AIR®
AMERICAN INSTITUTES FOR RESEARCH®

Study of ATLAS Use by Preservice and Early Career Teachers

SEPTEMBER 2018

David Manzeske | Kelly Reese | Feng Liu

MAKING RESEARCH **RELEVANT**

AMERICAN INSTITUTES FOR RESEARCH | AIR.ORG

Study of ATLAS Use by Preservice and Early Career Teachers

Two-Year Implementation and Effects on Teacher and Student Outcomes

SEPTEMBER 2018

David Manzeske | Kelly Reese | Feng Liu

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

www.air.org

5917_09/18

Acknowledgments

American Institutes for Research would like to thank those who gave their time and support for this study. In particular, we would like to thank the preservice and early career teachers and ATLAS facilitators who provided data; administrators from participating sites who facilitated ATLAS implementation and fulfilled archival data requests; and the National Board for Professional Teaching Standards, who coordinated efforts among partners and without whom this research would not have been possible.

Funding disclosure

This work was funded by the National Board for Professional Teaching Standards (NBPTS) through a 2012 U.S. Department of Education Investing in Innovation Fund grant (U411P120508) to NBPTS.

Abstract

In 2013, the National Board for Professional Teaching Standards received a 5-year Investing in Innovation Fund Development grant from the U.S. Department of Education to develop, implement, and study Accomplished Teaching, Learning, and Schools (ATLAS). ATLAS is an online case library that contains examples of "accomplished teaching" practice delivered by National Board Certified Teachers. A purpose of the grant activities was to expose preservice and early career teachers to ATLAS content, which was hypothesized to affect their teaching practice and the achievements of their students. This final report summarizes research on the first 2 years of the program's implementation at scale (during the grant's fourth and fifth years), and the effects of ATLAS use on the outcomes of preservice and early career teachers and on the mathematics and science achievements of students in Grades 3–6. The study team determined that ATLAS was implemented with fidelity at the institution of higher education and local education agency levels during the second study year but not the first study year. Regarding the effects of ATLAS use on the outcomes of preservice teachers, early career teachers, and students, the study did not identify any observable differences between ATLAS users and non-ATLAS users or their students.

Contents

Tables

Page

Exhibits

Page

Summary

In 2013, the National Board for Professional Teaching Standards received a 5-year, $3,000,000 Investing in Innovation Fund (i3) Development grant from the U.S. Department of Education Office of Innovation and Improvement to develop, implement, and study ATLAS: *Accomplished Teaching, Learning, and Schools*. ATLAS is an online case library that contains examples of "accomplished teaching" practice delivered by National Board Certified Teachers. The American Institutes for Research (AIR) is the external research organization that was contracted to study the i3 Development grant. This final report summarizes AIR's research on the program's first 2 years of implementation at scale (during the grant's fourth and fifth years), and its effects on preservice and early career teachers' outcomes as well as their students' achievement in mathematics and science.

About the Study

The AIR study team conducted a summative, external study of ATLAS implementation during the 2015–16 and 2016–17 school years. During this period, ATLAS was simultaneously implemented in two settings. In institutions of higher education, undergraduate students (preservice teachers) who were training to become classroom teachers of mathematics and science in Grades 3 through 6 used ATLAS with their teacher preparation course instructors. In local education agencies, first- and second-year teachers (early career teachers) of mathematics and science in Grades 3 through 6 used ATLAS as part of their job-embedded induction experience with mentor teaching staff.

The purpose of ATLAS use was to expose preservice and early career teachers to examples of "accomplished teaching" that would affect their teaching practice and their students' achievement in mathematics and science in Grades 3 through 6. The study team examined the extent to which ATLAS was implemented with fidelity, and it addressed research questions for each of three populations. For preservice teachers, the study team examined whether 2 years of ATLAS use had an effect on self-reported perceptions of preparedness, self-efficacy, and self-reflection and on a standardized measure of preservice teacher performance relative to other preservice teachers who did not use ATLAS. For early career teachers, the study team examined whether up to 2 years of ATLAS use had an effect on self-reported perceptions of preparedness, self-efficacy, and self-reflection relative to other early career teachers who did not use ATLAS. And for students of teachers who had used ATLAS for up to 2 years, the study team examined whether they performed better on mathematics and science state achievement tests relative to students of teachers who did not use ATLAS.

Summary of Findings

The study team generated findings related to implementation fidelity and examined the effects of ATLAS use on preservice teacher, early career teacher, and student outcomes. ATLAS was implemented with fidelity at the institution of higher education (IHE) and local education agency (LEA) levels during the second study year but not the first study year. Regarding the effects of ATLAS use on preservice teacher, early career teacher, and student outcomes, the study did not identify any observable differences between ATLAS users and non-ATLAS users. Moreover, the study team conducted exploratory analyses to examine the association between the amount of ATLAS use and preservice and early career outcomes; however, the study team did not identify any differences in measured outcomes among users who used ATLAS at different levels, nor relative to non-ATLAS users. Perhaps most important at this phase of the work, the National Board and its 13 implementation partners demonstrated—at scale—that high implementation fidelity of the ATLAS resource is possible.

Introduction

This report contains four sections. The first section (*ATLAS Overview*) contains information about the resource that is the focus of this work—ATLAS—and other contextual information regarding the grant-supported activities and ATLAS use decisions made by the grant partners that are relevant to the study. The second section (*Study Design and Methods*) contains information about how the study was conducted and related information regarding data collection and analysis. The third section (*Study Findings*) contains findings regarding ATLAS implementation fidelity and outcomes for preservice teachers, early career teachers, and students. The fourth and final section (*Discussion*) contains concluding remarks such as information that must be taken into consideration when interpreting the findings and considerations for future research and use of ATLAS.

ATLAS Overview

The **National Board for Professional Teaching Standards (National Board)** is a nonprofit organization that has established and periodically refines peer-reviewed standards defining what the National Board refers to as "accomplished teaching" practice. The standards exist for 16 content areas[1] and five developmental levels,[2] and the organization maintains an assessment process to certify teachers who have met those standards to receive the National Board Certified Teacher (NBCT) credential. To receive the NBCT credential, teachers complete four separate components that require them to demonstrate, analyze, and reflect upon their teaching performance as captured on video, in written analyses, in student work samples, and through standardized assessments of content knowledge.

In the ***Accomplished Teaching, Learning, and Schools*** **(ATLAS)** online resource, the National Board makes available in a digital case library a portion of the materials submitted during the NBCT certification process for those teachers who received the NBCT credential.[3] The available materials exist in the form of cases. Each case represents a single NBCT's classroom video and audio recording of instruction and written, reflective analyses referred to as commentary. Cases

[1] Art, Career and Technical Education, English as a New Language, English Language Arts, Exceptional Needs Specialist, Generalist, Health Education, Library Media, Literacy: Reading-Language Arts, Mathematics, Music, Physical Education, School Counseling, Science, Social Studies-History, and World Languages. For subcontent areas, see https://www.nbpts.org/wp-content/uploads/Certification-Areas-1.pdf

[2] Early Childhood, Middle Childhood, Early Adolescence, Adolescence, and Young Adulthood

[3] See https://www.nbpts.org/atlas/

are categorized, or "tagged," as defined by the National Board's standards as well as by various external standards and frameworks used in the teaching profession to define characteristics of practice.

In 2013, the National Board received a 5-year, $3 million Investing in Innovation Fund (i3) Development grant from the U.S. Department of Education Office of Innovation and Improvement to develop, implement, and study ATLAS.

As a part of ATLAS development and implementation, the National Board worked with eight local education agencies (LEAs) and five institutions of higher education (IHEs) in New York, Tennessee, and Washington State, as listed in Table 1. These LEA and IHE partners helped to collaboratively design the features of the ATLAS platform and create a facilitation process for use in teacher preparation courses and job-embedded teacher induction experiences.

Table 1. ATLAS Development and Evaluation Partners

State	Institutions of higher education	Local education agencies
New York	Niagara University	Niagara Falls City School District
Tennessee	Tennessee State University University of Memphis Vanderbilt University	Dyersburg City Schools Jackson-Madison County School System Metropolitan Nashville Public Schools Tipton County Public Schools
Washington State	Central Washington University	Educational Service District 105 Seattle Public Schools West Valley School District

As part of ATLAS's development and implementation and of the eventual research on grant activities, faculty and their students in teacher preparation programs within the partner institutions accessed the online ATLAS platform and its digital cases to supplement their course curriculum. Preservice teachers used ATLAS in a variety of course types and field experiences with various instructional focuses, such as content methods (i.e., methods for how to teach mathematics and science), assessment, learning environments, and responsiveness to students.

Under the supervision of their facilitating mentor, coach, or instructional support staff in elementary or middle school induction programs, early career teachers in their first or second year in the profession also accessed ATLAS cases as a part of their professional development sessions. Early career teachers used ATLAS with the guidance of district instructional coaches

and mentors in induction programs. They used ATLAS in ways that were directly aligned with areas in which school districts felt that new teachers needed the most support, as outlined by their district evaluation or performance rubric or framework. This knowledge of needed support was developed through prior, general experience with new teachers in their district or individual assessment of the participating teacher.

ATLAS is designed to be used with teachers of all grades and subjects in K–12 education. For the purposes of ATLAS development, partners from both elementary and secondary teacher preparation and induction programs participated. However, for the purposes of ATLAS implementation and the corresponding study of the implementation, only facilitators and teachers in early education or elementary preparation or in induction programs participated. The focus of the study was on ATLAS implementation fidelity and the effects of ATLAS on preservice and early career teachers of mathematics or science in Grades 3 through 6 and their students.[4]

ATLAS Features

ATLAS cases are available to users through an internet browser interface.[5] Each ATLAS case consists of a single NBCT's classroom video and audio recording (accessible by way of a streaming video) and accompanying written teacher commentary. The commentary explains the lesson being delivered and the teacher's rationale for the techniques, instructional choices, or student interactions being viewed, as well as additional context that may be helpful for understanding the classroom recording. Additional materials that support the instruction viewed in the recording are sometimes provided. The ATLAS platform includes search and tagging functionality that makes cases accessible according to different case characteristics (e.g., content, topic, grade level, relationship to academic standard or teaching and learning framework). Partner IHE and LEA faculty as well as National Board staff helped to determine relationships among prominent teaching frameworks so that users can search and access cases according to their state or local needs.

Those managing their institution's licensed use of the platform (in this case, IHE and LEA facilitators) also had the administrative capability to manage the individual users within their institutions and assign cases to specific users. In some institutions participating in the ATLAS

[4] Whereas the study focused on preservice teachers training to teach mathematics and science and on early career teachers of mathematics or science in Grades 3 through 6 (and their students), facilitators could choose to select ATLAS cases that were not exclusive to mathematics or science instruction. Facilitators could use their discretion to select cases that aligned with various curricular needs.

[5] https://atlas.nbpts.org/

study, only faculty at IHEs or leaders of job-embedded induction activities in school districts had individual accounts and ability to access cases for groups of teachers. In other sites, teacher candidates in preparation programs and teachers in school districts had individual accounts and could access ATLAS independently as a part of the curriculum.

Development Process and Implementation

Throughout ATLAS's development and piloting period (the 2013–14 and 2014–15 school years), the National Board hosted in-person and virtual meetings to support IHE faculty and LEA central office administrators in their testing and piloting of ATLAS. In fall 2013, IHE faculty and LEA administrators were developing protocols for embedding ATLAS cases in their courses and induction programming, respectively, and had started to use ATLAS with a sample of users. By the end of the 2014–15 school year, a group of preservice and early career teachers in every partner institution had been exposed to ATLAS cases and had provided early use, formative feedback.[6] Table 2 describes the development, implementation, and study activities taking place in each school year of the study.

Also, throughout the development and piloting phases, the AIR study team examined how IHE faculty, preservice teachers, district instructional coaches, and early career teachers interacted with the ATLAS platform and collected their feedback about its quality and utility in the process of professional preparation to teach mathematics and science in Grades 3 through 6. The formative feedback was compiled and delivered to the National Board and its development/implementation partners to make enhancements to the ATLAS resource and supporting materials. This process included making decisions regarding the ways that ATLAS use would be facilitated during the final 2 grant years (2015–16 and 2016–17), which is when implementation fidelity and study outcomes were assessed.

[6] Preservice or early career teachers exposed to ATLAS during this time period are not a part of study samples.

Table 2. Description of Partner Involvement in the ATLAS Development, Implementation, and Study Phases

School year	Description of development, implementation, or study phase
2013–14	**Early development:** IHE faculty and LEA administrative leaders experimented with ATLAS features and developed ways to facilitate ATLAS use in preparation for embedding case exposure within teacher preparation courses and school district induction or professional development.
2014–15	**Continued development and piloting:** Partners continued to work together to refine ATLAS facilitation processes in their curricula as well as to test and provide feedback to the National Board on platform functions, such as the logistics of video and commentary access, user management processes, search functionality, and video and commentary quality. ATLAS cases were tested with "live" participants in teacher preparation programs and school districts, and their feedback about quality and utility was gathered by the AIR study team.
2015–16	**Implementation with agreed parameters of use:** IHE and LEA partners agreed to embed ATLAS use in their programs according to specific guidelines, with a minimum number of cases and facilitation cycles used with each teacher. Implementation fidelity and preliminary outcome data were collected by the AIR study team to pilot test data collection instruments.
2016–17	**Continued implementation and outcome data collected:** IHE and LEA partners continued to use ATLAS with preservice and early career teachers. Implementation fidelity and outcome data on preservice and early career teachers were collected by the AIR study team.
2017–18	**Continued ATLAS implementation without study restrictions and without collection of additional outcome data:** Partners continued to use ATLAS in the ways they found to be most beneficial. The AIR study team collected preservice teacher and student outcome data that corresponded to the 2016–17 school year as it became available.

In April 2015, toward the end of the development and piloting phases, all IHE and LEA partners and National Board staff met in person to discuss and agree upon criteria for how ATLAS cases would be embedded within the partners' programs. This decision making was necessary to define implementation fidelity, which would be assessed during the study's 2 school years: 2015–16 and 2016–17. IHE and LEA partners decided on two separate sets of use parameters—one for IHEs and one for LEAs—while still adhering to a facilitation cycle with the same four cycle-of-use components: Motivate, Construct, Apply, and Follow Up (M-C-A-F; see Table 3).

Table 3. Four-Part Cycle of ATLAS Use Facilitation: Motivate, Construct, Apply, and Follow Up

Motivate: This first part of the cycle usually occurs before exposure to the ATLAS case materials. It is meant as an opportunity for the participant to engage in an activity that informs future learning and connects his or her own practice or prior knowledge to the topic.

Construct: This second part of the cycle is when exposure to the ATLAS case takes place. It is designed as an exploration of the performance evaluation framework component, criterion, or indicator as it relates to a particular teacher preparation curriculum or induction topic, and an exploration of the case materials as they relate to the topic.

Apply: This third part of the cycle is a time for the preservice or early career teacher either to plan how he or she will enact the learning from ATLAS exposure in the classroom or to try out the new learning in the classroom. Application can occur independently of the facilitator or group.

Follow Up: This fourth and final part of the cycle is an accountability mechanism to confirm that the preservice or early career teacher carried out the application. It includes reflection time for the teacher and facilitator to discuss the application (e.g., through a face-to-face meeting, e-mail, phone conversation, discussion board, etc.).

LEA partners agreed to the same frequency of ATLAS case use for all early career teachers, irrespective of being a first- or second-year teacher. However, IHE partners agreed to different frequencies of ATLAS case use for preservice teachers in their junior (or penultimate) and senior years of the teacher preparation program. Juniors were expected to use ATLAS more than seniors given time-related considerations while seniors completed student teaching experiences. Table 4 describes the implementation parameters for IHE and LEA partners.

Table 4. Parameters of ATLAS Use in Each School Year by Preservice and Early Career Teachers During the Study's 2 Years (2015–16 and 2016–17)

	IHE		LEA
Parameters	**Juniors (pre-internship / -student teaching year)**	**Seniors (internship / student teaching year)**	**First- and second-year teachers**

	IHE		LEA
Minimum number of unique ATLAS cases viewed/used	3	2	6
Minimum number of facilitated M-C-A-F cycles	6	2	4
Additional facilitation requirements	• Preservice teacher engagement with ATLAS will be anchored in the edTPA rubric(s). Although edTPA will be the primary framework of focus, partners may elect to integrate additional salient frameworks and/or standards along with edTPA in a given interaction. • Use of the instructional materials that accompany the cases is optional. • Number of cycles requiring use of both video (all or in part) and commentary (all or in part) is a minimum of one in the internship year and four in the pre-internship year. • Guided facilitation is required but can be synchronous or asynchronous. Facilitation can be conducted in person or virtually. Teacher candidates cannot function as facilitators. • Partner-created user guides can be used but are not required. • Teacher candidates should have individual access to ATLAS during their internship / student teaching year.		• When using ATLAS, the work of improving the practice of novice teachers must be aligned with the district's expectations for teacher performance (e.g., an indicator or component on an evaluation rubric, performance data). • Use of instructional materials with cases is optional. • Each teacher who uses ATLAS will be exposed to both video and commentary (not necessarily from the same case[s]). The teacher does not have to be exposed to both video and commentary during every occasion of ATLAS use. • Guided facilitation is required. Within an M-C-A-F cycle, there must be at least one in-person interaction with the facilitator. • Facilitation will be guided by a teacher leader, trained in ATLAS facilitation, and who has experience being effective with adult learners and in classrooms.

Collection of Implementation Data

Throughout the implementation period beginning with the 2015–16 academic year, AIR evaluators collected data on ATLAS implementation at the preservice or early career teacher level. AIR created a credential-secured website into which rosters of all the teachers in participating teacher preparation or induction programs were uploaded. Facilitators would then complete a form each time a cycle of ATLAS use was completed from 2015–16 through 2016–17 to report on which parts of the M-C-A-F cycle were completed, whether guided facilitation aligned to edTPA[7] or district rubrics occurred, and which unique ATLAS cases were used. To keep track of whether individual teachers were meeting the minimum parameters of use for the academic year, the website enabled a real-time summary of the unique ATLAS case exposure and M-C-A-F cycle completion at the teacher level for each institution. Because most institutions included multiple facilitators of ATLAS, this feature enabled facilitators to view whether teachers in their programs were on track to meet implementation exposure requirements through the collective facilitation of their colleagues. Implementation was monitored quarterly and reported to the National Board and partners.

Study Design and Methods

This 2-year study, which occurred during the 2015–16 and 2016–17 school years, was conducted in four parts. The first part corresponds to implementation fidelity over 2 years with preservice and early career teachers. The second, third, and fourth parts correspond to program effects on each of the three different study populations: preservice teachers in their final 2 years of their teacher preparation programs, early career teachers in their first and second teaching years, and students of the early career teachers who had been exposed to ALAS for up to 2 years. A description of each part, along with the associated research questions, follows.

Implementation Fidelity

Part 1: A descriptive, implementation study to assess the progress of the National Board and its partners in implementing ATLAS at scale with fidelity according to agreed-upon parameters of ATLAS use with preservice and early career teachers during the 2-year study period.

[7] The edTPA is a nationally recognized measure of teacher performance; see http://www.edtpa.com/PageView.aspx?f=GEN_About.html

1. Is ATLAS being implemented with preservice teachers relative to what is expected?
2. Is ATLAS being implemented with early career teachers relative to what is expected?

Program Effects

Part 2: A post-only quasi-experimental design to determine whether ATLAS participation had an effect on senior *preservice teachers'* perceptions of preparedness, self-efficacy, and self-reflection, as well as performance on an assessment of teaching readiness (the edTPA). Preservice teachers were first exposed to ATLAS as juniors in their preparation program in 2015–16, and the study looks at the effect of ATLAS on their outcomes in relation to their respective comparison group at the end of their senior year (and completion of the program), after 2 years of ATLAS exposure (in spring 2017).

3. What is the average effect of 2 years of ATLAS exposure on preservice teacher perceptions of preparedness?
4. What is the average effect of 2 years of ATLAS exposure on preservice teacher perceptions of self-efficacy?
5. What is the average effect of 2 years of ATLAS exposure on preservice teacher perceptions of self-reflection?
6. What is the average effect of tw2o years of ATLAS exposure on preservice teachers' edTPA performance?

Part 3: A post-only quasi-experimental design to determine whether ATLAS participation had an effect on first- and second-year *early career teachers'* perceptions of preparedness, self-efficacy, and self-reflection. Early career teachers were first exposed to ATLAS during their first teaching year in 2015–16, and the study looks at the effect of ATLAS on their outcomes in relation to their respective comparison group after up to 2 years of ATLAS exposure (in spring 2017).

7. What is the average effect of up to 2 years of ATLAS exposure on perceptions of preparedness of teachers who were within their first 2 years of teaching?
8. What is the average effect of up to 2 years of ATLAS exposure on perceptions of self-efficacy of teachers who were within their first 2 years of teaching?
9. What is the average effect of up to 2 years of ATLAS exposure on perceptions of self-reflection of teachers who were within their first 2 years of teaching?

Part 4: A pre-post quasi-experimental design using a propensity score matching approach at the student level to determine whether up to 2 years of early career teacher ATLAS usage had an effect on their *students'* academic outcomes in Grades 3 through 6 mathematics and science relative to the students' respective comparison group (in spring 2017).

10. What effect did up to 2 years of teachers' ATLAS exposure have on the mathematics achievement of students of teachers who were within their first 2 years of teaching?
11. What effect did up to 2 years of teachers' ATLAS exposure have on the science achievement of students of teachers who were within their first 2 years of teaching?

A description of the study samples and the selection of those samples can be found in Appendix A.

Methods for Examining Implementation Fidelity and Effects on Outcomes

The study team examined implementation fidelity of ATLAS use with preservice and early career teachers and effects of ATLAS use on preservice teacher perceptions and teaching readiness, early career teacher perceptions, and student academic outcomes. The subsequent sections describe the methods used to examine each of these measures by the corresponding research question.

Implementation fidelity. To answer research questions 1 and 2, the study team used implementation data that facilitators submitted each time an ATLAS M-C-A-F cycle was completed during the 2015–16 and 2016–17 school years. In particular, the study team determined whether relevant IHEs and LEAs met the high implementation standards regarding expected levels of facilitator *access* to ATLAS and *use* of ATLAS, for each study year. The high implementation standards were set by the study team in consultation with National Board staff. For both IHEs and LEAs, the high implementation standard for ATLAS access was met if at least 75% of facilitators were registered in the ATLAS system and at least 75% of teachers were exposed to ATLAS.

For IHEs, the high implementation standard for ATLAS use was different for the 2 study years. For the first study year, when preservice teachers were juniors (or in their penultimate year), the high implementation standard was met if, for at least 75% of IHEs, 75% of the preservice teachers were exposed to at least three cases and six M-C-A-F cycles. For the second study year, when preservice teachers were seniors (in the last year of their program), the high implementation standard was met if, for at least 75% of IHEs, 75% of the preservice teachers were exposed to at least two cases and two M-C-A-F cycles.

For LEAs, the high implementation standard for ATLAS use was the same for each of the 2 study years. When early career teachers were in their first and second years of teaching, the high implementation standard was met if, for at least 66% of LEAs, 75% of the early career teachers were exposed to at least six cases and four M-C-A-F cycles.

Effects on preservice and early career teacher perceptions. To answer research questions 3, 4, and 5 (for preservice teachers) and 7, 8, and 9 (for early career teachers), the study team used data from a survey that was administered in spring 2017 by the study team to ATLAS and non-ATLAS users. The outcomes assessed include perceptions on preparedness to teach, self-efficacy to teach, and self-reflection regarding teaching practices. The survey was piloted in spring 2016 and administered in spring 2017 to collect outcome data. The preservice teacher response rate was 33% (of 423 seniors), and the early career teacher response rate was 92% (of 131 first- and second-year teachers). A list of all survey items used to assess study outcomes is in Appendix B (see Exhibits B1, B2, and B3).

To assess preparedness, the study team modified eight items from the 2011–12 Schools and Staffing Survey (U.S. Department of Education National Center for Education Statistics, 2011; e.g., "When you enter your first year of teaching, how well prepared will you be to use a variety of instructional methods?"). The response options included the following: *not at all, somewhat prepared, well prepared,* and *very well prepared.* The observed reliability is reported in Appendix C.

To assess self-efficacy, the study team modified 12 items from the Ohio State teacher efficacy scale (Tschannen-Moran & Hoy, 2001; e.g., "When you enter your first year of teaching, to what extent will you be able to craft good questions for your students?"). The response options included the following: *not at all, minimal extent, moderate extent,* and *great extent.* These 12 items were primarily used to construct an overall self-efficacy measure, but the survey authors constructed the survey to have three factors: efficacy for instructional strategies, efficacy for classroom management, and efficacy for student engagement. The observed reliability is reported in Appendix C.

To assess self-reflection, the study team developed 11 items for the purpose of this study (e.g., "How often are the following statements true about you? I can identify the instructional practices that make me effective at helping students to learn.") The response options included the following: *never or almost never true, sometimes true, usually true,* and *always or almost always true.* The observed reliability is reported in Appendix C.

Additional information regarding the psychometric and analytic approaches is in Appendix C.

Effects on preservice teacher performance on the edTPA. To answer research question 6, the study team collected edTPA assessment data from participating IHEs. The edTPA was taken by ATLAS and non-ATLAS preservice teachers at the end of their graduating year. The edTPA has a standard protocol for administration and scoring, and it is used by each of the teacher preparation programs in the study.[8] The edTPA generates an overall score and three subtask composite scores: *planning, instruction,* and *assessment.* The edTPA was field tested in 2013 by Stanford University for interrater reliability using two kinds of reliability statistics: adjacent agreement rate and kappa-n. The adjacent agreement rate refers to the proportion of cases in which two independent scorers assign either the same score or scores within 1 point of each other. The statistic kappa-n is reported to account for scorer agreement by chance. The edTPA has an average adjacent agreement rate of .92 and an average kappa-n of .83, both of which are relatively high and are comparable to other well-established performance assessments (Stanford Center for Assessment, Learning, and Equity, 2013). The edTPA scores have high internal consistency with Cronbach's alpha ranging from 0.86 to 0.94 with an overall alpha of 0.91.[9] The observed reliability and additional information regarding the analytic approach are in Appendix C.

Effects on student mathematics and science outcomes. To answer research questions 10 and 11, the study team collected state-administered student achievement data from participating LEAs. The data correspond to achievement in Grades 3 through 6 mathematics and science from the 2016–17 school year. Data were collected for students of early career teachers who had been exposed to ATLAS for up to 2 years and matched students of early career teachers who had not been exposed to ATLAS. Additional information regarding the analytic approach is in Appendix C.

Study Findings

Findings are presented for each of the four study parts:

- Part 1 corresponds to implementation fidelity and addresses research questions 1 and 2.
- Part 2 corresponds to preservice teacher perception (2a) and edTPA performance (2b) outcomes and addresses research questions 3 through 6.
- Part 3 corresponds to early career teacher perception outcomes and addresses research questions 7 through 9.
- Part 4 corresponds to student achievement outcomes and addresses research questions 10 and 11.

[8] http://www.edtpa.com/PageView.aspx?f=GEN_About.html
[9] https://secure.aacte.org/apps/rl/res_get.php?fid=3621&ref=rl

Part 1: ATLAS was implemented with fidelity during the second study year

Implementation fidelity was separately assessed for each study year for participating IHEs and LEAs. For the second study year (2016–17), it was determined that IHE and LEA ATLAS users had access to ATLAS and that both preservice and early career teachers used ATLAS in such a way that the high implementation standard was met. However, for the first study year (2015–16), it was determined that IHE and LEA ATLAS users had access to ATLAS but that neither preservice nor early career teachers used ATLAS in such a way that the high implementation standard was met. Detailed tables that contain the Year 1 and Year 2 implementation fidelity findings for IHEs and LEAs are in Appendix D (see Tables D1 through D4).[10]

Part 2a: ATLAS use did not have a statistically significant effect on preservice teacher perceptions

Survey responses from preservice ATLAS and non-ATLAS users were compared to determine whether ATLAS use had an effect on perceptions of preparedness, self-efficacy, and self-reflection. The average differences between ATLAS and non-ATLAS users were small (Table 5), and none of these differences were statistically significant after controlling for preservice teacher and site differences (e.g., ACT score; Table 6).

Table 5. Average Scale Scores for Preservice ATLAS and Non-ATLAS Groups

Average scale scores				
Scale	ATLAS		Non-ATLAS	
	Average	*N*	Average	*N*
Preparedness	2.60	89	2.61	49
Efficacy for instructional strategies	5.04	89	4.98	49
Efficacy for classroom management	5.37	89	5.44	49
Efficacy for student engagement	4.88	89	4.78	49
Efficacy, total	3.56	89	3.73	49
Self-reflection	4.77	77	4.60	41

[10] These tables were created using the reporting template provided by the team that is conducting the national evaluation of the Investing in Innovation Fund program.

Table 6. Tests of Statistical Differences Between Preservice ATLAS and Non-ATLAS Groups

ATLAS versus non-ATLAS				
Scale	**Coefficient estimate**	**Standard error**	***p* value**	***N***
Preparedness	-0.09	0.52	.86	138
Efficacy for instructional strategies	-0.22	0.78	.78	138
Efficacy for classroom management	0.36	0.93	.70	138
Efficacy for student engagement	0.13	0.72	.86	138
Efficacy, total	-0.16	0.57	.78	138
Self-reflection	0.17	0.55	.76	118

Exploratory analyses revealed that there were three distinguishable implementation groups among preservice ATLAS users. These groups were identified based on the average number of ATLAS cases that preservice teachers used and the average number of M-C-A-F cycles that were facilitated (Table 7). On average, preservice teachers who were exposed to higher levels of ATLAS use did not report different levels of the measured outcomes relative to preservice teachers who were not exposed to ATLAS. Figures that illustrate descriptive results for each of the three main scales (preparedness, total self-efficacy, and self-reflection), by the three ATLAS implementation groups and for non-ATLAS users, are in Appendix D (see Figures D1 through D3).

Table 7. Preservice Teacher Implementation Groups

Type of implementation	**Mean no. unique ATLAS cases**	**Mean no. M-C-A-F facilitated cycles**	***N***
Group 1: High cases, high cycles	27.33	11.67	18
Group 2: Low cases, low cycles	3.95	3.91	56
Group 3: High cases, low cycles	18.67	5.47	15

Part 2b: ATLAS use did not have a statistically significant effect on preservice teacher edTPA performance

The edTPA scores from preservice ATLAS and non-ATLAS users were compared to determine whether ATLAS use had an effect on the total edTPA score and each of the three subtask composite scores. The average differences between ATLAS and non-ATLAS users were small

(Table 8), and none of these differences were statistically significant after controlling for preservice teacher and site differences (Table 9).

Table 8. Average edTPA Scores for Preservice ATLAS and Non-ATLAS Groups

Average edTPA scores				
Score	ATLAS		Non-ATLAS	
	Average	*N*	Average	*N*
Total edTPA score	46.68	89	46.16	97
Task 1 Planning subscore	13.78	89	13.64	97
Task 2 Instruction subscore	13.77	89	13.46	97
Task 3 Assessment subscore	14.91	89	14.99	97

Table 9. Tests of Statistical Differences Between Preservice ATLAS and Non-ATLAS Groups

ATLAS versus non-ATLAS				
Score	Coefficient estimate	Standard error	*p* value	*N*
Total edTPA score	0.52	1.28	.68	186
Task 1 planning subscore	0.14	0.49	.77	186
Task 2 instruction subscore	0.31	0.39	.43	186
Task 3 assessment subscore	-0.09	0.58	.88	186

Using the same implementation group classifications noted earlier (see Table 7), exploratory analyses revealed that preservice teachers who were exposed to higher levels of ATLAS use did not score differently on the edTPA assessment relative to preservice teachers who were not exposed to ATLAS. A table that contains descriptive results for each of the three main scales (preparedness, total self-efficacy, and self-reflection), by the three ATLAS implementation groups and for non-ATLAS users, is in Appendix D (see Table D5). A detailed table that contains the effect estimates and related information for the preservice teacher outcomes is in Appendix D (see Table D6).[11]

[11] These tables were created using the reporting template provided by the team that is conducting the national evaluation of the Investing in Innovation Fund program.

Part 3: ATLAS use did not have a statistically significant effect on early career teacher perceptions

Survey responses from early career ATLAS and non-ATLAS users were compared to determine whether ATLAS use had an effect on perceptions of preparedness, self-efficacy, and self-reflection. There were some average differences between ATLAS and non-ATLAS users (Table 10); however, none of these differences were statistically significant after controlling for early career teacher and site differences (e.g., ACT score; Table 11). Moreover, differences between first- and second-year teachers were examined for each of the scales; however, these groups were not found to report different perception levels. In other words, first- and second-year teachers reported similar levels of preparedness, self-efficacy, and self-reflection.

Table 10. Average Scale Scores for Early Career ATLAS and Non-ATLAS Groups

Average scale scores				
	ATLAS		Non-ATLAS	
Scale	Average	*N*	Average	*N*
Preparedness	3.05	61	3.09	58
Efficacy for instructional strategies	0.60	61	1.10	58
Efficacy for classroom management	2.45	61	1.61	58
Efficacy for student engagement	3.57	61	4.12	58
Efficacy, total	3.84	61	3.73	58
Self-reflection	3.85	61	4.01	58

Table 11. Tests of Statistical Differences Between Early Career ATLAS and Non-ATLAS Groups

ATLAS versus non-ATLAS				
Scale	Coefficient estimate	Standard error	*p* value	*N*
Preparedness (first- and second-year teachers)	0.47	0.70	.50	119
Preparedness (difference between first- and second-year teachers)	0.03	0.86	.97	119
Efficacy for instructional strategies (first- and second-year teachers)	1.08	0.93	.25	119
Efficacy for instructional strategies (difference between first- and second-year teachers)	-1.01	1.15	.38	119

ATLAS versus non-ATLAS				
Scale	Coefficient estimate	Standard error	*p* value	*N*
Efficacy for classroom management (first- and second-year teachers)	1.01	1.35	.46	119
Efficacy for classroom management (difference between first- and second-year teachers)	-0.35	1.67	.84	119
Efficacy for student engagement (first- and second-year teachers)	-0.18	1.33	.89	119
Efficacy for student engagement (difference between first- and second-year teachers)	-0.16	1.65	.92	119
Efficacy, total (first- and second-year teachers)	0.77	0.79	.33	119
Efficacy, total (difference between first- and second-year teachers)	-0.49	0.98	.61	119
Self-reflection (first- and second-year teachers)	-0.09	0.75	.90	119
Self-reflection (difference between first- and second-year teachers)	0.69	0.95	.47	119

Exploratory analyses revealed that there were two distinguishable implementation groups among early career ATLAS users based on the average number of ATLAS cases used and the average number of M-C-A-F cycles facilitated (Table 12). On average, early career teachers who were exposed to higher levels of ATLAS use did not report different levels of the measured outcomes relative to early career teachers who were not exposed to ATLAS. Figures that illustrate descriptive results for each of the three main scales (preparedness, total self-efficacy, and self-reflection), by the two ATLAS implementation groups and for non-ATLAS users, are in Appendix D (see Figures D4 through D6). A detailed table that contains the effect estimates and related information for the early career teacher outcomes is in Appendix D (see Table D7).[12]

[12] These tables were created using the reporting template provided by the team that is conducting the national evaluation of the Investing in Innovation Fund program.

Table 12. Early Career Teacher Implementation Groups

Type of implementation	Mean no. unique ATLAS cases	Mean no. M-C-A-F facilitated cycles	*N*
Group 1: High cases, high cycles	7.83	4.90	30
Group 2: Low cases, low cycles	2.84	2.22	32

Part 4: ATLAS did not have a statistically significant effect on student achievement outcomes

State student achievement data from students of ATLAS users and matched students of non-ATLAS users were compared to determine whether ATLAS use by early career teachers had an effect on Grade 5 and 6 student achievement in mathematics and science. The differences in achievement scores between students of ATLAS users and students of non-ATLAS users were small (Table 13) and not statistically significant after controlling for student and school differences (e.g., prior achievement; Table 14). Detailed tables that contain the effect estimates and related information for the student outcomes are in Appendix D (see Tables D8 and D9).[13]

Table 13. Average Achievement Scores for Students of ATLAS and Non-ATLAS Users

Score	ATLAS		Non-ATLAS	
	Adj. mean	*N*	Adj. mean	*N*
Mathematics	298.49	57	292.64	57
Science	712.17	197	731.08	197

Table 14. Tests of Statistical Differences Between Students of ATLAS and Non-ATLAS Users

ATLAS versus non-ATLAS				
Score	Standardized coefficient estimate	Standard error	*p* value	*N*
Mathematics	0.15	1.10	.89	114
Science	-0.47	0.47	.32	394

[13] These tables were created using the reporting template provided by the team that is conducting the national evaluation of the Investing in Innovation Fund program.

Discussion

During a 5-year period, the National Board and its five IHE and eight LEA partners developed, piloted, and implemented at scale a brand-new professional learning resource for educators. This study was conducted when the ATLAS resource was in its earliest phases of use. And, perhaps most important at this phase of the work, the National Board and its 13 implementation partners demonstrated—at scale—that high implementation fidelity of the ATLAS resource is possible.

We found no statistically significant effects of ATLAS use on the measured outcomes. However, it is common for interventions under development to not produce statistically significant effects. A recent analysis of i3 interventions found that 8% of Development grants produced a statistically significant effect on student outcomes versus 50% of scale-up grants and 40% of validation grants that implemented previously tested interventions (Boulay et al., 2018).

Although the study team did not identify any effects of ATLAS use on the measured outcomes, several factors might have influenced the study results. In particular, during the study period, IHE and LEA partners were in only their first 2 years of implementing ATLAS at scale. The implementation partners were not able to implement ATLAS with high fidelity until the second study year; therefore, preservice and early career teachers did not receive the full amount of ATLAS exposure that was intended. This element could have resulted in a downward bias in any implementation effects.

It is also possible that the ATLAS use parameters, as determined by the IHE and LEA partners before the start of the study, were not sufficient. For ATLAS use to result in effects on measured outcomes, ATLAS might need to be more deeply embedded into teacher training and induction programs. Perhaps ATLAS must be a primary resource for informing and enhancing classroom practice, as opposed to being used periodically throughout a school year. For ATLAS use to happen with such regularity, either ATLAS facilitators would need more availability to work with ATLAS users regularly or ATLAS use would need to become a practice that is less dependent on a facilitator.

There were other factors, beyond the control of ATLAS users and facilitators (and not directly assessed by this study), that could have influenced the study results. Over the course of the 5-year grant, structural changes occurred that could have affected implementation fidelity. There were IHE and LEA site leaders who had been involved from the start of the grant but later needed to shift their efforts and resources to other pressing matters (as is typical in such dynamic and demanding

settings). And there were school closings and reorganizations that coincided with ATLAS implementation, which could have distracted local leaders from implementing ATLAS as planned.

There were limitations to the study design as well. Foremost, the opportunity for preservice and early career teachers to be exposed to and use ATLAS did not occur by chance (random assignment). And although the study had access to a comparison group of non-ATLAS users, it is likely that the groups of ATLAS and non-ATLAS users differed from one another in unobservable ways (e.g., motivation, interest). If the groups were not equivalent on observable and unobservable factors, then such factors—rather than ATLAS use alone—could have accounted for any differences (or the lack thereof) between groups on the measured outcomes. This lack of equivalence, alone, should give the reader pause when interpreting the study findings.

The study sample size posed a limitation as well. Participating IHEs were experiencing declining enrollment in their teacher preparation programs prior to and during the study year. Programmatically, this decline made it difficult or infeasible for some sites to split already small cohorts into separate groups of ATLAS users and nonusers. Having smaller enrollment numbers was exacerbated by a low preservice teacher survey response rate (33%), which further reduced the sample and could have resulted in making the groups even less like one another than if the response rate were higher (e.g., above 80% or 90%). The study team attempted to adjust for such differences by accounting for pre-ATLAS exposure factors such as past performance on standardized assessments (e.g., Praxis I); however, participating institutions varied with respect to the types of such information that were collected, and some preparation programs did not have access to such data, which was not unexpected. As with the IHEs, LEAs had smaller than expected numbers of first- and second-year teachers during the study years. Given these factors, the study was not well powered to detect statistical differences between groups, and data limitations made it difficult to statistically adjust for differences between ATLAS and non-ATLAS users.

Future studies could expand the scope of measures used. In particular, preservice and early career teachers could be observed to examine differences in classroom practice. Measures of student experiences could be assessed as well to determine if students of ATLAS users have classroom experiences that are different from those of students whose teachers were not ATLAS users.

ATLAS use is still in the earliest phases of implementation and expansion; thus, there is more to be learned. Larger samples with additional years of data will provide more precise estimates of differences between groups of ATLAS users and non-ATLAS users on a number of outcomes. And when sites have matured in their ATLAS use, there will be more lessons to share regarding the quantity of ATLAS use that is needed and best practices for ATLAS implementation.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573.

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

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests* (Exp. ed.). Chicago, IL: University of Chicago Press.

Stanford Center for Assessment, Learning and Equity (SCALE). (2013). *edTPA field test: Summary report.* Palo Alto, CA: Author.

Tschannen-Moran, M., & Hoy, M. W. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17,* 783–805.

U.S. Department of Education National Center for Education Statistics. (2011). Schools and Staffing Survey, 2011–12. Washington, DC: Author. https://nces.ed.gov/surveys/sass/pdf/1112/SASS4A.pdf

What Works Clearinghouse. (2017). *Standards handbook (version 4.0).* Washington, DC: U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. Retrieved from https://ies.ed.gov/ncee/wwc/Docs/referenceresources/wwc_standards_handbook_v4.pdf

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis: Rasch measurement.* Chicago, IL: MESA Press.

Appendix A. Study Samples and Selection

This appendix contains the samples and selection description for the preservice teacher, early career teacher, and student samples, respectively.

Preservice Teacher Sample Selection

IHEs in the analytic sample for Part 2 of the study include the following: Central Washington University (WA), Niagara University (NY), Tennessee State University (TN), University of Memphis (TN), and Vanderbilt University (TN). Within each of the four IHE teacher education programs participating in the study, all preservice teacher candidates enrolled in the teacher preparation programs in 2015–16 at each IHE and pursuing majors related to teaching Grades 3 through 6 mathematics and science were included in the sample. Teacher candidates were in the treatment condition because of their enrollment in courses at campus locations in which faculty using ATLAS were instructing. In all IHEs, the teachers exposed to ATLAS cases were exposed in more than one course by more than one faculty member in both their junior and senior years.

Contemporaneous, within-IHE comparison groups were formed for three of the four IHEs. Vanderbilt University did not have a comparison group. A comparison group for this site was identified in 2015–16, but due to changes in their teacher preparation program beginning in 2016–17, the previous comparison teacher candidates were combined with ATLAS treatment teachers in courses, thus contaminating the group. Therefore, the former comparison teachers were excluded from the analytic sample entirely.

Comparison teacher candidates in each teacher preparation program were pursuing a similar certification/degree as those exposed to ATLAS and were taking a similar mix of required courses as treatment teacher candidates but at a different IHE campus location. They were not in any of the courses in which ATLAS cases were used.

Early Career Teacher Sample Selection

LEAs included in the analytic sample for Part 3 of the study include the following: Dyersburg City Schools, Jackson-Madison County School District, Metropolitan Nashville Public Schools, Niagara Falls City School District, Seattle Public Schools, Tipton County Schools, West Valley School District, and Yakima School District. For all LEAs except Metropolitan Nashville Public Schools, ATLAS and non-ATLAS teachers eligible for inclusion in the sample were all teachers in the participating school districts instructing mathematics or science in Grades 3 through 6 who were in their first year in the teaching profession in 2015–16. However, in Metropolitan

Nashville Public Schools, participating treatment and comparison schools were selected from the bottom 50% of schools in terms of academic performance of students during the previous 3 academic years. All ATLAS and non-ATLAS teachers instructing mathematics and science in Grades 3 through 6 who were in their first year of teaching in 2015–16 within the bottom 50% of schools were eligible to be included in the study. Outcomes were assessed for all early career teachers of Grades 3 through 6 mathematics and science in the analytic sample.

Student Sample Selection

Because of data availability, LEAs included in the analytic sample for Part 4 of the study include Jackson-Madison County School District, Metropolitan Nashville Public Schools, and Tipton County Schools. Students who are eligible for inclusion in the sample are all students of ATLAS and non-ATLAS teachers from Part 3 of the study. A propensity score matching approach was used to identify a group of students of non-ATLAS teachers that was similar to students of ATLAS teachers on a list of characteristics. These include student-level characteristics such as student prior academic performance, gender, race, special education status, English language learner status, and school-level characteristics such as school prior academic performance, enrollment, percentage of female students, percentage of students in different racial groups, percentage of students with disabilities, percentage of students with limited English proficiency, and percentage of economically disadvantaged students. See Appendix C for more details on propensity score matching.

Appendix B. Survey Outcome Measures

Exhibit B1. Preservice and Early Career Teacher Preparedness Items

[Stem for preservice teachers:] When you enter your first year of teaching, how well prepared will you be to . . .

[Stem for early career teachers:] How well can you . . .

Please mark one box for each row.

	Not at all	Somewhat prepared	Well prepared	Very well prepared
a. Handle a range of classroom management or discipline situations?	❑	❑	❑	❑
b. Use a variety of instructional methods?	❑	❑	❑	❑
c. Teach your subject matter?	❑	❑	❑	❑
d. Use computers in classroom instruction?	❑	❑	❑	❑
e. Assess students?	❑	❑	❑	❑
f. Differentiate instruction in the classroom?	❑	❑	❑	❑
g. Use data from student assessments to inform instruction?	❑	❑	❑	❑
h. Meet state content standards?	❑	❑	❑	❑

Exhibit B2. Preservice and Early Career Teacher Self-Efficacy Items

[Stem for preservice teachers:] When you enter your first year of teaching, to what extent will you be able to . . .

[Stem for early career teachers:] To what extent are you able to . . .

Please mark one box for each row.

	Not at all	Minimal extent	Moderate extent	Great extent
a. Use a variety of assessment strategies?	❑	❑	❑	❑
b. Provide an alternate explanation or example when students are confused?	❑	❑	❑	❑
c. Craft good questions for your students?	❑	❑	❑	❑
d. Implement alternative strategies in your classroom?	❑	❑	❑	❑
e. Control disruptive behavior in your classroom?	❑	❑	❑	❑
f. Get children to follow classroom rules?	❑	❑	❑	❑
g. Calm a student who is disruptive or noisy?	❑	❑	❑	❑
h. Establish a classroom management system with each group of students?	❑	❑	❑	❑
i. Get students to believe they can do well in schoolwork?	❑	❑	❑	❑
j. Help your students value learning?	❑	❑	❑	❑
k. Motivate students who show low interest in schoolwork?	❑	❑	❑	❑
l. Assist families in helping their children do well in school?	❑	❑	❑	❑

Note. Items a through d correspond to the efficacy for instructional strategies factor; items e through h correspond to the efficacy for classroom management factor; items i through l correspond to the efficacy for student engagement factor.

Exhibit B3. Preservice and Early Career Teacher Self-Reflection Items

[Stem for preservice and early career teachers:] How often are the following statements true about you?

	Never or almost never true	Sometimes true	Usually true	Always or almost always true
a. I can identify the instructional practices that make me effective at helping students to learn.	❑	❑	❑	❑
b. I can identify the areas of my instructional practice that I will need to improve over time.	❑	❑	❑	❑
c. I have a specific purpose for each choice that I make when designing my lessons.	❑	❑	❑	❑
d. I ask myself if a lesson or activity will work for all students.	❑	❑	❑	❑
e. If there is an ongoing problem in the classroom, I think of several ways to solve it and try more than one, comparing student responses or changes in behavior.	❑	❑	❑	❑
f. I ask myself questions about how well I am doing based on immediate student responses to me in the classroom.	❑	❑	❑	❑
g. I find myself analyzing whether I need to change course and adapt my lesson as I am delivering it in the classroom.	❑	❑	❑	❑
h. I find myself analyzing my instructional practice after delivering a lesson.	❑	❑	❑	❑
i. When I note progress or lack of progress after assessing my students, I try to connect it to specific actions I've taken in the past or things that I can do differently in the future.	❑	❑	❑	❑
j. If I want to improve something about my teaching practice, I search for sources of support or professional development that are known to help teachers improve that area of practice.	❑	❑	❑	❑
k. I can recognize the specific instructional skills and strategies that teachers are using if I have the opportunity to observe their teaching.	❑	❑	❑	❑

Appendix C. Psychometric and Analytic Methods

This appendix contains information regarding the psychometric methods used to analyze preservice and early career teacher survey data that were used in Part 2 and Part 3 of the study. It also contains analytic methods used to examine differences between ATLAS and non-ATLAS users (Part 2 and Part 3 of the study) and differences between students of ATLAS users and matched students of non-ATLAS users (Part 4 of the study).

Psychometric Survey Scaling Used in Part 2 and Part 3

Each survey included five scales: preparedness, efficacy for instructional strategies, efficacy for classroom management, efficacy for student engagement, total efficacy (an overall score with the items of the three subefficacy scales combined), and self-reflection.

The survey items were scaled using the Rasch model for ordered response categories (Andrich, 1978; Rasch, 1980; Wright & Masters, 1982) to determine whether the items reliably measure each overarching survey scale (or "construct"). Items that were designed to measure a single underlying construct such as "preparedness" were scaled together. The resulting scale scores provide a quantitative view of the frequency and intensity of respondents' answers across a set of items representing a given construct. This result differs from averaging the percentage of respondents endorsing each response option because the Rasch model considers the relative frequency with which each item and response option is used (i.e., item difficulty).[14]

Overall, the survey scales on each survey functioned well. As reported in Table C1, Cronbach's alphas reliabilities for the survey ranged from .85 to .94 for preservice teacher scales and from .75 to .92 for early career teacher scales, and Rasch reliabilities ranged from .69 to .89 for preservice teacher scales and from .66 to .87 for early career teacher scales.

[14] Item difficulty reflects how positively an item is endorsed. Items with low item difficulty will be frequently and positively endorsed (e.g., a high frequency of well prepared).

Table C1. Reliability Statistics by Scale for Each Respondent Group

Scale	Preservice teacher survey		Early career teacher survey	
	Cronbach's α	Rasch reliability	Cronbach's α	Rasch reliability
Preparedness	.92	.87	.85	.81
Efficacy for instructional strategies	.85	.69	.75	.66
Efficacy for classroom management	.91	.74	.89	.73
Efficacy for student engagement	.88	.77	.84	.73
Efficacy for instructional practice (overall)	.94	.89	.91	.87
Self-reflection	.93	.84	.92	.84

Scale scores were also transformed back onto the original rating scale metrics (e.g., *not at all, somewhat prepared, well prepared, very well prepared*) and can be interpreted as categorical summary responses across the range of items representing a given construct. See Appendix D, Tables D5, D6, D7, D10, D11, and D12.

Scale scores were then used to examine differences between ATLAS and non-ATLAS users to answer research questions 3 through 5 (Part 2) and 7 through 9 (Part 3).

Analytic Approach to Examine Teacher Survey Outcomes

The differences in teacher survey outcomes between ATLAS (treatment) and non-ATLAS (comparison) users were examined using a two-level statistical model with teachers nested within sites[15] that accounted for correlations among teachers from the same sites. This statistical technique allows for the control of other teacher characteristics to account for their possible influence on the outcomes. We conducted parallel analyses for preservice teachers (Part 2) and early career teachers (Part 3) separately.

A two-level statistical model is shown in the following equation (1), with teachers nested within sites for the survey outcomes.

[15] Sites are defined as universities for the preservice teacher sample and school districts for the early career teacher sample.

$$Y_{ij} = \beta_0 + \beta_1 Treatment_{ATLAS} + \beta_2 X_{TeacherCov} + \beta_3 State_k + \nu_j + \varepsilon_{ij} \quad (1)$$

In the model, Y_{ij} represents the teacher survey scale score (e.g., preparedness) of teacher *i* in school *j*, and $Treatment_{ATLAS}$ is a binary variable indicating whether the teacher *i* is an ATLAS teacher. The vector $X_{TeacherCov}$ represents teacher-level characteristics including ACT test score, SAT test score, WEST-B test score, education level, teaching experience, gender, race/ethnicity, and birth year. To eliminate bias in the estimate attributed to state (geographic) differences, we included state fixed effects, the vectors $State_k$. ε_{ij} and ν_j are the teacher- and site-level random error terms. The coefficient β_0 shows the mean scale score of the non-ATLAS (comparison) group, and β_1 (the parameter of interest) is the difference in mean scale score between ATLAS and non-ATLAS users or the overall effect of the ATLAS program.

K-Means Cluster Analysis. In addition to comparing outcomes of ATLAS and non-ATLAS users (preservice and early career teachers), the study team used a cluster analysis technique to group only ATLAS users into different clusters according to their ATLAS usage. Cluster analysis is a classification technique used to create groups (or "clusters") in a manner that minimizes the differences of the characteristics within those groups while maximizing the differences in characteristics between those groups. In this analysis, ATLAS teachers were classified according to patterns of their ATLAS usage, including the number of cases used and the number of M-C-A-F cycles to which they were exposed. Therefore, teachers within each group should be more similar to one another in their ATLAS usage and dissimilar to those teachers in other groups.

K-means cluster analysis creates groups using Euclidean distance between cluster centers. First, cluster centers are generated randomly, followed by several iterations whereby the values for each case (i.e., number of cases used and number of M-C-A-F cycles) are arranged based on the closest Euclidean distance to the center of the cluster. This process continues until either the iteration limit has been reached or the cluster center changes less than 2% from the previous iteration.

Using SPSS 15.0, k-means cluster analysis was employed to create groups of teachers according to the two grouping variables (number of cases used and number of M-C-A-F cycles). K-means cluster analysis allows the researcher to select a predetermined number of groups. In selecting an appropriate number of hypothesized groups, important considerations exist such as group size, interpretability, and meaningfulness. Given that two variables were used to create groups and that teachers were expected to have either a high or low usage for each of the two variables, four groups were initially created. Due to low group membership in some groups, the

three or two groups as shown in the following list were created. These group analyses resulted in a more robust distribution of group membership and, thus, better interpretability and meaningfulness of the groups. The cluster analysis revealed the following three groups for preservice teacher survey outcomes and edTPA scores:

1. High/High: high number of cases used and high number of cycles
2. Low/Low: low number of cases used and low number of cycles
3. High/Low: high number of cases used and low number of cycles

The cluster analysis revealed the following two groups for early career teacher survey outcomes:

1. High/High: high number of cases used and high number of cycles
2. Low/Low: low number of cases used and low number of cycles

We then estimated the differences between the clusters of ATLAS teachers and non-ATLAS teachers (the reference group) for the relevant study outcomes. Equation 2, following, was used to estimate the differences between each cluster of ATLAS teachers and the non-ATLAS teachers (the reference group).

$$Y_{ij} = \beta_0 + \beta_1 Cluster_n + \beta_2 X_{TeacherCov} + \beta_3 State_k + \nu_j + \varepsilon_{ij} \qquad (2)$$

Instead of having the binary variable $Treatment_{ATLAS}$ in the model, we included the vector $Cluster_n$ indicating the clusters of ATLAS teachers with non-ATLAS teachers as the reference group. The vector of coefficients β_1 (the parameters of interest) shows the difference between each cluster of ATLAS teachers and non-ATLAS teachers.

Analytic Approach to Examine edTPA Scores

Similar to the teacher survey analysis just described, the differences in teachers' edTPA scores between ATLAS and non-ATLAS teachers were examined using a two-level statistical model with teachers nested within sites that accounted for correlations among teachers from the same sites (Part 2). This statistical technique allows for control of other teacher characteristics to account for their possible influence on the outcomes. The two-level statistical model as shown in Equation 1 was used to analyze edTPA scores. All parameters are the same except the teacher-level characteristics included in the model, represented by the vector $X_{TeacherCov}$. The teacher-level characteristics used for edTPA score analysis include ACT test score, SAT test score, WEST-B test score, gender, race/ethnicity, whether a senior student, and the year of birth.

In addition, we used the same cluster analysis approach described earlier to group ATLAS teachers into three different clusters, and Equation 2 was used to estimate the difference between each of the three clusters of ATLAS teachers and the non-ATLAS teachers (the reference group).

Analytic Approach to Examine Student Achievement Outcomes

For Part 4 of the study, students of ATLAS teachers (treatment) were matched to students of non-ATLAS teachers (comparison). This was done because the quality of evidence about the effects of ATLAS use on achievement is dependent on identifying a sample of comparison students who closely resembled treatment students in their pre-achievement and background characteristics. Drawing on the data provided by the study districts, the study team used students' and school's pre-achievement and background characteristics that are known to correlate with student achievement, and possibly program participation, as a basis for choosing matched comparison groups.

The study team used propensity score matching to identify comparison students based on prior student and school achievement and other student and school characteristics. The matching process was implemented in three steps and was conducted separately for the two subjects, mathematics and science. First, a propensity score was generated using a logistic model for every student based on the pre-achievement and background characteristics.

The following equation was used during the matching process.

$$logit\big(P(ATLAS_j)\big) = \eta + \beta_0 Z_j + \sum_{\rho=1}^{\rho} \theta_\rho C_{\rho j}$$

where:

$ATLAS_j$ = an indicator of whether student *j* is in the treatment group: 1 if student *j* is in the treatment group, and 0 if not

$P(ATLAS_j)$ = the propensity of student *j* to be in the treatment group

η = intercept

Z_j = prior test score for student *j*

C_{pj} = a set of student demographics, school prior achievement, and school demographics for student *j* (e.g., gender, race/ethnicity, special education status, English proficiency level, school enrollment, school percentage of economically disadvantaged students, school percentage of students with disability, school percentage of students with limited English proficiency)

β_0 = the coefficient that represents the association between prior test score and the logit of the propensity score[16]

θ = a set of coefficients that represents the association between school prior achievement, each student and school demographic characteristic, and the logit of the propensity score

The propensity score for a student represents the probability that a student is in the treatment group, given the observed characteristics.

Second, each treatment student was matched without replacement[17] to one comparison student with the closest propensity score (within a caliper of 0.05 standard deviations)—that is, the one nearest "neighbor" with the closest propensity to being taught by an ATLAS teacher.

As a third and final step, we checked whether the matching produced two groups that were similar in pre-achievement and background characteristics. Consistent with the What Works Clearinghouse (WWC) standards (WWC, 2017), we considered the two groups to be balanced if the standardized mean difference (SMD) in prior student achievement (the pretest measure) and other characteristics between the two groups of students in the sample was less than or equal to 0.25 standard deviations. The matching process produced two, balanced analytic samples: one for the mathematics achievement outcome analysis and one for the science achievement outcome analysis (see Table C2 and Table C3, respectively).

[16] The logit of the propensity score is equal to the log of the ratio propensity score over (1 minus the propensity score).
[17] If a comparison student is matched to one treatment student, then this comparison student cannot be matched to other treatment students.

Table C2. Treatment and Comparison Group Characteristics for Student Achievement—Mathematics

Covariates	Before matching			After matching		
	Treatment (*N* = 189)	Comparison (*N* = 197)	SMD	Treatment (*N* = 57)	Comparison (*N* = 57)	SMD
Student characteristics						
Pre-score, mathematics	738.24	754.97	-0.37	742.67	747.82	0.12
Birth year	2005	2005	-0.08	2005	2005	0.05
Special education	8.5%	11.7%	-0.22	9.0%	7.0%	0.07
English language learner	31.2%	10.7%	0.81	19.0%	25.0%	0.13
White	19.6%	32.5%	-0.41	21.0%	23.0%	0.04
Black	46.6%	52.8%	-0.15	58.0%	53.0%	0.11
Hispanic	23.8%	10.7%	0.58	19.0%	19.0%	<0.01
Female	48.7%	47.7%	0.02	60.0%	53.0%	0.14
School characteristics in 2014–15						
Percentage proficient or above	45.2%	50.6%	-0.36	47.9%	48.0%	0.01
Enrollment	630	563	0.39	566	581	0.09
White	28.3%	38.4%	-0.39	27.9%	26.5%	0.06
Black	44.5%	44.7%	-0.01	51.9%	51.4%	0.02
Hispanic	22.3%	12.4%	0.52	16.9%	18.1%	0.06
Female	48.1%	49.0%	-0.39	49.0%	48.9%	0.07
Limited English proficiency	22.4%	12.5%	0.46	16.0%	17.6%	0.07
Student with a disability	13.0%	12.1%	0.27	13.1%	12.7%	0.13
Economically disadvantaged	85.2%	71.6%	0.75	82.6%	83.4%	0.06

Note. SMD = standardized mean difference

Table C3. Treatment and Comparison Group Characteristics for Student Achievement—Science

Covariates	Before matching			After matching		
	Treatment (*N* = 432)	Comparison (*N* = 397)	SMD	Treatment (*N* = 197)	Comparison (*N* = 197)	SMD
Student characteristics						
Pre-score, science	728.52	738.19	-0.22	730.38	730.21	0.00
Birth year	2005	2005	-0.18	2005	2005	0.04
Special education	11.1%	12.8%	-0.10	11.0%	11.0%	0.02
English language learner	23.6%	19.4%	0.15	19.0%	19.0%	<0.01
White	17.8%	34.3%	-0.53	22.0%	20.0%	0.05
Black	48.6%	43.1%	0.14	55.0%	53.0%	0.04
Hispanic	29.2%	18.9%	0.34	20.0%	24.0%	0.10
Female	51.4%	48.1%	0.08	47.0%	48.0%	0.03
School characteristics in 2014–15						
Percentage proficient or above	37.0%	49.6%	-0.82	39.6%	38.6%	0.07
Enrollment	565	612	-0.25	579	585	0.03
White	23.2%	36.7%	-0.69	25.3%	23.8%	0.08
Black	44.0%	42.0%	0.07	48.5%	51.8%	0.12
Hispanic	29.7%	17.1%	0.64	23.5%	21.9%	0.08
Female	49.1%	49.9%	-0.29	49.3%	49.6%	0.12
Limited English proficiency	27.6%	16.4%	0.57	21.1%	19.6%	0.08
Student with a disability	12.1%	12.1%	-0.01	12.4%	12.6%	0.05
Economically disadvantaged	89.3%	74.4%	1.03	86.3%	86.9%	0.06

Note. SMD = standardized mean difference

For each of the two analytic samples (mathematics and science), the difference in achievement between the treatment and comparison groups was estimated using a three-level statistical model with students nested within teachers nested within schools that accounted for correlations among students who were taught by the same teachers from the same schools, as well as students' pre-achievement and demographic characteristics, teachers' demographic characteristics, and schools' achievement and demographic characteristics. This method produced subject-specific estimates of the difference in average achievement between treatment and comparison students.

The three-level statistical model, as shown in the following equation (3), with students nested within teachers nested within schools was used. We conducted the analysis separately for the mathematics and science achievement outcomes.

$$Y_{ijk} = \beta_0 + \beta_1 Treatment_{ATLAS} + \beta_2 PreAchievement_{ijk} + \beta_3 PS_{ijk} + \beta_4 W_m + \beta_5 X_n + \beta_6 Z_o + \mu_k + \nu_{jk} + \varepsilon_{ijk} \quad (3)$$

In the model, Y_{ijk} represents the mathematics or science test score of student *i* of teacher *j* in school *k*, and $Treatment_{ATLAS}$ is a binary variable indicating whether the student *i* is taught by an ATLAS teacher. $PreAchievement_{ijk}$ is the pre-test score of student *i*, and PS_{ijk} is the propensity score of student *i*. The vector W includes student and school characteristics in the 2014–15 academic year (prior year) for student *i*.[18] The vector X represents teacher-level characteristics (e.g., education level, gender, race/ethnicity, etc.) for teacher *j*. Due to some district reorganizations, many students are not in the same schools between the outcome year (2016–17) and the pre-treatment, prior year (2014–15). We included the vector Z in the model representing school-level characteristics (e.g., percentage of students on track or above in mathematics or science, enrollment, percentage of economically disadvantaged students, percentage of students with limited English proficiency, etc.) in the 2016–17 school year for student *i*. ε_{ijk}, ν_{jk} and μ_k are the student-, teacher-, and school-level random error terms. The coefficient β_0 shows the mean test score of the comparison group, and β_1 (the parameter of interest) is the difference in mean test scores between treatment and comparison groups, or the overall effect of the ATLAS program.

[18] Although the propensity score matching was the primary method used to control for differences between treatment and comparison students, covariates with SMD that exceeded 0.05 between the two groups were included as additional controls in the respective outcome model (see Tables C2 and C3).

Appendix D. Supplemental Results Tables

This appendix contains supplemental results tables. Some of the tables are reported for express use by the team that is conducting the national evaluation of the Investing in Innovation Fund program, referred to as NEi3 (see Tables D1 through D4, D9, D13, D14, and D15).

Table D1. Implementation Fidelity Reporting for NEi3: Year 1 (2015–16), Institution of Higher Education (IHE)

Intervention component	Implementation measure (total number of measurable indicators representing each component)	Sample size at the sample level (no. of schools, districts, etc.)	Component level threshold for fidelity of implementation for the unit that is the basis for the sample level	Evaluator's criteria for "implemented with fidelity" at sample level	Component level fidelity score for the entire sample	Implemented with fidelity? (yes, no, N/A)
Access to ATLAS	2	4 IHEs	Score = 1 is high implementation (i.e., ≥75% of facilitators registered in ATLAS system by 12/1/16; *and* ≥75% of teachers access ATLAS)	At least 3 of 4 IHEs have high implementation (score = 1)	4 IHEs	Yes
Use of ATLAS	2	4 IHEs	Score = 1 is high implementation (i.e., ≥75% of teachers are exposed to at least 6 ATLAS cases; *and* ≥75% of teachers complete 4 cycles)	At least 3 of 4 IHEs have high implementation (score = 1)	0 IHEs	No

Table D2. Implementation Fidelity Reporting for NEi3: Year 1 (2015–16), Local Education Agency (LEA)

Intervention component	Implementation measure (total number of measurable indicators representing each component)	Sample size at the sample level (no. of schools, districts, etc.)	Component level threshold for fidelity of implementation for the unit that is the basis for the sample level	Evaluator's criteria for "implemented with fidelity" at sample level	Component level fidelity score for the entire sample	Implemented with fidelity? (yes, no, N/A)
Access to ATLAS	2	3 LEAs	Score = 1 is high implementation (i.e., ≥75% of facilitators registered in ATLAS system by 12/1/16; *and* ≥75% of teachers access ATLAS)	At least 2 of 3 LEAs have high implementation (score = 1)	3 LEAs	Yes
Use of ATLAS	2	3 LEAs	Score = 1 is high implementation (i.e., ≥75% of teachers are exposed to at least 6 ATLAS cases; *and* ≥75% of teachers complete 4 cycles)	At least 2 of 3 LEAs have high implementation (score = 1)	0 LEAs	No

Table D3. Implementation Fidelity Reporting for NEi3: Year 2 (2016–17), Institution of Higher Education (IHE)

Intervention component	Implementation measure (total number of measurable indicators representing each component)	Sample size at the sample level (no. of schools, districts, etc.)	Component level threshold for fidelity of implementation for the unit that is the basis for the sample level	Evaluator's criteria for "implemented with fidelity" at sample level	Component level fidelity score for the entire sample	Implemented with fidelity? (yes, no, N/A)
Access to ATLAS	2	4 IHEs	Score = 1 is high implementation (i.e., ≥75% of facilitators registered in ATLAS system by 12/1/16; *and* ≥75% of teachers access ATLAS)	At least 3 of 4 IHEs have high implementation (score = 1)	4 IHEs	Yes
Use of ATLAS	2	4 IHEs	Score = 1 is high implementation (i.e., ≥75% of teachers are exposed to at least 6 ATLAS cases; *and* ≥75% of teachers complete 4 cycles)	At least 3 of 4 IHEs have high implementation (score = 1)	3 IHEs	Yes

Table D4. Implementation Fidelity Reporting for NEi3: Year 2 (2016–17), Local Education Agency (LEA)

Intervention component	Implementation measure (total number of measurable indicators representing each component)	Sample size at the sample level (no. of schools, districts, etc.)	Component level hreshold for Fidelity of Implementation for the Unit that is the Basis for the Sample-Level	Evaluator's Criteria for "Implemented with Fidelity" at Sample Level	Component Level Fidelity Score for the Entire Sample	Implemented with Fidelity? (Yes, No, N/A)
Access to ATLAS	2	3 LEAs	Score=1 is high implementation (i.e., ≥ 75% of facilitators registered in ATLAS system by 12/1/16; and ≥75% of teachers access ATLAS)	At least 2 of 3 LEAs have high implementation (score=1)	3 LEAs	Yes
Use of ATLAS	2	3 LEAs	Score=1 is high implementation (i.e., ≥ 75% of teachers are exposed to at least 6 ATLAS cases; and ≥ 75% of teachers complete 4 cycles)	At least 2 of 3 LEAs have high implementation (score=1)	2 LEAs	Yes

Figure D1. Descriptive Survey Results by Preservice ATLAS Implementation Group, Preparedness

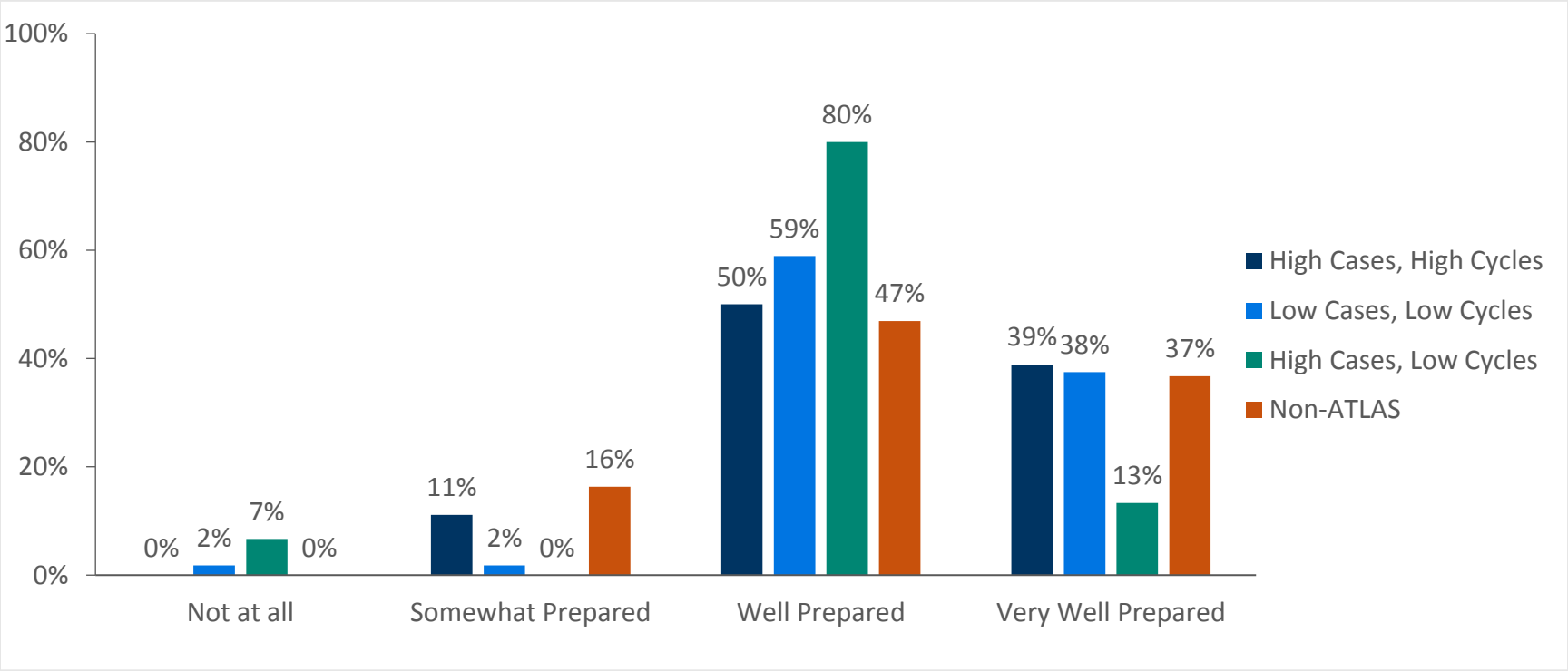

Figure D2. Descriptive Survey Results by Preservice ATLAS Implementation Group, Self-efficacy

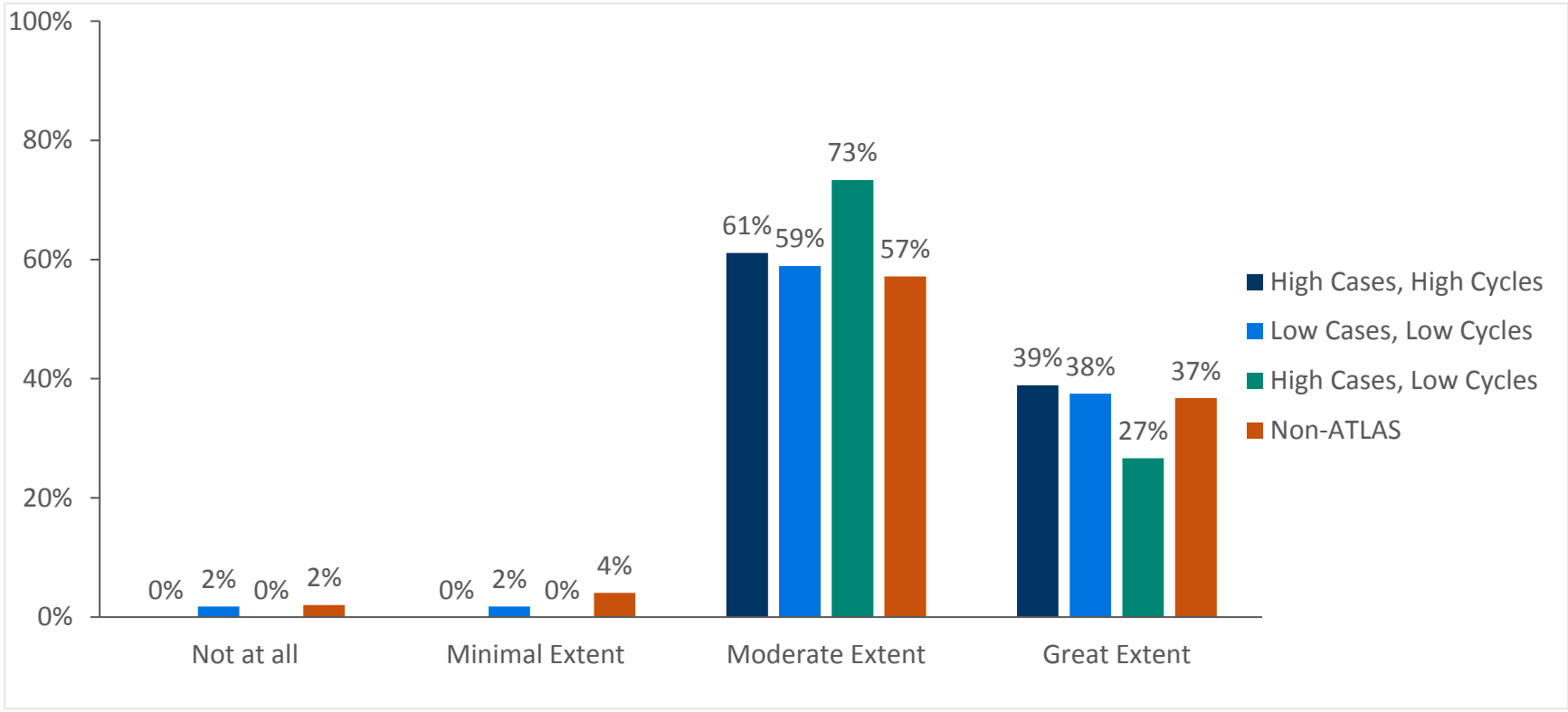

Figure D3. Descriptive Survey Results by Preservice ATLAS Implementation Group, Self-reflection

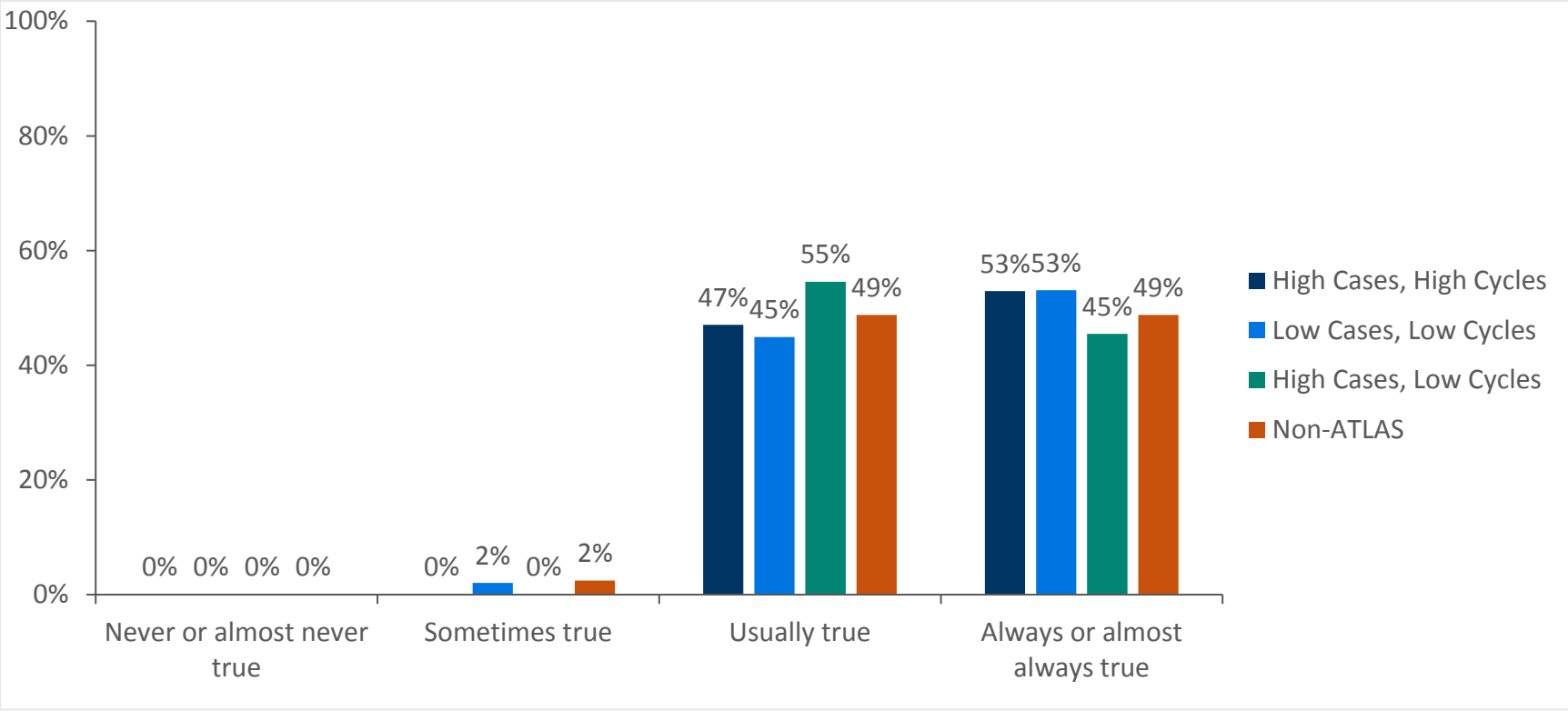

Table D5. Descriptive Survey Results by Preservice ATLAS Implementation Group, Preparedness

Type of Implementation	*N* Preservice Teachers	Mean No. Unique ATLAS Cases	Mean No. M-C-A-F Facilitated Cycles	Adj. Mean, total score	Adj. mean, Planning	Adj. mean, Instruction	Adj. mean, Assessment
Cluster 1: High Cases, High Cycles	16	27.63	11.63	45.21	13.52	13.30	13.97
Cluster 2: Low Cases, Low Cycles	54	4.24	4.30	47.29	14.16	13.95	14.92
Cluster 3: High Cases, Low Cycles	13	18.92	6.31	46.66	14.34	13.62	14.22
Non-ATLAS	97	—	—	46.50	13.98	13.57	14.84

Table D6. Effect Estimates for NEi3: Preservice Teacher Outcomes

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Contrast ID no.	Contrast name (optional)	Posttest measure name	Treatment group *N* of IHE clusters	Treatment group *N* of teachers	Comparison group *N* of IHE clusters	Comparison group *N* of teachers	Unadjusted treatment group SD	Unadjusted comparison group SD	Standard deviation source (code)	Comparison group mean (optional)	Impact estimate	Standardized effect size (optional)	Impact standard error	*p*-value	Code for impact model description	Degrees of freedom	Source of data (optional)	Level of inference (optional)
CS1.1		edTPA overall	5	89	4	97	8.53	7.30	A	46.21	0.52		1.28	0.68	A	165		
ES1.1		edTPA Task 1 Planning	5	89	4	97	2.22	2.67	A	15.43	0.14		0.49	0.77	A	165		
ES1.2		edTPA Task 2 Instruction	5	89	4	97	1.90	2.12	A	14.97	0.31		0.39	0.43	A	165		
ES1.3		edTPA Task 3 Assessment	5	89	4	97	2.54	3.19	A	15.20	-0.09		0.58	0.88	A	165		
CS1.2		Self-report survey: Preparedness	5	89	4	49	2.48	2.76	A	2.61	-0.09		0.52	0.86	A	117		
CS1.3		Self-report survey: Efficacy (overall)	5	89	4	49	2.62	3.25	A	3.73	-0.16		0.57	0.78	A	117		
ES1.4		Self-report survey: Efficacy, Inst. Strategies subscale	5	89	4	49	3.60	4.25	A	4.98	-0.22		0.78	0.78	A	117		

A	B	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S
Contrast ID no.	**Contrast name (optional)**	**Posttest measure name**	**Treatment group *N* of IHE clusters**	**Treatment group *N* of teachers**	**Comparison group *N* of IHE clusters**	**Comparison group *N* of teachers**	**Unadjusted treatment group SD**	**Unadjusted comparison group SD**	**Standard deviation source (code)**	**Comparison group mean (optional)**	**Impact estimate**	**Standardized effect size (optional)**	**Impact standard error**	***p*-value**	**Code for impact model description**	**Degrees of freedom**	**Source of data (optional)**	**Level of inference (optional)**
ES1.5		Self-report survey: Efficacy, Classroom Management subscale	5	89	4	49	4.20	4.87	A	5.44	0.36		0.93	0.70	A	117		
ES1.6		Self-report survey: Efficacy, Student Engagement subscale	5	89	4	49	3.45	3.94	A	4.78	0.13		0.72	0.86	A	117		
ES1.7		Self-report survey: Self-Reflection	5	77	4	41	2.30	2.42	A	4.60	0.17		0.55	0.76	A	97		

Figure D4. Descriptive Survey Results by Early Career ATLAS Implementation Group, Preparedness

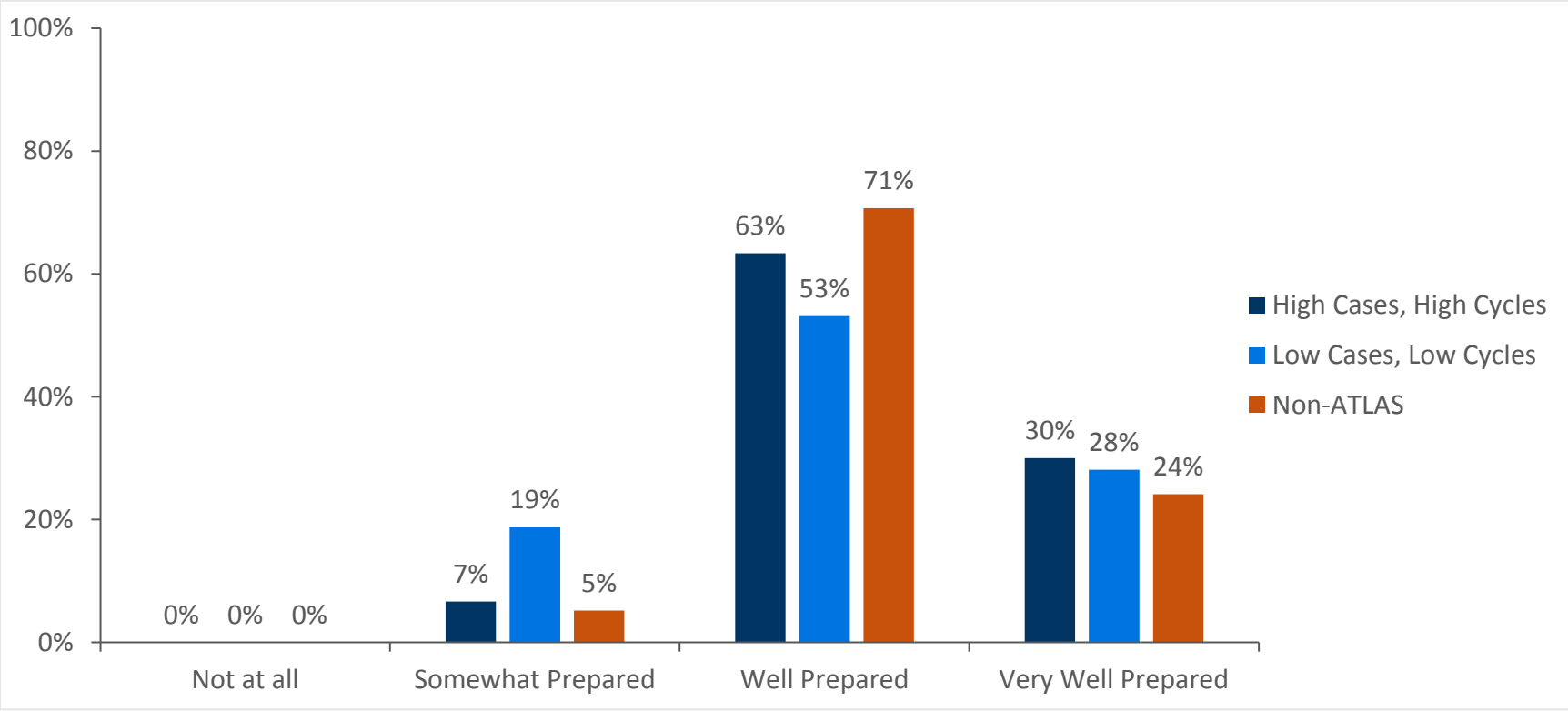

Figure D5. Descriptive Survey Results by Early Career ATLAS Implementation Group, Self-Efficacy

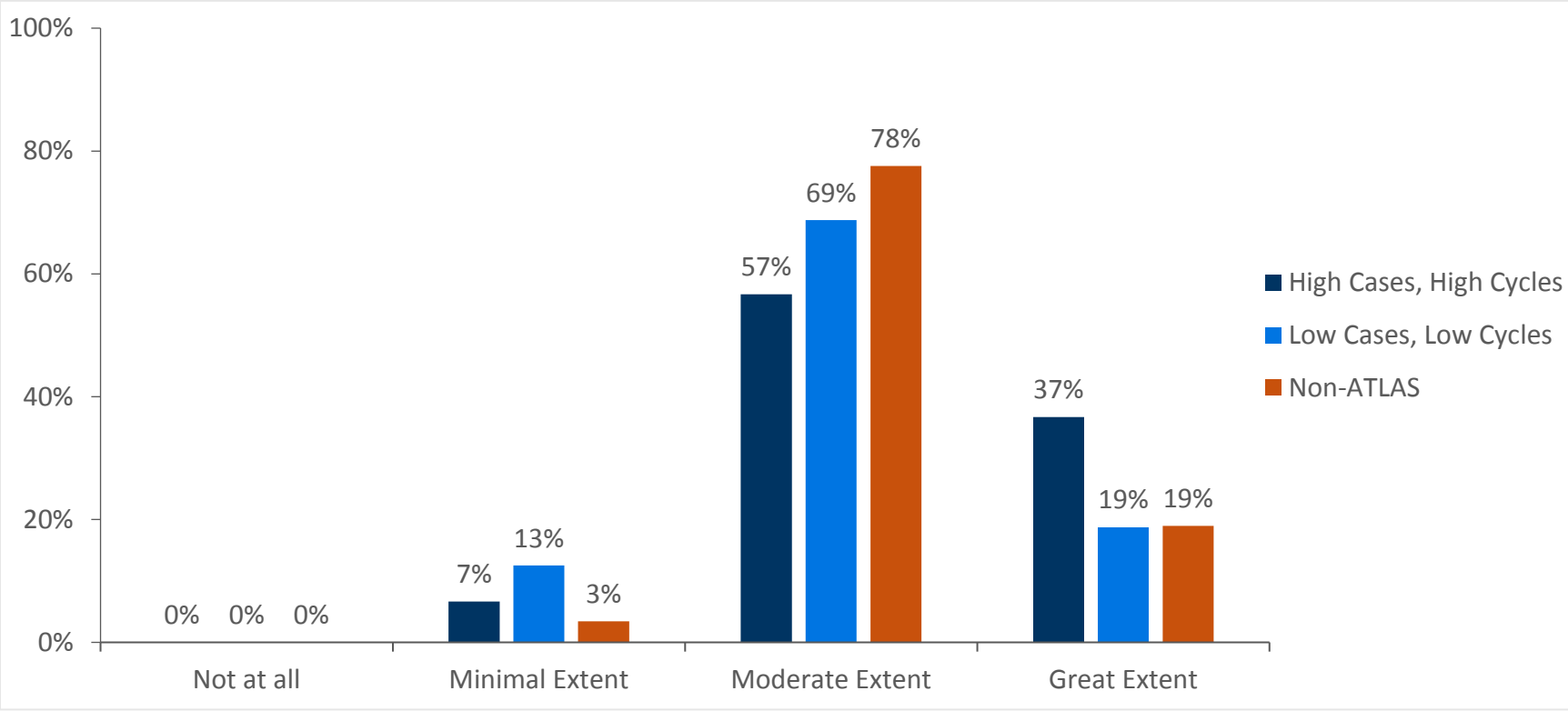

Figure D6. Descriptive Survey Results by Early Career ATLAS Implementation Group, Self-Reflection

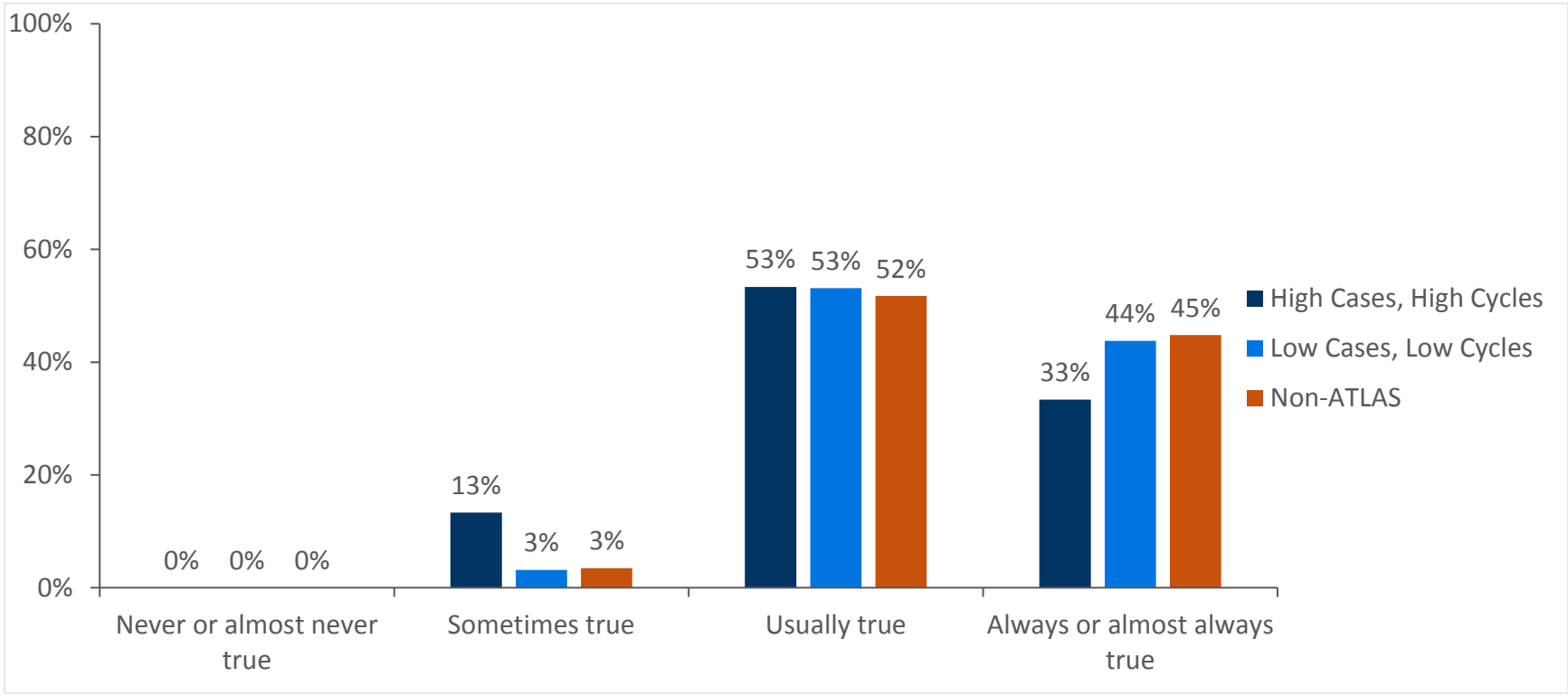

Table D7. Effect Estimates for NEi3: Early Career Teacher Outcomes

A	B	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S
Contrast ID no.	Contrast name (optional)	Posttest measure name	Treatment group *N* of schools	Treatment group *N* of teachers	Comparison group *N* of schools	Comparison group *N* of teachers	Unadjusted treatment group SD	Unadjusted comparison group SD	Standard deviation source (code)	Comparison group mean (optional)	Impact estimate	Standardized effect size (optional)	Impact standard error	*p*-value	Code for impact model description	Degrees of freedom	Source of data (optional)	Level of inference (optional)
CS2.1		Self-report survey: Preparedness	8	61	5	58	2.33	1.84	A	3.09	0.47		0.70	0.50	A	101		
CS2.2		Self-report survey: Efficacy (overall)	8	61	5	58	2.61	2.07	A	3.73	0.77		0.79	0.33	A	101		
ES2.1		Self-report survey: Efficacy, Instructional Strategies subscale	8	61	5	58	3.21	2.38	A	1.10	1.08		0.93	0.25	A	101		
ES2.2		Self-report survey: Efficacy, Classroom Management subscale	8	61	5	58	3.92	3.95	A	1.61	1.01		1.35	0.46	A	101		

A	B	C	D	E	F	G	H	I	J	K	L	M	N	0	P	Q	R	S
Contrast ID no.	Contrast name (optional)	Posttest measure name	Treatment group *N* of schools	Treatment group *N* of teachers	Comparison group *N* of schools	Comparison group *N* of teachers	Unadjusted treatment group SD	Unadjusted comparison group SD	Standard deviation source (code)	Comparison group mean (optional)	Impact estimate	Standardized effect size (optional)	Impact standard error	*p*-value	Code for impact model description	Degrees of freedom	Source of data (optional)	Level of inference (optional)
ES2.3		Self-report survey: Efficacy, Student Engagement subscale	8	61	5	58	4.22	3.77	A	4.12	-0.18		1.33	0.89	A	101		
ES2.4		Self-report survey: Self-Reflection	8	61	5	58	2.51	2.28	A	4.01	-0.09		0.75	0.90	A	101		

Table D8. Effect Estimates for NEi3: Student Achievement Outcomes

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
Contrast ID no.	Contrast name (optional)	Posttest measure name	Treatment group *N* of clusters	Treatment group *N* of students	Comparison group *N* of clusters	Comparison group *N* of students	Unadjusted treatment group SD	Unadjusted comparison group SD	Standard deviation source (code)	Comparison group mean (optional)	Impact estimate	Standardized effect size (optional)	Impact standard error	*p*-value	Code for impact model description	Degrees of freedom	Source of data (optional)	Level of inference (optional)
CS3.1		State standardized test: Mathematics	5	57	4	57	37.54	42.19	A	297.47	5.85		43.65	0.89	A	86		
CS3.2		State standardized test: Science	8	197	7	197	38.15	41.38	A	729.60	-18.91		11.97	0.11	A	363		

Table D9. Baseline Equivalence of Students for NEi3

A	B	C	D	E	F	G	H	I	J	K	L	M	N
Contrast ID no.	Contrast name (optional)	Pre-test measure name	Treatment group *N*	Comparison group *N*	Unadjusted treatment group SD	Unadjusted comparison group SD	Standard deviation source (code)	Comparison group mean (optional)	Treatment–Comparison difference	Standardized T-C difference (optional)	Pre-test shown in this row was used as a control in the impact model for this contrast? (Y/N)	Code for T-C difference calculation	Source of data (optional)
CS3.1		State standardized test: Mathematics	57	57	46.53	42.99	A	747.82	-4.15	-0.115	Y	B	
CS3.2		State standardized test: Science	197	197	36.57	58.85	A	730.21	0.17	0.004	Y	B	

AIR®

AMERICAN INSTITUTES FOR RESEARCH®

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

MAKING RESEARCH **RELEVANT**

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

LOCATIONS

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

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

5917_09/18