Building a Statewide Model for Scaling ASSISTments

Introduction

The ASSISTments Foundation (TAF) is delighted to submit an EIR Expansion Grant in partnership with Worcester Polytechnic Institute (WPI) and Abt Global (Abt), with support from the Maryland State Department of Education (MSDE), focused on developing, evaluating, and scaling ASSISTments Plus (A-Plus), a multi-level stakeholder support model for school districts implementing ASSISTments. At its heart is ASSISTments, a math practice and assessment platform used by teachers to conduct a four-step formative assessment routine (Appx. J.1.A.a-e). The design of ASSISTments is based on evidence demonstrating that digital tutoring, immediate feedback, and formative assessment improve math learning outcomes (see Conceptual Framework Exhibit 6). The impact of ASSISTments has been validated by two large randomized controlled trials, the most recent of which found that students of color and students experiencing poverty saw the strongest learning gains (Feng, Huang, & Collins, 2024).

To date, TAF has had great success engaging large numbers of individual teachers in using ASSISTments to improve their students' math achievement, with 2,000 free teacher users, reaching 90,000 students this past school year. However, to truly scale and sustain the impact of ASSISTments in classrooms, teachers need the support of their school, and districts need to invest in ASSISTments as a long-term priority. Especially since the pandemic, school districts want to play a role in selecting and monitoring the quality of EdTech adopted within their system so they can more effectively support learning recovery. As a free teacher-level solution, ASSISTments is often overlooked when districts decide what to adopt, limiting scalability.

In this project, through a partnership with MSDE, we will build multi-year district partnerships to implement A-Plus in a growing number of middle schools within the district (see MSDE Letter in Appx. C). As outlined in our logic model (Appx. G), the A-Plus model for schools has four key components: 1) Professional learning for teachers (replicating the support we provided in our previous studies), 2) Training and support for school champions, 3) the ASSISTments platform, and 4) Insights Hub, with real-time trend data across teachers within a school. The designated district champion will also have access to Insights Hub data and engage in planning and progress monitoring to support program success during the school year. The district champion and MSDE will also be critical in identifying additional partners to support continued expansion (see Appx. J.2.A). In developing A-Plus, we will address key barriers to scale, which include 1) the need to directly respond to the needs of students with high degrees of unfinished learning post COVID-19 pandemic and 2 & 3) a lack of capacity at the school level to support implementation.

The work in Maryland will provide a valuable model for place-based scale that we will bring to other states in Years 4 and 5 of the project. Through this project, we anticipate that A-Plus will reach 18,000 grade 6-8 students, the majority of whom are high-need and served by a school with Title I status (Appx. J.2.B).

The accompanying randomized controlled trial, conducted by Abt, will test whether ASSISTments improves student math achievement at scale. It will expand our evidence base to include impacts in more grades and contexts, directly and strategically supporting scale in Maryland and nationally. The current evidence base for ASSISTments is for 7th-grade students. For scalability, it is imperative we demonstrate impact across the middle school grades (6th, 7th,

and 8th), as schools are often seeking solutions that can be adopted holistically across grades. Moreover, past studies have focused on the use of ASSISTments to support homework. The proposed study will examine a broader range of use cases for ASSISTments, including for both classwork and homework. Lastly, past studies have primarily focused on teachers assigning grade-level practice from their curriculum. One of our scale strategies will make it easier for teachers to assign personalized content to address students' unfinished learning. As schools are seeking robust solutions that solve multiple instructional problems for teachers, and it will be beneficial for scalability to demonstrate the value of immediate feedback and actionable data from ASSISTments across these various use cases (see Appx. J.2.C). The study will also shed light on the impact of a multi-stakeholder support model that builds the capacity of designated school champions to support effective implementation. This will bolster our ability to scale ASSISTments as a school-level solution for improving middle school math achievement.

ABSOLUTE AND COMPETITIVE PRIORITIES

Absolute Priority 1 - Strong Evidence

Teachers' use of ASSISTments to support instruction has strong evidence of effectiveness in improving student achievement in 7th grade math, especially for students with lower prior achievement. The independent evaluation (Roschelle et al., 2016) found that students in schools randomly assigned to use ASSISTments for mathematics homework learned more than students in schools randomly assigned to business as usual (effect size $g = 0.18$). The researchers also found that students with lower prior mathematics scores experienced a greater benefit from the intervention (effect size $g = 0.29$). ASSISTments also reliably changed teacher practice. Teachers in the intervention group more effectively targeted class time to student difficulties and errors.

This study has been reviewed by the WWC (2019) and was designated with "positive effect without reservations" (see the Evidence Form for details as well as Appx. J.6).

More recently, researchers conducted a randomized control trial of ASSISTments involving 63 schools from 41 geographically diverse districts in North Carolina, with a demographically diverse sample of 5,991 7th grade students (Feng, Huang, & Collins, 2024). The researchers were unable to obtain the state End-of-Grade test (EOG), the immediate student learning outcome, because it was canceled for spring 2020. However, they obtained and analyzed the results of the following year's EOG test for all students in the randomized schools and found a significant positive impact on students' 8th grade mathematics achievement. The researchers found that ASSISTments benefitted non-white students, who started with significantly lower 6th grade EoG scores, significantly more than white students, and that the impact was stronger for Hispanic students than non-Hispanic students. Students in schools with a higher percentage of students from economically disadvantaged backgrounds also saw the strongest gains.

Absolute Priority 2 - Field-Initiated Innovations

ASSISTments is a true field-initiated innovation. The founders of ASSISTments met while teaching in a Title I school in Baltimore in 1995. They personally experienced the challenges of implementing a formative assessment routine, given how time consuming grading, giving feedback, and analyzing data trends for a class of 32 students can be. They had the foresight to know that with the inevitable proliferation of computers in classrooms, technology would be instrumental to making formative assessment scalable.

With this origin story, the ethos of valuing the teacher's voice is core to how TAF operates. We are a non-profit dedicated to scaling the impact of ASSISTments. In the past two years alone,

we have engaged 450 teachers in user research, 75% of whom work in Title I schools, to inform 23 significant feature improvements to our product. Through this grant, we will continue to conduct user-centered design and agile product development with a laser focus on meeting the needs of school districts.

Competitive Preference Priority 2 - Addressing the Impact of Covid-19

ASSISTments supports evidence-based formative assessment practices and is a proven strategy for Covid-19 recovery in 7th grade math, a core and high stakes subject area. The two randomized controlled trials showed that when teachers use ASSISTments consistently, it increases academic success in math for all students but also specifically supports learning acceleration for those who are furthest behind (Roschelle et al., 2016, Feng, Huang, & Collins, 2024). A cost-effectiveness analysis showed that the return on investment of ASSISTments is akin to a well-regarded and much more expensive high-dosage tutoring program (Feng, Weiser, & Collins, 2024). As of January 2024, school districts had $51 billion of the original $190 billion in federal relief funds remaining. As they look ahead, it is essential they identify cost-effective solutions (Fahle et al., 2024).

A. SIGNIFICANCE

As the nation recovers from the devastating impacts of the pandemic, mathematics education continues to be an urgent area of need for improvement. A recent analysis found that there is only one state where average student achievement in math exceeds pre-pandemic levels (Fahle et al., 2024). The average score for 13-year-old students on the most recent NAEP assessment declined by 9 points in math compared to 2019-2020 (larger than the decline in reading) (NAEP,

2023). The typical student in the poorest district had double the decline in math scores post-pandemic compared to students in the richest districts (Kane & Reardon, 2023, May 11).

Maryland is a compelling state to develop a proof point for reversing learning loss in middle school math. Maryland has 200,000 middle students across 23 school districts. It is a demographically diverse state, with 30% of students identifying as Black, 23% as Hispanic, 32% as White, and 9% as Asian (MSDE, 2023 Sept). The most recent state test data showed that 25% of grade 3-8 students were proficient in math. Only 17% of students were Algebra ready, a critical benchmark for future success, demonstrating the urgency for improvement in middle school (MSDE, 2023 Aug). The lessons learned and successes from implementing A-Plus in Maryland will be compelling to other states.

The demand for technology solutions for learning recovery continues to grow. Some estimate the pandemic increased spending on education technology by $50 billion (EdTech Evidence Exchange, 2021 March). School districts access an average of 2,591 products each month, 75% of which are teacher and student-facing (LearnPlatform by Instructure, 2023). Interest in Generative AI is growing alongside technology. RAND's American Educator Panel Omnibus Survey (RAND, 2023) found over 60% of teacher respondents expect to use GenAI either the same amount to support classroom instruction or more.

ASSISTments is the most rigorously studied and proven EdTech solution on the market. The ASSISTments Foundation is uniquely positioned to implement promising AI and tech-enabled strategies to benefit underserved students within the system. In a crowded market of for-profits, we are a mission-driven non-profit with a deep commitment to research and innovation dating back to our origins at Worcester Polytechnic Institute. One of our innovations in this project is to

enhance ASSISTments with personalized assignment recommendations for students to better address the high variability in student readiness for grade-level learning post COVID-19 pandemic.

To support this feature, we will need to increase the number of math problems within our content library. Traditionally, we have engaged teachers in developing content; in order to scale, we will use generative AI to provide crucial time and cost savings. For each of the 200 prerequisite learning standards we have identified for grades 3-8 (ANet, n.d.), we will create at least 30 practice problems, each with three intentionally sequenced hints. This kind of content development is typically time-consuming and costly. In the past three years, we have spent $160,715 engaging teachers in creating 14,796 “similar but not the same" math problems with embedded hint messages (see Appx. J.2.D.a. for an example). This is approximately $11 per problem, which includes the cost of training teachers, time spent creating the problems with hints, and quality control. We have long sought more scalable, cost-effective approaches to developing this content, including crowdsourcing hints (Patikorn & [redacted] 2020).

We will partner with [redacted] "Artificial Intelligence in Ed" Lab at Worcester Polytechnic Institute, where ASSISTments was first developed, to advance the quality and quantity of math problems with hints in ASSISTments using ChatGPT. [redacted] is globally recognized for his pioneering research in applying AI to education. A recent analysis of 4,447 papers accepted at the nine most prestigious venues for work in AI in education by Ahmed et al. (2024) revealed that [redacted] has the highest number of accepted papers over the past seven years. Additionally, this analysis by Ahmed shows that WPI ranks second behind CMU, outperforming other top AI institutions like MIT, Stanford, and Berkeley (U.S. News, 2024). Dr.

[redacted] lab discovered that GPT-generated explanations of math problems were rated higher by reviewers than those created by teachers (Prihar et al., 2023). Another study found that GPT-generated explanations can enhance student learning compared to no explanations and are as effective as teacher-created ones, suggesting the strong potential for GPT-generated content (Haim, Worden, & [redacted], 2024 submitted).

For this project, WPI will similarly refine an approach for creating math content using GPT, in this case, math problems with hints. [redacted] will develop prompts that support evidence-based practices from social psychology and learning science. Examples include making math problems more culturally relevant (Appx. J.2.D.b**)** and developing hints informed by well-regarded approaches to motivation, like Carol Dweck's growth mindset (Appx. J.2.D.c). His team of researchers will rigorously evaluate the effectiveness of ChatGPT-created hints on student learning within the platform, leveraging ASSISTments' backend research infrastructure (Ostrow et al., 2017). WPI will publish their findings as they have done previously (Patikorn & [redacted], 2020; Prihar et al., 2022) for the benefit of other learning platforms and educators.

B. STRATEGY TO SCALE

B.1 Strategies That Address Barriers To Scale

ASSISTments is a free EdTech solution that has primarily been adopted at the teacher-level. Being free supports organic teacher-by-teacher growth; however, we will continue to face barriers to scaling ASSISTments in a sustained and high quality way if we do not develop a more systems-based approach to adoption and implementation. A-Plus aims to scale ASSISTments in *partnership* with school districts and leverage local school capacity to ensure *high-quality* and

sustained adoption. Below are the three barriers we have identified and the strategies we will implement to overcome them.

Barrier	Strategy
Schools need solutions that help students from varied contexts and with different unlearned content recover from the COVID-19 pandemic	1. Automate personalized assignment recommendations for students aligned to prerequisite standards needed for learning acceleration
Schools and districts lack data to monitor implementation and student progress	2. Provide real-time data reporting for schools and districts
Limited capacity at the school level to support and sustain implementation	3. Develop professional learning for school champions

Exhibit 1: Barriers and Strategies to Scale ASSISTments

Strategy 1 - Automate personalized assignment recommendations for students

Addressing widespread learning loss is a top priority and an instructional challenge in school districts (Guryan & Ludwig, 2023). Districts are looking for solutions that help teachers address the wide ranging learner variability within their classrooms and monitor post-pandemic learning recovery. Based on learning data collected within our system, we will develop a feature that recommends practice problems to be efficiently assigned to individuals or subgroups of students. The problems will align to prerequisite standards that are precursors to grade-level standards where the student may be struggling (see our prerequisite map in Appx. J.2.D.d). All the problems will come with high quality hints students can access while they work. Practice in ASSISTments with feedback is a proven strategy for learning recovery. In our most recent study, there was a statistically significant relationship between the number of problems students completed and improvements in student learning (Feng, [redacted], Collins, 2023). To make these assignments as motivating for students as possible, we will add additional problem-solving tools for students, like interactive manipulatives. An aligned data report will allow teachers and

school and district administrators to monitor student progress on standards over time. As noted previously, through a partnership with WPI, we will develop a cost-effective approach to using ChatGPT to expand our item bank of problems to support this feature.

Strategy 2- Real-time data reporting for schools and districts

Access to real-time data at every level (student, class, teacher, school, and across schools) will allow leaders to effectively monitor implementation and have a clear understanding of student progress within ASSISTments assignments. In 2022, we developed the initial infrastructure for providing schools and districts access to data from ASSISTments across teachers, called the Insights Hub (Appx.J.2.E.a). This data reporting primarily focused on showing usage and student progress.

Through our user research with 20 school and district administrators across 8 school districts, we learned districts are interested in viewing data in a more standards-based way and not necessarily tied to their curriculum. In this project, we will develop a data view that shows student progress on standards over time (Appx. J.2.E.b), with the ability to quickly pinpoint students below proficiency by standard so they can proactively plan for intervention (Appx. J.2.E.c). We have also learned that schools and districts are eager to have real time data to monitor platform usage against implementation goals. Historically, we have provided static monthly usage reports. However, we will build another data view to enable real-time monitoring and customization to goals (Appx. J.2.E.d). Lastly, districts and schools would like to see the same assignment report as teachers see, but with the ability to look at average scores across teachers and schools (Appx. J.2.E.e).

Strategy 3 - Professional learning for school champions

With A-Plus, we will partner with school districts and be able to replicate the robust professional learning (PL) we provided teachers in our research studies, which is something we have been unable to provide to our free teacher users. This PL includes a 6-hour summer training and three to five in-person coaching visits per year. We will adapt the materials from prior studies to reflect product updates and the broader range of use cases and grade levels addressed in this project. Beyond teacher PL, we have learned from working with 8 school districts over the last two years that engaging multiple stakeholders at all levels in meaningful planning, as well as building capacity at the school level to support implementation, leads to greater implementation success. Based on this learning, we will designate a champion with the capacity to support the program's success in the district and within each school. Both will receive access and training on Insights Hub so they can access real-time data on teacher and student progress. Both will participate in structured quarterly reviews of data and progress against goals (Appx. J.2.F.a).

The school champion will be a math instructional leader and will be provided with implementation resources and professional learning from the A-Plus Coach to support teachers in two key stages: *Initial Launch* and *Sustained Success*. *Initial Launch* will include setting implementation goals, assessing readiness for launch, planning for stakeholder engagement, and building buy-in, guided by an implementation guide (Appx. J.2.F.b). *Sustaining Success* will focus on providing teachers with ongoing support. We will ensure the champions are well versed in ASSISTments and attend the summer training alongside teachers. They will also meet with the A-Plus Coach during each coaching visit to debrief and discuss implementation challenges and strategies for success. See Appx. J.2.F.c for more detailed descriptions.

B.2 Adequacy of the Management Plan

The project will take place over 5 years, from January 2025 through December 2029. Over these 5 years, the project will center on four core goals: Goal 1- Develop, Pilot, and Refine A-Plus within Maryland; Goal 2- Implement A-Plus in Study Schools with Fidelity; Goal 3- Rigorously Evaluate A-Plus, and Goal 4- Disseminate Results and Scale A-Plus within Maryland and Nationally. Exhibit 2 provides a high-level overview of key activities related to each goal aligned to the project's timeframe and how we will collectively reach 18,000 students through this project. The development activities of Goal 1 will focus on the key components of A-Plus as indicated in our logic model (Appx. G). We plan to leverage our Maryland partnership to pilot, study, and scale A-Plus and do intentional work adapting our intervention to meet the needs of partners outside of Maryland, paving the way for national scale.

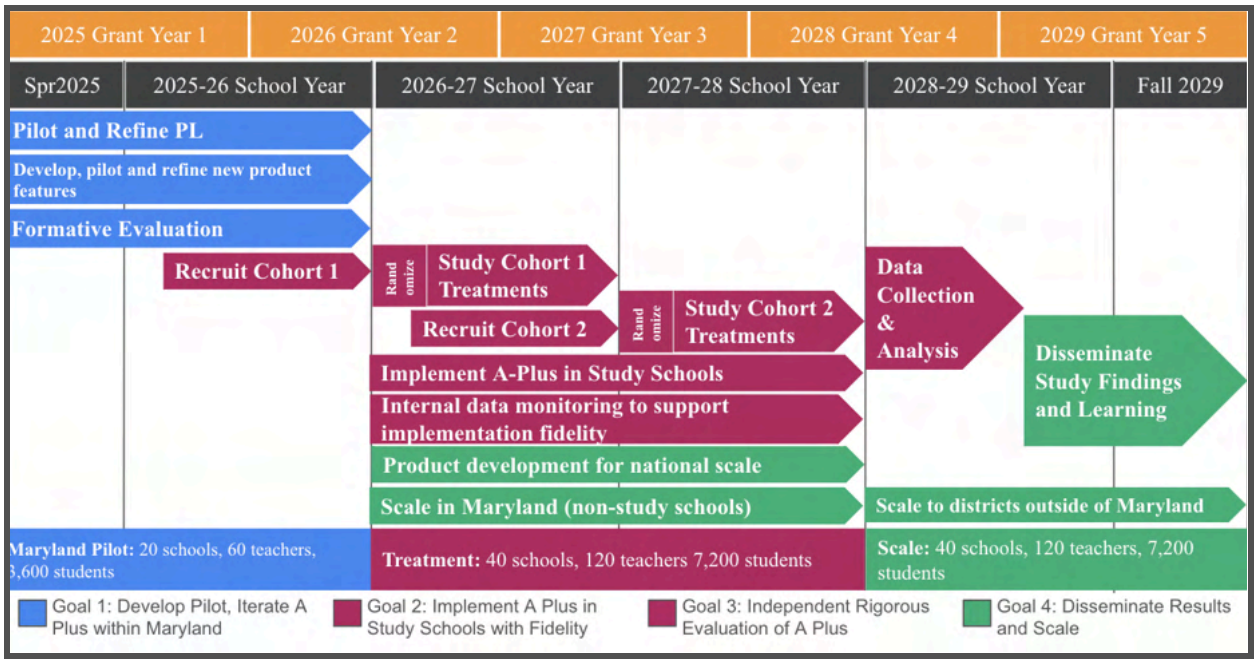

Exhibit 2: Overview of key activities related to each goal aligned to the project's timeframe.

To achieve project goals, we have developed a strong management plan with clear objectives (Exhibit 7) and a set of milestones that will be crucial to achieving each goal (Exhibit 3). We will successfully achieve our goals in collaboration with WPI and Abt. TAF already has an ongoing and successful partnership with WPI, overseen by a co-founder of ASSISTments, [redacted], having collaborated on 15 grant-funded projects to date. This includes well-established systems for co-developing new innovative features within ASSISTments, which will be critical to creating personalized assignment recommendations. TAF leadership has an existing relationship with Abt, who will conduct all aspects of the external evaluation through the technical assistance we have received through our mid-phase EIR grant. In developing this project proposal, we established a regular cadence of meetings and a strong understanding of the goals of this project. Lastly, this work will be successful due to an existing formal partnership with the MSDE (see MSDE Letter in Appx. C). MSDE is committed to helping us identify pilot, study, and scale district partners for the next four years. TAF meets regularly with the Director of Mathematics for the state, overseen by the Deputy Superintendent, as part of this $2M grant ($1M from the Arnold Foundation and $1M from the State).

Exhibit 3 delineates clear roles and responsibilities for each partner, as well as an indication of the specific team within TAF that will be involved: Data and Research, Product Development, Content Development, and Implementation. These milestones have been used to develop the budget, ensuring work is completed on time and within budget.

Milestones	Spring '25	SY 1	SY 2	SY 3	SY 4	Fall '29	Responsible Org/Teams	Involved
Goal 1: Develop, Pilot, and Iterate A-Plus within Maryland								
Partner with 20 schools from at least 3 districts in Maryland to pilot A-Plus	X	X					TAF (implementation)	MSDE
Implement and improve PL materials based on feedback from pilot participants	X	X					TAF (data, implementation)	PDs

Develop 3 new Insights Hub data views	X	X					TAF (product)	PDs, MSDE
Use ChatGPT to create at least 6,000 high-quality math problems	X						WPI, TAF (content)	CEs
Study the efficacy of AI-created hints and improve quality based on learning	X	X	X	X			WPI	TAF (content, product), PDs
Launch personalized assignment recommendations w/ ChatGPT problems	X	X					TAF (product, content), WPI	CEs, PDs
Collect quant + qual data to inform product improvements	X	X					TAF (product)	WPI, PDs
Final formative evaluation report with documented learning and improvements		X					TAF (data)	WPI, Abt
Goal 2: Implement A-Plus in Study Schools with Fidelity								
Implement A-Plus in 40 treatment schools			X	X			TAF (implementation)	SDs, MSDE
Monitor implementation fidelity and stakeholder experience to support success			X	X			TAF (data, implementation)	Abt
Goal 3: Rigorously Evaluate A-Plus								
Refine materials/procedures		X					Abt, TAF (data)	PDs
Recruit a total of 80 schools (40 treatment/40 control)		X	X				TAF (implementation)	Abt, MSDE
Random assignment			X	X			Abt	SDs
Prepare instruments and collect data		X	X	X			Abt	SDs
Measure implementation fidelity			X	X			Abt, TAF (data)	SDs
Assess intervention effects				X	X		Abt	SDs
Collect data and assess cost			X	X	X		Abt, TAF (data)	SDs
Goal 4: Disseminate Results and Scale A-Plus within Maryland and Nationally								
Scale A-Plus to 2 additional school districts in Maryland (non-study)			X	X	X		TAF (implementation)	MSDE, ScDs
Disseminate study and A-Plus resources					X	X	Abt, TAF, WPI	MSDE
Scale to districts in 2 different states				X			TAF (implementation)	ScDs
Assess needs of new partner sites and associated product development				X	X		TAF (product)	ScDs
Implement A-Plus, with needed product changes, in new partner sites					X	X	TAF (implementation, product)	ScD
Key Institutions and Teams: TAF - The ASSISTments Foundation; WPI - Worcester Polytechnic Institute; Abt - Abt; MSDE - Maryland Department of Education; PDs - Pilot Districts, SDs - Study Districts; ScDs - Scale Districts; CBs - Content Editors								

Exhibit 3: Milestones, Timeline, and Responsibilities

B.3. The Applicant's Capacity to Bring the Proposed Project to Scale on a National Level

Management capacity: This project will be led by the ASSISTments Foundation (TAF). TAF has sole intellectual property over the ASSISTments product and is charged with scaling ASSISTments nationally. We have a proven track record of successfully managing large scale, multi-partner projects; currently engaged in 12 projects funded through $17M in federal grants from NSF, IES, and EIR (including serving as the lead on two Mid-Phase EIR grants). Given this experience, we have a strong template for systems and structures that enable the success of large federal grants and for working with third-party research partners to execute on successful studies.

TAF has received a $200,000 grant from the Overdeck Family Foundation that will allow us to hire a skilled consultant to support planning for scale, adding invaluable capacity to the project activities of this grant. As part of that work, we will assemble an Advisory Board of state and district leaders, funders, and experts in EdTech scale to provide external guidance and feedback. The board will continue through the duration of this project and include funders, who can then be engaged and invest in future needs aligned to this project and scaling ASSISTments (see Overdeck Letter in Appx. C).

We are also extremely well-positioned to succeed in achieving the project's goals due to several invaluable partnerships. We have already named MSDE and the Arnold Foundation as key supporters contributing to our success in Maryland. Their support is part of a broader initiative called the Maryland Initiative to Scale Proven Programs, orchestrated by [redacted], Founder and President of the Coalition for Evidence-Based Policy. [redacted] sees Maryland as phase I of this work and seeks to launch similar initiatives in other states. [redacted] will be

championing ASSISTments in conversations with other state leaders and is providing significant ground softening for national growth (see [redacted] Letter in Appx. C).

Financial Resources: TAF has a $4.4M annual budget and is a financially healthy organization. We are a trusted partner and have succeeded in attracting grants from high-profile philanthropies (Schmidt Futures, Gates, CZI) and the U.S. DOE (NSF, IES, EIR). We have a number of supporters for this project. MSDE (see MSDE Letter in Appx. C) and the Arnold Foundation have each committed $1 million (see Arnold Match Committment in Appx. H) to support ASSISTments' scale in Maryland. The Arnold funding will be used as match funding and directly contribute to the activities of this project (see Budget Justification - non-federal funds).

Qualified personnel: TAF has 24 full-time employees, overseen by [redacted], who has successfully attracted top talent, established strong systems, and enabled ASSISTments to reach 1 million students in the past five years. Our team includes talented math and PL leaders, product designers, engineers, and data and research experts. In the past two years, the TAF implementation team, overseen by [redacted], has launched and provided services to 67 schools through 8 district partnerships (see Ann Arbor Letter in Appx. C). This team has already built a strong relationship with MSDE and successfully recruited Baltimore County and Prince George's County as partners for next year (see PGC Letter in Appx. C). The product team successfully built an initial version of the Insights Hub, demonstrating the ability to execute on ambitious product development. With a deep commitment to research, we have internal measurement and evaluation plans for all of our work, and that learning has enabled a clear understanding of stakeholder needs when it comes to developing the A-Plus model.

TAF is fortunate to have the expertise of [redacted] and his lab at WPI, as well as [redacted], a senior and seasoned evaluation director from Abt, on this project. Exhibit 4 provides an abbreviated list of key personnel for each partner written into the proposal budget, along with their experience and primary responsibilities on the project. A more detailed list of personnel for each partner is included in Appendix J.3.A-J.3.C.

Role in Project	Experience and Primary Responsibilities
The ASSISTments Foundation	
[redacted], **Executive Director** Principal Investigator 8.33% FTE in Yr 1-5	• Critical leader in the success of ASSISTments since developing it with [redacted] in 2005, including raising $3M to the launch of the ASSISTments Foundation in 2019 • 12 years of experience in middle school math education as a teacher, coach, and tutor • Master's Degree in Math Education working with [redacted], author of one of the best-selling books on instruction. • Senior personnel on dozens of NSF, IES, EIR, Gates, Hewlett, Schmidt, and other grants. TAF has an annual budget of over $4.3M. **For this project, she will:** • As the leader of TAF, she will oversee the successful execution of TAF responsibilities, write the annual report, and be the lead point of contact for Abt in the evaluation.
[redacted] **VP Programs and Partnership** 45% FTE in Yr 1, 50% FTE in Yr 2-3, 48% FTE in Yr 4, 45% FTE in Yr 5	• 17 years of experience in education as a teacher, school coach, district leader, program officer, and non-profit leader, with 13 years specifically in EdTech, research, and scale • 5 years developing and leading high quality PL programs for teachers • Currently oversees the success of 3 EIR Midscale projects, one in which TAF is a subgrantee and two where TAF is the lead. • As a Program Officer, oversaw investment and advised organizations in building the capacity to be data-driven and engage in formative evaluation of their programs **For this project, she will:** • As a cross-functional leader within the organization, [redacted] will oversee the success of the data and implementation teams and coordination across the three aligned to grant goals • Manages the relationship and all reporting with MSDE, [redacted] and other key stakeholders for the Maryland Initiative • Oversees the Data and Research team, ensuring high quality programmatic reporting and formative evaluation of A-Plus
[redacted] VP Product 25% FTE in Yr 1-5	• 30 years developing EdTech products

	● Proven ability to manage, lead and motivate diverse teams to deliver high-quality products in fast-moving and rapidly changing customer and market environments. ● Has led TAF's product development since the beginning, growing the team to 13 people. Adept at attracting and growing engineering talent **For this project, he will:** ● Manages all product related deliverables (as outlined in Goal 1). Includes overseeing product owners, UX designers, and engineers, as well as the content team ● Oversees success of product-related user research embedded within the formative evaluation around discovery, validation, usability and beta testing ● Oversee the successful collaboration with WPI on the development of personalized assignment recommendations
For more information on personnel from TAF, see Appx. J.3.A	
Worcester Polytechnic Institute	
PI [redacted] Project Director Creator of ASSISTments Computer Science Professor 7.56% FTE in Yr 1-5	● Known thought leader and researcher in AI, Learning at Scale, and Educational Data Mining, with over 300 published papers ● PhD in Computer Science from Carnegie Mellon University ● Computer Science Professor at WPI and Director of Learning Sciences & Technologies PhD program, including teaching AI. ● He was a former math teacher through Teach for America in Baltimore, where he met [redacted]. **For this project, he will:** ● Oversee using ChatGPT to develop high-quality problems with hints and the integration of ChatGPT into the ASSISTments content editors
For more information on personnel from WPI, see Appx. J.3.B.	
Abt	
[redacted] Evaluation Director 9% FT in Yr 1, 12% FTE in Y2-5	● 20 years of experience designing and conducting large-scale, multi-year evaluations of educational programs. ● Study Lead for the Comprehensive Center Implementation Evaluation and Principal Investigator for the evaluation of a Supporting Effective Educator Development grant. ● Has served as the evaluator on peer-reviewed studies for educational interventions, including for multiple grants for the Replication and Expansion of High-Quality Charter Schools ● Certified in WWC 5.0 standards and completed the Institute of Education Sciences Methods Training in Cost-Effectiveness and Benefit-Cost Analysis. **For this project, she will:** ● Direct the evaluation and the impact study, lead implementation fidelity measures and analysis, and the Cost-Effectiveness study
For more information on personnel from Abt, see Appx. J.3.C	

Exhibit 4: Roles, Experience, and Primary Responsibilities of Key Personnel

B.4 Dissemination for Further Development and Replication

TAF and its partners in this project will leverage multiple mechanisms to effectively disseminate information on this project, prioritizing school district leaders with the goal of generating strong awareness and demand for A-Plus. Within Maryland, we will develop and share implementation models, success stories, and best practices via a monthly newsletter to all partner schools. We will also leverage channels available to MSDE (e.g., newsletter to superintendents) to spotlight the work statewide, including testimonials from current partner district leaders.

We will also highlight and broadly disseminate the research findings and the implementation resources outlined in section B.5 nationally. We reach over 2,000 teachers via our monthly newsletter and will also send key updates to the over 500 school and district administrator contacts we have amassed across all 50 states over the last 5 years. TAF also has multiple active social media channels (Twitter- 1,151 followers, Linkedin- 1,891 followers). We also plan to spread awareness about A-Plus at 33 conferences over the 5 years of the grant, prioritizing conferences like NCTM with a strong presence of district leaders. We will also leverage the networks of Arnold Ventures and [redacted], in addition to partner organizations like Teaching Lab, who engage with over 30 system leaders nationally, and Open Up Resources, the providers of one of the curriculum we support. As we ramp up marketing efforts and sales for A-Plus, we will leverage Google Ads and other targeted strategies to build awareness, in addition to leveraging our existing networks.

WPI will be a critical partner in disseminating research findings on ChatGPT. In addition to their own network, TAF has a contact list of over 200 researchers and industry leaders with

whom we have a dedicated newsletter to share our cutting edge innovation work conducted as part of this project with WPI. Abt will disseminate learning from the study at conferences as well as in published papers. Lastly, we plan to present this research findings at the conferences we attend.

B.5 Utility of Products and Potential for Use in a Variety of Settings

The research, implementation resources, and related products that will result from this project will be published on our website and made readily accessible in a virtual format. The associated dissemination campaigns, as noted above, will elevate the profile of ASSISTments, benefitting our over 2,000 free teachers and attracting more teachers to our free tool. They will also elevate A-Plus, generating leads for additional partnerships.

Since ASSISTments is at the heart of all of this work, it is worth highlighting that ASSISTments will be a more effective learning solution for various school settings as a result of this project. The personalized assignment recommendations will allow teachers to be more responsive to learning variability. We will be able to provide teachers and schools better guidance on using ASSISTments to meet their needs because we are studying a wider range of use cases for ASSISTments in this project (Appx. J.2.C). The Implementation Planning Guide will encourage schools to start with their district's math goals to determine the best implementation use case (e.g., using ASSISTments for homework if the district has a no homework policy will not make sense.) The Readiness Assessment will help them strengthen the conditions for success (e.g., availability of laptops aligned to when teachers need to use the platform, ensuring dedicated time for teachers to review and plan with data). The Classroom

Observation Tool will allow school staff to know what to look for when observing teachers using ASSISTments, with consideration for different potential use cases.

As we scale (Years 3 and 4), we will continue to engage in user research and product development that supports adapting to local contexts. As one example, many states have adopted modified versions of state standards with varying degrees of alignment with Common Core State Standards (CCSS). Different curricula are also more prominent depending on the geography. Schools want this represented in their teacher tools. Integration of additional systems of standards (Appx. J.4) or expanding our content library to support additional curriculum are all potentially important as we engage with new district partners in other states. This is why user research and product development are essential throughlines of this project.

C. QUALITY OF THE PROJECT DESIGN

C.1 Conceptual Framework Underlying the Proposed Research

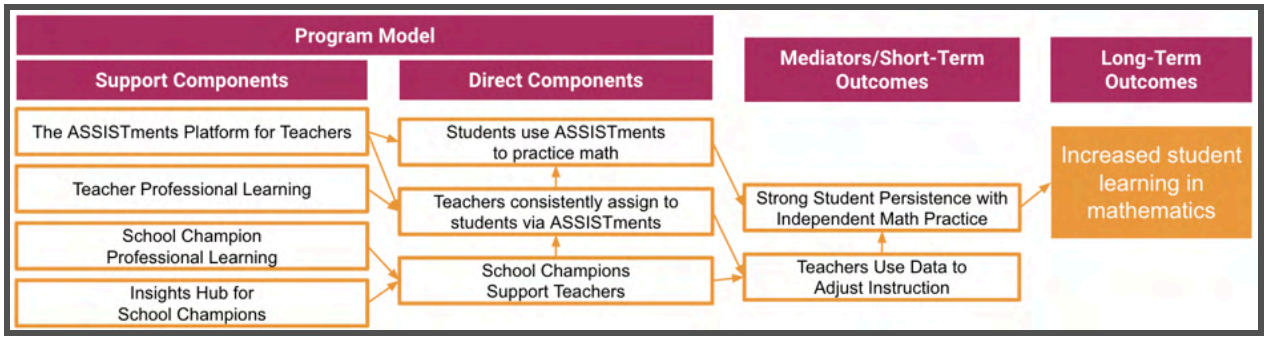

Exhibit 5: Abbreviated Logic Model for A-Plus

The above logic model (Exhibit 5, also Appx. G) highlights the conceptual framework underlying the proposed research, including the key components of the A-Plus program model and mediators for impact. The table below, Exhibit 6, highlights the core drivers of impact when adopting ASSISTments. Within the framework, we highlight how the A-Plus model further strengthens the intervention and potential for impact.

Conceptual Framework for ASSISTments	
Digital Tutoring	How Enhanced in A-Plus
Tutoring is an effective way to support student learning. Online platforms that provide tutoring in the form of hints and explanations have often been used to support K-12 students' mathematics learning (Mousavinasab et al., 2021). A series of meta-analyses reveal that online intelligent tutoring systems are effective in improving mathematics learning (Bowman-Perret, 2013; Feng et al., 2021; Koedinger et al., 1997; Kulik & Fletcher, 2016; Steenbergen-Hu & Cooper, 2013; VanLehn et al., 2005; Nickow, et al., 2020). As an online platform, ASSISTments provides a variety of tutoring strategies to support students' mathematics learning. The strategies include but are not limited to providing just-in-time support, such as hints and explanations.	- Through ChatGPT, we will scale the number of problems with high quality hints (an effective form of digital tutoring) available within the platform. - The addition of personalized assignments to supplement the existing content will increase student opportunities to practice math with hints, aligned to their learning needs.
Immediate feedback.	How Enhanced in A-Plus
ASSISTments immediately informs students on the correctness of their answers. Immediate feedback provides a quick confirmation of a student's correct understanding of a concept as well as a rapid indication of where a misunderstanding has occurred, and further clarification might be needed. Research suggests that immediate feedback benefits student learning, particularly in online learning environments with technology based systems (Cavalcanti et al., 2021).	Increasing the number of problems students complete through increased teacher assigning and automated assignments will provide students with more opportunities to practice with immediate feedback.
Using Data to Adjust Instruction part of a Formative Assessment Routine	How Enhanced in A-Plus
ASSISTments provides formative feedback to both teachers and students so they can receive timely information to adjust either instruction or response strategies (Black & Wiliam, 1998a, 1998b; Boston, 2002; Lee, et al., 2020; Roediger & Karpicke, 2006; Shute, 2008). Research has found that using formative assessment can improve student learning when the results are used to adjust instruction (Bergan et al., 1991; Lee et al., 2020; Speece et al., 2003). In ASSISTments, as students complete each problem, teachers receive feedback on student performance, which allows them to identify strengths and weaknesses across students and adjust instruction accordingly.	- Replicating the teacher training and coaching we provided in our studies will directly support teacher success with data-driven instruction - Adding data views for school and district champions will engage additional stakeholders in planning student intervention. - Professional learning for school champions will increase support for teachers with data-driven instruction - By partnering with schools formally and rolling out ASSISTments as a whole school (grade 6-8), intervention, review, and discussion of ASSISTments data for

	instructional planning will be more readily integrated into existing school meetings (e.g., grade-level planning meetings)

Exhibit 6: Conceptual Framework of A-Plus

C.2 The Extent to Which Goals, Objectives, and Outcomes Are Specified and Measurable

The four goals and 8 objectives of this project will guide the development, evaluation, and scaling up of A-Plus (see Exhibit 7). Goal 1 outcomes ensure we have a successful pilot and develop the strongest, most stakeholder-informed version of the A-Plus Program. Goals 2 and 3 support a strong evaluation that meets all rigorous requirements for EIR Expansion. Goal 4 outcomes align with our goal of further expanding the reach of A-Plus, broadly disseminating learning, research results, and open-source implementation tools, and achieving a pathway to sustainability through sales. It also acknowledges we will need to continue to evolve our product to meet the needs and requirements of different geographies as we scale (for example districts in other states may require specific additions to our content library).

Objectives	Outcomes/Measures
Goal: Develop, Pilot, and Iterate A-Plus in Maryland	
Objective 1: Develop and launch the key PL and product components of A-Plus	*By June 2026* ● 50 user interviews to inform product design ● 3 new data views added to the Insights Hub ● 6,000 ChatGPT problems with hints built into the platform ● 70% of pilot teachers assign personalized assignments ● 5 codified PL materials supporting school and district champions
Objective 2: Successfully pilot and improve A-Plus	*By June 2026* ● 80% of pilot participants (teachers, school and district champions) are satisfied with the A-Plus program ● 3 school districts, 20 schools and 3,600 students reached ● 4 formative evaluation reports complete
Goal: Implement A-Plus in Study Schools with Fidelity	
Objective 3: Deliver support to study schools with high fidelity	*By June 2027 and June 2028 (cohort 1 and 2)* ● 80% teacher attendance at summer training ● 80% school champion attendance at teacher training ● 100% Insights Hub users complete training ● 80% of teachers have minimum 3 coaching visits per teacher

	● 100% schools have Implementation Plan ● 100% of district partners attend 3+ quarterly progress meetings
Objective 4: High satisfaction with A-Plus program in study schools	*By June 2027 and June 2028 (cohort 1 and 2)* ● 80% average satisfaction on post training surveys ● 80% satisfaction with the A-Plus program overall (all stakeholders)
Goal: Rigorously Evaluate A-Plus	
Objective 5: Recruit a sufficiently large and diverse sample to examine subgroups of interest	*By July 2027, across cohorts 1 and 2* ● 80 schools with signed data-sharing agreements across cohorts 1 and 2 ● Successfully randomize 80 schools over two cohorts ● 2,700 students in the sample
Objective 6: Collect and analyze data on students in the study	*By June 2029* ● Average math achievement of 6th, 7th and 8th grade students in A-Plus schools compared to average math achievement in business as usual schools shows statistically significant gains ● Subgroup analysis is possible for students in Title I schools and by race/ethnicity ● Cost effectiveness study finds A-Plus to be cost-effective compared to similar interventions ● Release study by June 2029
Goal: Disseminate Results and Scale A-Plus within Maryland and Nationally	
Objective 7: Disseminate the results of the study, and associated implementation support materials broadly to our target audience of district leaders	*By June 2029* ● At least 5 high-quality resources highlighting the A-Plus model and the Study results released ● Presentations at 10 conferences in Years 3, 4, and 5 showcasing A-Plus and Study Results (Year 5)
Objective 8: Successfully scale A-Plus in Maryland and beyond	*By December 2029* ● Identify at least district partners in 2 different states. ● 3 product improvements aligned to new scale partner needs ● Launch A-Plus as a paid service for SY 29-30

Exhibit 7: Goals, Objectives, Outcomes, and Measures

C.3 The Design of the Proposed Project Addresses the Needs of the Target Population

A-Plus is designed to ensure that the implementation of ASSISTments is owned and appropriately adapted to local contexts and meets the needs of high-need students. We will work closely with the district champion to identify schools that will most benefit from the A-Plus intervention based on current math learning outcomes and the number of high-need students served. To build buy-in every step of the way, we will host information sessions with school

leaders and teachers to build excitement and awareness. The content of the sessions will be developed in partnership with our champions to ensure we have developed a plan and messaging that clarifies how ASSISTments solves authentic problems they experience and aligns with the math goals already in place within the school district.

The summer training will help teachers determine a plan for integrating ASSISTments successfully into their instructional routines and supporting student success in the initial month. This training will be high-quality and directly informed by our experiences from our two RCTs, which collectively trained approximately 144 teachers. Teachers' capacity to engage in data-driven instruction to meet the needs of their students will be further developed through 3-5 in-person coaching sessions provided by the A-Plus Coach (Appx. J.2.A). The A-Plus Coach will observe the teacher and then debrief what they observed alongside the school champion, with opportunities for the teacher to reflect on ways to improve and to plan the next steps based on their ASSISTments data reports. The A-Plus Coach will also meet with the school champion and review Insights Hub data together. During this time, the coach will guide the school champion in using the data to plan the next steps for student intervention and teacher support. The A-Plus Coach will also help the school champion identify and plan for upcoming meetings where ASSISTments data can add value for instructional planning (for example, an intervention or math team planning meeting).

Product development is a key workstream in this grant, and we will use best practices in user-centered design. This aligns to the ethos of how our product has been developed to date, as evidenced by the positive gains in math achievement made by students experiencing poverty and other high-need students in our studies. Through systematic user research with diverse teachers

and school and district stakeholders serving our target population, we will ensure any new product development will further strengthen that impact and align to the needs of teachers and students in high-poverty schools (Appx. J.5.A). This will be a critical part of our formative evaluation and part of the ongoing work as we continue to engage with district partners in Maryland and beyond.

D. PROJECT EVALUATION

The evaluation will consist of a formative and summative evaluation. The **formative evaluation** will focus on gathering critical data needed to effectively develop and refine the key components of the A-Plus model during the Spring of 2025 and SY 2025-2026 (Project Years 1 and 2) to maximize success with our target population. It will be led by the Director of Data and Research at TAF. They will work with key stakeholders to develop a priority set of research questions aligned to 1) the overall effectiveness of the program, 2) the design and quality of the PL and products, and 3) the capacity of the schools and districts to support effective adoption and sustain the work (see Appx. J.5.B for a more detailed plan). We will collect qualitative data via post-training surveys, a semesterly survey, 10 structured interviews a semester, and school visits from TAF staff, which include observing the A-Plus Coach and teacher implementation. The A-Plus Coach will maintain a log tracking time spent within and across schools. This will be further supplemented by user interviews with stakeholders (teachers, key school and district personnel) conducted by our Product Team to support a user-centered product design methodology (Appx. J.5.A). Qualitative data will be supplemented by analytics from our product (user behavior, feature utilization) and attendance/meeting data. The Formative Evaluation team, which also includes a data analyst, will synthesize learning quarterly to ensure learning informs

continuous improvement, and complete a final report in July 2026. These reports will be shared and discussed with MSDE, and customized for specific districts and schools as part of the quarterly progress check-ins.

The independent **summative evaluation** will estimate the impact of A-Plus on student math achievement and study the implementation and success of the strategies developed to scale. An independent team of highly experienced researchers from Abt, an evaluation firm with a long history of conducting rigorous implementation and impact studies, will conduct the summative evaluation. To ensure the summative evaluation captures key components of A-Plus, during the pilot phase of the grant, Abt will participate in regular meetings with the key TAF and WPI personnel; interview collaborators across institutions and co-developers and conduct a comprehensive review of all A-Plus program materials, and formative evaluation findings, as well as pilot implementation fidelity data, to fully understand the program as it develops and ensure the summative evaluation plans are strongly aligned. A complete table of Abt activities can be found in Appx. J.7.

The summative evaluation component will include a rigorous randomized controlled trial (RCT) to measure the impact of ASSISTments on mathematics learning, when implemented within the A-Plus model. The research will replicate previous evaluations conducted for 7th grade students using ASSISTments for homework, while also including grades 6 and 8, also use ASSISTments for classwork. There will be a total of 80 schools: 40 will receive the treatment and 40 will conduct business as usual. The summative evaluation will also measure the quality and fidelity of implementation of the A-Plus intervention in the 40 treatment schools with diverse populations and settings, including use of the Insights Hub and engagement with the PL

for administrators, and use of personalized assignment recommendations by teachers, as well as a cost analysis for the A-Plus intervention. Building on prior evaluations of ASSISTments, the study is designed to (1) yield impact estimates of ASSISTments on learning outcomes for a diverse population, (2) offer evidence of generalizability, (3) examine implementation fidelity in detail, (4) examine costs and feasibility of both implementing ASSISTments and scaling A-Plus to districts and states, and (5) provide guidance for the adoption and replication of A-Plus in other settings.

D.1. Methods Designed to Meet WWC Standards Without Reservations

Abt will conduct an independent study to address five research questions aligned with the A-Plus logic model (Appx. G) to assess the program's impact and implementation; RQ1 and RQ2 (Exhibit 8) focus on the impact of A-Plus on grade 6-8 students' mathematics achievement using a large-scale cluster RCT in 80 schools during the 2026-27 and 2027-28 school years; this design will meet WWC Standards without Reservations.

Research Questions	Data / Measures
Impact questions	
RQ1: Do students whose schools are assigned to implement A-Plus learn more than students whose schools are assigned to not implement A-Plus?	State assessment scores; student and school demographics
RQ2: Do the effects of A-Plus vary for students with different prior achievement, and for students of different grades, socio-economic status, race/ethnicity, or with other policy-relevant characteristics? Do effects vary by teacher experience or by the number of years schools have had a 1:1 computer-student ratio?	
Implementation questions	
RQ3: Are key components of A-Plus being implemented with fidelity? What factors hinder or facilitate implementation?	Computer records of usage; Interviews
RQ4: Do teachers perceive school champions as supporting the implementation of A-Plus? Are teachers' perceptions of the level of support from their school champions related to teachers' fidelity of implementation?	Teacher survey; State assessment scores

RQ5: What is the cost-effectiveness of A-Plus compared to the business as usual condition?	Cost template; interviews
Mediation questions	
RQ6: Do teachers in schools assigned to implement A-Plus report more data-driven classroom instruction? How is the impact of A-Plus on student learning mediated by data-driven teaching practice, implementation fidelity, and dosage?	Computer records of usage; Teacher survey; State assessment scores
RQ7: Does the level of student usage (such as the number of problems completed, and the number of minutes spent in the platform) mediate the effects of A-Plus on student learning?	Computer records of usage; State assessment scores

Exhibit 8. Summative Research Questions, Outcomes, & Measures**.**

Study design. The study will use a school-level clustered randomized controlled experimental design to meet What Works Clearinghouse standards without reservations (WWC, 2022). The RQs will be addressed with data collected from two consecutive cohorts of 40 schools each. Within each cohort, 20 schools will be randomly assigned to a treatment group that will implement the A-Plus model to support math instruction and 20 schools randomly assigned to a control group that will continue business-as-usual. In the treatment schools, all teachers in grades (6 through 8) will use ASSISTments to support math instruction and receive teacher training and in-person coaching from the Maryland-based A-Plus Coach. Teachers will also be supported by a designated champion within their school, who will receive PL. In the control condition, teachers will continue their current practices (business as usual), including any use of online tools other than ASSISTments. After the experiment is completed, control schools will also have the opportunity to participate in Summer training and use ASSISTments the following school year. Student rosters will be established prior to random assignment and Abt will collect these rosters two weeks prior to the start of each school year. The team will track both overall and differential school- and student-level attrition from both conditions (see Appx. J.8). Abt will minimize student-level attrition by relying on administrative data, which the WWC assumes will

have low rates of missing outcome data for individuals in clusters at follow-up (What Works Clearinghouse, 2022, p. 51). We will minimize differential attrition of schools by pairing schools in the treatment and control groups in each cohort such that if one attrits, the other will be omitted from the analysis; this approach minimizes risk of bias due to compositional change associated with cluster-level attrition. The risk of bias due to joiners is considered low due to random assignment of neighborhood schools to conditions, because switching schools is usually much more difficult than switching classrooms or teachers within schools (p. 44), and the risk of bias due to leavers is low since A-Plus is low burden (p. 47).

Sample and power. The team will recruit 80 public schools in Maryland. We will gather data on school characteristics to identify possible stratification variables (e.g., the percentage of students who achieved mathematical proficiency on prior tests, the percentage of low-income students) to improve precision and sample balance across groups. With support from Elizabeth Tipton and the webtool, "The Generalizer", we will develop a plan for recruiting a representative sample to increase the generalizability of the results of the study, with a focus on involving schools serving low-performing students and schools serving families with low income. A power analysis (using PowerUp! by Dong & Maynard, 2013) that assumes an average of three classrooms for each school grade, with 20 students per class, reveals that the proposed study has sufficient power to detect effects of 0.122 standard deviations on student achievement (see Appx. J.9).

Data Analysis Plan. The study team will estimate the impact of A-Plus on student achievement using an intent-to-treat (ITT) analysis with a two-level regression model (students nested within schools). Continuous scaled scores in grades 6 through 8 will first be transformed

into z-score units for comparability across grades. Analyses will test the overall impact of the intervention for students in grades 6 through 8, as well as differential effects on students and schools with different characteristics. To improve the precision of the impact estimate, each model will adjust for blocking (by cohort and district), a baseline measure, and other student, teacher, and school characteristics (see Appx. J.8).

D.2 Generation of Guidance About Effective Strategies Suitable for Replication

The proposed study is designed to generate insightful guidance for successful replication and scaling of A-Plus in other schools and districts. First, Abt will **strategically sample** and recruit a large number of geographically and racially diverse schools across Maryland. Second, Abt will also conduct interviews with school administrators to document the context of implementation, such as district priorities, adoption status of technology, classwork and homework policies, school-level support and cultures for data use, teacher collaboration and professional learning opportunities. Third, the evaluation will include **differential impact analysis** (RQ2, Exhibit 8) to assess to what extent the impact of A-Plus is moderated by the characteristics of students (e.g., race/ethnicity, eligibility for free or reduced-price lunch) teachers (e.g. years of experience), or schools (e.g., number of years with 1:1 ratio of computers to students) and to identify for which settings or populations the intervention is particularly effective. Fourth, the study includes two **mediator analyses**. RQ6 examines how the impact of A-Plus on teachers' classroom practices—such as using data to drive instruction, engaging students in targeted assignment review, and fostering a culture where mistakes are part of learning—mediates the effect on student learning. Similarly, RQ7 examines how the impact of A-Plus on student usage mediates the intervention's effect (Exhibit 8). To answer RQs 6 and 7, Abt will use Structural Equation

Models (SEM), which provides an appropriate inference framework for mediation analyses (Gunzler et al, 2013).

The study builds in **analysis of implementation data from multiple sources.** RQs 3 and 4 directly examines implementation fidelity, while RQs 6 and 7 focus on assessing how implementation fidelity and quality influence the impact of A-Plus (Exhibit 8). Accordingly, data will be collected on the four support components and three direct components that are features of implementation (see Appx. G for logic model). We will examine implementation fidelity based on PL records, computer system records of teacher and student usage, and teacher self-reported practices from pre- and post-surveys (see Appx. J.10 for implementation fidelity details and Appx. J.11 for data collection timeline). In particular, we will focus on the fidelity of the approaches identified to address barriers to scaling up proposed in B.1 and **identify facilitators and barriers to the implementation** of the intervention.

Last, policymakers and educational administrators require information on the cost of the resources needed to adopt, implement particular instructional interventions with fidelity, and to sustain the interventions. To provide information about whether A-Plus is more cost effective, as compared to business-as-usual practices (RQ5, Exhibit 8), Abt will conduct a **cost analysis** (see Appx. J.12). To fully capture what is needed to scale and replicate the impact of A-Plus in other settings, we will collect information on the amount and characteristics of all resources (such as personnel, materials, and equipment) using the ingredients method as described in Levin et al. (2018). Once all ingredients are identified, we will use CostOut - the CBCSE Cost Tool Kit (Hollands et al., 2015-22) to facilitate the estimation of costs and cost-effectiveness, or cost per standard deviation increase in the outcome measure.

D.3 Valid and Reliable Performance Data on Relevant Outcomes

As shown in the logic model (Appx. G), the key student outcome of A-Plus is mathematics achievement, and will be measured by grade 6 through 8 standardized state-mandated mathematics test scores from spring 2027 (Cohort 1) and spring 2028 (Cohort 2), with prior-year state mathematics tests scores from spring 2026 and spring 2027 serving as the baseline measure. The WWC assumes standardized state-mandated test scores to be reliable and are reviewable under the Measures of Academic Readiness, Knowledge, or Skills outcome domain. Abt will secure agreements with participating districts to obtain the extant state test scores in mathematics for grades 5 through 7 (as covariate) and grades 6 through 8 (as outcome), as well as student demographics (Exhibit 9).

Domain	Measure	Reliability/ Validity	Timing of Outcome Data	Baseline Measure
Math Achievement	Maryland Comprehensive Assessment Program (MCAP)	Standardized test	Spring 2027 (grades 6, 7, and 8) Spring 2028 (grades 6, 7, and 8)	Spring 2026 – prior year test score (grades 5, 6, and 7) Spring 2027 – prior year test score (grades 5, 6, and 7)

Exhibit 9: Measures and Timing of Outcome Data Collection

D.4 Components, Outcomes, and Measurable Threshold for Implementation

Components. The design of the evaluation is focused on the key components illustrated in the logic model (Appx. G). The key components include the ASSISTments Platform for Teachers, Teacher Professional Learning, School Champion Professional Learning, and the Insights Hub for School Champions. For both students and teachers, ASSISTments offers actionable data reports to which provide information on students' strengths and needs via diagnostic reports. For students, ASSISTments also offers student practice opportunities with immediate feedback and hints. For the analysis of implementation fidelity (RQ3, Exhibit 8),

treatment group teachers' classwork and homework assignments and students' use of ASSISTments will be assessed via computer system records. Based on prior findings (Fairman et al., 2015), we hypothesized that the intervention will lead to more targeted classwork and homework review practices and higher student engagement in classwork and homework.

To address RQ4, we will capture treatment group teachers' perceptions of school champions' support for the use of A-Plus though the School Implementation Climate Scale (SICS) (Thayer et al, 2022). SICS has strong reliability (alphas greater than 0.8 for all subscales). Additionally, we will examine whether perceptions of support are correlated with fidelity of implementation at the teacher level.

Short-Term Outcomes/Mediators. Teachers' increased use of data to adjust instruction and students' increased persistence on math assignments are hypothesized to mediate the effect of A-Plus on students' mathematics achievement (RQs 6 and 7, Exhibit 8). Students' persistence will be measured through computer system data. Teacher surveys will capture quality of delivery and program differentiation (whether critical features that distinguish the program from the comparison condition are present or absent during implementation). For each cohort, Abt will administer survey items drawn from the Teacher Data Use Survey (Wayman et al, 2016), which includes items on both teachers' use of data and school leader support for data use. All alpha reliabilities were greater than .80; most were greater than .90. Collecting measures of data use for both treatment and control group teachers will allow Abt to document the intervention contrast (Hill et al., 2023).

Appx. J.13 includes a table listing the survey items. Abt will also explore streamlining the survey in the pilot phase. All treatment and control teachers will take the survey online in the

summer (as baseline) before treatment teachers receive new user training and again at the end of the year. Teacher self-reported surveys will be triangulated with other sources of data (e.g., computer system data) for confirmation of teacher activities.

Measurable Threshold for Implementation. System records of all student and teacher usage of ASSISTments will serve as a primary source of data in assessing the extent to which it is used as intended (RQ3, Exhibit 8). Abt will also collect data on teachers' attendance at professional learning and views of the Insights Hub. Refining the implementation metrics developed in prior studies (Feng et al., 2014), Abt established metrics for **indicators of implementation compliance** for the support and direct components listed in the logic model (see Appx. J.10 for details). These metrics cover three categories of implementation fidelity: adherence (whether the components of the intervention are being delivered as designed), duration (the number, length, or frequency of sessions implemented), and participant responsiveness (the extent to which participants are engaged by and involved in the activities and content of the program) (O'Donnell, 2008). The minimum acceptable implementation thresholds are defined in Appx. J.10. During the school year, Abt will **calculate the implementation metrics** at the individual teacher and student level quarterly and share the results with TAF to inform continuous improvement.

References

Ahmad, K., Iqbal, W., El-Hassan, A., Qadir, J., Benhaddou, D., Ayyash, M., & Al-Fuqaha, A. (2024). Data-driven artificial intelligence in education: A comprehensive review. *IEEE Transactions on Learning Technologies. vol. 17,* pp. 12-31, 2024, https://doi.org/10.1109/TLT.2023.3314610

ANet (Achievement Network). (n.d.). Important Prerequisite Math Standards. https://docs.google.com/document/d/1YVlYSdzX1WHt4CYSfRtprf1102a4-IrMZ25ntuK25uk/edit

Bergan, J.R., Sladeczek, I.E., Schwarz, R., & Smith, A.N. (1991). Effects of a Measurement and Planning System on Kindergartners' Cognitive Development and Educational Programming. *American Educational Research Journal*. 1991;28(3):683-714. doi:10.3102/00028312028003683

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. Assessment in Education: Principles, Policy and Practice, 5, 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2): 139–149.

Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research & Evaluation*, 8(9).

Bowman-Perrott, L., Davis, H., Vannest, K., Williams, L., Greenwood, C., & Parker, R. (2013). Academic benefits of peer tutoring: A meta-analytic review of single-case research. *School psychology review, 42*(1), 39-55.

Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y. S., Gašević, D., & Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, *2*, 100027.

Dong, N., & Maynard, R. (2013). PowerUp!: A Tool for Calculating Minimum Detectable Effect Sizes and Minimum Required Sample Sizes for Experimental and Quasi-Experimental Design Studies. *Journal of Research on Educational Effectiveness, 6*(1), 24–67. 10.1080/19345747.2012.673143

EdTech Evidence Exchange. (2021 March). Overview: U.S. K-12 Public Education Technology Spending. https://EdTechevidence.org/wp-content/uploads/2021/07/FINAL-K12-EdTech-Funding-Analysis_v.1.pdf

Fahle, E., Kane, T. J., Reardon, S. F., & Staiger, D. O. (2024). The first year of pandemic recovery: A district-level analysis. *Education Recovery Scorecard. https://educationrecoveryscorecard. org/wp-content/uploads/2024/01/ERS-Report-Final-1.31*

Fairman, J., Porter, M., Fisher, S. (2015). *Principals Discuss Early Implementation of the ASSISTments Online Homework Tutor for Mathematics: ASSISTments Efficacy Study Report 2.* Menlo Park, CA: SRI International. Retrieved from: https://files.eric.ed.gov/fulltext/ED575978.pd

[redacted]

Feng, M., Huang, C., & Collins, K. (2023, June). Promising Long Term Effects of ASSISTments Online Math Homework Support. In *International Conference on Artificial Intelligence in Education*, pp. 212-217. Cham: Springer Nature Switzerland.

Feng, M., Huang, C., & Collins, K. (2024, April). *Supporting middle school math learning with a technology-based intervention: impacts, moderators, and usage.* A working paper. San Francisco, CA: WestEd.

[redacted]

Feng, M., Weiser, G., & Collins, K. (2024). Cost and cost effectiveness of ASSISTments online math support: Analysis from a randomized controlled study in middle school. WestEd. https://www.wested.org/resources/cost-and-cost-effectiveness-of-assistments-online-math-support/

Feng, S., Magana, A. J., & Kao, D. (2021, October). A Systematic Review of Literature on the Effectiveness of Intelligent Tutoring Systems in STEM. In *2021 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). IEEE.

Gunzler, D., Chen, T., Wu, P., & Zhang, H. (2013). Introduction to mediation analysis with structural equation modeling. *Shanghai archives of psychiatry, 25*(6), 390–394.

Guryan, J., & Ludwig, J. (2023). Overcoming Pandemic-Induced Learning Loss. In *Aspen Economic Strategy Group meeting draft*.

[redacted]

Hill, C. J., Scher, L., Haimson, J., & Granito, K. (2023). Conducting implementation research in impact studies of education interventions: A guide for researchers. (NCEE 2023-005). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. Retrieved from http://ies.ed.gov/ncee

Hollands, F.M., Hanisch-Cerda, B., Levin, H. M., Belfield, C.R., Menon, A., Shand, R., Pan, Y., Bakir, I., Cheng, H. & Kushner, A. (2015-22). *E$timator*. Teachers College, Columbia University. www.costtoolkit.org

LearnPlatform by Instructure. (2023). EdTech Top 40: A Look at K-12 EdTech Engagement During the 2022-23 School Year. https://drive.google.com/file/d/1horSGQq8RW8AM5rHcCrMgeBYfBAEjJGd/view

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, *33*(2), 124-140.

Kane, T. & Reardon, S. (2023, May 11). Parents Don't Understand How Far Behind Their Kids Are in School. *The New York Times.* https://www.nytimes.com/interactive/2023/05/11/opinion/pandemic-learning-losses-steep-but-not-permanent.html

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. *International Journal of Artificial Intelligence in Education*, *8*(1), 30-43.

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of intelligent tutoring systems: a meta-analytic review. *Review of educational research*, *86*(1), 42-78.

Lee, H., Chung, H. Q., Zhang, Y., Abedi, J., & Warschauer, M. (2020). The effectiveness and features of formative assessment in US K-12 education: A systematic review. *Applied Measurement in Education*, *33*(2), 124-140.

Levin, H.M., McEwan, P., Belfield, C., Bowden, A.B., Shand, R. (2018). *Economic Evaluation in Education: Cost-Effectiveness and Benefit Cost Analysis, 3rd Edition*. Sage Publications.

MSDE (Maryland State Department of Education). (2023, August 22). Maryland State Department of Education Provides 2023 Maryland Comprehensive Assessment Program (MCAP) Statewide Results. https://news.maryland.gov/msde/msde-2023mcap-statewide-results/#:~:text=Statewide%20results%20of%20the%20latest,recovery%20from%20pandemic%20learning%20loss

MSDE (Maryland State Department of Education). (2023, September 30). Enrollment by Race/Ethnicity and Gender and number of Schools; Maryland Public Schools. https://www.marylandpublicschools.org/about/Documents/DCAA/SSP/20232024Student/2023-2024-Enrollment-By-Race-Ethnicity-Gender-A.pdf

Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*, *29*(1), 142-163.

NAEP (National Assessment of Educational Progress). (2022). NAEP Report Card: 2022 NAEP Mathematics Assessment: Largest Score Declines in NAEP Mathematics at Grades 4 and 8 Since Initial Assessments in 1990. https://www.nationsreportcard.gov/highlights/mathematics/2022/

NAEP (National Assessment of Educational Progress). (2023). The Nation's Report Card: NAEP Long-Term Trend Assessment Results: Reading and Mathematics: Scores decline again for 13-year-old students in reading and mathematics. https://www.nationsreportcard.gov/highlights/ltt/2023/

Nickow, A., Oreopoulos, P., & Quan, V. (2020). The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence.

O'Donnell, C. L. (2008). Defining, Conceptualizing, and Measuring Fidelity of Implementation and Its Relationship to Outcomes in K–12 Curriculum Intervention Research. *Review of Educational Research, 78*(1), 33–84.

[redacted]

RAND. (2023, Fall). American Educator Panel (AEP) Omnibus Survey: Artificial Intelligence; School Assignment. https://bentobento.info/surveys/235

Roediger, H. III., & Karpicke, J. D. (2006). Test-enhanced learning: taking memory tests improves long-term retention, *Psychological Science*, *17*(3), 249–255.

Roschelle, J., Feng, M., Murphy, R. F., Mason, C. A. (2016). Online Mathematics Homework Increases Student Achievement. *AERA Open, 2*(4). DOI: 10.1177/2332858416673968

Puma, M. J., Olsen, R. B., Bell, S. H., & Price, C. (2009). What to do when data are missing in group randomized controlled trials (NCEE 2009-0049). National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. https://eric.ed.gov/?id=ED511781

Shute, V. J. (2008). Focus on formative feedback. *Review of educational research*, *78*(1), 153-189. https://myweb.fsu.edu/vshute/pdf/shute%202008_b.pdf

Speece, D. L., Molloy, D. E., & Case, L. P. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities: Research & Practice, 18*(3), 147–156.

Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of educational psychology*, *105*(4), 970.

Thayer, A. J., Cook, C. R., Davis, C., Brown, E. C., Locke, J., Ehrhart, M. G., Aarons, G. A., Picozzi, E., & Lyon, A. R. (2022). Construct validity of the school-implementation climate scale. *Implementation Research and Practice*, *3*. https://doi.org/10.1177/26334895221116065

U.S. News and World Report (2024) Best Artificial Intelligence Programs. https://www.usnews.com/best-graduate-schools/top-science-schools/artificial-intelligence-rankings

VanLehn, K., Lynch, C., Schulze, K., Shapiro, J. A., Shelby, R., Taylor, L., ... & Wintersgill, M. (2005). The Andes physics tutoring system: Lessons learned. International Journal of Artificial Intelligence in Education, 15(3), 147-204.

Wayman, J. C., Wilkerson, S. B., Cho, V., Mandinach, E. B., & Supovitz, J. A. (2016). *Guide to using the Teacher Data Use Survey* (REL 2017–166). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Appalachia. Retrieved from http://ies.ed.gov/ncee/edlabs

WWC (What Works Clearinghouse). (2019). WWC Review of This Study: Showing WWC "Without Reservation". https://ies.ed.gov/ncee/wwc/Study/86375

WWC (What Works Clearinghouse). (2022). *What Works Clearinghouse procedures and standards handbook, version 5.0.* U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (NCEE). Retrieved from: https://ies.ed.gov/ncee/wwc/Handbooks