Optional ERIC Coversheet — Only for Use with U.S. Department of Education Grantee Submissions

This coversheet should be completed by grantees and added to the PDF of your submission if the information required in this form **is not included on the PDF to be submitted**.

INSTRUCTIONS

- Before beginning submission process, download this PDF coversheet if you will need to provide information not on the PDF.
- Fill in all fields—information in this form **must match** the information on the submitted PDF and add missing information.
- Attach completed coversheet to the PDF you will upload to ERIC [use Adobe Acrobat or other program to combine PDF files]—do not upload the coversheet as a separate document.
- Begin completing submission form at https://eric.ed.gov/submit/ and upload the full-text PDF with attached coversheet when indicated. Your full-text PDF will display in ERIC after the 12-month embargo period.

GRANTEE SUBMISSION REQUIRED FIELDS

Title of article, paper, or other content

Investigating Efficacy, Moderators and Mediators for an Online Mathematics Homework Intervention

All author name(s) and affiliations on PDF. If more than 6 names, ERIC will complete the list from the submitted PDF.

Last Name, First Name	Academic/Organizational Affiliation	ORCID ID
Murphy, Robert	RAND Corporation	0000-0001-8747-9816
Roschelle, Jeremy	Digital Promise	0000-0003-2219-0506
Feng, Mingyu	WestEd	0000-0001-9635-1611
Mason, Craig	University of Maine	0000-0002-9690-1692

Publication/Completion Date—(if *In Press,* enter year accepted or completed) April 23, 2020

Check type of content being submitted and complete one of the following in the box below:

- ☑ If article: Name of journal, volume, and issue number if available
- ☐ If paper: Name of conference, date of conference, and place of conference
- ☐ If book chapter: Title of book, page range, publisher name and location
- ☐ If book: Publisher name and location
- ☐ If dissertation: Name of institution, type of degree, and department granting degree

Journal of Research on Educational Effectiveness, Vol. 13 No. 2

DOI or URL to published work (if available) https://doi.org/10.1080/19345747.2019.1710885

Acknowledgement of Funding— Grantees should check with their grant officer for the preferred wording to acknowledge funding. If the grant officer does not have a preference, grantees can use this suggested wording (adjust wording if multiple grants are to be acknowledged). Fill in Department of Education funding office, grant number, and name of grant recipient institution or organization.

"This work was supported by U.S. Department of Education **[Office name]** National Center for Educat through **[Grant number]** R305A12012 to **Institution]** SRI Internation. The opinions expressed are those of the authors and do not represent views of the **[Office name]** National Center for Educa or the U.S. Department of Education.

Journal of Research on Educational Effectiveness

ISSN: 1934-5747 (Print) 1934-5739 (Online) Journal homepage: https://www.tandfonline.com/loi/uree20

Investigating Efficacy, Moderators and Mediators for an Online Mathematics Homework Intervention

Robert Murphy, Jeremy Roschelle, Mingyu Feng & Craig A. Mason

JOURNAL OF RESEARCH ON EDUCATIONAL EFFECTIVENESS
2020, VOL. 13, NO. 2, 235–270
https://doi.org/10.1080/19345747.2019.1710885

INTERVENTION, EVALUATION, AND POLICY STUDIES

Investigating Efficacy, Moderators and Mediators for an Online Mathematics Homework Intervention

Robert Murphy[a], Jeremy Roschelle[b], Mingyu Feng[c] and Craig A. Mason[d]

[a]RAND Corporation, Santa Monica, California, USA; [b]Digital Promise, San Mateo, California, USA; [c]WestEd, Redwood City, California, USA; [d]College of Education and Human Development, University of Maine, Orono, Maine, USA

ABSTRACT

We report on a randomized controlled trial of an intervention that leverages the availability of laptops for all public-school students in the state of Maine. The intervention, called "ASSISTments," provides feedback to students as they solve mathematics homework problems and automatically prepares reports for teachers about student performance on daily assignments. Teachers received training and coaching on formative assessment. Data was collected from 43 schools, 87 teachers, and 2769 7th grade students. Planned analyses describe use of the intervention, analyze the impact of the intervention on an end-of-year standardized assessment, and explore variables that may moderate or mediate impacts. Findings indicate that students in the schools assigned to use ASSISTments learned more and the impact was greater for students with lower prior mathematics achievement. Although evidence shows that teachers used the intervention to target instruction to students' needs, the mediating role of this practice was unclear. We also examined the generalizability of the findings and found generalizability to be limited due to the setting in Maine. Implications for policy, practice, and future research are discussed.

ARTICLE HISTORY
Received 15 August 2018
Revised 7 November 2019
Accepted 11 December 2019

KEYWORDS
Efficacy; randomized controlled trial; mathematics; homework; formative assessment; technology; online

Introduction

We report an investigation of an intervention that addresses increasingly common and important policy contexts—the one-to-one technology initiative, the assignment of homework in mathematics classrooms, and the use of computers to provide students with automated, immediate feedback and to support teachers' use of data. Increasingly, districts and states are making technology available to students for one-to-one use in school and at home. For middle school mathematics, requiring teachers to assign and review homework is a common policy and yet there is also widespread dissatisfaction with the value of homework, as we discuss below. A common use of technology in mathematics education is to provide students with feedback to guide their learning (Roschelle et al., 2017). Further, educators are also increasingly called upon to use data to adapt instruction and improve student learning (US Department of Education, 2016),

CONTACT Jeremy Roschelle jroschelle@digitalpromise.org Digital Promise, San Mateo, CA 94403, USA.
This article was originally published with errors, which have now been corrected in the online version. Please see Correction (http://dx.doi.org/10.1080/19345747.2020.1787760)

and formative assessment is a recommended practice in mathematics (National Council of Teachers of Mathematics, 2013). However, to best of our knowledge, this is the first rigorous, large-scale experiment at the intersection of all three of these common policy contexts: (a) one-to-one technology, (b) mathematics homework practices, and (c) immediate feedback and formative assessment.

We observe that despite the controversies around homework (cf. Cooper, 2015), assigning homework is still a standard practice, particularly in math classes, with middle school and high schools students spending anywhere from 30 min to several hours per week on math homework depending on the grade level and course (Loveless, 2014; Pope et al., 2015). Although reviews of the research have shown mostly positive effects of homework on students grades and achievement (Cooper et al., 2006; Fan et al., 2017), parents, teachers, and students are often dissatisfied with the degree to which homework supports learning. Given that homework is both a widespread policy and also often unsatisfying as a learning experience, it is a sensible target for improvement.

A range of factors that could be important for improving homework is discussed in the literature. For example, Cooper et al. (2012) group the factors as (1) demographic and other "given" factors, (2) assignment characteristics such as frequency and length, (3) classroom facilitators for doing homework, (4) home-community factors, (5) classroom follow-up, and (6) outcomes. The present study does not attempt to address or collect data on every factor in the available literature. Instead, its logic is organized around (1) how homework is completed by students and followed up by teachers in classrooms and (2) new possibilities for feedback and formative assessment when technology is available both for students to take home and to help teachers to analyze student work. As such, one early precedent is Elawar and Corno's (1985) experiment in which 18 elementary school teachers were trained to provide constructive, written feedback on the mathematics homework of their sixth-grade students. In the intervention group, teachers provided written feedback to each student three times weekly for 10 weeks. Students in the intervention group learned more, and the intervention was associated with smaller achievement gaps between girls and boys. Yet today, teachers rarely provide constructive feedback to every student on each homework assignment, likely because doing so is burdensome. This burden is changing in the technology era.

This article reports on a homework study that leveraged the availability of personal computing devices at home and in school in Maine. As a result of the state's one-to-one (computers-to-student) policy, the Maine Learning Technology Initiative (Silvernail et al., 2011) enables all middle school students to take home computers.[1] The study investigated a homework intervention, ASSISTments, that provides relevant feedback to students as they solve mathematics problems on their take-home devices and automatically prepares reports for teachers about students' answers to the problems. For example, teachers can see which homework tasks led to frequent errors and can inspect common wrong answers. To support the use of these capabilities, teachers in the study were also provided with professional development aligned with formative assessment practices (Black & Wiliam, 2010; Scriven, 1967) and coached on how to use "short-cycle" learning

[1]In 2000 the state of Maine, through the Maine Learning Technology Initiative (MLTI), decided to provide laptop computers to every 7th and 8th grade student in the state. Maine now supplies computing devices to 80,000 teachers and students each year, including high school students.

data (Wiliam & Thompson, 2007) to make instructional decisions. Many of the instructional decisions were closely related to how to structure their daily homework review time during class, that is, how teachers follow up based on information about how their students are performing on homework tasks.

In this study of ASSISTments use, the intended use of the ASSISTments platform was to support assigning, completing, and reviewing of mathematics homework.[2] Teachers in treatment schools were asked to use ASSISTments to assign homework problems from their existing textbooks using the platform so that students could receive automated feedback and so they would receive automated reports on their students' responses to homework items. Teachers were not asked to change math content or homework policies more generally. Nor were they asked to change how they give course grades to students, involve parents, or incentivize homework compliance. Thus, the study was narrowly focused on presumed advantages of technology regarding feedback, automated reporting, and support for formative assessment.

In a prior publication (Roschelle et al., 2016), a research team reported that the ASSISTments intervention had a positive main effect on student learning of 0.18 standard deviation units (Hedges' g) and a larger effect for students with lower prior math achievement ($g = 0.29$) relative to peers with higher prior math achievement ($g = 0.12$) (Roschelle et al., 2016). The estimated impacts were modeled using a student-school hierarchical linear regression (HLM) model, a two-level HLM. The re-analysis presented in this article employs a three-level HLM (student-classroom-school) to account for clustering at both the school and teacher levels and to allow for exploring how teacher practices and behaviors (mediation variables) may contribute to and help explain impacts of the intervention on student learning. (It is not possible to explore teacher effects in the prior two-level student-school model.) This article also analyzes how the effects of ASSISTments might vary for a broader set of policy-relevant student subgroups (whereas Roschelle et al., 2016 included a subgroup analysis based on prior math achievement only). To provide readers with greater context for the intervention as deployed in this study and to help interpret the impact results, this article also analyzes data on teacher and student use of ASSISTments during the school year (this data was captured by the ASSISTments platform and was not previously reported). Finally, we report new findings from a generalizability analysis that compares the characteristics of school districts nationwide to school districts in the study in Maine.

Literature Review

To support framing and interpreting the study, we discuss the literature on homework, research on feedback and formative assessment, and prior literature on ASSISTments.

Research on Homework

The value of homework and how much nightly homework to assign has been widely debated amongst school administrators, teachers, researchers, and parents (Bennett &

[2]ASSISTments can be used for other purposes including as a supplemental instructional activity for the whole-class or targeted interventions with specific groups of students including students in need of remediation or enrichment.

Kalish, 2006; Buell, 2004; Cooper, 2015; Cooper et al., 2006; Gill & Schlossman, 2004; Kohn, 2000; Pressman et al., 2015). In particular, critics of homework emphasize the time pressures it places on students, particularly those students involved in extracurricular activities or who carry a heavy academic course load or who work or have caretaker responsibilities at home (Pressman et al., 2015). Homework has also been discussed as a source of potential tension between students and parents including homes where parents heavily monitor their children's homework, in households with multiple school-age children, and where parents are working one or more jobs or the night shift or might not have the subject matter knowledge to provide their children with support (Karbach et al., 2013; Kralovec & Buell, 2000; Núñez et al., 2017). When students cannot do or learn from homework, their negative academic attitudes may increase; also teachers or parents may develop overly general negative expectations about students based on homework (Cooper et al., 2012). Inequities between homes in the resources available to support homework, including access to a quiet study space and technology, have also been cited by authors as potentially contributing to achievement gaps between low- and high-income families (McDermott et al., 1984; Scott-Jones, 1984). On the topic of parental involvement in homework, some researchers have suggested that parent involvement may be more important in elementary schools than in middle school and high school where, on average, parents might be less proficient in the content to effectively support their children (Patall et al., 2008). Finally, as previously mentioned, there are the burdens that homework places on teachers who need to plan, organize, and distribute the assignments, collect the student work and, when time allows, provide students with constructive feedback on their work.

Despite the ongoing debate about the value of homework, the research suggests that when properly used and developmentally appropriate there are several potential benefits of homework to both students and teachers. Homework can be a time for students to practice new skills learned during class and to prepare for future assignments (Cooper, 2015; Cooper et al., 2006), and the added academic time on task can produce achievement gains (Cooper et al., 2012). Proponents also argue that homework can provide students with opportunity for autonomy and to develop self-regulation skills (Bembenutty, 2011; Ramdass & Zimmerman, 2011), such as self-efficacy as a learner and resilience in the face of challenge (Alleman & Brophy, 1991). Research has found associations between (a) teachers' emphasis on student autonomy and self-regulation (b) better student effect and achievement (Trautwein et al., 2009). Other authors have promoted homework's value to keep parents connected to their children's teachers, school work, and learning process (McDermott et al., 1984; Patall et al., 2008).

Scholars have studied factors that can enhance the benefits of homework. As previously mentioned, providing actionable feedback to students enhances learning (Elawar & Corno, 1985; Hattie & Timperley, 2007; Shute, 2008); feedback and formative assessment are the focus of this study and will be discussed in more detail shortly. Other factors were not explicitly manipulated in this study, but are useful for interpreting it. Students' perceptions of homework quality are important to how much effort they put into homework. Homework can be perceived by students as higher quality based on the assignment itself, but also on the degree to which the homework process supports their learning and is relevant to what happens in class, such as the nature of teacher follow-up (Dettmers et al., 2010). A student's perception of homework quality is linked to their motivation to do it, per the expectancy-value framework (Trautwein, Lüdtke, Kastens, et al., 2006). Putting effort into homework makes increased learning from it more likely (Trautwein, Lüdtke, Schnyder, et al., 2006). One study has suggested that student autonomy in doing homework is a better predictor than effort (Fernández-Alonso et al., 2015). A study based on PISA found that providing more homework support resources to students was positively associated with achievement (Kitsantas et al., 2011). Overall, researchers have recommended to paying attention to three clusters of factors, the learning value of doing homework to the student, the motivational factors (including expectancy and value), and changes in teacher practices related to homework that increase learning (Trautwein & Köller, 2003). Also, there is a literature on homework compliance and ways to incentivize compliance. Overall, this literature has shown mixed results regarding factors that influence homework (see discussion in Trautwein et al., 2009). This study did not aim to improve compliance.

Research has also been conducted on potential moderator variables, which address how the effects of homework on learning vary for students, for example, by gender, prior achievement, and family income. Some opponents of homework have argued that differences in the homework support resources between low- and high-income families may contribute to existing achievement gaps due to students from higher-income families benefiting more from homework as these students, on average, are more likely to live in homes where more effective forms of support are available (McDermott et al., 1984; Scott-Jones, 1984). The issue of the availability of resources in the home to support homework may be particularly relevant for low achieving students and students with special learning needs who receive special supports in the classroom but may lack similar supports in the home (Katz et al., 2012). In contrast to some other student characteristics, gender difference in both time and effort spent on homework has been widely documented (OECD, 2015; Mau & Lynn, 2000; Xu, 2006). Compared to boys, girls in the United States and internationally have been found to spend more time doing homework, including math homework, and using more effective self-regulating practices while doing so (Gershenson & Holt, 2015; Trautwein, 2007).

Homework as Source of Formative Assessment and Feedback

A central role of the use of ASSISTments for homework is to help teachers engage in effective formative assessment, a theory- and research-based practice (e.g., Black & Wiliam, 1998; Heritage, 2013). Leading researchers and practitioners in formative assessment define it as a process comprised of multiple components working together to make a difference in teacher practice and student learning (Bennett, 2011; Black & Wiliam, 2009; Brookhart, 2007; Guskey, 2007; Heritage, 2013). One definition that is relevant to this study states:

> Assessment refers to all those activities undertaken by teachers, and by the students in assessing themselves, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged. Such assessment become "formative assessment" when the evidence is actually used to adapt the teaching to meet the needs (Black & Wiliam, 1998, p. 2).

This definition emphasizes that data on the students' recent performance is used to guide teachers on how to adapt instruction (see also Marsh et al., 2006; Means et al., 2010). Technology can play a role in this process. For example, in one rigorous study that found a positive effect of technology-supported formative assessment, teachers collected answers to mathematics problems from students via networked handheld calculators and used this to data adapt their instruction to fit students' needs (Pape et al., 2012).

The definition also encompasses students taking an active role in their own learning, for example, by monitoring their own performance and potentially self-regulating further learning behaviors. A relevant literature at the student level is organized around the concept of formative feedback (Shute, 2008). Not all feedback is equally helpful to students; more effective forms of feedback are specific to the task at hand, are delivered in a timely manner, and direct students toward specific revisions to their approach (Hattie & Timperley, 2007).

Broadly speaking, evidence from prior research suggests formative assessment (with or without technology) improves learning outcomes (Black & Wiliam, 1998; Hattie & Timperley, 2007), although there has been criticism of the quality of the research (Dunn & Mulvenon, 2009; Kingston & Nash, 2011). A 2011 meta-analysis of 19 studies of formative assessment in mathematics found a mean effect size of 0.17, with a 95% confidence interval ranging from 0.14 to 0.20 (Kingston & Nash, 2011). The effects of teachers' formative assessment practices may differ depending on whether the cycle from gathering data to making an instructional decision is immediate (e.g., within days) or happens after a longer period of time (e.g., within months) (Wiliam & Thompson, 2007). Others have highlighted the critical roles of allocating sufficient time for teachers to review and reflect on the data and their instruction and the availability of meaningful professional development opportunities to help teachers learn how to make sense of different sources of learning data and use it to inform subsequent instructional decisions (e.g., Bennett, 2011). Sufficient time for review and planning and meaningful professional development is often lacking from many formative assessment interventions. More recent studies of formative assessment-based interventions have found no or mixed effects (Carlson et al., 2011; Cordray et al., 2013; Konstantopoulos et al., 2013; Quint et al., 2008). In many of these studies that reported a no-effects finding, the interventions involved the use of infrequent assessments, such as interim or benchmark tests and provide feedback across a broad set constructs, content areas, and skills. In contrast, the homework intervention considered in this current study uses technology to provide regular and immediate feedback (i.e., several times a week) to students and teachers at the individual task, problem, and topic level. It also provides teachers with professional learning opportunities and coaching geared toward helping teachers use data from nightly homework to inform instructional decisions. Through assigning and collecting student homework, teachers may also use homework as a source of formative feedback on their own instruction, including as an opportunity to identify students who are struggling with different areas of the content, as a gauge on the effectiveness of their lessons, and as a basis to modify subsequent instruction (Cooper, 1989; Kingston & Nash, 2011; Wiliam & Thompson, 2007).

Prior Research on ASSISTments

Broadly, this study is also concerned with how to effectively use educational technology for mathematics teaching and learning. By national and local policy, schools are increasingly encouraged to use educational technology for all school subjects (e.g., US Department of Education, 2016) and studies have generically reported small but positive effects of doing so (Kingston & Nash, 2011). Yet mathematics teachers are reported to be the least intensive users of educational technology (e.g., Becker, 2000); further the literature is replete with mixed results, depending on which specific technology and approach is implemented (Cheung & Slavin, 2013). This study explores one promising type of use of educational technology, which is for feedback and formative assessment (see Roschelle et al., 2017 for types of uses in mathematics).

ASSISTments is a Web-based platform that uses a cognitive tutoring approach developed at Carnegie-Mellon University to support mastery learning in mathematics (Bloom, 1971) and to provide formative feedback on learning progress to students and teachers (Feng & Heffernan, 2006). Consistent with Bennett's (2011) recommendation to provide teachers with support to implement more effective formative assessment practices, the developers of ASSISTments have also invested considerable resources on the development of a coaching model to support teachers use of student performance data provided by the platform to inform instructional decisions and better support their students.

Prior effectiveness research studies on ASSISTments had reported positive effects on student math learning, but, in contrast to the study in Maine, those studies were relatively short in duration and involved small numbers of teachers (Kelly et al., 2013; Mendicino et al., 2009; Singh et al. 2011). Recent research of the use of the platform in middle school has reported greater learning gains for students with lower prior mathematics achievement relative to others (Fyfe, 2016).

Current Study

The research results and findings reported in this article build off prior analyses of a randomized control trial study of ASSISTments in the state of Maine (Roschelle et al., 2016). In the study, the ASSISTments platform was used by 7th grade teachers to assign mathematics homework. Student received immediate feedback on their performance and teachers received performance reports at the class- and individual-level for each assignment and professional development and coaching on how to use the data to inform instructional decision making. The new analyses described below examine (1) how the effects of ASSISTments vary for different types of students; (2) whether a teacher formative assessment practice facilitated by the platform (focusing homework review on problems students' found the most difficult) can explain (or mediate) the effects of the use of the platform on student learning; (3) the association between levels of teacher and student use of the platform and student math achievement; and (4) the generalizability of the findings in the state of Maine to districts in other states across the United States.

Methods

Research Design

Following the design set out in a proposal funded by the U.S. Department of Education's Institute of Education Sciences, the evaluators implemented a school-level, randomized controlled trial design to test the efficacy of the use of ASSISTments to support math homework in 7th grade classrooms in the state of Maine. Forty-six schools, each with at least one 7th grade math teacher, were recruited. Schools were randomly assigned to an immediate use of ASSISTments condition (treatment) or a delayed use of ASSISTments condition (control). Prior to random assignment, schools were blocked on school type (K-8 or other). Within blocks, schools were matched on prior 6th grade state assessment scores. Three schools could not be paired and were treated as a single group for random assignment.

Schools remained in their assigned condition for two full academic years (see the research design in Figure 1). After the second year, the control schools were given access to ASSISTments and teacher professional development to support its use (i.e., delayed access). For schools assigned to the treatment condition, the first year of implementation was a pilot or "warm up" year, allowing teachers to become familiar with the ASSISTments' features and to practice using it. Treatments teachers used ASSISTments in the second implementation year with a new cohort of 7th grade students. The primary data collection activities (e.g., interviews, observations, surveys, logs) were conducted in the second year of implementation along with the administration of standardized math achievement assessments in the spring to assess the impacts of ASSISTments use on students' math learning.

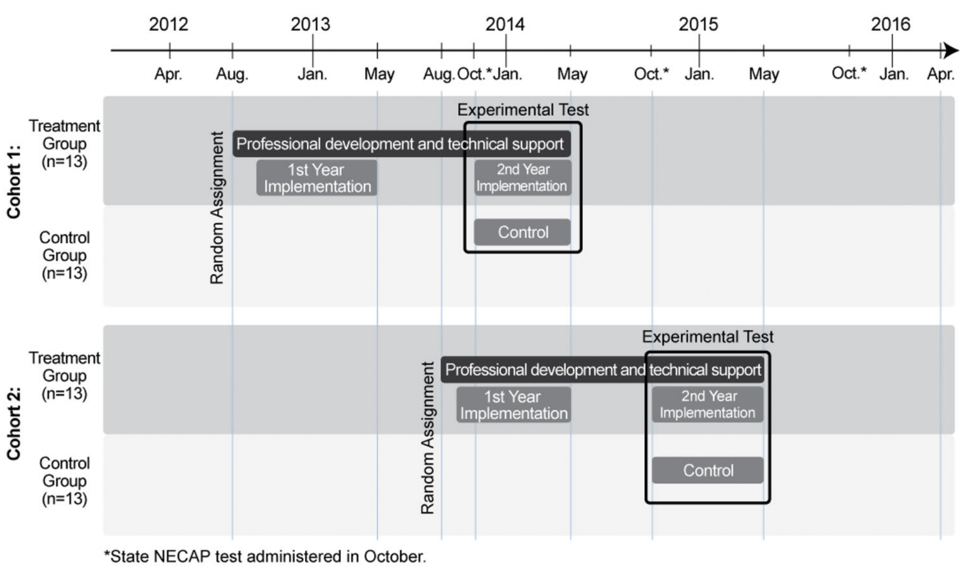

Figure 1. Study design. Two cohorts were recruited in consecutive years and schools in each cohort were randomly assigned to condition.

School and Teacher Sample

All schools that completed a participant application to confirm their commitment to meeting the requirements of the research, including participation in all research activities, were accepted into the sample. The 46 schools in the original sample served a mix of grade levels including K-8, 6-8, and 7-8. Schools were recruited in two cohorts with the 1st cohort of schools starting in the research in the summer of 2014 ($N = 17$) and the 2nd cohort ($N = 29$) the following summer, 2015. The research team documented details of the recruitment procedures in a separate technical report (Roschelle et al., 2014). Of the original 46 schools recruited, one dropped out after being matched, but prior to randomization. In this case, its pair school was re-matched and randomly assigned. A second school (treatment condition) dropped out after the start of the project. While its paired school remained in the project, data from both schools in this pair were excluded from analyses, resulting in a final analytical sample of 43 schools, 22 treatments, and 21 controls. Eighty-seven teachers participated from these schools; many schools had only one 7th grade mathematics teacher. The school-level overall attrition rate was 6.5%, and the school-level differential attrition rate was 3.8%. Details regarding student-level attrition are presented below.

The ASSISTments Intervention in Maine (Treatment Condition)

As previously discussed, the intervention consisted of assigning, doing, and reviewing student homework using ASSISTments, along with related professional development on formative assessment best practices. Seventh grade math teachers in treatment schools were expected to use the ASSISTments platform to assign math homework at least 3 nights per week. Students in the treatment schools accessed their math homework assignments by logging on to ASSISTments. Students worked on their assignments during and outside the regular school day. All seventh-grade students were allowed to take their school-issued laptop home at the end of the school day. If students did not have access to the Internet at home, students could download assignments while they were at school and upload their solutions to the problem set the next day. When working in off-line mode, students still receive immediate feedback on the correctness of their answers.

Using the ASSISTments platform, teachers could assign problems from their textbook, create their own problems to assign, or assign problems from an existing set of Skill Builder problems, a common set of problems available to all users of the platform that support mastery learning. With regard to problems from textbooks, WPI entered into the platform the item and page number and correct solution for each end-of-unit textbook problem for every textbook used by teachers in the treatment schools. When a homework assignment includes problems from the textbook, students refer to the actual problem in their textbook, can use paper and other resources (e.g., a calculator) to do their work, and enter their solutions in the ASSISTments platform. Students then receive immediate feedback on the correctness of their solution (correct or incorrect). If the solution is incorrect, the students make additional attempts and enter new solutions.

In contrast to the assignment of problems from the textbook, Skill Builder problems are available in a library on the platform and are indexed by state math standards. If they are struggling with a solution to a particular Skill Builder problem, students can request and receive step-by-step hints to the solution of the problem. Students receive similar problem types until they answer a string of three problems correctly (with a limit of ten unsuccessful attempts). To check on the retention of the acquisition of knowledge and skills, teachers also have the option of turning on a feature that reassesses students automatically on skills one week or two weeks after the successful completion of Skill Builder problem set (few teachers in the study used this feature). Ultimately, teachers decided how much and what type of homework was assigned, and they were asked to do so in accordance with their existing school homework policy. Further, as previously discussed, no attempt was made to intervene with regard to factors that have been found to have uncertain benefits to learning, such as addressing student's compliance on homework, how teachers gave student grades or incentives, or how parents were involved.

Nightly Reports on Homework Performance

Teachers receive reports via email or by logging in to the platform. The reports allow teachers to quickly review whether students completed the assignment, student performance on each problem, the average percentage of problems answered correctly for individual students as well as for the class, and the most common wrong answers per problem at the classroom level.

Teacher Professional Development and Technical Assistance

WPI hired a former math teacher and experienced ASSISTments user to provide professional development and technical assistance to teachers in treatment schools. Professional development consisted of a single two-day summer training and an additional day of technical assistance distributed across the school year via webinars, email, and in-person visits. The first summer training session introduced teachers to the platform and helped prepare them for using it with their students including how to create and deliver assignments. A second summer training session, after the first year of use, focused on advanced features of ASSISTments and the practice of formative assessment. Teachers also received specific guidance on how to use the information in the reports to facilitate their daily homework review time in the classroom including focusing attention on the homework problems that students had the most difficulty with, reviewing correct solution procedures for those problems as well as the common wrong answers to address underlying misunderstandings.

Homework Practices in the Control Condition

Participating teachers in control schools were expected to continue to implement any homework practices already in place or that were planned during the course of the study ("business as usual condition"). This included the use of the school's learning management system to assign homework or the use of an online homework resource. Teachers in both the treatment and control schools were expected to abide by any formal district homework policy. Previously reported findings from classroom observations in study classrooms (Fairman et al., 2016) found that teachers in the control group spent about the same amount of class time reviewing homework (23% of class time) as in the treatment group (25%).

Instruments and Measures

A variety of instruments were used to collect data relevant to the logic model, to monitor implementation of ASSISTments in the treatment classrooms, and assess the contrast in homework and homework review practices in classrooms across both experimental conditions.

Student Demographic Information

Student demographic information was collected directly from the Maine Department of Education's State Longitudinal Data System. This included student gender, free and reduced-price lunch status, and individualized education program (IEP) status. In addition, the evaluators collected classroom rosters from each teacher participating in the study. Rosters included information that identified the specific class, student names, grade level (for mixed grade classes), and students' special education and IEP status.

Student Prior Mathematics Performance

This experiment used students' 6th grade mathematics scores on the New England Common Assessment Program (NECAP) as a measure of prior achievement. The test was administered in the fall of 6th grade. Data for this statewide standardized assessment was also collected directly from the State Longitudinal Data System. Due to the state's transition from the NECAP to a Common Core-based state-wide standardized assessment, the last available NECAP scores were from Fall 2013.

Student 7th Grade Mathematics Performance

During the period of the study, Maine phased out the NECAP as their statewide assessment and did not administer any assessment during the 2014–2015 school year. As a result, the research team purchased and administered the TerraNova Common Core Assessment for mathematics at the end of 7th grade so that a consistent achievement measure would be collected across the participating schools in the two cohorts. The Terra Nova is published by Data Recognition Corporation and is nationally normed. Teachers were trained by the research team on how to administer the assessment. Completed test booklets were returned to the test publisher for scoring and scorers were blind to the assigned condition.

Teacher Homework Review Practices

Measures of teachers' homework review practices were based on data from teacher instructional logs. All teachers in both conditions were asked to complete daily 10 minute online logs across three different weeks selected by the research team and distributed equally across the second implementation year. At the end of each day during the selected weeks, teachers were asked to report if they reviewed the previous night's homework with their classes that day and how they chose the problems that they reviewed from the assignment (e.g., whether they reviewed all problems or reviewed only a select group of problems). Figure 2 shows the relevant item from the daily instructional log for teachers in the treatment group. We constructed separate measures based on the percentage of the total numbers of daily log days completed that a teacher reported the use of the practices represented in the sub-items of the question.[3]

System Use Data

The research team also used the electronic records collected by the ASSISTments system as a source of implementation data for the treatment group. For each classroom, the system collects time-stamped data (time and date) for each interaction with the platform. Data were extracted for the 23 treatment schools, 43 7th grade teachers, and 2,327 students. The data was extracted from the second implementation year (2013–2014 school year SY for schools in cohort 1, and 2014–2015 SY for schools in cohort 2). Teacher use measures computed from system use data included the number of weeks with one or more problem set assigned, total number of assignments, percent of student performance reports opened, and the percentage of problems assigned that were Skill Builder problems. The single student use measure included in the analysis was the total number of problems attempted on the ASSISTments platform.

6. How did you determine which problems to review with students? *(Mark all that apply.)*	
I reviewed all the problems	☐
I reviewed a sample of problems that I think my students might have difficulty on	☐
I used ASSISTments reports to identify difficult problems for students	☐
I used reports provided by other computer programs to identify the hard problems. Specify the program(s): ______________________	☐
I asked my students what they wanted to go over	☐

Figure 2. Relevant item from the daily instructional log (treatment group) used to construct measures of *homework review* practice (See related footnote for details on construction of the targeted homework review measure).

[3]Multiple sub-items were combined to construct the *targeted homework review* measure. In a prior item, if teachers indicated that they reviewed homework with their class that day, they were then asked to report how they determined which problems to review (see Figure 2). Teachers in the treatment condition were considered to use a *targeted homework review* practice if they selected one or more of the following responses—*I reviewed a sample of problems I thought teachers had the most difficulty on*, *I used ASSISTments reports to identify difficult problems for students*, and *I used reports from other computer programs to identify hard problems* (Teachers in the control condition did not see the sub-item about the use of ASSISTments). We then counted up the number of times a teacher self-reported they used a targeted review practice and divided this by the total number of daily instructional logs completed.

Data Analyses

We conducted four types of analysis:

1. A descriptive analysis of the sample and sample attrition, as well as the use of ASSISTments by teachers and students in the treatment condition.
2. An analysis of the impact of the intervention on student achievement in mathematics, including an analysis of student subgroups as potential moderators.
3. An analysis of a teacher homework review practice as a potential mediator between ASSISTments use and student learning.
4. An analysis of the generalizability of our findings to other settings in the United States.

Descriptive Analysis

An analysis of sample and sample attrition was conducted using the student demographic information and prior achievement measures (6th grade NECAP scores in reading and mathematics) provided by the state of Maine, classroom roster information provided by the districts, and records of students who completed a TerraNova assessment administered by the research team. We tracked both students who joined and left the study after the start of the school year (*joiners* and *leavers*), to examine differential attrition rates between conditions, and to see whether the population that dropped out of the study differed systematically from the students who remained. The descriptive analysis of classroom usage of ASSISTments was conducted using system use data collected automatically by the platform.

Impact Analysis

For the main impact analysis, we investigated the main effect as well as possible moderators and mediators of impacts. We used a three-level hierarchical model in which students were nested within classrooms and classrooms nested within schools. Student-level covariates used in the model included: prior 6th grade NECAP math score, student gender (0 = female, 1 = male), free and reduced-price lunch status (0 = not free/reduced eligible, 1 = free/reduced eligible), and individualized education program status (0 = no IEP, 1 = IEP). Classroom-level covariates included the mean prior 6th NECAP math score for all students in the classroom, the variance in prior 6th grade NECAP math scores for all students in the classroom, and the number of students in the classroom. In addition to the school-level treatment condition (control = 0, ASSISTments = 1), and 20 school matched-pair indicator variables, school-level covariates included the mean prior 6th grade NECAP math score for all current 7th grade students, the school-wide percentage of students receiving free and reduced-price lunch, and the size of 7th grade enrollment. All analyses were conducted with students' mathematics TerraNova scale scores as the dependent variable. Table 1 shows the Pearson correlations amongst the covariate, treatment status, and outcome variables.

The three-level hierarchical regression model used to estimate the main impact of assignment to the immediate use of ASSISTments on student TerraNova performance is shown below:

Table 1. Pearson correlations amongst the covariate (student-level), treatment status, and outcome variables in the three-level hierarchical linear model.

	1	2	3	4	5	6	7	8	9	10
Male	—	0.135**	−0.041*	−0.019	−0.015	−0.01	−0.226**	−0.218**	0.031	−0.037
IEP status	0.135**	—	0.094**	−0.004	−0.313**	−0.295**	−0.335**	−0.334**	0.007	−0.310**
Free/reduced-price lunch	−0.041*	0.094**	—	0.079**	−0.267**	−0.271**	−0.236**	−0.232**	−0.058**	−0.266**
Ethnic minority[a]	−0.019	−0.004	0.079**	—	−0.084**	−0.078**	−0.018	−0.027	−0.01	−0.073**
Grade 6 math (scaled score)[b]	−0.015	−0.313**	−0.267**	−0.084**	—	0.969**	0.633**	0.630**	0.007	0.749**
Grade 6 math (percentile)	−0.01	−0.295**	−0.271**	−0.078**	0.969**	—	0.629**	0.636**	0.01	0.753**
Grade 6 reading (scaled score)	−0.226**	−0.335**	−0.236**	−0.018	0.633**	0.629**	—	0.967**	−0.015	0.546**
Grade 6 reading (percentile)	−0.218**	−0.334**	−0.232**	−0.027	0.630**	0.636**	0.967**	—	−0.014	0.550**
Treatment	0.031	0.007	−0.058**	−0.01	0.007	0.01	−0.015	−0.014	—	0.115**
7th grade math score (TerraNova)	−0.037	−0.310**	−0.266**	−0.073**	0.749**	0.753**	0.546**	0.550**	0.115**	—

Note. [a]Ethnic Minority includes Asians, Blacks and Latinos with Asians representing 25% of the nonwhite sample.
[b]New England Common Assessment Program.
*$p < .05$.
**$p < .01$.
***$p < .001$.

Level 1 Model (Student)

$$TN_{ij} = \pi_{0jk} + \pi_{1jk}PriorMath_{ijk} + \pi_{2jk}Male_{ijk} + \pi_{3jk}IEP_{ijk} + \pi_{4jk}FRL_{ijk} + e_{ijk}$$

Level 2 Model (Classroom)

$$\pi_{0jk} = \beta_{00k} + \beta_{01k}MathCla_{jk} + \beta_{02k}ClaVar_{jk} + \beta_{03k}NCla_{jk} + r_{0jk}$$
$$\pi_{\bullet jk} = \beta_{\bullet 0k}, \quad \bullet = 1 \text{ to } 4, \text{ for } \pi_{1jk} \text{ through } \pi_{4jk}$$

Level 3 Model (School)

$$\beta_{00k} = \gamma_{000} + \gamma_{001}Trx_j + \gamma_{002}MathSch_k + \gamma_{003}FRLSch_k + \gamma_{004}NSch_k + \sum_{x=1}^{20}\gamma_{00(x+4)}SchPair_k + u_{00k}$$

$$\beta_{0\bullet k} = \gamma_{0\bullet 0}, \quad \bullet = 1 \text{ to } 3, \text{ for } \beta_{01k} \text{ through } \beta_{03k}$$

$$\beta_{\bullet 0k} = \gamma_{\bullet 00}, \quad \bullet = 1 \text{ to } 4, \text{ for } \beta_{10k} \text{ through } \beta_{40k}$$

The analyses were conducted in three steps. First, we tested for the main effect of the treatment on student math achievement using the model above ($\gamma 001$). Second, we tested for possible moderator effects associated with various student characteristics (prior Mathematics achievement, gender, special education status, and free and reduced-price lunch status) by adding an interaction term (treatment indicator by student characteristic) at the school level, testing the model separately for each variable and with all moderator interaction terms in the model (complete moderation model). Third, we examined the role of teacher-reported homework review behavior (the extent to which they focused their daily homework review activity on the homework problems students had the most difficulty solving) as a potential mediator of the effect of ASSISTments on TerraNova scores.

Mediation Analysis

Next, we analyzed whether a specific homework review practice that can be facilitated by the ASSISTments platform—using in-class time set aside for homework review to focus on those problems that proved to be the most difficult for students—might help explain (or *mediate*) the effect of ASSISTments use on students' math achievement. While there are several strategies for assessing mediation, each have their own limitations (Cheung, 2009; MacKinnon et al., 2002; MacKinnon & Pirlott, 2015; Pituch, et al., 2005; Schochet et al., 2014). In general, all the strategies share a common causal framework represented by Figure 3 for a single mediator mediation model. In the *unmediated model*, path *c* represents the total effect of the treatment on student test scores (γ001 in the Level 3 Model shown above). In the *mediation model*, the hypothesized mediation paths, *a* and *b*, represent the paths through which the treatment may have its effect on student achievement scores (i.e., the use of ASSISTments leads to a change in teacher practice and it is the change in teacher practice that leads to the gains in student achievement relative to the control group). Path *a* is the effect of the treatment on the teacher practice mediator variable (targeted homework review) while path *b* is the effect of the mediator on students' spring TerraNova score. The mediation effect (also known as the "indirect" effect) is the product of the two mediation paths (*ab*). Path *c*' is the direct of effect of the treatment on student test scores. The total effect of the treatment on the student achievement (*c*) is the direct effect (c') plus the mediation effect (*ab*).

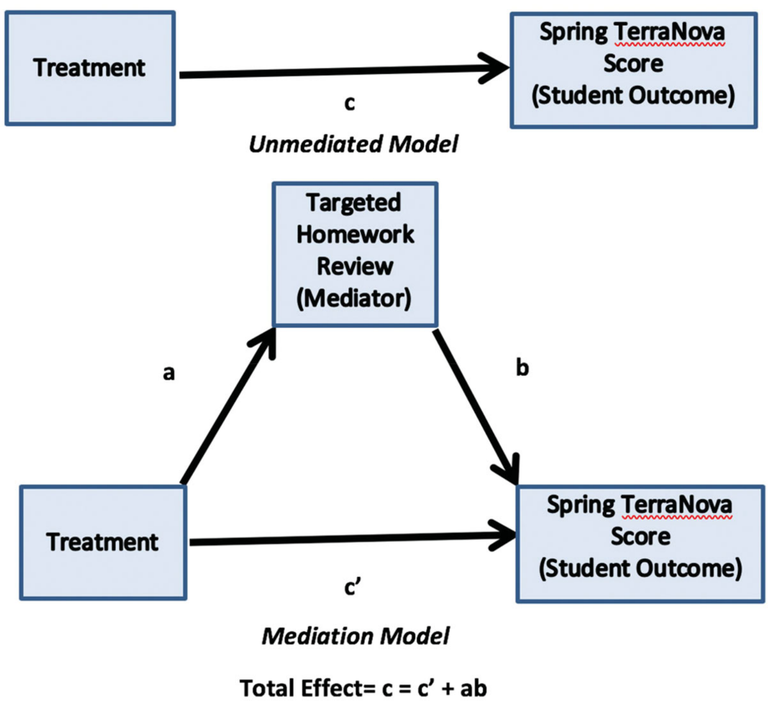

Figure 3. Single mediator mediation path diagram. Path diagram showing (1) the total effect (path *c*) of the treatment (ASSISTments) on the student outcome (math achievement) and (2) the indirect effect of the treatment on the student outcome (paths *a* and *b*) through the mediator (targeted homework review).

To estimate the mediation effect (*ab*) and test for its statistical significance we implemented two different strategies—a joint test of significance approach (MacKinnon et al., 2002) and the empirical M-test (MacKinnon et al., 2004).[4] The mediation paths *a* and *b* were estimated using hierarchical regression models that accounts for clustering and included covariates at each level of the models and the interactions of the covariates with the treatment status indicator. For the *joint test of significance* approach, the first step is to estimate the effect of the treatment on the mediator using a two-level regression model (teachers clustered within schools). If this treatment effect is not statistically significant, then the analysis is halted with a finding that no mediation is present since there is no statistically reliable evidence that the treatment changed the hypothesized teacher practice mediator. If the effect of the treatment on the mediator is statistically significant, the next step involves the use of a three-level regression model (students clustered within classrooms within schools) to estimate the effect of the mediator on the outcome variable controlling for the treatment condition. If the effect of the mediator on the outcome is statistically significant, mediation is indicated, without implying causality since the mediator was not experimentally manipulated (i.e., teachers were not randomly assigned to different levels of the mediator variable). While this approach is commonly used and its logic is easy to follow, the approach does have limitations including (1) a requirement that the path coefficients *a* and *b* are uncorrelated, (2) it doesn't provide confidence intervals for the indirect effect (*ab*) and (3) has lower statistical power to detect a mediation effect in the case of multi-level mediation (e.g., effect of a teacher practice on a student outcome) relative to the empirical M-test approach (MacKinnon et al., 2002; Pituch et al., 2005). For the empirical M-test approach (MacKinnon et al., 2004), the first step is to estimate the indirect effect by taking the product of the path coefficients from treatment to mediator (*a*) and from mediator to outcome (*b*). These parameters were estimated in steps 1 and 2 of the joint test of significance approach. The statistical significance and (asymmetric) confidence intervals for the product *ab* (indirect effect) are calculated using the PRODCLIN procedure in SAS and the values of the estimated path coefficients *a* and *b* along with their standard errors (MacKinnon et al., 2007). If the indirect effect is statistically significant (i.e., the confidence interval does not include 0), mediation is indicated, without implying causality.

Analysis of Relationship Between System Use and Math Achievement

We also examined how end-of-year math achievement varied with ASSISTments use within the treatment schools. Specifically, in a single three-level hierarchical regression model, we examined the relationship between several measures of teachers and students use of ASSISTments and student math achievement. These analyses were restricted to the treatment sample only. Specifically, we analyzed whether the following teacher use variables were significant predictors of student achievement: (1) the number of weeks in which at least one problem set was assigned, (2) the number of total homework problem sets assigned, (3) the percent of problems assigned that were Skill Builders, and (4) the percent of problem sets for which a teacher reviewed a report. In addition, in the same model, we also analyzed whether the number of

[4]The *joint test of significance* approach is a variation of the widely used causal steps approach (Baron & Kenny, 1986).

homework problems completed by a student on ASSISTments was a significant predictor of math achievement. All of the use measures analyzed were based on data automatically captured by the ASSISTments platform.[5] The total number of weeks used, total problem sets assigned, and the rate at which teachers reviewed the student performance reports are proxy measures of teachers' commitment to complying with the research team's expectation for the use of ASSISTments and its features to support instruction. Including a measure of the extent to which teachers assigned problems from the system's Skill Builder problem library allows us to analyze whether greater exposure to mastery-based problem sets is associated with higher math achievement scores. Finally, to examine the association between students with more direct exposure to the ASSISTments platform and immediate feedback on homework and student math achievement we included in the model a measure of the total number of problems completed by each student.

Generalizability Analysis

Since the schools recruited into the sample were not randomly selected from the population of schools in the state Maine, we used the Generalizer software tool (thegeneralizer.org) to assess the extent to which the impact findings from the study's school sample might apply to all 547 schools in the state of Maine as well as to schools in other states in the United States (Tipton & Miller, 2016). The tool computes a generalizability index between 0 and 1 using pretreatment covariates (school size, characteristics of the school population, urbanicity, district size, and percent unemployment) and a propensity score that summarizes the degree of similarity between the study sample and the inference population and the extent to which the study's Average Treatment Effect (ATE) is unbiased for the inference population (Tipton, 2014). The Generalizer software tool utilizes school data from the National Center for Education Statistics' Common Core of Data and the U.S. Census Bureau's American Community Survey dataset.

Study schools were compared to the population of schools in Maine and the other 49 states in the U.S. meeting the inclusion criteria based on a propensity score that estimates the probability that a school in the population would be selected into the study given its value on the covariates. The larger the generalizability index, the more similar are the sample and inference population schools and, thus, the higher the probability the impact estimates from the study generalize to the inference population. An index above 0.90 indicates "very high generalizability"; 0.70–0.90 "high generalizability"; 0.50–0.69 "medium generalizability"; and below 0.50 "low generalizability."[6]

[5]For our measure of the rate at which teachers reviewed the performance reports generated by the ASSISTments system, for each problem set assigned, we counted the number of occasions teachers "opened" a report by clicking on a link in the online teacher interface that takes the teacher to a presentation of class and student performance on each problem in the set.

[6]A generalizability index in the high and medium ranges indicate statistical adjustments are needed to remove potential bias due to difference between the study sample and inference populations (Tipton, 2014). While statistical adjustments are expected to perform well when an index is in the high range, statistical adjustments may not remove all of bias when the index is in the medium range. When the index is less than 0.50, statistical adjustments are not expected to perform well and generalization of the results to the inference population is unjustified.

Results and Findings

Sample, Attrition Rates, and Baseline Equivalence

As reported above, 43 schools and students in classroom taught by 87 teachers were included in the analytical sample. In finalizing the student-level analytic sample, we first analyzed the students who left the study (leavers) and joined the study (joiners) after the study schools were randomly assigned to condition. With regard to *leavers*, student-level attrition from the study was based on student's enrollment in 6th grade, the grade the students were entering when the study commenced (1st Year Implementation). The initial sample based on 6th grade enrollment was 3,035 students, of whom 2,653 remained enrolled in their schools during the course of the study and had a valid TerraNova assessment score at the end of 7th grade. Of the 387 students who did not complete the study, 221 were in the control condition and 161 were in the ASSISTments condition. The resulting overall student-level attrition was 12.6% with a differential attrition between Treatment and Control groups of 6.9%, a level which is at the boundary between (a) *acceptable* and (b) *acceptable under optimistic assumptions* in relation to potential bias associated with individual-level attrition (What Works Clearinghouse, 2017).

Table 2 shows the prior achievement scores and demographics for students in the original sample at the start of the school year by experimental condition. The sample was overwhelmingly White (92%) and just under 40% received free or reduced-price lunch (39%). Thirteen percent of the student sample had an individualized education program indicating that these students potentially received some type of specialized instruction and related services during their mathematics instruction. We found small statistically significant differences ($p < 0.01$) between conditions on two baseline variables including percentage of free and reduced-price lunch and percentage ethnic minority (Latinos, African-Americans, Asians, American-Indian, and 2 or more races) with students in treatment schools being less likely to receive free and reduced-price lunch (37% versus 42%) and less likely to be a member of an ethnic minority (7% versus 10%) compared to students in the control schools. There were no statistically significant differences

Table 2. Mean and standard deviation of demographic characteristics and prior test scores for all students enrolled in the study schools at the start of the school year and separately for treatment and control group students along with results of tests for group differences.

	All ($N = 3{,}035$)		Treatment ($N = 1{,}689$)		Control ($N = 1{,}346$)		
Variables	Mean	(SD)	Mean	(SD)	Mean	(SD)	DIfference (T-C)
Male	0.50	0.50	0.51	0.50	0.48	0.50	0.03
IEP status	0.13	0.34	0.13	0.34	0.13	0.34	0
Free/reduced-priced lunch status	0.39	0.49	0.37	0.48	0.42	0.49	−0.05**
Ethnic minority[a]	0.08	0.27	0.07	0.25	0.10	0.30	−0.03**
Grade 6 math (scaled score)[b]	644.25	11.20	644.59	10.78	643.83	11.71	0.76
Grade 6 math (percentile)	54.03	27.40	54.74	26.87	53.13	28.03	1.61
Grade 6 Reading (scaled score)	648.10	11.60	647.94	11.25	648.30	12.04	−0.36
Grade 6 reading (percentile)	54.99	27.74	54.60	27.48	55.49	28.07	−0.89

Note: [a]Ethnic minority includes Asians, Blacks and Latinos with Asians representing 25% of the nonwhite sample.
[b]New England Common Assessment Program.
**$p < .01$.

between the conditions for 6th grade standardized test scores in reading or mathematics (Table 2).

We also tracked the students who joined a study school after the start of the school year. One hundred and sixteen students (57 in the control group and 59 in the treatment group) joined a study school after the start of the school year and completed the study by taking the TerraNova assessment in 7th grade. Since joiners in each condition were equivalent on the baseline 6th grade reading and math scores, joiners were included in the final analytical sample (see Table 3).[7] Thus the total student sample included in the impact analyses was 2,769 students comprised of the 2,653 students who were initially enrolled and completed the study plus the 116 joiners (Table 4).

Table 3. Mean and standard deviation of demographic characteristics and prior test scores for students who joined the study after the start of the school year (joiners) by treatment and control groups along with results of tests for group differences.

	Treatment ($N = 59$)		Control ($N = 57$)		Difference
Variables	Mean	(SD)	Mean	(SD)	(T-C)
Male	0.34	0.48	0.49	0.5	−0.15
IEP status	0.19	0.39	0.16	0.37	0.03
Free/reduced-priced lunch status	0.61	0.49	0.75	0.43	−0.14
Ethnic minority[a]	0.19	0.39	0.04	0.19	0.15**
Grade 6 math achievement (scaled score)[b]	639.42	11.73	642.4	10.56	−2.98
Grade 6 math achievement (percentile)	43.52	26.23	48.14	26.03	−4.62
Grade 6 reading achievement (scaled score)	646.02	12.93	644.51	10.54	1.51
Grade 6 reading achievement (percentile)	50.25	28.33	45.82	26.71	4.43

Note: [a]Ethnic Minority includes Asians, Blacks and Latinos with Asians representing 25% of the nonwhite sample.
[b]New England Common Assessment Program.
$^{**}p < .01$.

Table 4. Mean and standard deviation of demographic characteristics and prior test scores for all students in the final analytical sample and separately for the treatment and control group students along with results of tests for group differences.

	All ($N = 2,769$)		Treatment ($N = 1,587$)		Control ($N = 1,182$)		Difference
Variables	Mean	(SD)	Mean	(SD)	Mean	(SD)	(T-C)
Male	0.49	0.50	0.50	0.50	0.47	0.50	0.03
IEP status	0.11	0.32	0.11	0.32	0.11	0.31	0.00
Free/reduced-price lunch status	0.38	0.49	0.35	0.48	0.41	0.49	−0.06**
Ethnic minority[a]	0.07	0.26	0.07	0.25	0.07	0.26	0.00
Grade 6 math achievement (scaled score)[b]	645.01	10.55	645.07	10.44	644.93	10.69	0.14
Grade 6 math achievement (percentile)	55.73	26.54	55.96	26.26	55.42	26.91	0.54
Grade 6 reading achievement (scaled score)	648.56	11.34	648.41	11.11	648.75	11.66	−0.34
Grade 6 reading achievement (percentile)	56.07	27.16	55.75	27.04	56.51	27.33	−0.76

Note: [a]Ethnic Minority includes Asians, Blacks and Latinos with Asians representing 25% of the nonwhite sample.
[b]New England Common Assessment Program.
$^{**}p < .01$.

[7]Students joining the treatment schools were more likely to be a member of an ethnic minority—19% of the joiners in the treatment group compared to 4% in the control group—or 22 of the 59 joiners in the treatment group and 5 of the 57 students in the control group.

Teacher and Student Use of ASSISTments

Measures of teachers' use of ASSISTments were compiled by aggregating data from the platform's user log files at the section or class level. Many teachers taught multiple sections or classes. There was variation in how teachers between schools used ASSISTments, as well as within schools and, often, how the same teachers used the platform across different sections. Thus, for the purpose of describing teachers and students use of ASSISTments, for each use variable, we computed a separate use statistic for each section a teacher taught and then aggregated these measures across sections. Table 5 shows the descriptive statistics for the teacher and student ASSISTments use variables.

To assess the frequency to which ASSISTments was used for homework in treatment classrooms, we counted the number of weeks in which a teacher assigned a class at least one ASSISTments assignment. Although there are 38 weeks in a school year, teachers and students only had access to ASSISTments for a portion of this time. In Maine, it takes schools up to 6 weeks from the start of the school year to distribute laptops to students. Further, teachers typically did not use ASSISTments during weeks in which students were reviewing for or taking state-mandated tests. Overall, we found that the median classroom used ASSISTments for 22 weeks or little more than half of the instructional weeks available (55%) and the median section received a total of 78 problem sets (assignments) during the school year or between three and four assignments per week in a week when ASSISTments was used. This level of use was consistent with the study leaders' expectations. During the initial teacher training sessions, teachers were told to use ASSISTments to assign homework 2–3 times per week.

A majority of the homework problems assigned by teachers were from their textbook (about 99% of problems in the median section); the rest were Skill Builder problems. However, based on the difference between the percentage of Skill Builders assigned for the median and average sections (1% for the median compared to 13% for the average section), some teachers relied heavily on the assignment of this problem type, while others rarely assigned Skill Builders. In fact, slightly more than half of all teachers in the treatment schools (53%) assigned Skill Builders. We also examined whether assignments were made to a whole classroom or to a subset of students (this data is not included in Table 5). Although the predominant use pattern was to assign the whole class the same problem sets for homework, about 20% of assignments were issued to a smaller subgroup of students, suggesting some teachers were using ASSISTments to customize and differentiate homework assignments for different groups of students.

We also analyzed the proportion of ASSISTments reports that a teacher *opened* a report by clicking on a link in an email sent to the teacher or on the ASSISTments platform. We use the opening of reports as an indicator that a teacher is using

Table 5. Median, mean and standard deviation of teacher and student use of ASSISTments variables by section.

Variables	*N*	Median	Mean	SD
Total number of problems attempted by students	1,335	733	846	620
Percent of assignments that teacher reviewed reports	86	0.75	0.73	0.25
Total number of assignments assigned	86	78	93	71
Percent of problems assigned that were skill builders	86	0.01	0.13	0.20
Number of weeks with at least one assignment	86	22	21	11

ASSISTments to review student homework performance, which is a precondition to using homework data to adjust instruction. In the median section, teachers opened 75% of the reports available to them.

In terms of students use of ASSISTments at the class level, students in the median class attempted a total of 733 problems on ASSISTments across the entire school year or approximately 33 problems per week in the weeks ASSISTments was used to assign homework problems.

We also investigated whether student use of ASSISTments differed for different groups of students, including by gender, prior achievement, free and reduced-price lunch status, and IEP status (see Table 6). For each group we found statistically significant differences in the amount of problems completed on the ASSISTments platform. Males attempted 10% more problems than females. Students with higher prior achievement (at or above the median) attempted 23% more problems than students with lower prior achievement (below the median). Students not receiving free and reduced-price lunch attempted 16% more problems than those students who do. Finally, students without an IEP attempted 32% more problems than student with an IEP.

Main Effect of Intervention

Results from our three-level hierarchical linear main effects model are presented in Table 7 (see results for *Main Effects*). Controlling for student, classroom, and school--level covariates, we found a significant positive treatment effect of approximately 8.5 scale points for those students who used ASSISTments (γ_{001}=8.501, $p < .001$). This corresponded to an effect size (Hedge's g) of 0.22 standard deviation units and a 95% confidence interval (CI) for the effect size of 0.15 to 0.30.[8] This is a slightly greater effect than was reported in the prior publication ($g = 0.18$) using a two-level model (Roschelle et al., 2016).

Table 6. Mean, standard deviation tests for differences in teacher and student use of ASSISTments by gender, prior math achievement, free and reduced-price lunch participation, and individualized education program status.

Variables	*N*	Total Problems Attempted (Mean)	Standard deviation	Difference
Male	760	895	657	87**
Female	761	808	638	
Grade 6 math achievement (above median)	766	937	713	173***
Grade 6 math achievement (below median)	741	764	566	
Do not receive free/reduced-price lunch	977	896	688	125***
Receive free/reduced-price lunch	534	771	566	
No individualized education program (no IEP)	1,337	876	654	214***
Indiviualized education program (IEP)	174	662	587	

Note. $^{**}p < .01$.
$^{***}p < .001$.

[8]We also ran the main effects analysis with the 116 *joiners* removed from the analysis sample. The results were very similar to the results for the full analytical sample. With the joiners removed, the estimated main effect of ASSISTments use was 8.612 points (t(18) = 4.525, $p < 0.001$) on the 7th grade TerraNova math assessment.

Table 7. Results of three-level hierarchical regression model to estimate main effect of treatment (ASSISTments).

Parameters	Unconditional (standard errors)	Main effects	All interactions
Regression coefficients (fixed effects)			
Intercept (γ000)	688.666 (2.287)***	685.239 (1.473)***	685.144 (1.473)***
Treatment status (Ttx) (γ001)	—	8.492 (1.952)***	8.505 (1.952)***
Mean grade 6 math achievement[a]-school (γ002)	—	0.624 (0.624)	0.602 (0.623)
Mean free/reduced-price lunch-school (γ003)	—	0.073 (0.160)	0.101 (0.160)
7th Grade enrollment-school (γ004)	—	0.013 (0.030)	0.022 (0.031)
Mean grade 6 math achievement-class (γ010)	—	0.743 (0.134)***	0.828 (0.135)***
Grade 6 math achievement variance-class (γ020)	—	0.030 (0.019)	0.035 (0.019)
Class size-class (γ030)	—	−0.010 (0.132)	−0.009 (0.132)
Grade 6 math achievement (γ100)	—	2.257 (0.055)***	2.393 (0.076)***
Male (γ200)	—	−2.455 (0.886)**	−4.209 (1.359)**
Individulized education program (IEP) (γ300)	—	−8.571 (1.522)***	−11.002 (2.311)***
Free/reduced-price lunch (FRPL) (γ400)	—	−4.043 (0.988)***	−5.055 (1.491)**
Grade 6 math achievement × Trx (γ101)	—	—	−0.284 (0.102)**
Male × Trx (γ201)	—	—	3.192 (1.788)
IEP × Trx (γ301)	—	—	4.142 (3.069)
FRPL × Trx (γ401)	—	—	2.049 (1.969)
Variance components (random effects)			
Residual (σ^2)	919.22	505.01	501.38
Level 2 Int ($\tau_\pi{}^2$)	533.41	56.63	56.22
Level 3 Int ($\tau_\beta{}^2$)	57.98	8.71	8.82
Model summary			
Deviance statistic	27,149.52	24,988.23	24,968.71
Number of estimated parameters	4	35	39

Note: Treatment Status (Trx) = 0 = Control, 1 = Treatment.
[a]New England Common Assessment Program.
$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

We also note that 6th grade scores on the NECAP mathematics assessment was positively related to TerraNova mathematic scores in spring of 7th grade ($\gamma_{100} = 2.257$, $p < .001$). Lower TerraNova math scores were associated with being male ($\gamma_{200} = -2.455$, $p < .01$), students receiving special education services ($\gamma_{300} = -8.571$, $p < .001$), and students receiving free and reduced-price lunch ($\gamma_{400} = -4.043$, $p < .001$).

Moderator Effects

We then analyzed whether a set of exogenous student characteristics moderate the effects of ASSISTments use on student math achievement by testing for interaction between the student-level characteristics and treatment condition. The student characteristics investigated include gender, prior mathematics achievement, free and reduced-price lunch status, and IEP status. Results are shown in Table 7. A separate analysis was conducted for each moderator variable and the final model included all interaction terms in the model by adding the interaction terms to the school-level of the *Main Effects* model described above (see results for *All Interactions* model).

As shown in Table 7, when all interaction terms are included in the same model, we found a statistically significant and negative effect between prior math scores (6th grade

NECAP math scores) and the treatment condition favoring students entering the study with lower prior math schools ($\gamma_{101} = -0.284$, $p < 0.01$). We found no statistically significant evidence that the other student-level characteristics moderated the effects of ASSISTments use on math achievement.[9] To provide readers with a sense of the size of the moderation associated with prior math achievement, the treatment effect for students with higher prior math achievement (one SD above the mean on 6th grade NECAP) was 4.87 points, while the treatment effect for students with lower prior math scores (one SD below the mean) was 12.13 points.

Test for Mediation

Two approaches were used to examine whether teacher behavior (i.e., reviewing targeted homework problems with the class) was a mediator of the effect of ASSISTments on student math achievement: the joint test of significance (MacKinnon et al., 2002) and the empirical M-test (MacKinnon et al., 2004). Results from the application of both strategies found no support for a mediational effect of the targeted homework review measure.

Joint Test of Significance

Based on the results of the joint test of significance approach, we found no evidence to support a finding that the targeted homework review practice measured mediates the effect of ASSISTments use on student math achievement (see Table 8). Specifically, while treatment had a statistically significant effect on the mediator variable—treatment teachers reported using targeted homework review in 33% more of the instructional logs relative to control teachers (Model A; $a = 33.454$ $t(13) = 3.414$, $p < .01$)—the teacher review practice variable was not a statistically significant predictor of 7th grade math achievement scores ($b = -0.062$, $t(103) = -1.521$, $p > .05$) when the variable was added to the classroom-level of the *All Interactions* model (Model B). This is inconsistent with a finding of mediation.

Empirical M-Test

To apply the empirical M-test, we first computed the indirect effect by taking the product of the mediation path coefficients a and b from the hierarchical regression models A and B in Table 8 ($ab = (33.454)\ (-0.062) = -2.074$). We then computed the asymmetric 95% confidence interval for the indirect effect (the product of two normally distributed random variables) using the PRODCLIN computer program and the values of a and b and their standard errors (Mackinnon et al., 2007). The results are summarized in Table 9. The resulting 95% confidence interval is $[-51.825, 8.7737]$. Since the interval includes the value of 0, we can't reject the possibility at a 95% confidence level that there is no mediation effect due to the impact of the treatment on teachers' use of a targeted homework review practice. Thus, the finding from the empirical M-test replicates the finding from the joint test of significance.

[9]Note Table 7 shows that when the student variable by treatment status interaction terms were analyzed separately, we found statistically significant and positive moderation of the effects of ASSISTments on math achievement of being male and having an individualized education program. However, these results are no longer statistically significant when all the interaction terms are included in same model.

Table 8. Results of the mediation analysis examining whether a measure of teacher targeted homework review mediates the effect of the treatment (ASSISTments) on student learning.

Model A		Model B	
Parameters	Coefficients (standard errors)	Parameters	Coefficients (standard errors)
Regression coefficients (fixed effects)		Regression coefficients (fixed effects)	
Intercept (γ000)	43.147 (6.979)***	Intercept (γ000)	685.625 (1.850)***
Treatment status (Trx) (γ001)	33.454 (9.798)**	Treatment status (Trx) (γ001)	10.157 (2.423)**
Mean 6th grade math-school (γ002)	0.132 (2.449)	Mean 6th grade math-school (γ002)	0.721 (0.650)
Mean free/reduced-price lunch-school (γ003)	−0.124 (0.830)	Mean free/reduced-price lunch-school (γ003)	0.194 (0.165)
7th Grade enrollment-school (γ004)	0.031 (0.147)	7th grade enrollment-school (γ004)	0.012 (0.029)
Mean 6th grade math-class (γ010)	0.261 (0.256)	Mean 6th grade math-class (γ010)	0.752 (0.156)***
6th Grade math variance-class (γ020)	0.029 (0.045)	6th Grade math variance-class (γ020)	0.036 (0.024)
Class size-class (γ030)	−0.645 (0.374)	Class size-class (γ030)	0.160 (0.175)
		Targeted homework review-class (γ040)	−0.062 (0.041)
		6th Grade math (γ100)	2.413 (0.088)***
		Male (γ200)	−4.445 (1.545)**
		Individualized education program (IEP) (γ300)	−8.754 (2.650)**
		Free/reduced-price lunch (FRPL) (γ400)	−4.960 (1.676)**
		6th grade math × Trx (γ101)	−0.328 (0.113)**
		Male × Trx (γ201)	4.190 (1.964)*
		IEP × Trx (γ301)	2.076 (3.384)
		FRPL × Trx (γ401)	1.724 (2.154)
Variance components (random effects)		Variance components (random effects)	
Residual (σ^2)	310.027	Residual (σ^2)	479.061
Level 2 Int ($\tau_\pi{}^2$)	554.088	Level 2 Int ($\tau_\pi{}^2$)	63.403
		Level 3 Int ($\tau_\beta{}^2$)	0.021
Model summary		Model summary	
Deviance statistic	1,117.72	Deviance statistic	20,491.37
Number of estimated parameters	2	Number of estimated parameters	40

Note: *$p < .05$, **$p < .01$, ***$p < .001$.
Treatment Status (Trx) = 0 = Control, 1 = Treatment.

Table 9. Results of the M-test mediation analysis approach, examining whether a measure of teacher targeted homework review mediates the effect of the treatment (ASSISTments) on student math achievement.

Path coefficients	Coefficients (standard errors)
Path *c* (total effect of treatment on student math achievement)[a]	8.505 (1.952)***
Path *a* (effect of treatment on mediator)[b]	33.454 (9.798)**
Path *b* (effect of mediator on student math achievement)[c]	−0.062 (0.041)
Path *c*′ (direct effect)[d]	10.157 (2.423)**
Mediation effect (*ab*) [confidence interval][c]	−2.074 [−51.825,8.7737]

Note: [a]Estimated treatment status (Ttx) (γ001) parameter in Table 7.
[b]Estimated treatment status (Ttx) (γ001) parameter from Model A in Table 8.
[c]Estimated targeted homework review-class (γ040) parameter from Model B in Table 8.
[d]Estimated treatment status (Trx) (γ001) from Model B in Table 8.
[e]Mediation effect = the product of path coefficients *a* and *b*. Confidence interval computed using PRODCLIN (MacKinnon et al., 2007).
*$p < .05$, **$p < .01$, ***$p < .001$.

Limitations

There are several methodological limitations to the mediation analysis. First, since the analysis was considered a secondary analysis, the study's sample size was not designed to be sufficiently large to find small, yet perhaps meaningful, mediation effect. In addition, we tested mediation of a single formative assessment practice mechanism—a measure of whether teachers targeted their homework review time to those problems that proved to be the most difficult for students. We can't rule out the possibility that important unmeasured teacher and student mediators may exist that might help explain the effect of the treatment on student math achievement including increases in students' perceptions of homework quality, their expectancy and value regarding homework, or other indicators of their autonomy, self-regulation or effort. We must also acknowledge the strong possibility of the existence of unobserved confounding variables, including unobserved mediators, that are correlated with both the selected mediator—in our case *targeted homework review*—and student achievement. Since teachers were not randomly assigned to the different levels of the mediator, by not accounting for the existence of confounding variables in the analytical models, the strength of the relationship between the hypothesized mediator and outcome variable may be overestimated and spurious if the unobserved variables are the true source of mediation. A final limitation we need to consider is bias due to measurement error associated with the mediator.[10] The regression-based approaches used in the mediation analysis assume the mediator is perfectly measured without measurement error. However, in practice, all variables have some degree of unreliability. The presence of measurement error will lead to an underestimation of the strength of the association between the mediator and outcome variable, lower statistical power, and reduce the study's ability to detect a true mediational relationship.[11] Although we expect some degree of unreliability in our self-reported measure of the mediator, we have no reason to believe that accounting for the unreliability of the measure would have significantly changed the findings of the mediation results.

Relationship Between System Use and Math Achievement

We found positive relationship between two use variables and end-of-year 7th grade math achievement—the percentage of homework problems assigned by teachers that were Skill Builders and total number of problems completed by students. Table 10 shows the results for three-level hierarchical linear regression model examining the relationship between teacher and student use of ASSISTments and student math achievement within the treatment sample. Controlling for a range of individual, classroom, and school characteristics, students in classroom whose teachers assigned a greater percentage of Skill Builder problems ($t(66) = 2.073$, $p < .05$) and students who completed more problems through the ASSISTments platform ($t(1{,}312) = 4.913$, $p < .001$) had higher 7th grade math achievement scores. These analyses are exploratory and cannot be used to make definitive claims about causal relationships between different types and levels of use of ASSISTments and student math achievement. While it is possible that higher achievement scores result from assigning more mastery-based homework problems (in

[10]One other possible limitation of mediation analysis cited in the literature is the possibility of *reverse causation*, i.e., the possibility that the changes in the level of outcome variable caused by the treatment cause changes in the proposed mediator variable. However, since in the present study the measure of *targeted homework review* happened weeks prior to the assessment of student math achievement, reverse causation is not possible and was not a considered a threat to the interpretation of the mediation analysis results.

[11]Note that the effects of bias on the size of the mediator effect (path *b*) due to the presence of (1) unobserved confounding variables and (2) measurement error will be in opposite directions and will offset each other to some extent.

Table 10. Results of three-level hierarchical model to test the relationship between teacher and student use of ASSISTments and student math achievement (treatment group only).

Parameters	Coefficients (standard errors)
Regression coefficients (fixed effects)	
Intercept (γ000)	694.110 (1.849)***
Mean grade 6 math achievement[a]-school (γ001)	0.451 (0.812)
Mean free/reduced-price lunch-school (γ002)	−0.074 (0.178)
7th Grade enrollment-school (γ003)	−0.003 (0.039)
Mean grade 6 math achievement-class (γ010)	0.762 (0.176)***
Grade 6 math achievement variance-class (γ020)	0.048 (0.031)
Class size-class (γ030)	−0.055 (0.161)
Report review rate-class (γ040)	8.656 (5.180)
Total number of assignments-class (γ050)	0.025 (0.029)
Percent of skill builders assigned-class (γ060)	13.030 (6.285)*
Number of weeks of use-class (γ070)	−0.287 (0.173)
Grade 6 math achievement (γ100)	2.057 (0.076)***
Male (γ200)	0.449 (1.138)
Individualized education program (γ300)	−6.152 (1.968)**
Free/reduced-price lunch (γ400)	−2.321 (1.279)
Total number of problems attempted (γ400)	0.009 (0.002)***
Variance components (random effects)	
Residual (σ^2)	433.503
Level 2 Int ($\tau_\pi{}^2$)	48.639
Level 3 Int ($\tau_\beta{}^2$)	40.865
Model summary	
Deviance statistic	12,890.54
Number of estimated parameters	19

Note: [a]New England Common Assessment Program.
*$p < .05$, **$p < .01$, ***$p < .001$.

the form of Skill Builders) and students attempting more homework problems with immediate feedback, it is also possible that teacher and student factors that are confounded with these use variables explain some or all of the relationship between these use variables and higher math achievement scores (and despite our use of covariates in our model to adjust for potential observable confounding factors). For example, we know from the results of analyses described above that students who attempted more ASSISTments problems were also more likely to have higher prior math achievement than others, less likely to receive free and reduced-price lunch (a proxy for family income) and less likely to have an individualized education program. We also know that just over half of the teachers assigned Skill Builder problems to their students (52%). Thus, it is possible that teachers who assigned Skill Builders (1) were more likely to assign these problems to students who were more advanced math learners or (2) were, in general, more likely than their peers to use effective instructional strategies in the classroom.

Generalizability of Study Findings

Of the 43 schools involved in the study, the Generalizer program was able to identify 41 schools in its database using the National Center of Education Statistics school identifier number. Generalizability was assessed for each of the 50 states and the United States as a whole. The population was restricted to schools that included 7th grade, the target population of this study. As described above, the Generalizer program produces a generalizability index that can be used to summarize the strength of the similarity between the study sample and the inference population. For the current study sample, the estimated generalizability index was above 0.90 for the state of Maine, indicating that the study sample generalizes to the inference population in Maine and that the estimated impact is unbiased for this population. In contrast to the generalizability findings for Maine, when we look across the United States, four states had a generalizability index between 0.70 and 0.90; fourteen between 0.50 and 0.69; and for the remaining thirty-one states the index was below 0.50—indicating the results from this study have low generalizability to districts in a majority of states. Figure 4 below provides a visual representation of the results (high and medium generalizability states are combined into a single category in the graphic, 0.50–0.90).

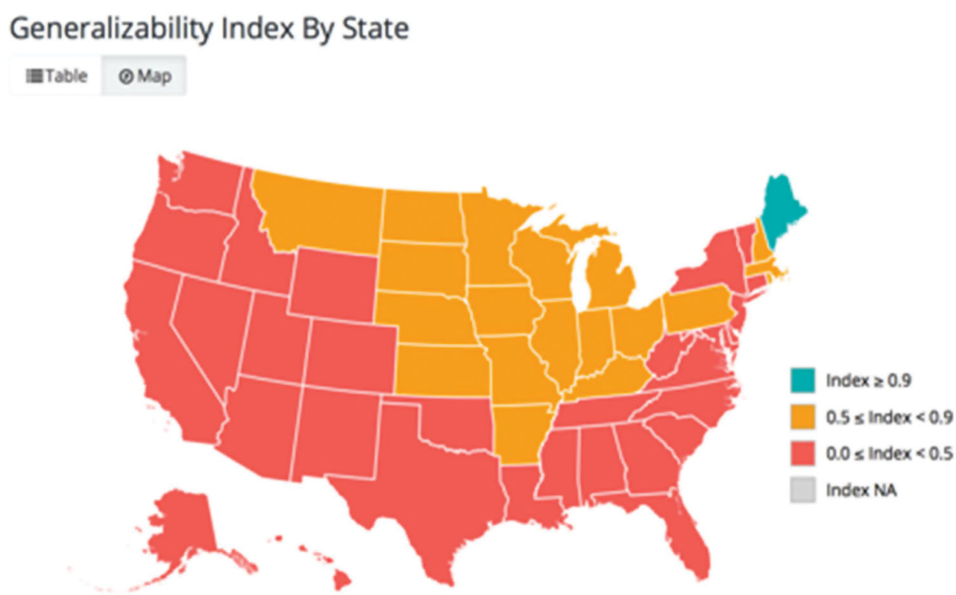

Figure 4. The results of a generalizability analysis of the sample indicated the study findings might apply *well* throughout the state of Maine (green) and *somewhat* in 18 other states (orange). The study does not generalize well in the 31 other states (red), primarily in the West, South, and some parts of the Northeast.

Discussion

The evaluation team planned and successfully implemented a randomized control trial to evaluate the impact of an online homework intervention, ASSISTments, on student mathematics achievement. They recruited and maintained a sufficient number of schools to statistically power the study, randomly assign them to condition and avoid contamination between conditions. Student attrition, overall and between groups, was within guideline levels for a cluster-random assignment design study. The quality of the original study as well as the prior published results (Roschelle et al., 2016) motivated this deeper investigation of findings, particularly with regard to moderators, mediators, and generalizability.

Use of the ASSISTments platform to assign homework by teachers generally fit the expectations established prior to the experiment: teachers assigned student work in ASSISTments three to four times per week and they opened reports (a precursor to formative assessment) 75% of the time. The median classroom used ASSISTments for 22 weeks of a 38-week school year and students in the median classroom completed almost 733 problems. Given the lag between the start of the school year and when schools in Maine typically distribute the laptops to students and extracurricular activities and test preparation that take away from regular instruction, 22 weeks of ASSISTments use is considered a meaningful duration.

Although the research team and trainers encouraged treatment teachers to assign homework using ASSISTments at least three nights per week, variation in how and how often the platform used was expected. A "one-size-fits-all" approach was not the intent of ASSISTments' developers; ASSISTments was designed to be used in a multitude of ways. In fact, during this study some teachers clearly used ASSISTments to assign customized assignments to small groups of students and individuals as well as the whole class.

Consistent with the findings of Roschelle et al. (2016) in a previous analysis using a two-level hierarchical linear regression model, using a three-level model we found a statistically significant main effect of the intervention on the standardized math assessment administered by the research team (Hedges' $g = 0.22$); 7th grade students in schools that used ASSISTments had higher scores on the study's measure of math learning, the 7th grade TerraNova math assessment To help put the size of this effect of ASSISTments' use in Maine into context, the research team published a technical report (Roschelle et al., 2017). In terms of *conventional benchmarks* associated with the findings from relevant, published meta-analyses, an effect size of 0.22 standard deviation units is about 30% greater than the average effect size reported in a meta-analysis of formative assessment interventions (Hedges' $g = 0.17$, reported in Kingston & Nash, 2011) and two and one-half times larger than the average effect size reported in a meta-analysis of rigorous studies of computer-based interventions (Hedges' $g = 0.09$, reported in Cheung & Slavin, 2013). In relation to measures of *annual expected progress* for 7th grade math learning, an effect size of 0.22 represents an additional two-thirds of a year of gain in math learning for the average 7th grade student in the treatment group when compared to the 0.30 expected gain due to a full year of 7th grade math instruction (Lipsey et al., 2012). In terms of *policy relevant performance gaps*, the main effect of 8.5 points on the 7th grade TerraNova math assessment is two times larger than the difference in scores between the average student in the sample who receives and does not receive free-and-reduced-price lunches (−4.0 points) and similar in size to the difference between students with and without an individualized education program (−8.6 points).

We also found a differential positive effect of the intervention for students with different prior math achievement upon entering the study. Specifically, we found a statistically significant interaction effect between prior mathematics scores and the treatment: 7th grade students with lower prior mathematics scores experienced a greater benefit from the ASSISTments intervention. As in the earlier Elawar and Corno (1985) study, providing more and better feedback to students may be especially beneficial to lower performing students (higher performing students are more likely to get math problems right without feedback and thus may experience less need for feedback). As lower performing students are more likely to struggle with homework, they may have benefited more from treatment teachers' greater propensity to target their homework review around common errors or deeper discussion around solutions to math problems that are challenging. These discussions were enabled in the treatment condition through the nightly homework performance reports the ASSISTments platform sends to teachers. Also, the intervention may have influenced student perceptions of homework quality; plausibly, giving a student feedback at home along with the opportunity to try problems again could increase their ability to learn from homework and their sense that doing homework is a worthwhile activity. As discussed in the literature, perceptions of homework quality relate to students' expectancy and value judgement; these correlate to student effort; and student effort drives learning (Dettmers et al., 2010; Trautwein, Lüdtke, Kastens, et al., 2006). Alternatively, one might interpret the results in terms of providing students more autonomy to do homework by giving them immediate feedback (Fernández-Alonso et al., 2015) or more support while doing homework (Kitsantas et al., 2011)—both factors are consistent with the prior literature. It is also reasonable to interpret the autonomy and support as supporting student self-regulation. Self-regulation is associated with more learning from homework (Ramdass & Zimmerman, 2011). (We also noticed that students who attempted more problems in ASSISTments had greater learning, but we are reluctant to interpret this as more student effort per cautions in the literature). We do not have specific measures of each of these individual student factors that may mediate the effect of the intervention on student learning for students with lower prior math achievement—autonomy, support, self-regulation, expectancy-value motivation—and thus cannot model whether these factors may explain why ASSISTments homework-based intervention was more beneficial for students with lower prior achievement in mathematics.

We do not expect that other factors discussed in the literature are explanatory. For example, we did not intervene with respect to a teacher's homework compliance policy. Surveys given to teachers in both the treatment and control teachers show that a few teachers in each group perceived homework compliance as slightly higher, but overall there was no major shift. Overall, most teachers in both groups reported compliance was above 50%. We also have no reasons to suspect that parents were involved differently in the two conditions as a result of the intervention.

Teachers homework practices were also found to be influenced and impacted by the intervention. A previous analysis of interviews and observations from this experiment had identified a distinctive change in teaching practice (Fairman et al., 2016). In particular, teachers in the ASSISTments condition used the reports from ASSISTments to target specific problems during their classroom review of homework, focusing on those problems that most students had difficulty solving on the nightly homework assignment. ASSISTments provided teachers with information about common wrong answers for each assignment and teachers were observed to address these specific wrong answers. These teacher behaviors would be visible to students and might also have influenced student perception of homework quality (as discussed in the previous paragraph) and student's feeling of teachers' paying attention to their learning. From our analysis of the system use logs we know that a majority of nightly homework performance reports were opened by teachers (75% of all reports). Further, data collected from both treatment and control teachers through teacher instructional logs indicated statistically significant differences in homework review practices with treatment teachers more likely to shift toward a targeted review of homework. However, the present analysis did not find a mediating effect of a self-reported measure of the extent to which teachers focused their review of homework in the classroom on those problems that students experienced the most difficulty solving.

We also found a statistically significant association between spring test scores and teachers who assigned a greater percentage of Skill Builder problems and students who attempted more problems on the ASSISTments platform. While this analysis was correlational and exploratory, it does suggest that it might of value if future research on the ASSISTments platform included a rigorous test of whether a blended mastery-learning homework approach (a mix of textbook and Skill Builder problems) might be a particularly effective implementation model for the use of the platform.

Finally, we considered the generalizability of our findings from the state of Maine to other states across the nation. One advantage of conducting the study in Maine was the equitable distribution of one-to-one computing devices to all students under the states' one-to-one laptop program, making it possible for all students to have access to ASSISTments when they were away from school. However, one corresponding limitation of siting the study in Maine is the potential lack of generalizability of the study's findings to states, regions and districts where computing devices are not equitably available for homework. In addition, the generalizability of the findings from this study are limited due to the fact that the student population in Maine is more homogeneous and the population densities lower than in many other regions of the country. Our generalizability analysis, which examined the degree of similarity between the sample and populations in other regions of the United States and the nation overall, found that generalizability was medium to high for only 19 states (including the state of Maine). Consequently, we would recommend caution in extrapolating these findings to regions of the country with different population demographics. A replication in North Carolina, a state with a more diverse student population, is now underway. Finally, we highlight that implementation quality has often been found to be an important variable in other studies of the impact of formative assessment and educational technology interventions. In this study, the ASSISTments team provided ample training and coaching to teachers that was well-received by participants. Our findings might not generalize to weaker or shorter-term implementations.

Conclusions

We conducted a rigorous, large-scale experiment at the intersection of all three relevant policy contexts: (a) one-to-one technology, (b) mathematics homework practices, and (c) formative assessment. The findings show that the area of overlap can be fruitful point of intervention in the enduring quest to improve middle school mathematics achievement.

Although schools could design interventions with only two of these factors, we believe there are good reasons to continue to explore all three factors. For example, it is possible to take a formative assessment approach to improving mathematics homework but without technology. Yet, in this case, we see potential advantages to using technology, such as providing immediate feedback directly to students as they do their homework and making homework a more productive learning opportunity for all students, particularly those with limited access to supportive adults and peers outside of school. Further, the ASSISTments intervention was designed to be easy for teachers to use including reducing the burden of giving students immediate feedback on their homework. For example, the homework problems in their existing textbooks were available in ASSISTments to be assigned to students; teachers did not have to find and align new homework items to their instructional plans. Also, teachers reported that the automatic scoring and reporting that makes students' homework performance immediately visible to teachers enabled them to re-focus their energies away from grading and toward adapting their instruction. Hence, we suggest that while educators could see this study as validating their investment in only two of the three factors, there are reasons why they may want to focus efforts on using technology to improve mathematics homework via formative assessment practices.

With regard to research, there is much more to be done. Although we found a promising impact of the intervention on teaching practice (the switch to targeted homework review), our model of mediation did not show a relationship between change in practice and increased student learning. In future research, it would make sense to investigate a range of theory-based teacher and student mediators, and design experiments and strengthen measurements so as to be better able to analyze mediation. Also, we pointed out the limited generalizability of this work, due to the specifics of the setting in Maine. This work will inform an ongoing replication study in North Carolina, also funded by U.S. Department of Education, which will further explore for whom and under what conditions and practices this type of intervention can yield positive student impacts. Finally, there is much more that could be explored at the three-way intersection of homework, formative assessment, and technology. We did not yet explore parent involvement for example, but this is a much-discussed factor in the literature, and it is plausible that technology could help with positive parent engagement and involvement in homework. There is also more that an online homework platform like ASSISTments could provide in terms of resources to students as they do homework, such as an online support community and guides to additional learning resources.

Funding

This material is based on work supported by the Institute of Educational Sciences of the U.S. Department of Education under Grant R305A120125. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of Institute of Educational Sciences.

ORCID

Jeremy Roschelle http://orcid.org/0000-0003-2219-0506

References

Alleman, J., & Brophy, E. J. (1991). *Reconceptualizing homework as out-of-the-school learning opportunities* (Occasional Paper No. 135). Michigan State University, Institute for Research in Teaching.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Becker, H. J. (2000). Findings from the teaching, learning, and computing survey. *Education Policy Analysis Archives*, *8*, 51. https://doi.org/10.14507/epaa.v8n51.2000

Bembenutty, H. (2011). Meaningful and maladaptive homework practices: The role of self-efficacy and self-regulation. *Journal of Advanced Academics*, *22*(3), 448–473. https://doi.org/10.1177/1932202X1102200304

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Bennett, S., & Kalish, N. (2006). *The case against homework: How homework is hurting our children and what we can do about it*. Crown.

Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *92*(1), 81–90. https://doi.org/10.1177/003172171009200119

Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles Policy and Practice*, *5*(1), 7–73.

Black, P. J., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31.

Bloom, B. S. (1971). Mastery learning. In J. H. Block (Ed.), *Mastery learning: Theory and practice* (pp. 47–63). Holt, Rinehart and Winston.

Brookhart, S. M. (2007). Expanding views about formative classroom assessment: a review of the literature. In J. H. McMillan (Ed.), *Formative classroom assessment: theory into practice* (pp. 43–62). Teachers College Press.

Buell, J. (2004). *Closing the book on homework: Enhancing public education and freeing family time*. Temple University Press.

Carlson, D., Borman, G. D., & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform on reading and mathematics achievement. *Educational Evaluation and Policy Analysis*, *33*(3), 378–398. https://doi.org/10.3102/0162373711412765

Cheung, M. W. L. (2009). Comparison of methods for constructing confidence intervals of standardized indirect effects. *Behavior Research Methods*, *41*(2), 425–438. https://doi.org/10.3758/BRM.41.2.425

Cheung, A. C. K., & Slavin, R. E. (2013). The effectiveness of educational technology applications for enhancing mathematics achievement in K-12 classrooms: A meta-analysis. *Educational Research Review*, *9*, 88–113. https://doi.org/10.1016/j.edurev.2013.01.001

Cooper, H. (2015). *The battle over homework: Common ground for administrators, teachers, and parents* (4th ed.). Simon and Schuster.

Cooper, H. M. (1989). Synthesis of research on homework. *Educational Leadership*, *47*(3), 85–91.

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research*, *76*(1), 1–62. https://doi.org/10.3102/00346543076001001

Cooper, H., Steenbergen-Hu, S., & Dent, A. L. (2012). Homework. In K. R. Harris, S. Graham, T. Urdan, A. G. Bus, S. Major, & H. L. Swanson (Eds.), *APA educational psychology handbook, Vol. 3. Application to learning and teaching* (pp. 475–495). American Psychological Association. http://dx.doi.org/10.1037/13275-019

Cordray, D. S., Pion, G. M., Brandt, C., & Molefe, A. (2013, March 7). *The impact of the measures of academic progress (MAP) program on student reading achievement* [Paper presentation]. Paper presented at the Society for Research on Educational Effectiveness, Washington, DC.

Dettmers, S., Trautwein, U., Lüdtke, O., Kunter, M., & Baumert, J. (2010). Homework works if homework quality is high: Using multilevel modeling to predict the development of achievement in mathematics. *Journal of Educational Psychology*, *102*(2), 467–482. https://doi.org/10.1037/a0018453

Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, *14*(7), 1–11.

Elawar, M. C., & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: Changing teacher behavior a little rather than a lot. *Journal of Educational Psychology*, *77*(2), 162–173. https://doi.org/10.1037/0022-0663.77.2.162

Fairman, J., Feng, M., & Roschelle, J. (2016). *Investigating online homework: How do teaching practices change?* [Paper presentation]. NCTM 2016 Research Conference, San Francisco, CA.

Fan, H., Xu, J., Cai, Z., He, J., & Fan, X. (2017). Homework and students' achievement in math and science: A 30-year meta-analysis, 1986–2015. *Educational Research Review*, *20*, 35–54. https://doi.org/10.1016/j.edurev.2016.11.003

Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the ASSISTments system. *Technology, Instruction, Cognition & Learning*, *3*(1/2), 6.

Fernández-Alonso, R., Suárez-Álvarez, J., & Muñiz, J. (2015). Adolescents' homework performance in mathematics and science: Personal factors and teaching practices. *Journal of Educational Psychology*, *107*(4), 1075–1085. https://doi.org/10.1037/edu0000032

Fyfe, E. (2016). Providing feedback on computer-based algebra homework in middle-school classrooms. *Computers in Human Behavior*, *63*, 568–574. https://doi.org/10.1016/j.chb.2016.05.082

Gershenson, S., & Holt, S. (2015). Gender gaps in high school students' homework time. *Educational Researcher*, *44*(8), 432–441. https://doi.org/10.3102/0013189X15616123

Gill, B., & Schlossman, S. (2004). Villain or savior? The American discourse on homework, 1850–2003. *Theory into Practice*, *43*(3), 174–181. https://doi.org/10.1207/s15430421tip4303_2

Guskey, T. R. (2007). Using assessments to improve teaching and learning. In D. B. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning* (pp. 15–29). Solution Tree.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, *77*(1), 81–112. https://doi.org/10.3102/003465430298487

Heritage, M. (2013). *Formative assessment in practice: A process of inquiry and action*. Harvard Education Press.

Karbach, J., Gottschling, J., Spengler, M., Hegewald, K., & Spinath, F. M. (2013). Parental involvement and general cognitive ability as predictors of domain-specific academic achievement in early adolescence. *Learning and Instruction*, *23*, 43–51. https://doi.org/10.1016/j.learninstruc.2012.09.004

Katz, I., Buzukashvili, T., & Feingold, L. (2012). Homework stress: Construct validation of a measure. *The Journal of Experimental Education*, *80*(4), 405–421. https://doi.org/10.1080/00220973.2011.610389

Kelly, Y., Heffernan, N., Heffernan, C., Goldman, S., Pellegrino, G., & Soffer, D. (2013). Estimating the effect of web-based homework. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *The artificial intelligence in education conference* (pp. 824–827). Springer-Verlag.

Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, *30*(4), 28–37. https://doi.org/10.1111/j.1745-3992.2011.00220.x

Kitsantas, A., Cheema, J., & Ware, H. W. (2011). Mathematics achievement: The role of homework and self-efficacy beliefs. *Journal of Advanced Academics*, *22*(2), 310–339. https://doi.org/10.1177/1932202X1102200206

Kohn, A. (2000). *The homework myth: Why our kids get too much of a bad thing*. Da Capo Press.

Konstantopoulos, S., Miller, S. R., & van der Ploeg, A. (2013). The impact of Indiana's system of interim assessments on mathematics and reading achievement. *Educational Evaluation and Policy Analysis*, *35*(4), 481–499. https://doi.org/10.3102/0162373713498930

Kralovec, E., & Buell, J. (2000). *The end of homework: How homework disrupts families, overburdens children, and limits learning*. Beacon Press.

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., Roberts, M., Anthony, K. S., & Busick, M. D. (2012). *Translating the statistical representation of the effects of education interventions into more readily interpretable forms* (NCSER 2013-3000). National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Loveless, T. (2014). *Homework in America. Part II, 2014 Brown Center report of American education*. Brookings Institution.

MacKinnon, D. P., Fritz, M. S., Lockwood, C. M., & Williams, J. (2007). Distribution of the product confidence limits for the indirect effect: Program PRODCLIN. *Behavior Research Methods*, *39*(3), 384–389. https://doi.org/10.3758/BF03193007

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods*, *7*(1), 83–104. https://doi.org/10.1037/1082-989X.7.1.83

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4

MacKinnon, D. P., & Pirlott, A. G. (2015). Statistical approaches for enhancing causal interpretation of the M to Y relation in mediation analysis. *Personality and Social Psychology Review*, *19*(1), 30–43. https://doi.org/10.1177/1088868314542878

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making sense of data driven decision making in education: Evidence from recent RAND research* (OP-170). RAND Corporation.

Mau, W., & Lynn, R. (2000). Gender differences in homework and test scores in Mathematics, Reading and Science at tenth and twelfth grade. *Psychology, Evolution & Gender*, *2*(2), 119–125. https://doi.org/10.1080/14616660050200904

McDermott, R. P., Goldman, S. V., & Varenne, H. (1984). When school goes home: Some problems in the organization of homework. *Teachers College Record*, *85*, 391–409.

Means, B., Chen, E., DeBarger, A., & Padilla, C. (2010). *Teachers' ability to use data to inform instruction: challenges and supports. (Part I. An exploratory study of teachers' data literacy tools and Part II. Resources for data literacy professional development)*. U.S. Department of Education.

Mendicino, M., Razzaq, L., & Heffernan, N. T. (2009). A comparison of traditional homework to computer-supported homework. *Journal of Research on Technology in Education*, *41*(3), 331–359. https://doi.org/10.1080/15391523.2009.10782534

National Council of Teachers of Mathematics. (2013). *Formative assessment: A position of the National Council of Teachers of Mathematics.*

Núñez, J. C., Epstein, J. L., Suárez, N., Rosário, P., Vallejo, G., & Valle, A. (2017). How do student prior Achievement and homework behaviors relate to perceived parental involvement in homework? *Frontiers in Psychology*, *8*, 1217. https://doi.org/10.3389/fpsyg.2017.01217

OECD. (2015). *The ABC of gender equality in education: Aptitude, behavior, confidence*. OECD Publishing.

Pape, S. J., Irving, K. E., Owens, D. T., Boscardin, C. K., Sanalan, V. A., Abrahamson, A. L., Kaya, S., Shin, H. S., & Silver, D. (2012). Classroom connectivity in Algebra I classrooms: Results of a randomized control trial. *Effective Education*, *4*(2), 169–189. https://doi.org/10.1080/19415532.2013.841059

Patall, E. A., Cooper, H., & Civey Robinson, J. (2008). Parent involvement in homework: A research synthesis. *Review of Educational Research*, *78* (4), 1039–1101. https://doi.org/10.3102/0034654308325185

Pituch, K. A., Whittaker, T. A., & Stapleton, L. M. (2005). A comparison of methods to test for mediation in multisite experiments. *Multivariate Behavioral Research*, *40*(1), 1–23. https://doi.org/10.1207/s15327906mbr4001_1

Pope, D., Brown, M., & Miles, S. (2015). *Overloaded and underprepared: Strategies for stronger schools and healthy, successful kids*. Jossey-Bass.

Pressman, R. M., Sugarman, D. B., Nemon, M. L., Desjarlais, J., Owens, J. A., & Schettini-Evans, A. (2015). Homework and family stress: With consideration of parents' self confidence, educational level, and cultural background. *The American Journal of Family Therapy*, *43*(4), 297–313. https://doi.org/10.1080/01926187.2015.1061407

Quint, J. C., Sepanik, S., & Smith, J. K. (2008). *Using student data to improve teaching and learning: Findings from an evaluation of the formative assessments of students thinking in Reading (FAST-R) Program in Boston elementary schools*. MDRC.

Ramdass, D., & Zimmerman, B. J. (2011). Developing self-regulation skills: An important role of homework. *Journal of Advanced Academics*, *22*(2), 194–218.

Roschelle, J., Feng, M., Gallagher, H. A., Murphy, R., Harris, C., Kamdar, D., & Trinidad, G. E. (2014). *Recruiting participants for large-scale random assignment experiments in school settings*. SRI International.

Roschelle, J., Feng, M., Murphy, R., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open* *2*(4), 233285841667396. https://doi.org/10.1177/2332858416673968

Roschelle, J., Murphy, R., Feng, M., & Bakia, M. (2017). *How big is that? Reporting the effect size and cost of ASSISTments in the Maine homework efficacy study*. SRI International.

Schochet, P. Z., Puma, M., & Deke, J. (2014). *Understanding variation in treatment effects in education impact evaluations: An overview of quantitative methods* (NCEE 2014–4017). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Analytic Technical Assistance and Development.http://ies.ed.gov/ncee/edlabs.

Scott-Jones, D. (1984). Family influences on cognitive development and school achievement. *Review of Research in Education*, *11*, 259. https://doi.org/10.2307/1167237

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *AERA monograph series on curriculum evaluation*, No. 1. Rand McNally.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Silvernail, D. L., Pinkham, C. A., Wintle, S. E., Walker, L. C., & Bartlett, C. L. (2011). *A middle school one-to-one laptop program: The Maine experience*. Maine Educational Policy Research Institute, University of Southern Maine.

Singh, R., Saleem, M., Pradhan, P., Heffernan, C., Heffernan, N., Razzaq, L., Dailey, M. (2011). *Improving K-12 homework with computers*. Proceedings of the Artificial Intelligence in Education Conference, Auckland, New Zealand.

Tipton, E. (2014). How generalizable is your experiment? An index for comparing experimental samples and populations. *Journal of Educational and Behavioral Statistics* *39*(6), 478–501. https://doi.org/10.3102/1076998614558486

Tipton, E., & Miller, K. (2016). *The Generalizer: A webtool for improving the generalizability of results from experiments*. http://www.thegeneralizer.org

Trautwein, U. (2007). The homework-achievement relation reconsidered: Differentiating homework time, homework frequency, and homework effort. *Learning and Instruction*, *17*(3), 372–388. https://doi.org/10.1016/j.learninstruc.2007.02.009

Trautwein, U., & Köller, O. (2003). The relationship between homework and achievement–still much of a mystery. *Educational Psychology Review*, *15*(2), 115–145. http://dx.doi.org/10.1023/A:1023460414243

Trautwein, U., Lüdtke, O., Kastens, C., & Köller, O. (2006). Effort on homework in grades 5–9: Development, motivational antecedents, and the association with effort on classwork. *Child Development*, *77*(4), 1094–1111. https://doi.org/10.1111/j.1467-8624.2006.00921.x

Trautwein, U., Lüdtke, O., Schnyder, I., & Niggli, A. (2006). Predicting homework effort: Support for a domain-specific, multilevel homework model. *Journal of Educational Psychology*, *98*(2), 438–456. https://doi.org/10.1037/0022-0663.98.2.438

Trautwein, U., Niggli, A., Schnyder, I., & Lüdtke, O. (2009). Between-teacher differences in homework assignments and the development of students' homework effort, homework emotions, and achievement. *Journal of Educational Psychology*, *101*(1), 176–189. https://doi.org/10.1037/0022-0663.101.1.176

US Department of Education. (2016). *Future ready learning: Reimagining the role of technology in education*. US Department of Education, Office of Educational Technology.

What Works Clearinghouse. (2017). *What Works Clearinghouse standards handbook* (Version 4.0). Institute of Education Sciences.

Wiliam, D., & Thompson, M. (2007). Integrating assessment with learning: what will it take to make it work? In C.A. Dwyer (Ed.), *The future of assessment: Shaping teaching and learning* (pp. 53–82). Lawrence Erlbaum Associates.

Xu, J. (2006). Gender and homework management reported by high school students. *Educational Psychology*, *26*(1), 73–91. https://doi.org/10.1080/01443410500341023