (Face-to-face schools)

Tested 3+ times
- 86.5 (Mixed-mode schools)
- 87.2 (Face-to-face schools)

Year 2

Tested 2+ times
- 83.2 (Mixed-mode schools)
- 81.8 (Face-to-face schools)

Tested 3+ times
- 63.0 (Mixed-mode schools)
- 61.7 (Face-to-face schools)

Tested 2+ times
- 83.5 (Mixed-mode schools)
- 81.6 (Face-to-face schools)

Tested 3+ times
- 63.6 (Mixed-mode schools)
- 62.1 (Face-to-face schools)

Year 3

Tested 3+ times
- 58.0 (Mixed-mode schools)
- 48.3 (Face-to-face schools)

Tested 3+ times
- 56.5 (Mixed-mode schools)
- 46.0 (Face-to-face schools)

■ Mixed-mode schools   ■ Face-to-face schools

SOURCE: A2i assessment data from the A2i platform, collected throughout the 2018-2019, 2019-2020, and 2020-2021 school years.

NOTE: None of the estimated differences between the mixed-mode and face-to-face schools is statistically significant at the 0.05 level.

Appendix Figure C.2. Distribution of Percentage of Students Tested at Adequate Frequency in a School, by Year, Assessment, and Program Condition

Year 1

Letters2Meaning

Word Match Game

Year 2

Letters2Meaning

Word Match Game

Year 3

Letters2Meaning

Word Match Game

■ Mixed-mode schools   ■ Face-to-face schools
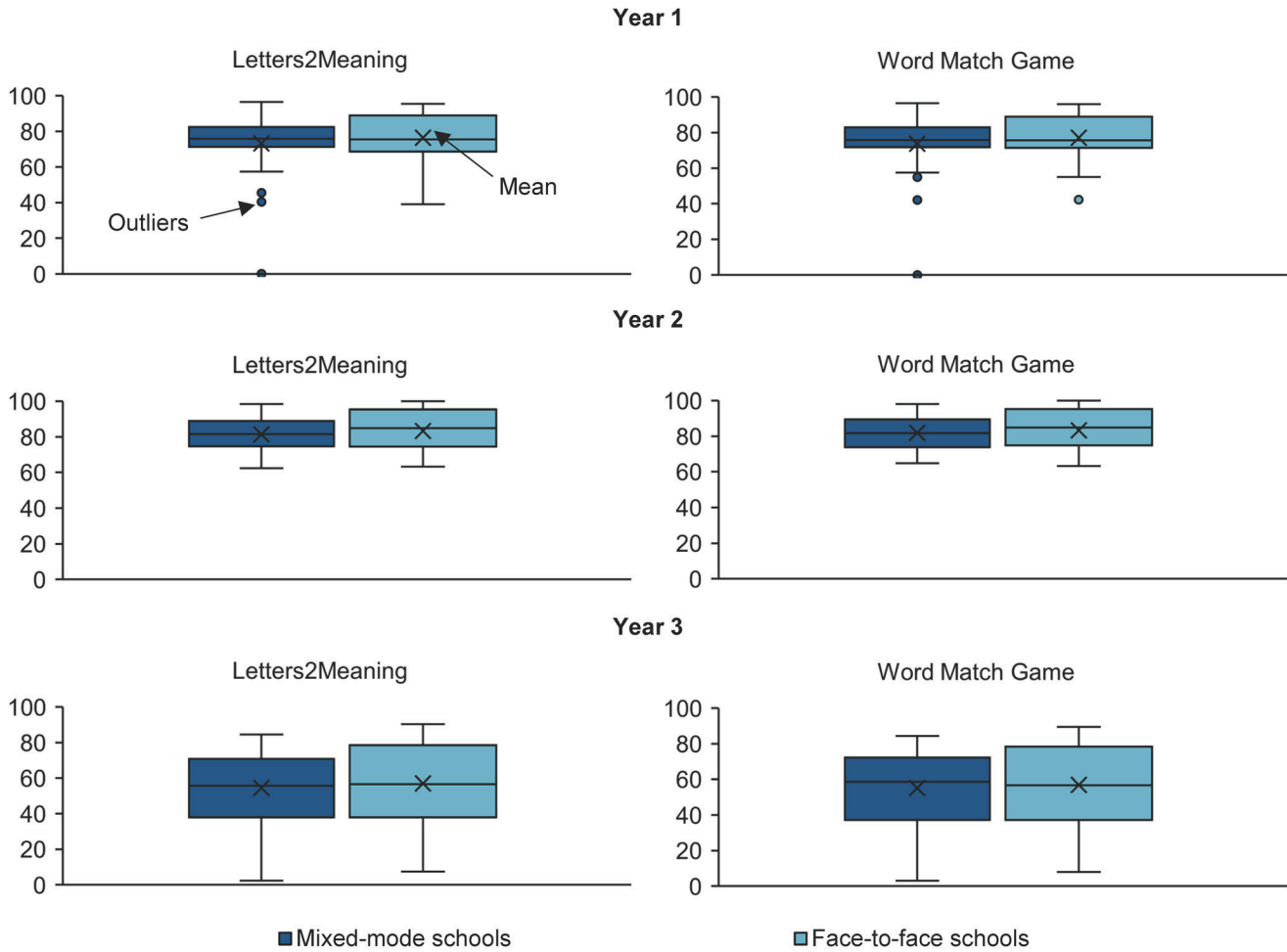
SOURCES: Authors' calculation based on A2i assessment data from the A2i platform, collected throughout the 2018-2019, 2019-2020, and 2020-2021 school years.

NOTES: For Year 1 and Year 3, the figure shows the distribution of the percentage of students who took a given A2i assessment three times or more within the school year. The threshold is adjusted to two times for Year 2 to account for the school closures due to COVID.

Table C.4. Teachers' Responses to Assessment-Related Questions in Survey, by Year and Program Condition

Outcome	Year 1			Year 3		
	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference
Percentage of teachers who agreed with the statement						
My students know how to take a computer-based A2i assessment.	86.9	92.3	-5.3	86.6	87.7	-1.1
My students have the self-regulation to sit for an A2i assessment.	73.3	79.2	-5.9	78.2	73.5	4.7
I had enough help from school staff members during administration to give A2i assessments.	53.0	60.0	-7.0	69.3	61.5	7.8
I had the training I needed to administer the A2i assessments	84.6	86.4	-1.8	86.8	90.7	-3.9
I can administer A2i assessments efficiently.	54.0	58.4	-4.4	89.0	89.8	-0.9
Amount of adult support for A2i assessment administration received (%)						
None	51.0	63.5	-12.4	82.3	71.2	11.1
Part of the time	23.7	11.8	11.8 *	12.4	8.9	3.4
All the time	25.3	23.8	1.5	5.3	18.7	-13.4 *
Number of schools	30	29		19	18	
Number of teachers	153	168		124	149	

SOURCE: Teacher survey data collected in the spring of 2019 and spring of 2021.

NOTE: The estimated differences are regression-adjusted, controlling for the blocking of random assignment and teachers' teaching experience in general, in current school, and in current grade, as well as their highest degree earned and types of teaching certification. Rounding may cause slight discrepancies in calculating means and differences. *indicates the estimated differential impact is statistically significant at the 0.05 level.

III. SUPPLEMENTARY FINDINGS FOR TEACHER USAGE OF THE A2i PLATFORM

A key part of teachers' usage of the A2i platform is the amount of time they spent on the platform's web pages. The study grouped these web pages into four categories based on their content:

- Student information pages, including student information, test scores, graphs, progress, and classroom information

- Resource pages, including resources and a video library

- A2i tools pages such as lesson plans, reports, and administrative information

- A2i login pages, including the system portal, account, and login pages

Appendix Table C.5 presents the amount of time teachers spent on these different types of web pages by year and by program condition. These findings show that, while there were some variations across years, teachers spent the largest proportion of their time on the student information pages.

The findings presented in Figure 5 and Figure 10 in the report suggest that teachers from the face-to-face schools spent more time on the A2i web pages than their counterparts in the mixed-mode schools. One hypothesis for this difference is that teachers in the face-to-face schools might spend more time on the system during or soon after the in-person PD event because face-to-face interactions with the Literacy Outcomes Specialists (LOSs) might have encouraged or induced them to dig deeper into the available resources. To test this hypothesis, the study checked if teachers from face-to-face schools spent more time on the platform during PD days (days when a PD event took place, in the school or virtually) than those from mixed-mode schools. Findings in Appendix Table C.6 show that teachers from both groups spent about the same amount of time on the platform during PD days, except for Year 1 when teachers in the face-to-face group spent more time on A2i on non-PD days. By and large, more in-person interactions with the LOSs did not seem to be related to greater usage of the platform, and it cannot explain why teachers from face-to-face schools used the platform more.

The amount of time teachers spent on the A2i platform also seems to vary by grade level. Appendix Table C.7 presents findings on teachers' usage by grade level and shows that teachers in the focal grades in each implementation year tended to spend more time on the platform than teachers from the non-focal grades.

Another important feature of the A2i platform is the lesson planning tools and resources it provides to teachers. These tools are designed to help teachers plan their lessons based on A2i-generated recommendations and the reading curriculum used in the school. Overall, teachers did not make full use of such tools. Appendix Table C.8 provides more details.

Appendix Table C.5. Amount of Time Teachers Spent on A2i Platform, by Web Page Type, Year, and Program Condition

Outcome	Year 1			Year 2			Year 3		
	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference
Percentage with usage above threshold	10.6	20.1	-9.5 *	35.0	51.7	-16.7 *	23.8	23.7	0.0
Total minutes	108.74	142.41	-33.67 *	169.48	193.32	-23.85	111.86	137.11	-25.26
Student information	80.65	106.89	-26.25 *	118.10	127.35	-9.24	78.80	93.82	-15.01
Resource	2.05	3.05	-1.00	2.07	2.68	-0.60	2.54	3.68	-1.15
Tools	20.00	25.06	-5.06	17.34	24.70	-7.36	5.43	7.70	-2.26
Login	6.06	7.69	-1.63	31.96	38.66	-6.70	25.08	32.08	-7.00
Number of schools	30	29		25	25		20	20	
Number of teachers	199	215		222	254		225	274	

SOURCE: A2i usage data obtained for this study.

NOTE: The estimated differences are regression-adjusted, controlling for the blocking of random assignment and teachers' teaching experience in general, in current school, and in current grade, as well as their highest degree earned and types of teaching certification. Rounding may cause slight discrepancies in calculating means and differences. *indicates the estimated differential impact is statistically significant at the 0.05 level.

Appendix Table C.6. Amount of Time Teachers Spent on A2i Platform, by Timing, Year, and Program Condition

Outcome	Year 1			Year 2			Year 3		
	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference
Total minutes	108.74	142.41	-33.67 *	169.48	193.32	-23.85	111.86	137.11	-25.26
On days with any professional development (PD) activity	32.35	34.86	-2.51	35.02	48.20	-13.18	22.70	22.84	-0.14
On days without PD activity	76.39	107.38	-30.99 *	134.46	144.77	-10.31	89.16	114.28	-25.12
Number of schools	30	29		25	25		20	20	
Number of teachers	199	215		222	254		225	274	

SOURCE: A2i usage data obtained for this study.

NOTE: The estimated differences are regression-adjusted, controlling for the blocking of random assignment and teachers' teaching experience in general, in current school, and in current grade, as well as their highest degree earned and types of teaching certification. Rounding may cause slight discrepancies in calculating means and differences. *indicates the estimated differential impact is statistically significant at the 0.05 level.

Appendix Table C.7. Amount of Time Teachers Spent on A2i Platform, by Grade, Program Condition, and Year

Outcome	Year 1			Year 2			Year 3		
	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference
Grade K									
Total minutes	115.66	131.66	-16.00	136.60	158.77	-22.17	44.68	105.93	-61.26
% above threshold	11.3	18.7	-7.4	29.4	45.5	-16.1	8.5	20.7	-12.2
Grade 1									
Total minutes	103.58	173.56	-69.97 *	164.82	203.61	-38.78	68.46	99.95	-31.49
% above threshold	9.4	24.5	-15.1 *	32.4	50.4	-18.0 *	15.9	22.2	-6.3
Grade 2									
Total minutes				212.36	222.01	-9.66	185.43	220.11	-34.68
% above threshold				42.5	60.9	-18.4 *	36.5	35.4	1.1
Grade 3									
Total minutes							170.12	187.71	-17.59
% above threshold							39.6	33.9	5.7
Number of schools	30	29		25	25		20	20	
Number of teachers	199	215		222	254		225	274	

SOURCE: A2i usage data obtained for this study.

NOTE: The estimated differences are regression-adjusted, controlling for the blocking of random assignment and teachers' teaching experience in general, in current school, and in current grade, as well as their highest degree earned and types of teaching certification. Rounding may cause slight discrepancies in calculating means and differences.
*indicates the estimated differential impact is statistically significant at the 0.05 level.

Table C.8. Teachers' Responses to the Use of Lesson Planning Tools and Related Questions in Survey, by Year and Program Condition

Outcome	Year 1			Year 3		
	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference
# of days teachers' used the system-suggested lesson plan in a week	0.68	1.05	-0.37	0.78	0.90	-0.12
Used A2i-generated lesson plan for at least some groups (%)	18.8	26.3	-7.4	26.0	19.6	6.4
Time spent on lesson planning per week (%)						
Under 1 hour	6.9	15.1	-8.2 *	10.5	11.6	-1.1
Between 1 and 2 hours	53.1	48.2	5.0	29.7	44.8	-15.1
Between 2 and 3 hours	22.9	18.3	4.5	28.3	20.6	7.7
Between 3 and 4 hours	6.6	4.2	2.4	15.9	10.1	5.7
More than 4 hours	10.5	13.6	-3.1	15.6	12.9	2.7
Number of schools	30	29		19	18	
Number of teachers	153	168		124	149	

SOURCE: Teacher survey data collected in the spring of 2019 and spring of 2021.

NOTE: The estimated differences are regression-adjusted, controlling for the blocking of random assignment and teachers' teaching experience in general, in current school, and in current grade, as well as their highest degree earned and types of teaching certification. Rounding may cause slight discrepancies in calculating means and differences. *indicates the estimated differential impact is statistically significant at the 0.05 level.

IV. SUPPLEMENTARY FINDINGS FOR TEACHER BELIEFS ABOUT A2I COMPONENTS AND USE OF DIFFERENTIATED SMALL-GROUP INSTRUCTION

The theory of action for the A2i system posits that implementing the key A2i components could lead to changes in teachers' knowledge and perception of A2i and encourage teachers to carry out differentiated small-group instruction in their classrooms. Figures 6 and 10 in the report show no differences between mixed-mode and face-to-face schools in these measures based on teachers' survey responses. Appendix Table C.9 provides supplementary information for these findings.

V. TEACHERS' REPORT ON HOW COVID-19 AFFECTED THEIR TEACHING IN YEAR 3

Schools and teachers spent the 2020–2021 school year coping with the disruptions caused by the COVID-19 pandemic. Many aspects of teaching were fundamentally changed during the school year. The Year 3 survey asked teachers about such changes; their responses are summarized in Appendix Table C.10.

Appendix Table C.9. Teachers' Responses to Questions Related to Knowledge, Opinions, and Differentiation in Survey, by Year and Program Condition

Outcome	Year 1			Year 3		
	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference
On teacher knowledge						
Total score (0 to 6 points)	4.78	4.51	0.27	4.95	4.83	0.12
Percentage of teachers who answered correctly						
Scenario A	88.7	84.5	4.2	89.1	91.0	-1.9
Scenario B	89.3	84.9	4.4	87.7	90.1	-2.5
Scenario C	84.7	80.5	4.2	96.2	87.3	8.9
Scenario D	87.0	78.6	8.4	93.4	82.8	10.6
Scenario E	59.6	58.2	1.4	65.3	64.7	0.6
Scenario F	68.8	64.6	4.2	77.4	81.2	-3.7
On A2i assessment						
Percentage of teachers who agreed with the statement:						
I believe that the data collected on A2i assessments is valid if taken at school.	44.6	54.4	-9.8	85.1	81.7	3.4
I believe that the data collected on A2i assessments is valid if taken at home.				19.1	16.5	2.6
I found the A2i assessment results useful for understanding my students' academic needs.	52.5	56.4	-3.8	63.9	65.3	-1.4
I found the A2i recommended amounts of each type of instruction useful.	57.2	59.8	-2.7	65.4	60.2	5.2
On A2i-generated grouping recommendations						
Created small groups based on recommendation (Y/N)				43.2	53.8	-10.7
Agreement with recommendation (%)						
Less than half of the students	16.1	7.8	8.3 *	15.6	19.7	-4.1
About half of the students	32.3	25.8	6.5	21.6	16.0	5.6
More than half of the students	51.6	66.7	-15.1 *	62.9	64.4	-1.5
Changed groupings for students for whom you disagree with recommendation (%)						
None of the students	11.3	13.5	-2.2	15.3	16.9	-1.6
Some of the students	48.1	53.3	-5.3	71.0	61.2	9.7
All students	40.6	33.2	7.4	13.7	21.9	-8.2

(continued)

Appendix Table C.9 (continued)

Outcome	Year 1			Year 3		
	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference
On A2i-generated instruction time recommendations						
Gave instruction according to A2i time recommendation (%)				41.0	37.0	4.0
Agreement with recommendation (%)						
Less than half of the students	17.8	11.9	6.0	19.4	19.1	0.3
About half of the students	28.5	22.3	6.2	28.3	19.6	8.7
More than half of the students	53.6	65.8	-12.2	52.3	61.4	-9.1
Changed instruction time for students for whom you disagree with recommendation (%)						
None of the students	16.3	23.2	-6.9	22.4	16.9	5.4
Some of the students	58.6	64.9	-6.3	70.9	74.0	-3.1
All students	25.2	12.0	13.2 *	6.8	8.9	-2.1
On individualized small-group instruction						
Do you use this practice in your reading block? (%)	98.2	98.5	-0.3	72.7	76.2	-3.5
How many days do you use it in a typical week (number of days)	4.18	4.24	-0.06	3.01	2.69	0.32
Percentage of reading block time dedicated to small groups						
Up to a quarter	26.4	31.0	-4.6	80.7	80.8	-0.1
About half	42.9	34.9	8.0	11.5	14.6	-3.1
More than half	25.8	30.0	-4.2	6.4	3.9	2.5
All	4.8	4.5	0.3	1.4	0.6	0.7
Number of schools	30	29		19	18	
Number of teachers	153	168		124	149	

SOURCE: Teacher survey data collected in the spring of 2019 and spring of 2021.

NOTE: The estimated differences are regression-adjusted, controlling for the blocking of random assignment and teachers' teaching experience in general, in current school, and in current grade, as well as their highest degree earned and types of teaching certification. Rounding may cause slight discrepancies in calculating means and differences. *indicates the estimated differential impact is statistically significant at the 0.05 level.

Appendix Table C.10. Teachers' Responses to Survey Questions Related to the Impact of COVID-19, by Program Condition, Year 3 Survey

Outcome	Mixed-Mode Schools	Face-to-Face Schools	Estimated Difference
Has COVID-19 negatively affected the following (Yes/No)			
The length of your literacy block	84.2	81.9	2.3
Class management	58.2	65.2	-7.0
Teaching resources/ materials available	61.9	59.6	2.3
Availability of co-teachers, teachers' aides, volunteers	77.3	71.6	5.7
Core curriculum	45.4	49.3	-3.9
Administrators' expectations/ support	38.5	29.5	9.0
Number of children in the classroom or online	62.4	59.0	3.4
Professional development and coaching from the Literacy Outcomes Specialist	47.1	39.6	7.5
Student engagement in virtual lessons	79.2	80.0	-0.8
Student ability to use technology for virtual lessons	69.6	67.9	1.7
Student ability to follow instructions in virtual lessons	78.0	75.8	2.2
Available time to plan individualized lessons	69.7	80.7	-11.1
Number of schools	19	18	
Number of teachers	124	149	

SOURCE: Teacher survey data collected in the spring of 2021.

NOTE: The estimated differences are regression-adjusted, controlling for the blocking of random assignment and teachers' teaching experience in general, in current school, and in current grade, as well as their highest degree earned and types of teaching certification. Rounding may cause slight discrepancies in calculating means and differences.

*indicates the estimated differential impact is statistically significant at the 0.05 level.

Supplementary Information on Student Outcomes

This appendix provides detailed information about the student reading outcomes used in the study. It also presents detailed differential impact estimates on these student outcomes for various samples.

I. STUDENT OUTCOME MEASURES

The study used three reading measures to track and capture students' short-term and longer-term reading performance during the study period.

First, the study used the Gates-MacGinitie Reading Test (fourth edition, Form S and Form T, Level 1) to measure first grade students' general reading skills and serve as a common general reading measure across all study schools after one year of exposure to A2i in spring 2019. This test generally takes about 100 minutes to administer with paper and pencil. The test yields a single overall performance score as well as separate scores measuring students' decoding and comprehension skills. The Gates-MacGinitie is a nationally normed assessment. Normative scores were developed in 2005–2006 with a sampling plan based on geographic region, family income, enrollment size, parents' years of schooling, and other factors. Studies have shown total test and subtest internal consistency levels with coefficient values at or above 0.90 for the total tests and around 0.88–0.89 for the subtests at all levels except Adult Reading.[1]

Second, as part of the A2i system, two assessments, Word Match Game and Letters2Meaning, measure three constructs required by the A2i recommendation algorithms—vocabulary, decoding, and comprehension. The Word Match Game assessment includes a semantic matching task where students are asked to click on two of the three or four words that are semantically related. The Letters2Meaning assessment includes letter and letter-sound identification, word reading, spelling, and sentence construction (students make sentences from a pool of words). The online assessments are adaptive inasmuch as subsequent items are automatically selected based on whether students' previous responses were correct or incorrect and the difficulty level of the item. Students take these assessments independently on either laptop or desktop computers with a mouse and headphones. Each assessment provides scores as Age Equivalents, Grade Equivalents, and Developmental Scaled Scores. Prior research has shown that the assessments are psychometrically strong and are highly correlated with other standardized measures of language and reading.[2]

Lastly, as the confirmatory outcome of the study, the team collected third grade students' scores on state or district standardized reading tests in the spring of Year 3. This measure captures students' general reading skills at the end of third grade and is of policy relevance to district and school decision makers. Due to COVID-19-related disruptions, not all study

---

1. Johnson (2005).
2. Connor et al. (2017).

districts took the state standardized tests. Instead, these districts used district-wide benchmark tests to measure students' reading levels.[3] To pool the findings across districts, all test scores were standardized within the district using the within-sample district means and standard deviations. The study also assessed the differential impact of the two professional development (PD) models on the percentage of students who performed at or above proficiency level in these tests.

II. DETAILED DIFFERENTIAL IMPACT FINDINGS

The following tables provide detailed findings for the differential impacts of the two PD delivery models on students' reading levels. Appendix Table D.1 presents such findings for the primary cohort of students who started the program as first-graders in the fall of Year 1. Appendix Table D.2 presents differential impact findings for a subset of primary cohort students who stayed with the program for the pre-COVID-19 period. This sample includes 50 schools that stayed in the study for Year 1 and Year 2. Appendix Table D.3 shows the differential impact findings for primary cohort students from the 40 schools that stayed with the study for all three years. Lastly, Appendix Table D.4 shows the differential impacts between the two PD models for students from all grade levels that were ever exposed to A2i. Note that across all these tables, the estimated differential impacts between the two PD delivery models are small in magnitude, ranging from 0 to 0.08 standard deviations in absolute value in effect size, and none of them are statistically significant at the 0.05 level.

---

3.   Such tests include the STAR reading assessment (https://www.renaissance.com/products/star-assessments/), the MAP Reading test (https://www.nwea.org/map-growth/), and the iReady Diagnostic assessment for reading (https://i-readycentral.com/familycenter/what-is-i-ready/).

Appendix Table D.1. Estimated Differential Impacts for the Primary Cohort of Students

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Impact		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Impact	Estimated Impact in Effect Size	P-Value
State/district ELA test (collected for Year 3)											
Standardized score	28	1,607	0.04	1.13	27	1,525	0.01	1.00	0.03	0.03	0.530
Percentage scored at or above proficiency	28	1,607	50.8	50.0	27	1,525	50.5	50.0	0.2	0.00	0.931
A2i assessment tests (scaled scores)											
Letters2Meaning											
Year 1	30	2,182	525.63	90.73	29	2,150	528.04	90.97	-2.41	-0.03	0.673
Reading at grade level (%)	30	2,182	55.7	49.8	29	2,150	55.5	49.8	0.2	0.00	0.951
Year 2	25	1,562	584.02	64.72	25	1,613	583.41	69.49	0.60	0.01	0.876
Reading at grade level (%)	25	1,562	69.2	46.7	25	1,613	67.6	47.5	1.6	0.03	0.467
Year 3	20	1,133	631.48	73.29	20	1,075	627.70	80.38	3.78	0.05	0.386
Reading at grade level (%)	20	1,133	50.2	50.0	20	1,075	49.8	50.0	0.4	0.01	0.874
Word Match Game											
Year 1	30	2,213	480.04	13.51	29	2,149	480.42	13.79	-0.39	-0.03	0.637
Reading at grade level (%)	30	2,213	48.3	49.9	29	2,149	50.4	50.0	-2.1	-0.04	0.287
Year 2	25	1,564	486.79	11.66	25	1,613	486.55	12.09	0.24	0.02	0.724
Reading at grade level (%)	25	1,564	51.3	50.0	25	1,613	51.2	50.0	0.1	0.00	0.967
Year 3	20	1,133	492.46	11.26	20	1,073	492.63	11.84	-0.17	-0.01	0.732
Reading at grade level (%)	20	1,133	41.5	49.0	20	1,073	40.5	49.0	1.0	0.02	0.666

(continued)

Appendix Table D.1 (continued)

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Impact		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Impact	Estimated Impact in Effect Size	P-Value
GMRT (scaled score, collected for Year 1)											
Total											
Scaled score	29	1,888	394.73	44.58	29	1,949	394.77	45.17	-0.03	0.00	0.988
At grade level (%)	29	1,888	53.0	50.0	29	1,949	52.3	50.0	0.7	0.01	0.731
Comprehension											
Scaled score	29	1,900	390.18	42.69	29	1,960	392.47	44.22	-2.29	-0.05	0.390
At grade level (%)	29	1,888	49.1	50.0	29	1,949	51.4	50.0	-2.3	-0.05	0.401
Vocabulary											
Scaled score	29	1,908	406.74	55.98	29	1,962	404.83	54.78	1.91	0.03	0.511
At grade level (%)	29	1,920	61.3	49.1	29	1,973	57.7	49.5	3.6	0.07	0.126

SOURCES: The Gates-MacGinitie Reading Test (GMRT) data collected in spring 2019. A2i assessment data collected by the A2i platform from summer 2018 to spring 2021. District records data for the 2017–2018 to 2020–2021 school years.

NOTES: The analysis sample for each outcome includes all students who were in Grade 1 in the study schools at the start of Year 1 and had valid outcomes. Sample size varies by outcome. ELA = English Language Arts.

The differential impacts are estimated using two-level hierarchical linear models to account for the nested structure of the data, with students nested within schools. The models control for the blocking of random assignment and for the following baseline variables: grade, age, race and ethnicity, district-provided poverty indicator, English learner status, Individualized Education Plan status, and baseline A2i assessment scores. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model. None of the differences is statistically significant at the 0.05 level.

Appendix Table D.2. Estimated Differential Impact for the Primary Cohort of Students in a Consistent Sample as of Spring Year 2

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Impact		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Impact	Estimated Impact in Effect Size	P-Value
State/district ELA test (collected for year 3)											
Standardized score	25	1,445	0.04	1.10	25	1,440	0.01	1.00	0.04	0.04	0.523
Percentage scored at or above proficiency	25	1,445	51.4	50.0	25	1,440	50.5	50.0	0.9	0.02	0.753
A2i assessment tests (scaled scores)											
Letters2Meaning											
Year 1	25	1,819	530.52	88.48	25	1,891	529.28	92.79	1.24	0.01	0.796
Reading at grade level (%)	25	1,819	57.4	49.7	25	1,891	56.8	49.8	0.6	0.01	0.844
Year 2	25	1,602	583.30	64.49	25	1,637	582.47	69.69	0.83	0.01	0.830
Reading at grade level (%)	25	1,602	68.9	46.7	25	1,637	67.2	47.5	1.7	0.04	0.452
Year 3	20	1,160	629.98	73.27	20	1,109	626.09	80.99	3.89	0.05	0.385
Reading at grade level (%)	20	1,160	50.0	50.0	20	1,109	49.9	50.0	0.2	0.00	0.951
Word Match Game											
Year 1	25	1,821	480.73	13.17	25	1,891	480.70	13.95	0.03	0.00	0.966
Reading at grade level (%)	25	1,821	49.8	50.0	25	1,891	51.1	50.0	-1.4	-0.03	0.462
Year 2	25	1,605	486.64	11.73	25	1,637	486.45	12.11	0.20	0.02	0.786
Reading at grade level (%)	25	1,605	51.2	50.0	25	1,637	51.3	50.0	-0.1	0.00	0.965
Year 3	20	1,161	492.20	11.30	20	1,106	492.38	11.84	-0.18	-0.01	0.741
Reading at grade level (%)	20	1,161	41.2	49.0	20	1,106	40.3	49.0	0.9	0.02	0.705

(continued)

Appendix Table D.2 (continued)

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Impact		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Impact	Estimated Impact in Effect Size	P-Value
GMRT (scaled score, collected for year 1)											
Total											
Scaled score	25	1,624	396.21	44.80	25	1,715	396.10	45.27	0.11	0.00	0.965
At grade level (%)	25	1,624	54.4	50.0	25	1,715	54.1	50.0	0.4	0.01	0.875
Comprehension											
Scaled score	25	1,632	391.68	42.67	25	1,724	393.57	44.45	-1.89	-0.04	0.512
At grade level (%)	25	1,624	50.2	50.0	25	1,715	52.5	50.0	-2.2	-0.04	0.449
Vocabulary											
Scaled score	25	1,642	408.18	56.05	25	1,726	406.45	55.06	1.73	0.03	0.581
At grade level (%)	25	1,650	62.5	49.0	25	1,735	59.4	49.5	3.2	0.06	0.199

SOURCES: The Gates-MacGinitie Reading Test (GMRT) data collected in spring 2019. A2i assessment data collected by the A2i platform from Summer 2018 to spring 2021. District records data for the 2017-2018 to 2020-2021 school years.

NOTES: The analysis sample for each outcome includes students who were enrolled in the 50 schools that stayed in the study for Year 1 and Year 2, were in Grade one at the start of Year 1 and had valid outcome.
  The differential impacts are estimated using two-level hierarchical linear models to account for the nested structure of the data, with students nested within schools. The models control for the blocking of random assignment and for the following baseline variables: grade, age, race and ethnicity, district-provided poverty indicator, English learner status, Individualized Education Plan status, and baseline A2i assessment scores. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model. None of the differences are statistically significant at the 0.05 level.

Appendix Table D.3. Estimated Differential Impacts for the Primary Cohort of Students in a Consistent Sample, as of Spring Year 3

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Impact		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Impact	Estimated Impact in Effect Size	P-Value
State/district ELA test (collected for year 3)											
Standardized score	20	1,025	0.07	1.08	20	974	0.06	0.96	0.01	0.01	0.915
Percentage scored at or above proficiency	20	1,025	48.2	49.9	20	974	50.0	50.0	-1.9	-0.04	0.606
A2i assessment tests (scaled scores)											
Letters2Meaning											
Year 1	20	1,025	536.69	81.64	20	974	540.30	82.60	-3.61	-0.04	0.497
Reading at grade level (%)	20	1,025	59.8	49.5	20	974	61.0	49.1	-1.2	-0.03	0.747
Year 2	20	1,025	586.14	59.57	20	974	587.32	63.43	-1.18	-0.02	0.775
Reading at grade level (%)	20	1,025	69.8	46.7	20	974	70.4	46.4	-0.6	-0.01	0.813
Year 3	20	1,025	632.48	72.75	20	974	631.85	78.04	0.63	0.01	0.888
Reading at grade level (%)	20	1,025	51.3	50.0	20	974	51.0	50.0	0.3	0.01	0.926
Word Match Game											
Year 1	20	1,025	481.21	12.70	20	974	481.63	13.63	-0.42	-0.03	0.550
Reading at grade level (%)	20	1,025	49.79	50.00	20	974	53.16	49.97	-3.37	-0.07	0.108
Year 2	20	1,025	486.54	11.61	20	974	486.71	11.65	-0.17	-0.01	0.853
Reading at grade level (%)	20	1,025	50.7	50.0	20	974	51.6	50.0	-0.9	-0.02	0.774
Year 3	20	1,025	492.81	11.24	20	974	492.84	11.53	-0.04	0.00	0.948
Reading at grade level (%)	20	1,025	42.4	49.1	20	974	41.3	49.0	1.1	0.02	0.639

(continued)

Appendix Table D.3 (continued)

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Impact		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Impact	Estimated Impact in Effect Size	P-Value
GMRT (scaled score, collected for year 1)											
Total											
Scaled score	20	1,012	395.17	43.13	20	963	396.72	45.27	-1.55	-0.03	0.612
At grade level (%)	20	1,012	55.8	50.0	20	963	54.5	50.0	1.3	0.03	0.652
Comprehension											
Scaled score	20	1,018	390.96	40.76	20	968	394.55	44.50	-3.59	-0.08	0.276
At grade level (%)	20	1,012	51.3	50.0	20	963	53.0	50.0	-1.7	-0.03	0.666
Vocabulary											
Scaled score	20	1,019	406.93	55.15	20	969	406.97	54.24	-0.04	0.00	0.992
At grade level (%)	20	1,025	63.9	49.1	20	974	60.0	49.4	3.9	0.08	0.266

SOURCES: The Gates-MacGinitie Reading Test (GMRT) data collected in spring 2019. A2i assessment data collected by the A2i platform from summer 2018 to spring 2021. District records data for the 2017-2018 to 2020-2021 school years.

NOTES: The analysis sample for each outcome includes students who were enrolled in the 40 schools that stayed in the study for all three years, were in Grade 1 at the start of Year 1 and had valid outcomes. ELA = English Language Arts.

   The differential impacts are estimated using two-level hierarchical linear models to account for the nested structure of the data, with students nested within schools. The models control for the blocking of random assignment and for the following baseline variables: grade, age, race and ethnicity, district-provided poverty indicator, English learner status, Individualized Education Plan status, and baseline A2i assessment scores. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model. None of the differences is statistically significant at the 0.05 level.

Appendix Table D.4. Estimated Differential Impacts for All Students

Outcome	Mixed-Mode Schools				Face-to-Face Schools				Estimated Impact		
	Number of Schools	Number of Students	Mean	Standard Deviation	Number of Schools	Number of Students	Model-Adjusted Mean	Standard Deviation	Estimated Impact	Estimated Impact in Effect Size	P-Value
A2i assessment tests (scaled scores)											
Letters2Meaning											
Year 1	30	4,524	492.17	90.97	29	4,693	492.34	89.24	-0.17	0.00	0.964
Reading at grade level (%)	30	4,524	52.9	50.0	29	4,693	52.6	50.0	0.3	0.01	0.909
Year 2	25	5,227	506.37	101.28	25	5,594	504.78	102.55	1.59	0.02	0.681
Reading at grade level (%)	25	5,227	56.5	49.7	25	5,594	54.7	49.9	1.8	0.04	0.388
Year 3	20	4,577	563.40	103.28	20	5,000	564.13	106.18	-0.73	-0.01	0.872
Reading at grade level (%)	20	4,577	51.8	50.0	20	5,000	51.8	50.0	0.0	0.00	0.994
Word Match Game											
Year 1	30	4,560	474.82	14.32	29	4,692	474.75	14.64	0.07	0.01	0.910
Reading at grade level (%)	30	4,560	42.1	49.2	29	4,692	42.8	49.3	-0.6	-0.01	0.725
Year 2	25	5,231	476.32	15.39	25	5,601	476.14	15.61	0.18	0.01	0.762
Reading at grade level (%)	25	5,231	40.9	48.8	25	5,601	40.5	48.9	0.5	0.01	0.804
Year 3	20	4,576	484.52	14.83	20	4,998	484.85	15.23	-0.32	-0.02	0.535
Reading at grade level (%)	20	4,576	44.8	49.6	20	4,998	44.7	49.8	0.1	0.00	0.929

SOURCES: A2i assessment data collected by the A2i platform from summer 2018 to spring 2021. District records data for the 2017–2018 to 2020–2021 school years.

NOTES: The Year 1 sample includes all Kindergarten and Grade 1 students with valid A2i assessment scores. The Year 2 sample includes all Grade K-2 students with valid A2i assessment scores. The Year 3 sample includes all Grade K-3 students with valid A2i assessment scores.

The differential impacts are estimated using two-level hierarchical linear models to account for the nested structure of the data, with students nested within schools. The models control for the blocking of random assignment and for the following baseline variables: grade, age, race and ethnicity, district-provided poverty indicator, English learner status, Individualized Education Plan status, and baseline A2i assessment scores. All missing values in these covariates are imputed with zero and missing indicators for all covariates are also included in the model.

None of the differences is statistically significant at the 0.05 level.

Cost Analysis

This appendix provides details of the sources, assumptions, and approaches the team used to calculate the fixed and variable costs per student for these two models. It also discusses the limitations with the cost analysis approach.

I. CALCULATING PROGRAM COST PER STUDENT

The total cost per student was obtained through a series of calculations. First, the team determined the number of students who received the mixed-mode and face-to-face models of A2i, as well as those attending expansion schools. The total number of students was obtained through the annual financial report for the grant and the number of students in the study schools was obtained from A2i system data and student records data that was collected for this evaluation. The number of students attending expansion schools is the difference between the total number of students and the number of students in the study.

Next, the team determined the total cost, fixed cost, and variable cost. The costs were obtained through the annual report and from Learning Ovations. The total cost considered was total grant dollars that Learning Ovations received for program development and implementation. Learning Ovations then provided the total cost of Learning Outcomes Specialist (LOS) salaries, as well as an estimate of the cost of travel per in-person LOS school visit. LOS salaries and travel costs combined are the variable costs; in other words, this is where resources are expected to be used differently between the mixed-mode and face-to-face models. The remaining Learning Ovations costs are considered fixed costs; in other words, these funds are assumed to have been used evenly across schools regardless of professional development (PD) delivery model.

The price per student for the fixed cost was determined by dividing the total fixed cost by the number of students by model. The price per student for the variable cost considered the different amounts of time that the LOSs spent with and traveled to each school. First, the total number of minutes spent by the LOS doing PD activities in each model by year was calculated, using data from the evaluation. Next, the total number of minutes the LOSs spent preparing for each school by model was calculated; the LOSs spent more time preparing for in-person events than virtual events. Adding the time an LOS spent preparing for and in PD events gives a total accounting of all LOS time by model. The total LOS salary divided by the hours the LOS spent in each model is the cost of PD by hour. That total plus the travel costs for each type of school provide the full variable costs by model. Dividing the total costs by the number of students by model gives the per-student cost by model.

Finally, the total cost per student is thus the fixed cost per student plus the variable cost per student. This was calculated for each year of the study.

II. LIMITATIONS OF THE APPROACH

While these cost calculations provide an informed estimate of the cost per student to implement A2i, there are several limitations due to the unavailability of data. First, travel costs relied upon estimates. Each trip varied in cost from $5 to $60+ in driving expenses on the expense reports. Trips varied in length and may have been paired with additional associated costs like tolls, multiple school trips (driving between multiple destinations), or covered longer distances due to traffic. Although the research team was not able to trace individual receipts for LOS travel costs and match them to each visit, Learning Ovations helped the team estimate travel costs to be $50 per in-person visit.

Next, the total number of minutes in PD was also an estimate. In order to be consistent in approach across all three years, the team used an estimate of 30 minutes for each Literacy Huddle and 45 minutes for each Individualized Classroom Coaching session. Though these PD events were designed to be this duration, the reality may have varied and that cannot be accounted for in these calculations.

Finally, the preparation time for each type of PD is also based on estimates. The difference in the time LOSs spent in preparation for face-to-face, mixed mode, and expansion schools was provided by Learning Ovations as an average of the total minutes spent by LOSs across schools in each model throughout the school year. Therefore, the average number of minutes spent was used in cost calculations rather than the exact number of minutes each LOS spent with schools in each model.

These estimates all create limitations on the calculations of differentiated costs. However, these numbers do reflect the best cost calculations possible with the available data.

Expansion Efforts

The United2Read project included several strategies to facilitate the national expansion of the A2i program. Throughout the study, Learning Ovations explored the implementation of these strategies, finding varying levels of feasibility.

First, the project sought to improve the cost effectiveness and efficiency of the A2i Professional Development Support System by applying technology to lower the cost of on-site professional development (PD) and the ongoing coaching required to maintain effective implementation. The study found that technology did lower the cost of on-site PD and ongoing coaching. The main text of the report explores in detail the number of PD sessions by delivery model and the differential cost. While there were differences in the PD experiences and costs by model, the differences were minimized in the second half of Year 2 and in Year 3 due to the effects of COVID-19, as virtual interaction became the standard for everyone. Per the Digital Promise Global qualitative work, it became hard to reach teachers because there were so many demands on them during the transition to virtual teaching. As a way to be sensitive to teachers' situations, mandatory A2i Literacy Huddles and Individualized Classroom Coaching sessions varied by school, largely becoming optional. Small-group instruction was also hard to implement during this time, because of the difficulties surrounding remote instruction.

Next, the project explored using technology and IBM Watson cognitive computing to leverage existing district investments in curriculum resources. The use of IBM Watson to index district curricula in order to align A2i lesson plan suggestions was explored in Year 1 but was deemed not necessary and Learning Ovations stopped pursuing this activity. As of January 2018, it was clear that Watson/artificial intelligence (AI) methods were not going to be able to perform the necessary function; investigation into the use of IBM's Watson technology for indexing revealed that the indexing process was too complex for AI. The indexing of curricula was also not the bottleneck for Learning Ovations that had been originally predicted, and thus was no longer a barrier to scale. In addition, Learning Ovations had become much more efficient in this process and had expanded the indexing team to a point where, even if Watson worked, it might have been too costly to pursue further.

Finally, Learning Ovations adjusted their plan to apply Education, Innovation, and Research (EIR) grant funding to scale to multiple regions across the country. It decided to move away from plans for physical centers, and instead to being a virtual organization with regional staff members hired to support the four states locally. It currently has staff members located on both the East and West coasts. The Literacy Outcome Specialists (LOSs) on both coasts are assigned to schools in their respective areas, which facilitated travel before the pandemic and allowed them to work during the same school hours. The cost of many physical regional hubs was not feasible, or necessary, since the LOSs spent significant time travelling. In hindsight, it was fortuitous that Learning Ovations did not have to carry facilities costs during the pandemic. Functionally, they have a staffing model tailored to each region, with staff members that have regional knowledge and ease of access, just not a colocation facility.

REFERENCES

Bloom, Howard S., Carolyn J. Hill, Alison Rebeck Black, and Mark W. Lipsey. 2008. "Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions." *Journal of Research on Educational Effectiveness* 1, 4: 289–328. Website: doi.org/10.1080/19345740802400072.

Brighton, Catherine M., Holly L. Hertberg, Tonya R. Moon, Carol A. Tomlinson, and Carolyn M. Callahan. 2005. *The feasibility of high-end learning in a diverse middle school* (RM05210). Storrs: University of Connecticut, The National Research Center on the Gifted and Talented.

Carr, Peggy G. 2022. "Commissioner's Statement on the Upcoming Release of NAEP Long-Term Trend Assessment Results." What's New (blog), August. Website: nces.ed.gov/whatsnew/commissioner/remarks2022/8_31_2022.asp.

Connor, Carol McDonald, Lara J. Jakobsons, Elizabeth C. Crowe, and Jane Granger Meadows. 2009a. "Instruction, and Student Engagement in Reading First Classrooms." *Elementary School Journal* 109, 3: 221–250.

Connor, Carol McDonald, Frederick J. Morrison, Barry Fishman, Elizabeth C. Crowe, Stephanie Al Otaiba, and Christopher Schatschneider. 2013. "A Longitudinal Cluster-Randomized Controlled Study on the Accumulating Effects of Individualized Literacy Instruction on Students' Reading From First Through Third Grade." *Psychological Science* 24, 8:1,408–1,419.

Connor, Carol McDonald, Fredrick J. Morrison, Barry Fishman, Sarah Giuliani, Melissa Luck, Phyllis S. Underwood, Aysegul Bayraktar, Elizabeth C. Crowe, and Christopher Schatschneider. 2011a. "Testing the Impact of Child Characteristics × Instruction Interactions on Third Graders' Reading Comprehension by Differentiating Literacy Instruction." *Reading Research Quarterly* 46, 3: 189–221.

Connor, Carol McDonald, Frederick J. Morrison, Barry J. Fishman, Claire Cameron Ponitz, Stephanie Glasney, Phyllis S. Underwood, Shayne B. Piasta, Elizabeth Coyne Crowe, and Christopher Schatschneider. 2009b. "The ISI Classroom Observation System: Examining the Literacy Instruction Provided to Individual Students." *Educational Researcher* 38, 2: 85–99.

Connor, Carol McDonald, Frederick J. Morrison, Barry J. Fishman, Christopher Schatschneider, and Phyllis S. Underwood. 2007. "Algorithm-Guided Individualized Reading Instruction." *Science* 315, 5,811: 464–465.

Connor, Carol McDonald, Frederick. J. Morrison, Leslie E. Katch. 2004. "Beyond the reading wars: Exploring the effect of child-instruction interactions on growth in early reading." *Scientific Studies of Reading* 8, 305–336.

Connor, Carol McDonald, Frederick J. Morrison, Jocelyn N. Petrella. 2004. "Effective Reading Comprehension Instruction: Examining Child x Instruction Interactions." *Journal of Educational Psychology* 96, 682–698.

Connor, Carol McDonald, Frederick J. Morrison, Christopher Schatschneider, Jessica Toste, Erin Lundblom, Elizabeth C. Crowe, and Barry Fishman. 2011b. "Effective Classroom Instruction: Implications of Child Characteristics by Reading Instruction Interactions on First Graders' Word Reading Achievement." *Journal of Research on Educational Effectiveness* 4, 3: 173–207.

Connor, Carol McDonald, Frederick J. Morrison, and Phyllis S. Underwood. 2007. "A Second Chance in Second Grade: The Independent and Cumulative Impact of First- and Second-Grade Reading Instruction and Students' Letter-Word Reading Skill Growth." *Scientific Studies of Reading* 11, 3: 199–233.

REFERENCES *(CONTINUED)*

Connor, Carol McDonald, Karen Taylor, Taffeta Wood, and Sarah Siegal. 2017. "Using Fidelity Data to Elucidate the Results of a Design Study on Assessment-2-Instruction Technology." Paper presented at the Annual Conference of the Society for Research in Education Effectiveness, Washington, DC, March 3.

Curriculum Associates. 2021. *Academic Achievement at the End of the 2020–2021 School Year*. North Billerica, Massachusetts: Curriculum Associates.

Johnson, Kathleen M. 2005. "Test Review of Gates-MacGinitie Reading Tests, Fourth Edition, Forms S and T." In Robert A. Spies and Barbara S. Plake (eds.), *The Sixteenth Mental Measurements Yearbook*. Lincoln, Nebraska: University of Nebraska Press.

Kasad, Zareen, Mai Chou Vang, and Viki M. Young. 2021. "United2Read Implementation Study 2020-21 Findings Brief." Unpublished Paper. Washington, DC: Digital Promise.

Kuhfeld, Megan, James Soland, Karyn Lewis, Erik Ruzek, and Angela Johnson. 2022. "The COVID-19 school year: Learning and recovery across 2020-2021." Aera Open 8: https://doi.org/10.1177/23328584221099306

Lewis, Karyn, Megan Kuhfeld, Erik Ruzek, and Andrew McEachin. 2021. "Learning during COVID-19: Reading and math achievement in the 2020–2021 school year." Portland, Oregon: NWEA. Website: www.nwea.org/research/publication/learning-during-covid-19-reading-and-math-achievement-in-the-2020-2021-school-year/.

National Assessment of Educational Progress. 2019. "NAEP Report Card: Reading." Washington, DC: U.S. Department of Education. Website: www.nationsreportcard.gov/reading/nation/achievement?grade=4.

National Assessment of Educational Progress. 2022. "NAEP Long-Term Trend Assessment Results: Reading and Mathematics." Washington, DC: U.S. Department of Education. Website: www.nationsreportcard.gov/highlights/ltt/2022/.

National Center for Education Evaluation. 2022. *What Works Clearinghouse Procedures and Standards Handbook, Version 5.0*. Washington, DC: U.S. Department of Education.

National Institute of Child Health and Human Development. 2000. *Teaching Children to Read: An Evidence-Based Assessment of the Scientific Research Literature on Reading and its Implications for Reading Instruction*. Bethesda, MD: U.S. Department of Health and Human Services.

Puma, Michael J., Robert B. Olsen, Stephen H. Bell, and Cristofer Price. 2009. "What to Do When Data are Missing in Group Randomized Controlled Trials. NCEE 2009-0049." National Center for Education Evaluation and Regional Assistance.

Renaissance. 2021. "How Kids Are Performing – Spring 2021." Wisconsin Rapids, Wisconsin: Renaissance. Website: doc.renlearn.com/KMNet/R63444.pdf.

Reynolds, Arthur J., and Suh-Ruu Ou. 2004. "Alterable Predictors of Child Well-Being in the Chicago Longitudinal Study." *Children and Youth Services Review* 26, 1: 1–14.

Seeking Alpha. 2022. "Scholastic Acquires Learning Ovations, Creator of A2i Literacy Assessment and Instructional System." Website: seekingalpha.com/pr/18927065-scholastic-acquires-learning-ovations-creator-of-a2i-literacy-assessment-and-instructional.

Tomlinson, Carol Ann. 1995. "Deciding to Differentiate Instruction in Middle School: One School's Journey." *Gifted Child Quarterly* 39, 2: 77–87.

Tomlinson, Carol Ann, Catherine Brighton, Holly Hertberg, Carolyn M. Callahan, Tonya R. Moon, Kay Brimijoin, Lynda A. Conover, and Timothy Reynolds. 2003. "Differentiating Instruction in Response to Student Readiness, Interest, and Learning Profile in Academically Diverse Classrooms: A Review of Literature." *Journal for the Education of the Gifted* 27, 2–3: 119–145.

REFERENCES *(CONTINUED)*

Young, Viki M., Zareen Kasad, and Mai Chou Vang. 2019. "United2Read Implementation Study First-Year Findings Brief, 2018-19." Unpublished Paper. Washington, DC: Digital Promise.

Young, Viki M., Mai Chou Vang, and Zareen Kasad. 2020. "United2Read Implementation Study 2019-2020 Findings Brief." Unpublished Paper. Washington, DC: Digital Promise.

YouthTruth. 2021. "Students Weigh In, Part II: Learning and Well-Being During COVID-19." San Francisco: YouthTruth. Website: youthtruthsurvey.org/students-weigh-in-part2/.