Ph.D. Thesis

Personalized Goal Support to Enhance Technology-Supported Learning

Conrad Borchers
Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213, USA
cborcher@cs.cmu.edu
CMU-HCII-26-104

June 22, 2026

Thesis Committee:

Vincent Aleven, Carnegie Mellon University, Co-Chair
Kenneth R. Koedinger, Carnegie Mellon University, Co-Chair
Geoff Kaufman, Carnegie Mellon University
Roger Azevedo, University of Central Florida
Emma Brunskill, Stanford University

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

This work was supported by the Learning Engineering Virtual Institute. The opinions, findings, and conclusions expressed herein are those of the authors and do not necessarily reflect the views of the Institute. This research was funded by the Institute of Education Sciences (IES) of the U.S. Department of Education (Award No. R305A220386). C.B. was additionally supported by a Siebel Scholarship.

Abstract

Sustaining independent student practice in classrooms using adaptive learning software remains a central challenge in realizing the potential of AI-enhanced learning environments. Goal setting, that is, defining clear, quantifiable objectives, can increase student effort. However, goal setting is rarely integrated with the data-driven feedback and orchestration capabilities of modern learning systems. This thesis investigates how goal setting can be extended into *adaptive goal support*: routines that embed personalized targets, feedback, and adjustment into regular classroom practice to improve student effort and learning.

This dissertation leverages fine-grained platform data to generate weekly practice targets, support brief goal review, and provide adaptive recommendations grounded in prior performance. Empirical studies trace the evolution of this approach from paper-based goal contracts to fully data-driven implementations with automated feedback and classroom-level reward structures independently sustained by teachers.

Across quasi-experimental studies in middle school mathematics, adaptive goal support is robustly associated with increased student practice and skill mastery. In a large-scale Fall 2025 deployment across six middle schools (with four schools including a no-goal baseline), interrupted time series models show substantial positive shifts in weekly minutes practiced and skills completed following implementation. Implementation-intention scaffolding further predicts practice, with effects concentrated among students with lower baseline practice time and attenuated at higher levels of baseline practice time. Baseline intrinsic motivation for mathematics shows a modest positive association with intervention benefit, though motivational measures explain limited variance overall. Gains are also modestly associated with increases in math self-efficacy. Qualitative analysis of students' implementation-intention responses further indicated that students most often aimed to translate goals into environmental regulation strategies involving managing distractions, social regulation, and help-seeking. Finally, students' selected goals closely track system recommendations, explaining 61% of the variance in selected effort goals and 75% of the variance in selected progress goals.

Overall, the findings demonstrate that analytics-informed adaptive goal support is a scalable and effective complement to adaptive middle school mathematics practice software. More broadly, this dissertation argues that educational AI systems should adapt to the behavioral and organizational processes that shape whether sustained practice occurs. The findings further motivate a three-component theoretical model of effort-aware adaptivity in classroom learning environments. Adaptive goal support appears to improve outcomes by (1) increasing the amount of time students are willing to spend practicing, (2) increasing the productivity and focus of engagement during already allocated practice time, and (3) balancing these benefits against the opportunity costs introduced by scaffolding routines themselves within constrained instruction time budgets. Together, the studies position effort regulation as an important and underdeveloped dimension of educational adaptivity and point toward future educational AI systems that personalize support for student persistence.

Keywords: Educational Technology, Learning at Scale, Intelligent Tutoring Systems, Hybrid Tutoring, Human-Computer Interaction, Learning Analytics, Educational Data Mining, Goal Setting, Effort Regulation, Motivational Scaffolding.

Contents

Chapter 1

Introduction

Active learning environments, such as intelligent tutoring systems and teachable agents, hold promise for improving educational outcomes by engaging students in sustained, effortful problem-solving activities (Koedinger, Carvalho, Liu, & McLaughlin, 2023). However, these benefits are highly dependent on students' ongoing motivation and commitment to practice. Although adaptive instructional systems effectively personalize learning experiences based on individual needs, their overall impact is constrained when students fail to invest sufficient time and effort (R. S. Baker, 2007; Koedinger et al., 2023).

Practice goal setting, the process of establishing clear, quantifiable objectives related to effort and performance, often linked to contingent rewards, has been shown to be an effective method for increasing student effort during learning (Alwahbi, 2020). Traditionally, such interventions have relied on paper-based contracts involving parents or teachers (Alwahbi, 2020; Peacock, Ervin, Daly, & Merrell, 2009). However, these approaches face limitations in terms of scalability and integration with digital learning environments (Peng, Borchers, & Aleven, 2024). **Goal support** refers to the set of interventions that help learners establish, monitor, adjust, and attain effort and progress goals. In this dissertation, goal support includes explicit goal setting, progress tracking, adaptive feedback, personalized goal recommendations, and supporting human accountability structures.

Thesis statement: *This dissertation argues that adaptive goal support, comprising explicit goals, progress tracking, feedback, goal recommendations, and supporting human infrastructure, can increase student effort and learning in technology-supported environments by making effort visible, actionable, and adaptive. Through the integration of learner data, personalized feedback, and scalable support structures, intelligent learning environments can extend adaptivity beyond mastery and task performance to include students' progress toward effort goals.*

There are compelling reasons to conjecture that embedding practice goal support within digital learning environments introduces unique affordances. First, these environments provide objective measures of student activity, allowing for continuous, automated feedback on effort and self-regulation—features that have proven critical in prior research (Roll, Wiese, Long, Aleven, & Koedinger, 2014). Second, when well-designed, adaptive learning systems can improve outcomes across student groups and thereby help close opportunity gaps in learning. This means that per unit of additional active practice learners engage in through goal setting, they may benefit more from technology-supported instruction than from traditional instruction (Koedinger et al., 2023), and structured goal support may help sustain effort across student groups (Alwahbi, 2020).

Third, such systems offer scalability, allowing human resources—such as teacher attention—to be distributed more efficiently, which is known to improve or at least modulate in-tutor learning (Jin, Borchers, Fancsali, & Aleven, 2025; Karumbaiah, Borchers, Falhs, et al., 2023; Karumbaiah, Borchers, Shou, et al., 2023).

Aligned with the principle of human-AI complementarity in educational technology design (Holstein, Aleven, & Rummel, 2020), the interventions examined in this dissertation extend beyond traditional technology support for self-regulated learning (SRL) by focusing on more foundational self-regulation processes, particularly effort regulation. In contrast, most AI-driven learning technologies designed to support SRL have concentrated narrowly on cognitive and metacognitive strategies during task execution (Azevedo et al., 2022; Li et al., 2025). However, foundational SRL processes, such as managing effort and setting or calibrating goals, remain underexplored within intelligent learning environments, despite their critical role in sustaining student engagement and achievement (Locke & Latham, 2019). Notably, these processes often occur outside the immediate context of tutoring systems and learning sessions, taking place before learning begins and after it concludes (Gurung et al., 2025). The theoretical contribution of this work is to conceptualize and investigate how learning technologies can support these peripheral yet essential phases of self-regulation. By leveraging adaptive, data-driven insights, this dissertation develops interventions that guide students in managing their effort and goals beyond the confines of task-level support, thus expanding the scope and impact of intelligent learning environments and their adaptivity.

1.1 Research Objectives

This dissertation embedded adaptive goal-setting support into active learning systems. By leveraging learner performance data, we facilitated feedback delivery, personalized goal recommendations, and accountability structures that have traditionally relied on human facilitation (Borchers, Houk, Aleven, & Koedinger, 2025; Borchers, Peng, et al., 2025; Peng et al., 2024). Building on prior work in adaptive behavioral feedback (Adams et al., 2017) and the calibration of self-regulated effort through feedback (Hadwin & Webster, 2013), we examined how AI-driven goal-setting interventions promote sustained student engagement and improved learning outcomes, culminating in scaled deployment through a human-AI hybrid tutoring program and dashboard, including targeted support for students not meeting their goals. The central objective of this thesis was to explore the design, implementation, and evaluation of intelligent practice goal setting support in active learning environments. Specifically, this dissertation addressed the following aims:

- Evaluate the impact of practice goal-setting support on student effort and skill mastery in real-world educational settings using educational technologies based on active problem solving practice.
- Analyze how adaptive feedback loops influence student effort calibration, goal adjustment, achievement, and learning.
- Develop scalable means that automate feedback and provide personalized goal recommendations.

1.2 Contributions

This research contributes to both theoretical and practical advancements in AI-supported active learning. The contributions include:

- Frameworks for integrating goal-setting support into AI-driven learning environments, bridging research on goal attainment, self-regulation, and adaptive learning technologies.
- Empirical evidence on the effectiveness of adaptive, data-driven goal recommendations in improving student engagement and learning outcomes.
- Scalable methods for practice goal setting that reduce teacher workload while maintaining student autonomy and personalized learning pathways.
- Insights into the role of adaptive feedback in supporting foundational self-regulation activities necessary for effective goal pursuit in AI-enhanced learning systems.

1.3 Structure of the Thesis

The remainder of this thesis is organized as follows:

- Chapter 2 provides a comprehensive review of related work in goal setting, self-regulated learning, and active learning systems.
- Chapter 3 summarizes early design research and a pilot study with parents, traditionally stakeholders of goal contracts, paving the way for interventions that followed.
- Chapter 4 presents an empirical evaluation of goal setting with rewards in personalized learning through quasi-experimental methods.
- Chapter 5 focuses on differential outcomes of the intervention, including those predicated on adaptive goal setting.
- Chapter 6 describes the culminating study implementing adaptive goal support at scale with minimal researcher facilitation and analysis of intervention mechanisms based on student interaction data with goal scaffolds and assessment of motivational variables such as intrinsic motivation, goal orientation, and self-efficacy.
- Chapter 7 describes a general discussion of the dissertation's findings, their contributions, and broader theoretical and practical significance.

Chapter 2

Background and Related Work

This chapter was adapted from my published doctoral consortium article:

> Conrad Borchers, Kenneth R. Koedinger, and Vincent Aleven. 2025. Intelligent Support for Practice Goal Setting to Enhance Learning. In *Proceedings of the 26th International Conference on Artificial Intelligence in Education (AIED '25)*. Palermo, Italy.

Summary Statement in Relation to the Thesis

This chapter reviews existing research on goal setting, self-regulated learning (SRL), and their integration into AI-driven educational technologies. It identifies gaps in current approaches and motivates the need for intelligent, scalable goal-setting support in active learning environments. The chapter establishes a theory of change in adaptive practice goal setting motivated by empirical findings and theoretical considerations grounded in prior research.

2.1 Principles Guiding Data-Driven Goal Support

The chapter begins by outlining literature-derived principles that could make data-driven goal support effective. Below, we outline each of these principles and relate them to this thesis research. We then derive specific hypotheses and a theory of change building on these principles.

2.1.1 Principle 1: Provide Goal Achievement Feedback

Feedback is a cornerstone of effective learning, benefiting both domain-specific knowledge acquisition and broader self-regulation skills (Roll et al., 2014). Regular, targeted feedback improves students' ability to monitor and adjust their effort toward achieving learning goals. In the context of AI-supported education, adaptive feedback—tailored in response to learner performance—has demonstrated effectiveness across various domains (Adams et al., 2017; Roll et al., 2014). However, most adaptive interventions in educational technology have focused narrowly on providing in-task support, such as hints or scaffolding for problem-solving (Azevedo et al., 2022; Li et al., 2025), while neglecting foundational self-regulation activities like effort regulation and goal calibration

(Gollwitzer, 2012; Locke & Latham, 2019). Similarly, seminal models of adaptivity in technology-enhanced learning systems, such as intelligent tutoring systems, do not consider adaptivity in relation to student effort regulation beyond adaptivity to affect and engagement once learning is already underway (Aleven, McLaughlin, Glenn, & Koedinger, 2016).

What could feedback and adaptivity look like in the context of goal setting where students observe and potentially adjust goals over time? And how effective is that for learning? This question is a central one in this thesis. Recent studies highlight the potential of integrating adaptive feedback with data-driven goal recommendations to enhance students' capacity for self-regulation (Borchers, Peng, et al., 2025; Peng et al., 2024). Such interventions could not only guide students in adjusting their goals based on performance (as described next) but also sustain engagement by fostering reflective practice. Despite their promise, adaptive goal-setting mechanisms remain underexplored in AI-supported learning systems, presenting a critical opportunity for future research and design innovation.

2.1.2 Principle 2: Adapt Goal Recommendations to Past Performance

Goal-setting theory asserts that specific, challenging goals enhance performance by directing attention, mobilizing effort, and fostering persistence (Locke & Latham, 2019). In educational contexts, this translates to practice goals framed around quantifiable objectives, such as problem completion or time allocation, which are proven to improve student engagement and outcomes (Alwahbi, 2020; Kahle & Kelley, 1994). To maximize effectiveness, goal recommendations should be dynamically adapted to students' historical performance averages, based on evidence from behavioral health interventions (Adams et al., 2017). This adaptive approach ensures that goals remain attainable yet sufficiently challenging, thereby optimizing the probability of success and amplifying engagement benefits (Wäschle, Allgaier, Lachner, Fink, & Nückles, 2014).

Traditional implementations of goal-setting, including paper-based contracts involving parents or teachers, leverage extrinsic rewards and social accountability to sustain effort (Alwahbi, 2020). However, these methods face limitations in scalability and responsiveness, particularly in digital learning environments where real-time data can inform more nuanced adaptations. By automating goal-setting processes and embedding them within educational technologies, it becomes feasible to support large student populations without increasing the workload for teachers and caregivers (Peng et al., 2024).

Therefore, a central objective of this thesis is to devise interventions that can leverage learner data to adjust and recommend goals over time to improve longitudinal performance in addition to just providing feedback on performance as such. This could give learners the opportunity to gradually improve their learning achievement. Yet in the context of K-12 education such interventions have not yet been studied, and methods for scaling such support, for example, through dashboards, are also underexplored.

2.1.3 Principle 3: Retain Learner Autonomy for Extrinsic Goals

Self-determination theory distinguishes between intrinsic motivation, driven by inherent interest, and extrinsic motivation, driven by external outcomes (Deci & Ryan, 1985; Ryan & Deci, 2000). Though often seen as oppositional, intrinsic and extrinsic motivators can jointly enhance performance, including in education (Cerasoli, Nicklin, & Ford, 2014). Crucially, extrinsic motivation varies in autonomy—from externally imposed (e.g., avoiding punishment) to self-endorsed (e.g., pursuing meaningful goals) (Priniski, Hecht, & Harackiewicz, 2018; Ryan & Deci, 2000). This thesis defines autonomy as the degree of personal choice in motivation, consistent with SDT (Deci & Ryan, 1985).

Autonomy has been shown to support goal progress by increasing effort, reducing conflict, and fostering readiness for change (Koestner, 2008), with benefits confirmed in adolescents and young adults (Koestner, Otis, Powers, Pelletier, & Gagnon, 2008). Yet, some students may not view math practice as worthwhile, perceiving classroom goals as imposed—beliefs that predict effort and achievement (Greene, DeBacker, Ravindran, & Krows, 1999). Whether student-involved goal setting improves outcomes remains an open question, particularly given the limited research on autonomy in naturalistic educational settings (Legaspi, Xu, Konishi, Wada, & Ishikawa, 2022).

Given the notable empirical evidence to retain autonomy in goal setting with extrinsic rewards (Cerasoli et al., 2014; Patall, Cooper, & Robinson, 2008) as well as my own design research demonstrating that students prefer final control over goals when they are recommended through AI (Borchers, Peng, et al., 2025), I build on these principles to study how effective learner-centered goal achievement in tandem with data-driven recommendations is in authentic K-12 classrooms. This high level of learner oversight over the goal-setting process is also true to the original intervention I built on, where parents or teachers jointly set and negotiate goals with the student (Alwahbi, 2020; Kahle & Kelley, 1994).

Yet, scaling such learner-centric interventions that retain autonomy within digital learning environments requires thoughtful integration. Embedding social accountability structures—such as tutor-mediated goal reviews or digital contracts—into AI-supported systems could sustain student motivation while preserving scalability.

2.2 Theory of Change Overview

In sum, data-driven goal support in AI-enhanced learning, as conceived in this thesis, is founded on the seamless integration of adaptive feedback, performance-based goal adjustments, and learner-mediated autonomy. By addressing these foundational principles, learning technologies could move beyond task-specific scaffolding toward fostering sustained student engagement and achievement. Addressing this gap is the primary contribution of the present thesis.

The presented research investigates how intelligent practice goal-setting support can improve student engagement and learning in active learning environments. Building on the principles outlined above, and grounded in foundational goal-setting and effort regulation theories (Locke & Latham, 2019), the study proposes and tests three central hypotheses that inform both the design and evaluation of the interventions.

The first hypothesis (H1) posits that *actively involving students in goal setting, while preserving their control over final goal selection, enhances effort regulation and engagement during online practice activities.* This claim builds on evidence from autonomy benefits in goal progress toward extrinsic rewards (Koestner, 2008; Koestner et al., 2008) and past practices of goal-setting contracts in non-digital contexts, where learners are setting goals jointly with a teacher or parent (Alwahbi, 2020; Peacock et al., 2009). Does the same hold true when students independently set goals using historical performance data to negotiate practice goals with limited human oversight?

The second hypothesis (H2) focuses on the dynamic aspect of goal pursuit, hypothesizing that *regular feedback and adaptive goal recommendations based on students' historical performance improve the likelihood of goal achievement, thereby sustaining engagement and learning over time.* Feedback plays a dual role: it not only facilitates general domain learning but also strengthens students' self-regulation by improving their ability to monitor and adjust effort toward goal attainment (Roll et al., 2014). Furthermore, aligning goal recommendations with students' historical performance averages draws from past successful behavioral science interventions in adult populations (Adams et al., 2017). These adaptive adjustments are expected to enhance the probability of goal achievement and, consequently, maintain student engagement, given the known motivational benefits of meeting goals, such as reduced procrastination (Wäschle et al., 2014).

The third hypothesis (H3) asserts that *achieving practice goals, especially when reinforced through rewards, promotes further engagement and learning.* Empirical studies have demonstrated that goal-setting contracts with teachers, tutors, or caregivers can significantly boost student effort and learning outcomes (Alwahbi, 2020; Kahle & Kelley, 1994). Such interventions leverage external accountability and contingent rewards to motivate students, yet, again, have seen little evidence when delivered at scale through data, AI, and limited human supervision (as is common in large classrooms or at-scale remote tutoring). Past research predicts that goal achievement boosts motivation longitudinally, while lack of achievement can have the opposite effect (Wäschle et al., 2014). To investigate this hypothesis, we model both the overall intervention effect as well as its relationship to these fine-grain, longitudinal achievement events.

Scalability of past goal interventions remains limited due to the need for continuous adult involvement. The dissertation addresses that limitation by automating goal recommendations and feedback mechanisms while preserving the motivational benefits of accountability provided by teachers or human tutors. By embedding goal-setting support within digital learning environments, data-driven goal setting can promote sustained engagement at scale, complementing human facilitation with scalable technological solutions (Holstein et al., 2020).

Together, these hypotheses form the basis of this dissertation's **theory of change**, illustrated in Figure 2.1. The model links student-controlled goal setting (H1), adaptive goal feedback and adjustments (H2), and the impact of goal achievement (H3) to key outcomes such as practice engagement and skill mastery. As we discuss below, we assume a general link whereby more practice engagement (most commonly measured in time across educational technologies) leads to more learning opportunities to practice cognitive skills and, hence, content mastery. The empirical relationship between opportunities and mastery in educational technologies providing practice opportunities with feedback and as-needed instruction through hints is backed up by large-scale empirical evidence from tutoring system data (Koedinger et al., 2023; Simpson, Norberg, & Fancsali, 2024). While time might not always lead to productive learning opportunities, and the efficiency of time use may differ between students (R. Baker et al., 2008; R. S. Baker, 2007), in general, we expect more time to lead to more practice.

Each hypothesis is systematically examined across the studies presented in this dissertation. Chapter 5 tests **H1** by evaluating the effectiveness of student-led goal setting on engagement and learning outcomes through an experiment in contrast to teacher assigned goals. Chapter 5 also addresses **H2** by comparing static goals to adaptive, data-driven goal recommendations and investigating how goal achievement trajectories influence future success. The general effectiveness of goal setting with contingent rewards (H3) is both analyzed in Chapter 4 as well as through analyses of how rates of goal achievement relate to the general intervention effect (also in Chap-

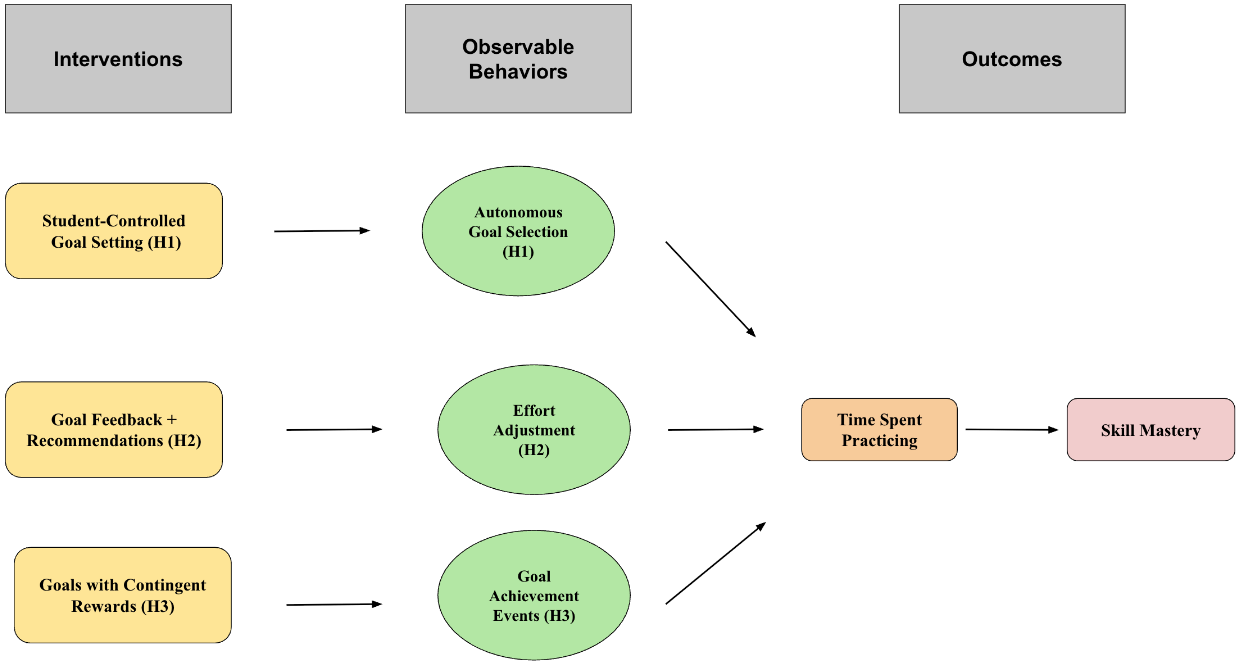


Figure 2.1: Theory of change diagram with three hypotheses: H1 (Active Goal Setting), H2 (Adaptive Goal Adjustments), and H3 (Impact of Goal Achievement). These hypotheses are examined in relation to practice engagement and skill mastery outcomes.

ter 5). Finally, Chapter 6 extends these findings under scaled implementation, testing whether tutor-mediated goal support generalizes across schools and platforms, examining process mechanisms such as implementation intentions, and relating motivational measures to individual-level intervention benefit.

To operationalize these hypotheses, goal-setting support interventions were designed to integrate with learning platform data that track student activity. Leveraging real-time performance data, the interventions dynamically inform personalized goal recommendations while maintaining student autonomy in the final goal selection (see Chapter 5).

Chapter 3

Goal Setting Contracts: A Low-Cost Intervention in Tutored Practice

This chapter was adapted from the published conference article:

Conrad Borchers, Ha Tien Nguyen, Paulo F. Carvalho, Kenneth R. Koedinger, and Vincent Aleven. 2025. Goal Setting Engages More Caregivers in Online Math Homework than Instructional Support. In *European Conference on Technology Enhanced Learning* (pp. 61-75). Cham: Springer Nature Switzerland.

Summary Statement in Relation to the Thesis

In this chapter, we report on a pilot study of a lightweight goal-setting contract implemented in a middle school classroom and integrated into a math tutoring system. The contract was compared to a more involved intervention that had been co-designed with caregivers and required hands-on support via in-system collaboration. Despite its simplicity, the goal-setting contract achieved substantially higher uptake among students, suggesting that low-friction goal support may offer a more scalable alternative to resource-intensive designs. These findings motivate the broader integration of classroom-based goal contracts into intelligent learning systems to support self-driven and personalized student practice, as is explored in the ensuing chapters of this thesis.

3.1 Introduction and Related Work

Caregiver (i.e., a parent or member of a child's caring community) involvement can be critical for academic achievement, contributing motivational and instructional homework support. Caregiver involvement has been linked to higher grades and better test performance (Jeynes, 2007). However, many caregivers face barriers, such as limited time, resources, and content knowledge, hindering homework support (Pelemo, 2022), and exacerbating opportunity gaps.

What motivates caregivers to support their students with homework? Seminal work (Hoover-Dempsey & Sandler, 1997) identifies three key factors: (a) perceptions of responsibility, (b) beliefs that their help is effective, and (c) perceptions of whether their student and school expect their involvement. However, since this framework was established nearly three decades ago, the homework environment has become increasingly digital (Mrazek et al., 2021), potentially altering caregiver roles. Recent research highlights new opportunities for caregiver involvement brought by the increasing availability of educational technologies in K-12 education, such as improving teacher-student communication through instant messaging and supporting student self-regulation during online homework through learning analytics (Gonzalez-DeHass & Willems, 2024; Nguyen, Borchers, Xia, & Aleven, 2024; Peng et al., 2024).

Yet, these studies do not fully explore how novel technologies and interventions can encourage caregiver involvement in ways aligned with Hoover-Dempsey and Sandler's model (Hoover-Dempsey & Sandler, 1997). For instance, technology could enhance caregivers' perceived efficacy by helping them better understand their student's homework, particularly in challenging subjects like mathematics (Nguyen et al., 2024). Alternatively, it could strengthen motivational support as another key caregiver homework role (Peng et al., 2024). While promising, the adoption rates of such interventions remain unclear. This study addresses this gap by comparing two novel interventions designed to support caregiver involvement in digital math homework and their adoption rates, including for known-to-be effective interventions such as goal setting (Alwahbi, 2020; Borchers, Nguyen, Carvalho, Koedinger, & Aleven, 2025a). As a secondary contribution, we study the obstacles to caregiver support in students' tutoring system practice that technology may introduce or solve.

3.1.1 Instructional vs. Motivational Caregiver Tools and Roles

The current study examines the feasibility and affordances of two distinct approaches to caregiver involvement in helping students engage in tutoring system practice. The first approach, a "hands-on" design, encourages caregivers to collaborate with students and bridge knowledge gaps, argued to address curricular challenges caregivers often report (Nguyen et al., 2024; Pelemo, 2022). The second approach, a "hands-off" goal-setting model, leverages caregiver-student goal setting, which has been found effective in past research but has not been studied in tutoring systems (Alwahbi, 2020).

As an instructional, hands-on tool, we study intelligent tutoring systems (ITS) that enhance student learning across K–12 and early college settings (Kulik & Fletcher, 2016). While ITS have shown promise in digital homework (Feng, Huang, & Collins, 2023), their potential to help caregivers support their students in ITS practice remains underexplored. Caregivers often excel at providing emotional and motivational encouragement but often lack confidence in instructional roles (Hasan, Noor, Rahman, & Rahman, 2020). This raises the following key questions: How can an ITS, suitably enhanced, complement caregivers' strengths? Can these systems enhance caregivers' instructional capabilities (Hoover-Dempsey & Sandler, 1997)? Research has highlighted caregivers' need for tools that help bridge content knowledge gaps (Nguyen et al., 2024). Nguyen et al. (Nguyen et al., 2024) found that caregivers favor conversational support systems that offer step-level guidance during homework when integrated with ITS, underscoring the importance of integrating instructional aids into ITS.

In contrast, as a motivational, hands-off tool, we study a goal-setting approach that draws from research on homework behavioral contracts (Peacock et al., 2009). Cooper (Cooper, Heron, & Heward, 2007) defines contingency contracts as written agreements between two parties that specify roles, assigned tasks, designated rewards for the student, and the conditions required to earn rewards. Lab studies indicate that student practice time increases when parents and students jointly establish goals (Kahle & Kelley, 1994). More recent literature reviews further support the effectiveness of contingency contracting in fostering cognitive and non-cognitive skill development in students (Alwahbi, 2020).

While tutoring systems can help address instructional barriers faced by caregivers, goal-setting contracts may be more accessible due to requiring less active engagement. We conducted technology probe studies in two middle schools with distinct student populations. This setting allowed us to examine differences in the feasibility, adoption, and perceived effectiveness of these tools. We ask:

> **RQ1:** To what extent do caregivers engage with (hands-on) intelligent tutoring and (hands-off) goal-setting probes for homework support?
>
> **RQ2:** What are strategies to overcome common obstacles to caregiver engagement in computer-supported homework?

3.2 Methods

This study investigates the feasibility and desirability of novel designs for middle school caregiver homework support in mathematics ITS. Technology probes allow us to (1) collect information about real-world use and users, (2) field-test the novel caregiver tool, and (3) generate design insights into homework support tools in middle school math supported by learning technologies (Hutchinson et al., 2003).

3.2.1 Participants and Recruitment

The study involved two American middle schools: East School, a suburban school in the North-eastern US, and West School, a suburban school in the Pacific Northwest. Both schools were recruited through previously established research partnerships with schools and school districts. Permission to conduct research was obtained from both schools following the approved IRB protocol.

We initiated contact with participating mathematics teachers by obtaining research permissions from school administrators. Collaborative discussions with administrators and teachers helped determine the classes involved in the study and set adequate data collection time frames and mathematics content.

Students of the East School participated in the study across two eighth-grade classes, taught by the same teacher and totalling 44 students. One class, Class 1, consisting of 24 students, was taught a standard eighth-grade math curriculum, while the other class, Class 2, with 20 students, followed an advanced math track based on state test scores. The student population of East School is predominantly Caucasian (90%) and 45% of the student population was classified as low-income based on Free/Reduced Lunch (FRL) statistics available. Less than 5% of the students are English Language Learners (ELL). Overall, 35% of the total school-wide student population met state math standards, representing curricular proficiency expectations for students of that grade level.

At West School, a classroom of 31 students participated. The classroom comprised standard eighth-grade math students, special education students, and English language learners (ELL). One teacher acted as the primary instructor. A second teacher provided additional support, particularly for the special education students. Based on national statistics, the student population at West School is 20% Caucasian and 75% of the students were classified as low-income based on Free/Reduced Lunch (FRL) and official state government resources, and 25% were ELL. Overall, 30% of the total school-wide student population met math standards assessed by the Smarter Balanced Assessments. All school demographics are rounded to 5 percentage points to preserve anonymity.

3.2.2 Tutoring System Probe

3.2.2.1 Recruitment Process

Student-caregiver pairs in both schools were recruited to receive access to the caregiver support probe, which involved caregivers using the tutoring system "hands-on." First, a study strategy was determined with the participating teachers. After securing approval from the school and the teacher, the participating mathematics teacher distributed an email to all caregivers containing information about the study and its objectives. These details were provided as an email attachment containing an official letter about ten days before the study was to start. This letter explained that caregivers had the opportunity to participate in a study trialing a novel homework support tool designed for caregivers, emphasizing that participation was voluntary and detailing the data collection process. Based on teacher discussions, translated study materials were provided to multilingual households at West School so that no household was excluded from participating in the study.

In addition to the letter, caregivers completed an informal, 30-minute onboarding Zoom session, which walked them through the tool's functionalities. One research team member walked participants through the system using a slideshow, as well as live demos of the tool. Participants were given an opportunity to have any questions answered and to test the system live through test accounts during the session.

Across both schools, students and caregivers could sign up for the study and continue to do so until the last day. Caregiver consent and student assent were obtained for tool access. One caregiver per household could participate. Caregivers signed up for the study by filling out an online consent form in Qualtrics, followed by a brief survey on their household's demographics. Finally, all caregivers received a third letter that allowed them to opt out of having their student's practice log data retained for research purposes. No caregiver, however, opted out of that option.

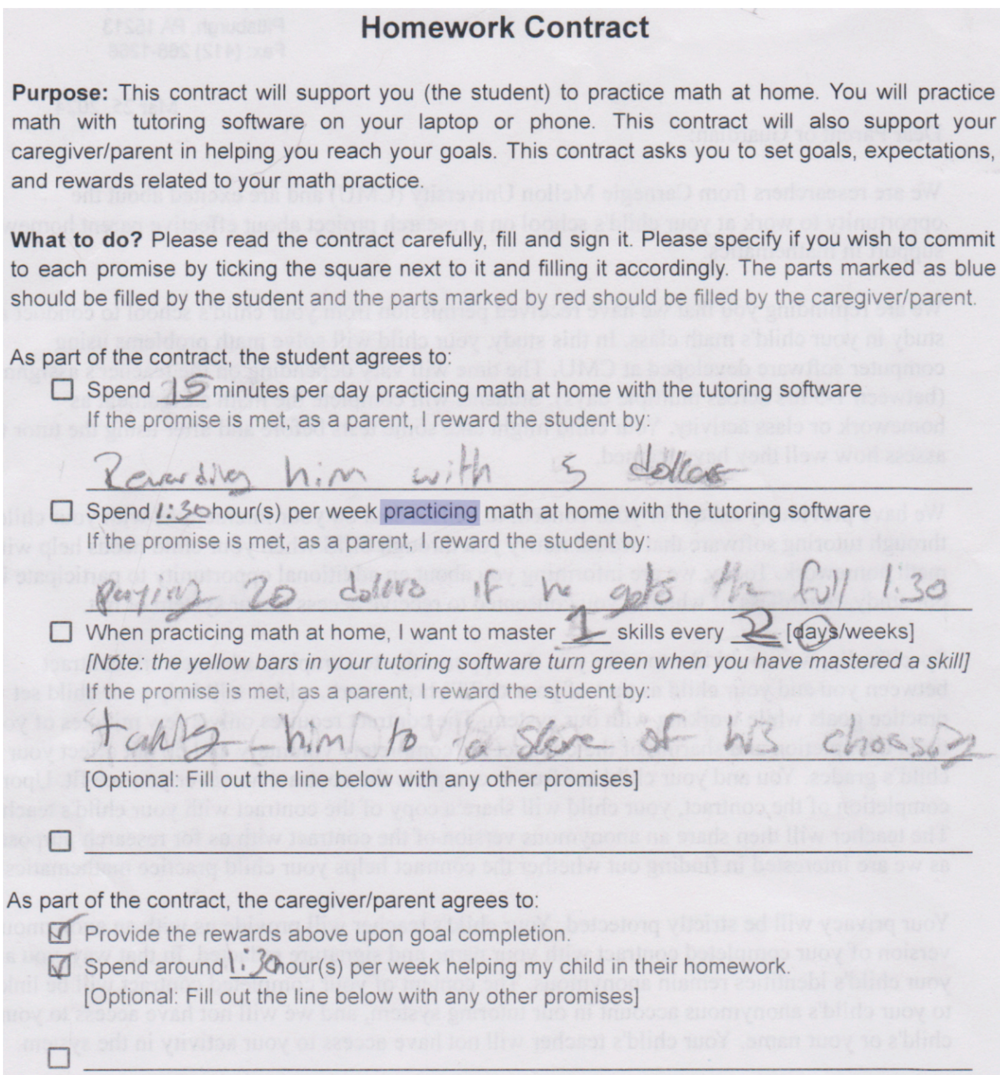
3.2.2.2 Tutoring System Technology Probe Design

Details on the hands-on caregiver system design and its implementation can be found in past research and are not described in detail in this thesis as it is not the focus of its inquiry (Borchers, Nguyen, Carvalho, Koedinger, & Aleven, 2025b; Nguyen et al., 2024; Venugopalan, Yan, Borchers, Lin, & Aleven, 2025).

3.2.3 Homework Contract Probe

A second design probe was introduced at West School only, focusing on caregiver accountability beyond problem-solving support. This probe, and the decision to only trial it at West School, emerged from preliminary thematic analyses of interviews at East School, which revealed caregivers' inability to provide accountability support through the tool. The probe aimed to assess caregivers' perceived usefulness and response rates, aligning with the broader research goal of developing an intervention accessible to diverse caregivers. To that end, this probe did not require a formal sign-up process but was handed out to students similar to a homework assignment requiring caregiver sign-off at home, as detailed below.

The decision to implement a goal-setting contract was also informed by design research indicating that students prefer goal-setting support over other evidence-based, offline homework interventions when using tutoring systems (Peng et al., 2024). Adapted from recommendations in (Peacock et al., 2009), the homework contract was performed on paper and encouraged caregivers and students to negotiate practice goals and, optionally, rewards. Caregivers could also commit to providing tutoring or other forms of support. Fig. 3.1 shows a contract example. To ensure accessibility, the contract was translated into multiple languages for non-English-speaking caregivers.



Homework Contract

Purpose: This contract will support you (the student) to practice math at home. You will practice math with tutoring software on your laptop or phone. This contract will also support your caregiver/parent in helping you reach your goals. This contract asks you to set goals, expectations, and rewards related to your math practice.

What to do? Please read the contract carefully, fill and sign it. Please specify if you wish to commit to each promise by ticking the square next to it and filling it accordingly. The parts marked as blue should be filled by the student and the parts marked by red should be filled by the caregiver/parent.

As part of the contract, the student agrees to:

☐ Spend 15 minutes per day practicing math at home with the tutoring software
If the promise is met, as a parent, I reward the student by:
Rewarding him with 5 dollars

☐ Spend 1:30 hour(s) per week practicing math at home with the tutoring software
If the promise is met, as a parent, I reward the student by:
Paying 20 dollars if he gets the full 1:30

☐ When practicing math at home, I want to master 1 skills every 2 [days/weeks]
[Note: the yellow bars in your tutoring software turn green when you have mastered a skill]
If the promise is met, as a parent, I reward the student by:
taking him to a store of his choosing

[Optional: Fill out the line below with any other promises]

☐

As part of the contract, the caregiver/parent agrees to:

☑ Provide the rewards above upon goal completion.
☑ Spend around 1:30 hour(s) per week helping my child in their homework.
[Optional: Fill out the line below with any other promises]

☐

Figure 3.1: Example redacted homework contract.

On the first day of the study at West School, a researcher introduced and passed out homework contracts for students to complete at home. A participating teacher reminded the students to complete and submit the contracts to the researcher in class daily. Twelve students returned homework contracts in the first week, and five students returned homework contracts in the second week, with 17 out of 31 (54.8%) homework contracts returned by the end of the study.

3.2.4 Procedure

The study took place over four-week periods. The first two weeks involved classroom practice without caregiver involvement, followed by two weeks with the intelligent caregiver support module probe to support caregiver engagement. Students practiced linear equations using the Lynnette ITS (Long, Holstein, & Aleven, 2018) for about 30 minutes daily in class. Lynnette offers immediate feedback and step-level hints. Daily assignments included 12 problems, with unfinished problems completed at home. For students who finished early, Lynnette integrated additional units on graph interpretation. Log data captured student practice, caregiver engagement, and tool interactions, including system usage time, hint requests, problem completion rates, and caregiver notifications. These metrics were used to assess engagement and identify usage patterns, logged in the standard DataShop format for tutoring systems (Koedinger et al., 2010). A researcher was present in the classroom throughout the study to address technical issues, observe student engagement, and document factors influencing tool use. Teachers maintained their regular classroom routines, offering occasional support during student practice.

3.2.5 Data Analysis Methods

3.2.5.1 Engagement Rates (RQ1)

were defined by the extent of caregiver and student interaction with the probes. Caregiver engagement was measured through: **Caregiver Tool Signup and Usage:** Caregivers were considered engaged if they signed up and recorded at least one session in the tutoring system's support module based on log data. **Caregiver Tool Signup but No Usage:** Dropout rates reflected caregivers who signed up but did not use the system, based on the absence of logs. **Homework Contract Return:** Engagement with the homework contract probe was measured by the return of a signed contract, whether signed by a caregiver or another household member acting in a caregiver role.

For an exploratory analysis featuring 31 West School students, we measured engagement through average weekly math practice time based on session start and end timestamps in log data. To compare engagement between students who completed the contract and those who did not, we conducted a t-test, with students as the analysis unit. We replicated the t-test for in-class and out-of-class usage as separate outcomes, as we predicted that out-of-school student engagement might be more sensitive to caregiver support differences. For robustness, we also report Wilcoxon rank-sum tests as we observed skew in our data.

3.2.5.2 Observation Notes (RQ2)

Analyses related to RQ2 derived strategies to overcome common obstacles to caregiver involvement in digital homework. Specifically, researchers observed students during homework practice sessions in classrooms, noting factors that influenced engagement or lack of participation—such as barriers, challenges, or personal preferences—through brief informal discussions with students (and, at times, with participating teachers). Some student responses reflected caregivers' perspectives on tutoring system support and reasons for non-participation or probe signups. These observations were triangulated with pre- and post-interview data and log data analyses to provide a comprehensive understanding of caregiver involvement challenges.

We used an open-ended thematic approach and consensus-based analysis (Corbin & Strauss, 1990; Hammer & Berland, 2014). Two researchers reviewed class observation notes and separately developed themes related to RQ2. Themes were grouped according to forms of caregiver engagement, involvement styles, and household obstacles. The researchers then discussed the generated themes iteratively until they reached a consensus.

3.3 Results

3.3.1 RQ1 Engagement rates

We analyzed system log data to gain quantitative insight into caregiver signup rates. Signup rates were generally low (12%). The dropout rate, defined as signing up to use the tool but not using it, was 20% in East School and 50% in West School. While caregiver engagement with

Table 3.1: Comparison of Caregiver Signup Rates for the Instructional, Hands-On Tool between East and West Schools.

	East School	**West School**	**Overall**
# Students (Periods)	44 (2)	31 (1)	75 (3)
# Signups (Dropouts)	5 (1)	4 (2)	9 (3)
Signup Rate	11.4%	12.9%	12.0%
Dropout Rate	20.0%	50.0%	33.3%

the caregiver support module was comparatively low at West School, households were relatively receptive to the homework goal-setting contract probe. All households were sent the contract, and 54.8% returned a signed homework contract. Among those, 34.5% agreed to provide help with homework, and 27.6% agreed to combine help with offering rewards. Overall, adoption rates of the hands-on instructional tool (12.0%, N=75) and hands-off, goal-setting tool (54.8%, N=31) were significantly different based on a two-sample test for equality of proportions, $\chi^2(1) = 19.49$, $p < .001$).

3.3.1.1 Engagement with Tutoring System after Goal Contract Completion

An exploratory, associational analysis using a sample of 31 West School students was conducted. Completing the homework contract was marginally associated with increased practice in and out of school settings. System log data indicated that students who returned the homework contract practiced more than twice as much per week with the computer-based tutoring system (M = 67.65 mins, SD = 62.83 mins) compared to those who did not (M = 34.33 mins, SD = 28.07 mins), $t(29)$ = 2.03, $p = .052$. Due to apparent skew in the data, we confirmed that this difference was also marginally significant using a Wilcoxon rank-sum test, U = 162.00, $p = .054$. The difference was even more pronounced when filtering data outside of regular classroom hours, where students who returned a contract practiced about three times as much (M = 34.32 mins, SD = 95.14 mins) compared to those who did not (M = 11.54 mins, SD = 57.46 mins), $t(26)$ = 1.91, $p = .067$. A similar trend was observed for in-class practice (M = 13.76 mins, SD = 4.81 mins for contract returners vs. M = 10.25 mins, SD = 6.45 mins for non-returners), $t(29)$ = 1.62, $p = .117$. Both differences were similarly marginal when tested using a Wilcoxon rank-sum test (all $p > .054$).

3.3.2 RQ2: Obstacles to Caregiver Involvement

3.3.2.1 Student Perspectives on Non-Participation

Based on informal interviews with students during classroom practice at West School (Section 3.2.5.2), several students (25.8%) mentioned language barriers as a significant obstacle; five students (16.1%) reported that their non-English-speaking caregivers could not understand the study requirements or interact with the tool. Caregiver availability and resources also played a role, as three students (9.7%) cited busy caregivers who "worked both day and night shifts." Notably, three West School students independently came forward to say that their caregiver was not available to sign contracts; two of them then asked if another family member could sign the contract and one later returned a contract signed by another family member.

At East School, students might have declined to participate because of the prospect of getting notified about their struggling with a skill and asking their caregiver for help (which the participating teacher communicated to students before the study). This mechanism was changed between Studies I and II so that students could always ask for caregiver help through the system, independent of struggle. The East School teacher phrased it the following way:

TeacherEast: *"[Other students] were a little turned off about their parent being notified when they got work wrong/what content they struggle with while doing the parent study. I told them not to fear that!"*

This perspective was also echoed by an East School student, sharing a preference for requesting help from their caregiver at any point in time, instead of that option being open only after the system detected that they were struggling:

Stu3East: *"I guess I was a little I mean, cause once you like you do it, cause you have to do it like a certain amount of times to get the option. So it's kind of So at that point it was a little frustrating, I guess, because you'd have to like, you know, you've already gotten it wrong so many times."* **Stu3East:** *"I wish it would have been like an option with like the hint thing"*

3.3.2.2 Teacher Perspectives on Non-Participation

Some caregivers perceived the signup process and probe as "extra work," with the East School teacher remarking that broad caregiver engagement has been challenging for the district:

TeacherEast: *"I had a few students tell me that their parents didn't want them doing the study due to it being extra work."* and *"Parent involvement at a middle school level is quite difficult. Our district especially has this struggle as it has a wide range of socioeconomic status."*

A similar insight emerged from one of the West School teachers, who highlighted that caregiver engagement for additional education-related activities is challenging for them, in part due to language barriers. Even though translated documents explaining the study procedure and goal contract translations were provided, they were likely insufficient for participation:

Teacher1West: *"[West School] is a unique place...I think doing it earlier would have gotten more people because students check out more after Christmas in 8th grade. So if they aren't engaged, then the parents are much less likely to be engaged. The diversity we had brings issues with trust of unknowns and the translation doesn't always get the message through in the way we want. And it being a core 8 special education and Multi-language learner class usually has more apathy towards school and extra work. Like I said, in my [geometry] class we would have probably had almost 100% involvement just for bonus points. But that isn't the reality of school."*

In summary, language barriers and caregiver availability impacted participation, particularly at West School, where students struggled to engage. At East School, while students were more engaged, some were initially hesitant about notifying caregivers when struggling with content. Teacher feedback from both schools highlighted that caregiver involvement was an ongoing challenge.

3.4 Discussion

Caregiver involvement is crucial for student success, yet factors like time constraints and limited content knowledge often hinder support (Pelemo, 2022). Tutoring systems present opportunities to enhance caregiver homework support, but effective technology designs remain an open challenge (Gonzalez-DeHass & Willems, 2024; Nguyen et al., 2024). We examined two caregiver engagement modes—a real-time intelligent support module and a goal-setting contract—in middle school math classrooms. Findings illustrate how both approaches relate to participation and engagement, offering insights for future designs.

3.4.1 RQ1: How Much Did Caregivers Engage With Each Probe?

3.4.1.1 Lowering Barriers to Active Caregiver Participation

Across schools, the proportion of caregivers who signed up for the tutoring system support was similar (12.9% at West vs. 11.4% at East School), but dropouts were notably higher at West School (50% vs. 20%). Observations and informal student comments suggest that some families found the formal signup and login process burdensome; dropout after enrollment could signify that some students were able to navigate the tutoring system well on their own (or because no opportune moment to help live through the system arose). In contrast, the caregiver homework contract provided less friction, as it was completed on paper and handed out to students, similar to regular homework assignments.

Future research could systematically explore technology access and competing family responsibilities in contexts where caregivers speak multiple languages or hold multiple jobs (Nguyen et al., 2024; Pelemo, 2022). Automated and multimodal translation assistance may be explored to improve inclusivity in future tool design iterations (Yang et al., 2024).

3.4.1.2 Aligning Support Tools with Household Routines

Caregivers often engage in motivational support through encouragement and accountability, yet many lack confidence in tutoring (Hoover-Dempsey & Sandler, 1997; Nguyen et al., 2024). Accordingly, significantly more households engaged with the goal-setting contract (55%) compared to the tutoring system involvement tool (12%). Caregivers were more willing to engage in accountability mechanisms than real-time instructional guidance. Such accountability benefits students based on past research on contingency contracts in education, which have been shown to improve student persistence and practice (Alwahbi, 2020; Peacock et al., 2009). Integrating such motivational support into ITS—potentially through digital contracts or automated progress reminders—could enhance student engagement without overburdening caregivers (Borchers, Houk, et al., 2025; Peng et al., 2024). A limitation of our study is that we cannot gauge how much caregivers engaged with the homework contract beyond completing or signing off on it, which is subject to future research. Our log data analysis showed that students who completed the contract with their caregiver or household member tended to practice more out of class than those who did not, though this difference was only marginally significant ($t(26) = 1.91$, $p = .067$). This associational analysis, based on 31 West School students, could be explained by a selection effect (whereby more motivated students are more likely to complete a practice contract with their caregiver) or by the previously documented effectiveness of goal setting in non-digital homework (Alwahbi, 2020; Peacock et al., 2009). Yet, from response rates in this study alone, it is evident that these contracts are likely to be adopted by a broader range of caregivers and households than involving caregivers directly in tutoring system instruction. Recent evidence offers quasi-experimental support for the effectiveness of practice contracts on learning and engagement in educational technology (Borchers, Houk, et al., 2025).

3.4.2 RQ2: Opportunities to Overcome Involvement Obstacles

3.4.2.1 Recognizing the Broader Household Support Network

Observations at West School revealed a need to design for peer and broad household support beyond the student-caregiver dyad. Three students at West School asked if family members other than their caregivers could complete the homework contract with them. These students requested to involve their older siblings. Bronfenbrenner's Ecological Systems Theory emphasizes that multiple layers of environmental resources (e.g., families and schools) influence a student's development (Bronfenbrenner, 2000). The theory has been successfully used to understand student developmental outcomes (Newland, 2015). Accordingly, tool design should integrate the student's broader family context and other surrounding mesosystems, such as extended school and community, that connect the family microsystem to other student experiences.

3.4.2.2 Integrating Goal Setting Into Tutoring Systems

Our current goal-setting intervention presents a low-friction, low-resource tool for caregiver involvement in digital homework. It is possible, though subject to future research, that better integration of the goal-setting contract with the tutoring system could lead to better student learning. For instance, systems could integrate lightweight progress notifications in relation to goals, which has shown promise in past research (Aurino, Tsinigo, & Wolf, 2022). Further, goal-setting tools for tutoring systems (Peng et al., 2024) could be merged with data-driven insights for caregivers and students on goal completion based on log data. Such analytics have been previously successfully used in teacher-facing dashboards (van Leeuwen, Knoop-van Campen, Molenaar, & Rummel, 2021). Integration with other household rituals and media for providing accountable support (e.g., through calendar applications or reminders) may further increase the ease with which caregivers can exercise accountability.

3.4.3 Limitations

The design recommendations we offered assume that caregivers and students will have access to learning technologies, digital devices running them, and the internet, which is not always the case (Kamenetz, 2019). It is important to note that fundamental needs in caregiver-student relationships are prerequisites for effective relationships between educational institutions and caregivers, such as trust (Schweizer, Niedlich, Adamczyk, & Bormann, 2017). The technology designed here is not intended to displace these needs.

Our study was conducted in two schools with unique institutional contexts, demographics, and resources. Although these sites provided valuable contrasts in socioeconomic and linguistic diversity, the findings may not be generalizable to all K–12 settings, given differences in technology adoption and school-home communication, including rural, charter, or private schools.

Moreover, while we captured student practice time based on log data and qualitative indicators of caregiver involvement, we did not measure actual academic improvements. Our focus was on feasibility and adoption rather than learning gains. Log data and our interviews are also limited in telling the story of how exactly students and caregivers engaged with goal setting and technology at home. Analyses of different caregiver use strategies, which could be studied through at-home observations or richer diary logs and experience sampling, could further enhance the design and effectiveness of the tools studied here.

3.5 Conclusion

We studied two design probes that allow caregivers to support their students in practicing with tutoring systems. We contrasted motivational and instructional aids that aim to overcome time, knowledge, and resource constraints. While caregiver access to a conversational support module in a tutoring system for live interactions saw low signup and usage rates (12% overall), significantly more households chose to engage with a goal-setting probe (55%).

These findings contribute novel evidence that tools enhancing motivational caregiver roles may be more widely adopted than tools focused on enhancing instructional support that caregivers provide to students. Even when caregivers lack content knowledge or have limited availability, goal-setting, and flexible accountability supports may foster meaningful increases in student practice. Observational data further highlight the broader "caring community" that surrounds the student—siblings, other relatives, and trusted adults—who may step in where a traditional parent–child homework partnership is difficult to establish. Future research may focus on improving the integration between goal-setting contracts and tutoring systems, which can track goal progress through log data.

Acknowledgments

This research was funded by the Institute of Education Sciences (IES) of the U.S. Department of Education (Award #R305A220386). We thank Sina Shahmoradi for contributions to early data analyses and literature review.

Chapter 4

Learning Benefits of Goal Setting with Rewards in Hybrid Tutoring

---

This chapter was adapted from the published conference article:

> Conrad Borchers, Alex Houk, Vincent Aleven, and Kenneth R. Koedinger. 2025. Engagement and Learning Benefits of Goal Setting with Rewards in Human-AI Tutoring. In *Proceedings of the 26th International Conference on Artificial Intelligence in Education (AIED '25)*. Palermo, Italy.

Summary Statement in Relation to the Thesis

In this chapter, we present a quasi-experimental study examining the effects of goal-setting contracts with rewards on student engagement and learning in a hybrid human-AI tutoring context. In hybrid tutoring classrooms, learners are supported by technology and human tutors through remote video conferencing software (D. R. Thomas et al., 2024). The intervention was implemented in collaboration with school partners and integrated into weekly classroom routines, requiring minimal teacher involvement. Results showed that students who participated in goal setting spent more time on math practice (about 25%) and demonstrated substantially higher skill mastery (38-54% depending on the mastery metric; see Results). The findings suggest that simple, reward-linked goal-setting can be effectively embedded into hybrid tutoring programs to boost both engagement and learning (H3 of this dissertation's theory of change; see Chapter 2). This work supports the broader thesis aim of developing scalable, intelligent goal-support tools that motivate sustained practice in AI-enhanced learning environments.

4.1 Introduction and Related Work

Goals are *"object or aim of an action, for example, to attain a specific standard of proficiency, usually within a specified time limit"* (Locke & Latham, 2002). Setting goals as a motivational and performance-enhancing strategy has been extensively studied. Seminal work by Locke and Latham (Locke & Latham, 2019) identified key factors contributing to goal achievement: providing goal achievement feedback, fostering goal commitment through rewards, ensuring requisite knowledge and skills to achieve goals, and accounting for situational support, such as teacher or parent involvement (Kahle & Kelley, 1994). *Homework behavioral contracts*, or contingency contracts, are one common method for implementing goal setting in classroom and homework settings (Peacock et al., 2009). Cooper (Cooper et al., 2007) describes contingency contracting as a written agreement between two parties. This contract outlines the tasks assigned to each participant, the rewards designated for the student, and the conditions required to earn rewards. Empirical studies have shown that student practice time increases when parents and students jointly set specific goals (Kahle & Kelley, 1994). More recent literature reviews demonstrate that contingency contracting is generally an effective strategy to support the acquisition of cognitive and non-cognitive skills in students (Alwahbi, 2020).

Integrating goal setting with personalized learning remains underexplored (DeMink-Carthew, Olofson, LeGeros, Netcoh, & Hennessey, 2017; Shogren et al., 2024). Goal-setting contracts are typically performed on physical paper, limiting integration with the feedback and data-driven capabilities of learning systems (Peng et al., 2024). The combination of goal setting and personalized learning may be particularly effective for several reasons. First, it may outperform traditional goal-setting methods due to the continuous performance feedback provided by technology, which helps students calibrate their efforts and manage overconfidence or underconfidence in their ability to achieve goals (Hadwin & Webster, 2013; Muis, Winne, & Ranellucci, 2016). Second, compared to pen-and-paper homework, students may achieve greater learning gains per unit of effort when goal setting is combined with intelligent tutoring, thereby amplifying the engagement benefits traditionally associated with goal setting (Kahle & Kelley, 1994; Koedinger et al., 2023). Third, prior research suggests that regular goal setting, feedback, and evaluation cycles can strengthen students' self-regulated learning (SRL) skills, which may transfer to other learning tasks, enhancing overall learning effectiveness (Chang, Lin, Hajian, & Wang, 2023).

Although past AIED systems have supported learners in setting goals (e.g., (Duffy & Azevedo, 2015)), these studies focused on process goals and improving metacognition *during* learning rather than performance goals typical for goal-setting contracts (Peacock et al., 2009). Additionally, prior research on SRL goal support has usually been limited to short-term instructional interventions (e.g., of a few hours (Duffy & Azevedo, 2015)). In contrast, the present study examines interventions over several weeks, allowing for observing engagement changes and persistence, which we model through linear time trends.

The present study adopts a goal-setting approach that minimizes the need for teacher involvement. We worked with teachers integrating goal setting and rewards into their hybrid tutoring classrooms. We observed whether integrating goal-setting classroom practices into hybrid tutoring is feasible and has tangible student learning benefits. Using an interrupted time series design (McDowall, McCleary, & Bartos, 2019), we estimated the influence of our intervention on practice time. Further, to validate whether students learn more, or merely engage more (Koedinger et al., 2023; Simpson et al., 2024), we monitor estimated skill mastery during the goal-setting. Finally, to confirm if intervention benefits are long-lasting rather than short-term, we model linear time-related engagement trends. We investigate the following research questions:

RQ1: Do students engage in more practice during hybrid tutoring sessions after completing goal-setting contracts compared to before?

RQ2: Does student engagement remain stable over time during goal support in hybrid tutoring?

RQ3: Are changes in engagement during hybrid tutoring with goal setting reflected in skill acquisition?

This study contributes to the growing research body on human-AI-supported learning (Holstein et al., 2020). It examines the immediate impact of goal-setting contracts, support, and rewards on practice time and their effects on skill mastery, providing insights into strategies for enhancing learning in AIED learning settings.

4.2 Methods

4.2.1 Sample and Recruitment

We analyzed data from a hybrid tutoring program running for 12 weeks between October and December 2024. Data from Thanksgiving break (week of November 27th) was excluded, as no school activities occurred. Students (N=110) were from a charter school in the Mid-Atlantic United States, serving grades 6-9. All students enrolled at the school were invited to participate in the tutoring program, and those who provided consent were included in the sample. Students were all male, nearly all African American, and all from low-income backgrounds.

As part of the hybrid tutoring program, online human tutoring is available to students during one class session per week, during which they engage in math practice with the IXL software. A researcher and the school's two math teachers facilitated the goal-setting activities in person. Each classroom session was 43 minutes. Eight classes were served, four working with the 6th and 7th-grade teacher and four with the school's 8th and 9th-grade teacher.

The researcher and one tutor supervisor facilitated tutor participation and training. The tutors were fifteen university student workers who had worked with the student participants since the beginning of the 2024-25 school year as part of the hybrid tutoring program. They provided as-needed mathematics support and goal progress check-ins via the Pencil video conferencing software. Human tutor support was initiated by either the student or the tutor based on student learning needs based on standardized test scores. Tutors varied in how frequently they tutored, from one period per week to five periods per week ($M = 2$ periods per week).

4.2.2 Materials

Learners practiced using the IXL Math software, an adaptive online learning platform designed to support personalized math practice. The effectiveness of the IXL math curriculum has been documented in past research, finding significant improvements in learning relative to comparable non-IXL schools throughout a three-year intervention in grades 3-8 (Bashkov, 2021). IXL is often used by teachers in the United States (Matthews, 2025) and provides a comprehensive curriculum covering a wide range of topics aligned with the Common Core State Standards. The platform uses real-time analytics to adapt problem difficulty based on a student's performance using knowledge tracing and mastery learning with a proprietary algorithm, ensuring a tailored learning experience that matches their skill level.

Students receive immediate feedback on their answers to math problems, typically requiring a single solution step. Brief motivational messages are displayed after correct attempts. The system provides an explanation or a step-by-step problem walkthrough for incorrect responses. IXL may occasionally offer a worked example to support students in mastering concepts. Further, IXL tracks skill progress and generates reports, enabling educators to monitor individual performance and identify areas needing support. Motivational elements, such as achievements and milestones, are integrated to sustain student engagement.

4.2.3 Procedures

During all hybrid tutoring sessions, before and after the goal-setting phase (six weeks each), students attended their regular 43-minute math class and signed into IXL and Pencil, a video conferencing program, on their assigned Chromebooks. Students were greeted by a tutor and sent to individual breakout rooms, where they worked through assigned problems in IXL Math and could ask for assistance from tutors. On Pencil, students would share their screen with tutors so that tutors would view the students' IXL work and provide content support as needed. Tutors also provided motivational support, such as praising the students' efforts. Each class period was attended by approximately the same set of tutors each week. Students were introduced to all tutors supporting their class session at the beginning of the school year in order to build rapport. On a weekly basis, a student may have interacted with any of the tutors in their class session's set of tutors or another tutor if one of the regular tutors was absent. The classroom teacher and a research assistant helped students resolve any technical issues during the lesson. Additionally, we observed that teachers usually shared an IXL leaderboard with students during practice, which continued during the goal-setting phase, which allows students to see how many minutes they practiced in the current week. We observed that some students would occasionally check the leaderboard to gauge if they had reached their goal during practice.

Students completed a math practice goal contract in the first week of goal setting, handed out by a research assistant at the beginning of their math practice period (Fig. 4.1). Students completed the contract independently and chose their goal for minutes practiced and, optionally, the number of skills mastered per week and a third, custom goal. The contract was a design adapted from research recommendations described in Peacock et al. (Peacock et al., 2009). Specifically, it actively engaged the student in setting goals and detailed rewards associated with reaching them every week. Finally, it highlighted the importance and purpose of the goal with detailed instructions at the top and highlighted that remote tutors would help the student achieve their goals and learn math at the bottom.

Students received a biweekly printout goal achievement report, *if* they over- or underachieved by a margin of 33% or more, handed out by a research assistant (e.g., "Average: 30 minutes (150% goal achievement)") and were able to adjust their goal. Further, remote tutors had access to their set goals and occasionally reminded them about their goals. The two goal categories (i.e., achieving a certain number of practice minutes or the number of skills mastered) were determined through discussions with the school's STEAM coordinator and math teachers. If students met one of their goals, they would get a reward the next week, which was distributed by a research assistant at the beginning of the tutoring sessions. Students were free not to choose any goals, and 24 (22%) did choose not to do so or were absent on the first week of goal setting, though the class research assistant offered all students the opportunity to complete a goal contract in the first subsequent week they were in class. All students continued participating in the hybrid tutoring program as usual. To avoid selection bias, students without a goal contract were not excluded from the analysis.

Math Practice Contract

Purpose: This contract will support you (the student) to practice math. You will practice math with IXL on your laptop or tablet. This contract will also support PLUS tutors in helping you reach your goals.

What to do? This contract asks you to set goals related to your math practice. Please read the contract carefully, fill and sign it. Please specify if you wish to commit to each promise by ticking the square next to it and filling it accordingly. The parts marked as blue should be filled by you, the student.

As part of the contract, I, the student, agree to:

- [x] Spend 30 minutes per week practicing math with IXL. (Recommendation: 20 minutes).
  Note: I will practice the relevant content my class is currently learning. I will also make sure to be logged into Pencil when practicing math with IXL if I am in class.

- If the promise above is met, Conrad or your teacher will reward you, the student, by:
  >> Providing a fruit snack at the end of each week.
  >> Providing multiple fruit snacks if the promise is met three weeks in a row (streak).

- [x] Optional: When practicing math in IXL, I want to master 5 skills every 1 week.

[Optional: Fill out the line below with any other goals for IXL]

- [x] I want to try harder

As part of the contract, PLUS tutors will be available to:

- Help you reach your goals in IXL
- Help you practice math in IXL

Figure 4.1: Illustration of the goal-setting process with an example contract.

During the goal-setting phase, hybrid tutoring continued in the same format with the addition of students' individual goal contracts. Students who met their goal would receive a reward at the start of the next tutoring session, distributed by the teacher and a research team member. To integrate goal setting into the hybrid tutoring program effectively, we followed recommendations and requests from the school's STEAM coordinator and math teachers. Accordingly, each week of goal achievement was rewarded with a packet of healthy fruit snacks.

To support student goal achievement, tutors could view whether a student had met their goal in the previous week via a dashboard (Fig. 4.2). The dashboard included recommendations for which students tutors should initiate tutoring interactions based on their inferred support needs (which is not the subject of this study). Tutors were trained to praise students who achieved their goals for their efforts. If a student had not met their goal, tutors were trained to provide advice tailored to why the student felt they did not meet it. For example, if a student expressed inadequacy in math, the tutors were asked to empathize with their struggles and motivate them to persevere. These evidence-based practices are a regular and effective part of hybrid tutoring beyond math support (D. Thomas et al., 2023).

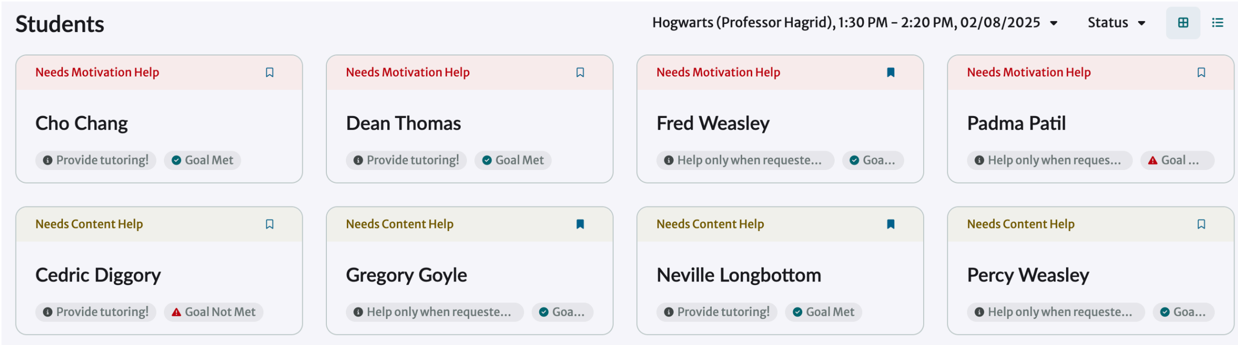


Figure 4.2: Example student data dashboard as seen by remote tutors.

4.2.4 Measures and Data Preprocessing

A researcher compiled a weekly report detailing the number of minutes each student practiced and the number of skills each student mastered, practiced, and was proficient in. In IXL, the threshold for estimated knowledge mastery of skills was 80% for proficient and 100% for mastered, referred to as "SmartScore" in the student- and teacher-facing application. This dataset was used to populate the student data dashboard used by tutors and for data analysis (Section 4.2.5).

We computed the following measures: a **time** variable denoted the number of weeks (to estimate the general trend in the outcome variable over the entire observation period), a **goal setting time** week indicator (a separate week count to model time trends specifically after goal setting had been introduced, capturing changes in practice trends compared to before goal setting), and a binary **goal setting indicator** distinguishing pre- and during-goal setting phases.

4.2.5 Data Analysis

To investigate our three research questions, we employed an interrupted time series design paired with linear mixed-effects modeling to analyze trends in practice time and skill acquisition. The unit of analysis was individual student weeks. Each row presents a student's weekly practice outcome regarding the number of minutes practiced in IXL and the number of practiced and mastered skills.

Given that goal setting was introduced in the middle of the school term, an interrupted time series design naturally emerged in the study data. Interrupted time series analysis is a quasi-experimental approach used to assess the effect of an intervention by analyzing data trends before and after its implementation (McDowall et al., 2019). This method is particularly suited for evaluating interventions where randomized control trials are not feasible or desirable. In our context, a student-level randomized assignment of the goal-setting intervention was undesirable as it could cause students to get frustrated about not getting the opportunity to earn rewards, potentially weakening the control condition and student morale. By comparing learning before and during the goal-setting phase, the design helps identify immediate and sustained changes attributable to goal-setting while accounting for underlying longitudinal trends in student effort and learning. As shown in Fig. 4.3, the interrupted time series design of this study segmented the data into two phases: (1) a pre-goal-setting baseline phase and (2) a during-goal setting phase, each spanning six weeks. By examining practice and learning differences between these phases within students (i.e., we compare each student's engagement and outcomes before and after), we sought to infer the effectiveness of the goal-setting activity regarding student practice behaviors.

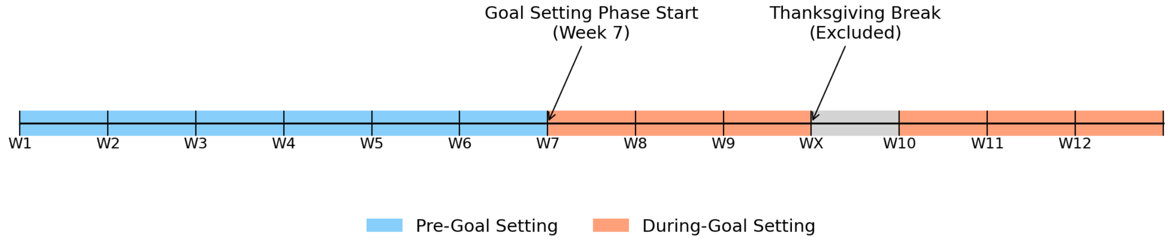


Figure 4.3: Schematic Study Timeline

To evaluate the impact of goal setting in hybrid tutoring, we fit a linear mixed-effects model with scaled weekly practice time as the dependent variable. This method is considered quasi-experimental because the introduction of goal setting was not randomized but implemented naturally as part of the program, allowing for comparison of outcomes before and after the intervention without random assignment (McDowall et al., 2019). Since the practice outcomes (e.g., minutes practiced, skills mastered) are not independent within students, potentially biasing p-values of model coefficients toward significance, we included a random student intercept in the models. This adjustment accounted for each student's baseline practice behavior (Bolker, 2015). The model further included fixed effects for time, time since the introduction of goal setting, and a main effect of the goal setting (see Section 4.2.4). The model of weekly student engagement (Y_{ij}) was specified as follows:

$$Y_{ij} = \beta_0 + \beta_1(\text{Week}_j) + \beta_2(\text{Week Goal}_j) + \beta_3(\text{Goal (Yes/No)}_j) + u_i + \varepsilon_{ij} \tag{4.1}$$

where u_i represents the random intercept for student i, index j denotes week counts, and ε residual error. Linear time trends in engagement are separately modeled via week counts for the overall period (Week variable) and the intervention period (Week goal variable, starting at Week 7) to detect changes in engagement trends related to the intervention. Finally, the binary coefficient of `Goal (Yes/No)`, set to 1 during the intervention, estimates the intervention effect on the outcome Y after adjusting for time trends and individual student differences through u_i. This effect is important because it captures the immediate level shift in student engagement attributable to the onset of goal setting.

To address **RQ1**—whether students engage in more practice during hybrid tutoring sessions with goal-setting contracts than without—we examined the main effect of goal setting (`Goal (Yes/No)`) using the mixed-effects model. This effect, estimated in standard deviation units after standardizing the outcomes, quantifies the average difference in practice time between the pre- and during-goal setting phases. We visualized weekly practice time trends using a centrality plot to complement this analysis. The plot displayed mean practice minutes per week, with reference lines indicating pre- and during-goal setting averages.

For **RQ2**, which examines whether students' practice achievement increases over time following the introduction of goal setting, we interpreted the time trends after goal setting was introduced using the separate week count (`Week_Goal`). This allowed us to evaluate whether goal-setting contracts led to any significant longitudinal increase or decrease in practice time, captured as linear trends.

To study **RQ3**—whether changes in practice time during goal setting are reflected in skill acquisition—we descriptively analyzed aggregated summary statistics of skill mastery and compared them to results for **RQ1**. This analysis leveraged multiple mastery measures (described in the Measures Section 4.2.4) to examine patterns in skill acquisition before and during goal setting.

4.3 Results

4.3.1 Descriptive Differences in Time Spent

Fig. 4.4 illustrates the average number of minutes practiced per week, with dashed lines representing the average minutes practiced before and during goal setting. During goal setting, students exhibited a higher average practice time ($M = 21.9$, $SD = 20.7$) than before ($M = 17.6$, $SD = 25.1$). This difference reflects a 24.43% increase in mean practice time during goal setting. The blue line represents weekly fluctuations in the mean minutes practiced across all participants.

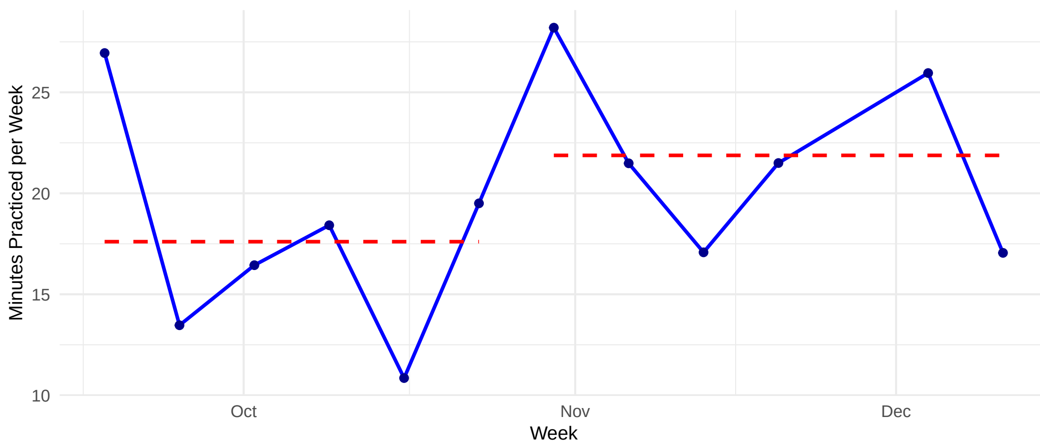


Figure 4.4: Average number of practice minutes across weeks with dashed reference lines showing the average outcomes before and during goal setting.

Table 4.1: Fixed effects estimates from the mixed effects interrupted time series model of the weekly practice time outcome (scaled to standard deviations).

Predictors	**Estimates**	**95% CI**	**p-value**
Intercept (β_0)	0.65	0.30 - 0.99	<.001
Week (β_1)	-0.06	-0.10 - -0.02	.008
Week Goal (β_2)	0.02	-0.03 - 0.08	.404
Goal (Yes/No) (β_3)	0.48	0.28 - 0.69	<.001

4.3.2 Interrupted Time Series Modeling

RQ1 related to whether students engaged in more weekly math practice during hybrid tutoring during the intervention. The fixed effects estimates in Table 4.1 summarize the key findings of the interrupted time series analysis. In line with the descriptive data above, goal setting was associated with a significant increase in the number of minutes practiced per week after adjusting for time trends ($\beta_3 = 0.48$, $p < .001$), corresponding to an average effect of about 0.5 *SD*.

Answering whether student engagement remained stable throughout the goal-setting phase (RQ2), weekly practice time showed a small but significant decline ($\beta_1 = -0.06$, $p = .008$), suggesting a downward trend prior to goal setting. The post-time effect was not significant ($\beta_2 = 0.02$, $p = .404$), indicating that the rate of change during goal setting did not significantly differ from the previous trend. These results suggest that goal-setting led to a significant improvement, though it could not disrupt the existing downward trend in engagement. It also suggests that the goal-setting effect was robust across time and did not recede.

4.3.3 Did Students Also Learn More During the Intervention?

One concern is that students will not learn more when setting goals primarily related to effort (e.g., minutes spent), which could be achieved by practicing with less effort (R. Baker et al., 2008). Therefore, we investigated skill mastery estimated from the IXL learning software before and during goal setting (RQ3). Table 4.2 summarizes the descriptive statistics for skills practiced, proficient, and mastered across two 6-week periods: pre-goal setting and during-goal setting.

Table 4.2: Descriptive statistics of average skills practiced, proficient, and mastered over two 6-week periods: pre-goal and during-goal settings.

Metric	**Pre-Goal Setting**	**During-Goal Setting**
Total Skills Practiced	7.46 (SD: 6.75)	11.49 (SD: 9.01)
Total Skills Proficient	3.42 (SD: 4.08)	4.84 (SD: 5.57)
Total Skills Mastered	2.97 (SD: 3.64)	4.12 (SD: 5.18)

The results show a substantial increase in all skill mastery metrics following the introduction of goal setting into hybrid tutoring. The total skills practiced increased by 53.97%, from an average of 7.46 (SD: 6.75) pre-goal setting to 11.49 (SD: 9.01) during goal setting. Similarly, the total skills proficient increased by 41.52%, from 3.42 (SD: 4.08) to 4.84 (SD: 5.57), while the total skills mastered rose by 38.72%, from 2.97 (SD: 3.64) to 4.12 (SD: 5.18).

4.4 Discussion

This study investigated the impact of goal-setting contracts, goal support, and rewards on students' practice behaviors and skill acquisition in a hybrid tutoring program. Motivated by the challenge of sustaining student engagement in AIED systems, this study aimed to explore whether structured goal-setting mechanisms could address motivational gaps and enhance learning outcomes.

4.4.1 Engagement in Practice Improved Through Goal Setting (RQ1)

The analysis revealed a significant increase in average weekly practice time during the goal-setting period. This finding highlights the potential of goal-setting contracts with rewards to enhance student engagement. The observed increase in practice time suggests that students responded positively to the intervention, aligning with prior research on the motivational benefits of goal setting (Alwahbi, 2020; Kahle & Kelley, 1994). However, unlike past research, our results are notable in large part because of the novel context of hybrid tutoring and personalized learning, which enables data-driven forms of goal support that are more scalable (Adams et al., 2017; Alwahbi, 2020). Specifically, we advance the theory of who can function as an accountable partner in goal-setting interventions: hybrid tutors and, to a lesser extent, the personalized learning technology itself (which reminded students of their goal progress through a data-driven leaderboard, for instance, enabling goal monitoring).

From a practical standpoint, this result implies that hybrid tutoring systems can incorporate goal-setting mechanisms to boost engagement without necessitating constant teacher or parental oversight, unlike past research (Kahle & Kelley, 1994). Schools and educators may consider embedding similar contracts into digital learning platforms to foster sustained student participation to the degree that resources to monitor, support, and reward goal completion are available.

4.4.2 Changes in Practice Time Remained Stable Over Time (RQ2)

Students maintained higher practice levels during the intervention. This boost in engagement did not significantly change over time. This finding suggests that while contracts may effectively sustain engagement, they do not automatically promote continuous improvement (although one past study found such virtuous cycles in goal achievement (Wäschle et al., 2014)). Still, variation in achievement could lead to a separation of performance levels within a class, subject to future research, and may be studied through student-level performance histories. Finally, we observed a significant trend whereby student practice time faded throughout the 12 weeks, possibly due to unobserved factors such as increased testing.

4.4.3 Skill Acquisition Benefits Exceed Engagement Benefits (RQ3)

The goal intervention yielded substantial increases in skill mastery, with gains in skills practiced (54%), proficient (42%), and mastered (38%) exceeding those in practice time (25%). This disproportionate improvement suggests that the intervention may have enhanced not only the quantity but also the quality of practice. Our finding is contrary to the idea that students would simply maximize unproductive practice time to achieve engagement goals with little effort (R. Baker et al., 2008).

Notably, most students chose to set practice time goals, not skill mastery goals, aligning with past research, noting that it is easier for middle school students to express goals in the former, more familiar metric (Peng et al., 2024). Hence, initially, expressing goals in terms of the number of minutes worked may be sufficient for students to achieve mastery learning goals in AIED systems, as considerable design research in explainable AI is required to make mastery-based problem selection intuitive for middle school students (Borchers, Ooge, Peng, & Aleven, 2025).

4.4.4 Limitations and Future Work

Several limitations warrant consideration. First, our analysis cannot disentangle the effects of goal-setting mechanisms, reward structures, and hybrid tutor goal support on extrinsic and intrinsic motivation. Meta-analytic evidence demonstrates that intrinsic motivators (e.g., goal-directed achievement) and extrinsic incentives (e.g., performance-contingent rewards) simultaneously influence performance, rather than necessarily undermining each other, including in education (Cerasoli et al., 2014). Therefore, future work may focus on *isolating* the distinct contributions of extrinsic and intrinsic motivation, for example, via modeling of local effects of intrinsic goal-setting and adjustment events from extrinsic reward events.

Second, the six-week goal-setting period may have been insufficient to capture long-term effects, including potential impact of receding engagement if goal-setting were to be taken away. Third, given the observational nature of the study design, which closely works around existing teacher and program practices, evidence remains quasi-experimental and the intervention partially confounded with the (personalized) learning content. Fourth, the chosen linear mixed-effects model does not account for potential non-linear time trends in practice behavior.

Future research could study differential goal trajectories in and out of classrooms. Past research predicts that students' self-efficacy benefits of meeting goals may propel them into a virtuous cycle of increased achievement (Wäschle et al., 2014). Lastly, a research assistant facilitated the goal-setting procedure in this study. While we believe that human tutors could independently implement goal setting and monitoring in the future, thus requiring neither extra teacher effort nor classroom support, this remains an untested assumption. The minimized teacher involvement observed in this study was feasible due to the hybrid tutoring context, where remote tutors assumed responsibility for monitoring and discussing goal progress with students. These responsibilities may not translate directly to traditional classroom settings without hybrid support, as the same workload would otherwise fall on classroom teachers. Future research should examine whether human tutors can effectively manage goal-setting responsibilities without additional support and explore the scalability of such interventions across contexts.

4.5 Conclusion

Although active learning in AIED systems is widely recognized as effective for improving educational outcomes, research on supporting students to initiate, sustain, and achieve practice efforts is limited. Drawing from goal-setting research on non-digital homework, we studied the integration of goal-setting contracts with contingent rewards into a middle school hybrid tutoring program, where human tutors supported students working with a personalized learning system. The intervention significantly boosted weekly practice time and skill mastery based on interrupted time series modeling over 12 weeks. Students spent about 25% more time on task during the goal-setting phase, offsetting a prior downward trend in engagement. Skill development rose disproportionately: total skills practiced jumped by about 50%, while proficiency and mastery increased by about 40%. Notably, the intervention's immediate impact did not change over time despite an overall semester-wide decline in engagement.

The results of this study have practical implications for designing scalable improvements to learning and engagement in resource-constrained educational environments, such as hybrid tutoring with as-needed human support in AIED learning systems. By leveraging goal-setting contracts and integrating goal support into existing tutoring programs, schools can enhance student engagement and learning outcomes with minimal additional burden on teachers.

Acknowledgments

This work was made possible with the support of the Learning Engineering Virtual Institute. The opinions, findings and conclusions expressed in this material are those of the authors. This research was funded by the Institute of Education Sciences (IES) of the U.S. Department of Education (Award #R305A220386).

Chapter 5

Differential Effects of Adaptive Goal Setting and Achievement

---

This chapter was adapted from the under-review manuscript:

> Conrad Borchers, Kenneth R. Koedinger, and Vincent Aleven. Adaptive Goal Calibration Through Feedback Improves Goal Attainment and Persistence, Especially for Lower-Effort Students. *Manuscript under review.*

Summary Statement in Relation to the Thesis

This chapter presents evidence that adaptive, student-centered goal setting with data-driven calibration can improve goal achievement in hybrid tutoring environments (particularly for students with lower baseline practice time), replicating patterns of results seen in the previous chapter. Drawing on data from two schools in a randomized crossover design, we find that students in the Adaptive Skills condition achieved their weekly goals more frequently than in the static, teacher-assigned condition.

These benefits were most pronounced after mid-intervention goal adjustments based on students' personalized goal-achievement feedback, suggesting that personalized performance feedback fosters better calibration and sustained success. Momentum effects were descriptively stronger in the adaptive condition: students who met their goals in one week were substantially more likely to meet them again in the following week. Furthermore, intervention effects were significantly stronger for students who met their goals more frequently, after adjusting for prior effort. This underscores the dual advantage of helping students adjust their goals using data: practice and calibration benefits.

Both goal-setting conditions significantly increased time-on-task, but only adaptive, self-set goals significantly improved skill proficiency. While students with adaptive goals also demonstrated greater gains in proficiency and time on task than under static goals, these differences did not reach statistical significance. Notably, students with higher baseline practice levels benefited less from either intervention. Still, as much as 84% of all students showed positive practice-related improvement during goal setting. Goal difficulty was uncorrelated with intervention benefits.

Given that students with high prior effort tended to benefit less, differences in motivational profiles may play a critical role. Students with high prior effort may already possess stronger intrinsic motivation, whereas those who benefited most from goal setting may have begun with lower levels of intrinsic motivation. In such cases, extrinsic motivators (e.g., rewards) can serve as an entry point for fostering intrinsic motivation. For highly motivated students, more effective support may involve engaging with their goal hierarchy—clarifying both the reasons for pursuing a goal and the strategies for achieving it.

In relation to this dissertation's theory of change (see Chapter 2), these findings support H1 and H2 regarding the benefits of student-centered, adaptive goals, and H3 by showing that goal achievement is a predictor of intervention benefit. Chapter 6 extended this work by deploying tutor-mediated goal support across multiple schools, measuring intrinsic motivation and goal orientations at scale, and testing implementation intentions as an additional scaffold. Those analyses found that baseline motivational traits explained limited variance in individual gains, whereas process features such as implementation intentions were more strongly associated with outcomes, especially among lower-effort students (see Chapter 7).

5.1 Introduction and Related Work

Active learning, central to AI in education (AIED) systems such as intelligent tutoring systems, teachable agents, and inquiry-based platforms, requires students to engage in problem-solving instead of passive instruction (Koedinger, Kim, Jia, McLaughlin, & Bier, 2015). When instructional support is sufficient and not excessive, active learning yields benefits across all levels of prior knowledge (Koedinger et al., 2023). Its effectiveness nonetheless hinges on students' motivation to persist through effortful practice, a finding consistently replicated across demographic groups and real-world classrooms using cognitive modeling methods (Koedinger et al., 2023; Simpson et al., 2024). How can research-based interventions reduce effort-related disparities so that all students benefit from personalized learning technologies?

Goal setting, especially when paired with extrinsic rewards, has emerged as a promising strategy to promote student engagement in adaptive learning environments (Alwahbi, 2020; Borchers, Houk, et al., 2025). An open question is whether such interventions work equally well for all students. Effort is shaped by multiple factors, including students' expectancy-value beliefs and motivational orientations (Trautwein, Lüdtke, Schnyder, & Niggli, 2006). In particular, the impact of extrinsic motivators depends on perceived autonomy: voluntarily pursuing a self-set goal versus being compelled to pursue an externally set goal (Koestner, 2008; Koestner et al., 2008). Adaptive, data-driven goal-setting supports could personalize learning while scaffolding and maintaining student autonomy. Such adaptive goals have so far been investigated mainly in health behavior interventions, where recommendations are typically based on historical performance averaging (Adams et al., 2017). We examine this approach in the context of computer-based education, which can provide objective data on students' goal attainment and progress. Such data may support students in setting goals, particularly given that many students are not well calibrated when establishing realistic goals (Hadwin & Webster, 2013).

The present study extends a long line of goal-setting research demonstrating that specific, proximal goals improve persistence and performance (Locke & Latham, 2002, 2019). Although K-12 self-regulated learning interventions have shown that strategy instruction and related supports can improve student achievement based on seminal meta-analyses (Dignath & Büttner, 2008; Xu, Zhao, Zhang, Liew, & Kogut, 2023), this literature has concentrated mainly on cognitive and metacognitive regulation. Much less attention has been given to effort regulation as a target of adaptive support. In particular, existing work has seldom leveraged educational technology log data to provide personalized goal recommendations, help students calibrate goals through performance-based feedback, and do so in ways that preserve learner autonomy (Roll et al., 2014). Consequently, adaptive learning systems such as intelligent tutoring systems have only rarely addressed effort regulation as an object of personalization (Aleven et al., 2016). The present study contributes to this literature by testing a data-driven, autonomy-supportive goal-setting approach that uses students' own learning data to support the setting and revision of effort- and mastery-focused goals.

Comparing adaptive goal-setting support (learners guided by system-generated recommendations to adjust self-set goals) with static, teacher-assigned goals has theoretical and practical relevance. Theoretically, it contributes to research on how autonomy-supportive environments enhance motivation and self-regulated learning under extrinsic incentives (Alwahbi, 2020; Borchers, Houk, et al., 2025; Borchers, Peng, et al., 2025; Cerasoli et al., 2014; Patall et al., 2008) and on how learners' agency in setting and calibrating goals influences the learning benefits of goal setting (Alwahbi, 2020; Borchers, Peng, et al., 2025). Practically, it informs the design of interventions that provide personalized scaffolds for goal calibration for students who struggle to set appropriate goals. Tailoring motivational supports to individual needs may reduce effort-based disparities in outcomes and help all students benefit from AI-enhanced learning environments.

We conducted a randomized crossover experiment with middle school students in a tutoring program, comparing teacher-assigned static goals to student-selected goals with data-driven recommendations. Both conditions received weekly extrinsic rewards. In the following, we review the theoretical and empirical foundations for autonomy-supportive goal setting, data-driven calibration, and motivational support in educational technology.

5.1.1 The Role of Student Autonomy Under Extrinsic Rewards

Self-determination theory (SDT) distinguishes between intrinsic motivation, which refers to doing something because it is inherently interesting or enjoyable, and extrinsic motivation, which refers to doing something because it leads to a tangible outcome or reward (Deci & Ryan, 1985; Ryan & Deci, 2000). While these forms of motivation are often viewed as opposing forces, meta-analytic evidence suggests that intrinsic and extrinsic motivators can work in tandem to influence performance, particularly in educational contexts (Cerasoli et al., 2014). Importantly, the benefit of extrinsic rewards varies with the degree of autonomy it allows (Cerasoli et al., 2014; Patall et al., 2008). For example, a student might complete homework to avoid punishment (low autonomy) or choose to do it in pursuit of a valued outcome, such as a desirable career or academic success (high autonomy) (Ryan & Deci, 2000). In this study, we use the term autonomy to refer to this degree of personal choice, consistent with the framework established by SDT (Deci & Ryan, 1985).

Prior research indicates that autonomy supports goal-directed behavior and achievement when extrinsic rewards are present (Cerasoli et al., 2014; Patall et al., 2008). These meta-analytic findings explain why some studies have found no loss of intrinsic motivation through rewards and praise (Bear, Slaughter, Mantz, & Farley-Ripple, 2017), while others have argued the contrary

(Deci, Koestner, & Ryan, 1999). Autonomy contributes to greater goal progress by enabling individuals to exert more effort, experience less internal conflict, and feel more prepared to change their behavior (Koestner, 2008). These benefits have also been confirmed in meta-analytic studies involving young adults and high school students (Koestner et al., 2008). Not all students perceive mathematics practice as personally meaningful; in classroom settings, goals may be viewed as externally imposed instead of self-endorsed. Such perceptions predict effort and achievement in mathematics (Greene et al., 1999). An open empirical question is whether having students set their own goals leads to better learning outcomes than assigning goals. Empirical research on how autonomy affects daily goal pursuit in naturalistic educational settings remains limited (Legaspi et al., 2022).

5.1.2 Data-Driven Goal Support

Prior work has explored open learner models and dashboards that use learner performance data to support self-regulated learning (Borchers, Ooge, et al., 2025; Bull, 2020), but few studies have examined how such tools scaffold students' goal-setting processes in the pre-actional and post-actional phases of learning (Gollwitzer, 2012). One exception is Long and Aleven (Long & Aleven, 2017), who noted learning benefits of students who were able to choose practice problems in a math tutoring system based on their estimated knowledge proficiency. This form of goal setting, however, did not include tangible goal performance standards and rewards.

Indeed, most personalized learning systems support regulation of metacognitive processes, focusing on monitoring progress or delivering help in response to errors (Aleven, Mclaren, Roll, & Koedinger, 2006; Azevedo, Witherspoon, Chauncey, Burkett, & Fike, 2009) or responding to affect (Arroyo et al., 2016). In contrast, there is a lack of data-driven systems that help students with foundational effort regulation processes, such as setting and adjusting effort goals. This leaves a critical gap in the development of students' self-regulation. Recent design work has begun to address this need by developing systems that not only track students' performance but also generate personalized goal recommendations based on log data (Borchers, Peng, et al., 2025; Peng et al., 2024). For instance, adaptive dashboards have been designed to display mastery progress, visualize effort, and provide dynamic goal suggestions that reflect a student's prior achievement level. In a prototyping study with middle school students, Borchers et al. (Borchers, Peng, et al., 2025) found that students responded positively to such recommendations, especially when they retained final control over their goals. Students perceived the guidance as helpful for pacing and calibrating effort, aligning with broader findings in the behavioral sciences showing the effectiveness of adaptive goal-setting in other domains such as health (Adams et al., 2017).

5.1.3 The Present Study

The current study evaluates data-driven goal-setting supports in an authentic, longitudinal classroom context. In a randomized crossover experiment with 184 middle school students in a tutoring program, we tested whether adaptive, student-set goals supported by learning analytics-based recommendations improve students' effort, goal calibration, goal achievement, and skill mastery relative to static, teacher-assigned goals. In doing so, the study examines whether effort regulation can be treated as an additional target of adaptivity in educational technology, alongside the more commonly personalized dimensions of content, feedback, and pacing (Aleven et al., 2016).

By comparing autonomy-supportive, data-informed goal revision with conventional fixed goals, the study aims to clarify both the effectiveness of this approach and the conditions under which it is most beneficial. We address the following research questions:

- **RQ1a**: Does setting adaptive, student-set weekly goals increase students' likelihood of achieving their goals compared to static, teacher-assigned goals?
- **RQ1b**: Does goal achievement in a given week predict higher rates of goal achievement in the following week, and does this momentum effect differ by goal condition?
- **RQ2**: Does setting adaptive, student-set weekly goals improve students' skill proficiency and practice time compared to static, teacher-assigned goals?
- **RQ3**: How are students' pre-intervention practice time and their goal achievement during the intervention related to their individual-level benefits from goal support?

The study contributes three concrete findings: (1) adaptive, student-set goals with one calibration opportunity increase goal attainment over static goals, (2) goal achievement exhibits week-to-week persistencewith descriptively stronger carryover under adaptive goals, (3) benefits are largest for students with lower baseline practice time.

5.2 Methods

5.2.1 Sample and Study Context

We conducted a study in the context of a K-12 tutoring program. The study lasted for 11 weeks between January and April 2025. Students (N=184) were from two charter schools in the Mid-Atlantic United States, serving grades 5-9. All students enrolled at the schools were invited to participate in the tutoring program, and those who provided consent were included in the sample. The study followed an institutionally approved IRB protocol, and school permissions were obtained. In School 1 (N=101), all students were male. The school includes about 80% African American students, with over 95% being economically disadvantaged according to state records. In School 2 (N=83), genders were approximately equally represented, with about 90% being African American and about 80% economically disadvantaged.

As part of the tutoring program, online human tutoring was available to students during one (School 1) or two (School 2) class sessions per week, during which they engaged in math practice with the IXL software. Tutors were trained college students who completed online training on evidence-based tutoring practices (e.g., responding to errors and providing effective praise) before joining the program. Tutors interacted with students via the Pencil video-conferencing platform, typically supporting about five students each per session, and provided content-focused and socio-emotional support as needed, responding to student requests and checking in throughout the class period. A researcher and the school's two math teachers (three total) facilitated the goal-setting activities in person. Each classroom session was about 43 minutes, though occasionally varied based on the instructor's needs. Twelve classes were served. At School 1, this included two sixth-grade classes and two fifth-grade classes. At School 2, this included two 6th-grade and two 7th-grade classes, as well as three 8th and one 9th-grade classes.

All learners used IXL Math, an adaptive online platform for personalized math practice aligned with Common Core standards. Prior research shows that IXL use leads to significant learning gains compared to non-IXL schools over a three-year period in grades 3–8 (Bashkov, 2021). Widely used in U.S. classrooms (Matthews, 2025), IXL adapts problem difficulty through real-time analytics and mastery-based algorithms. Students receive immediate feedback and motivational messages after correct answers, with explanations or walkthroughs provided for errors. The platform also offers occasional worked examples and tracks skill progress, enabling teachers to monitor performance and support learning. Students can select from a set of teacher-recommended skills to work on during classroom time and practice a skill until they demonstrate mastery. Students can stop at any time, switch to another skill, or continue practicing until mastery is reached. Middle school IXL skill examples include adding and subtracting integers, solving one-step equations, and finding unit rates.

5.2.2 Experimental Design

At the outset, all students completed a three-week baseline period during which no experimental manipulation occurred. This baseline period served as a quasi-experimental contrast to students' achievement levels during the two goal-setting interventions. It lacked goals, rewards, and structured monitoring present during intervention weeks. Students were then randomly assigned to one of two experimental groups. In the first four weeks of the intervention, Group 1 pursued *adaptive, student-set goals* supported by algorithmically generated recommendations, while Group 2 pursued *static, teacher-assigned goals*. In both conditions, students received weekly extrinsic rewards (i.e., fruit snacks; in discussion with participating teachers and school administrators) for meeting their goals, with reward contingency held constant across conditions. In the static goal condition, students were assigned a sheet with a teacher-assigned goal at the beginning of the four-week period. In the adaptive goal condition, students set a weekly goal at the beginning of the four-week period. They then received a single goal feedback and adjustment opportunity after two weeks of practice, allowing us to examine the benefits of mid-intervention calibration and its interaction with goal-setting autonomy. A condition crossover occurred after Week 4. Group 1 transitioned to static goals, and Group 2 began receiving adaptive goals with data-informed recommendations. Students in the static goals condition set a goal regarding the number of minutes they aim to practice math with IXL each week, while the adaptive goals condition set a goal regarding the number of IXL skills they aim to master each week. To minimize carryover confusion in the crossover design, we used distinct goal metrics across conditions (minutes practiced for static goals; skills mastered for adaptive goals). This reduced the likelihood that students would inadvertently pursue the prior phase's target. As a manipulation check, weekly minutes and skills mastered were strongly rank-correlated in our sample ($\rho = .58$), suggesting that the two metrics capture strongly related aspects of student practice.

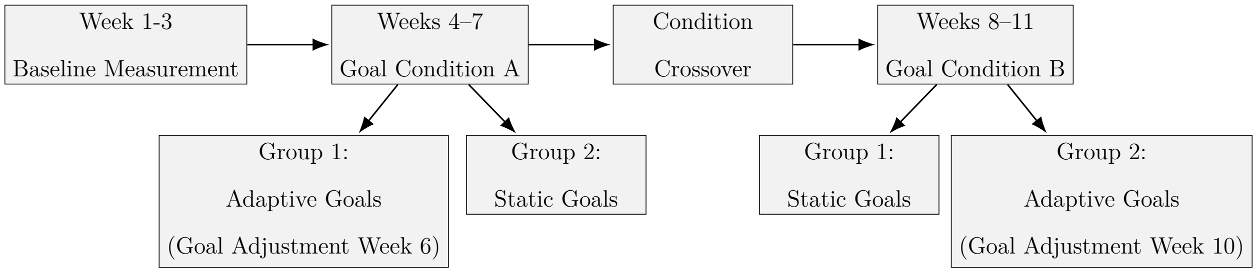


Figure 5.1: Study design schematic. Students were assigned to either adaptive or static goal-setting conditions, with a crossover after four weeks. Students received midpoint goal calibration in the adaptive goals condition. These midway points divide the four-week interventions into Phase 1 and Phase 2.

5.2.3 Measures

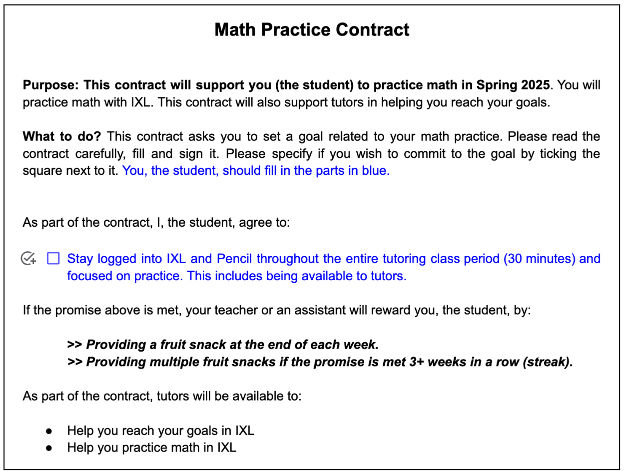
We extracted weekly indicators of student engagement and progress from IXL reports and tutor program logs. *Goal achievement* was a binary measure indicating whether a student met their assigned or self-set weekly goal and therefore earned the weekly reward. *Time-on-task* was operationalized as total minutes practiced in IXL per week. *Skill proficiency* was measured as the count of IXL skills marked “proficient” each week (IXL’s internal proficiency flag). We also computed *baseline effort* as each student’s average weekly minutes practiced during the three-week pre-intervention baseline period for moderation analyses.

5.2.4 Procedures and Intervention

All students participated in a tutoring model integrated into their regular math class (see Sample and Study Context Section). During these sessions, students logged into the IXL adaptive math platform and the Pencil video conferencing platform on their assigned Chromebooks. Upon login, they were greeted by a remote tutor and placed into individual breakout rooms. Tutors observed students’ IXL practice via screen sharing, provided content support as needed, and offered motivational encouragement. Tutoring groups were staffed by consistent tutor teams introduced to students at the start of the academic year to promote familiarity and rapport. Classroom teachers and research assistants supported students on-site by resolving technical issues and helping to distribute goal-setting materials, as well as goal contracts and reports.

During the eight-week intervention phase, students were assigned to goal-setting conditions (adaptive or static; see Figure 5.1). In the first week of each condition, students were handed a paper-based math practice contract, distributed by a research assistant. Adapted from recommendations by Peacock et al. (Peacock et al., 2009), the contract prompted students to commit to weekly practice goals and offered clear reward contingencies. Students could accept a teacher-assigned goal for practice time or a self-set goal for skills mastered per week, depending on their condition (see Figure 5.2). All students who met their weekly goals received a fruit snack the following week at the beginning of the period, with two rewards per week for sustained streaks of goal achievement (of at least three weeks). Rewards were typically distributed by research assistants on site, and occasionally by teachers who had access to a shared weekly goal report.

Static versus Adaptive Goals. Students in the static condition were assigned a teacher-set goal, typically defined as staying logged into IXL and Pencil for the duration of the class period



Math Practice Contract

Purpose: This contract will support you (the student) to practice math in Spring 2025. You will practice math with IXL. This contract will also support tutors in helping you reach your goals.

What to do? This contract asks you to set a goal related to your math practice. Please read the contract carefully, fill and sign it. Please specify if you wish to commit to the goal by ticking the square next to it. You, the student, should fill in the parts in blue.

As part of the contract, I, the student, agree to:

☐ Stay logged into IXL and Pencil throughout the entire tutoring class period (30 minutes) and focused on practice. This includes being available to tutors.

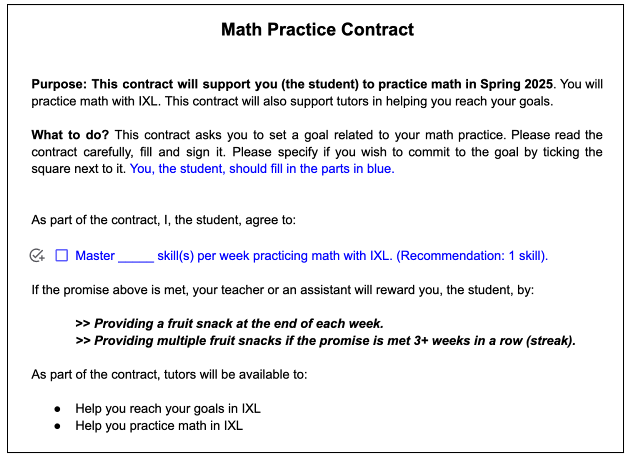
If the promise above is met, your teacher or an assistant will reward you, the student, by:

>> Providing a fruit snack at the end of each week.
>> Providing multiple fruit snacks if the promise is met 3+ weeks in a row (streak).

As part of the contract, tutors will be available to:

- Help you reach your goals in IXL
- Help you practice math in IXL

(a) Static, teacher-assigned goal contract



Math Practice Contract

Purpose: This contract will support you (the student) to practice math in Spring 2025. You will practice math with IXL. This contract will also support tutors in helping you reach your goals.

What to do? This contract asks you to set a goal related to your math practice. Please read the contract carefully, fill and sign it. Please specify if you wish to commit to the goal by ticking the square next to it. You, the student, should fill in the parts in blue.

As part of the contract, I, the student, agree to:

☐ Master _____ skill(s) per week practicing math with IXL. (Recommendation: 1 skill).

If the promise above is met, your teacher or an assistant will reward you, the student, by:

>> Providing a fruit snack at the end of each week.
>> Providing multiple fruit snacks if the promise is met 3+ weeks in a row (streak).

As part of the contract, tutors will be available to:

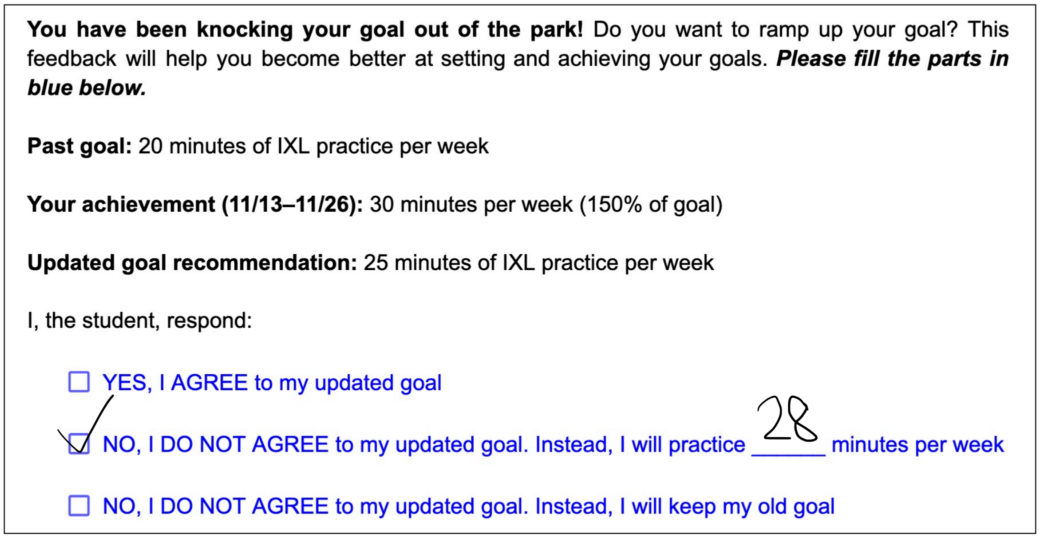
- Help you reach your goals in IXL
- Help you practice math in IXL

(b) Adaptive, student-set goal contract

Figure 5.2: Paper-based math practice contracts used during the intervention. Students either received a teacher-assigned time-on-task goal (static condition) or selected a skill-based mastery goal (adaptive condition), both paired with weekly rewards for goal attainment.

and practicing (30–40 minutes per period, depending on the class). In this condition, all students of the same class received the same goal. In contrast, students in the adaptive condition selected their own skill goals (e.g., mastering one or more grade-level skills per week). Contracts in the adaptive goal condition mentioned an initial recommendation of one skill per week to anchor student choice, based on past program averages. To support goal calibration, these students also received data-informed goal recommendations and performance feedback generated from their past practice data at the midpoint of the intervention segment, after two weeks, via reports that were handed to them (see Figure 5.1). These reports showed their average skills mastered and the percentage of goal completion in previous weeks. Students then independently read the goal feedback and adjustment sheet and adjusted (or kept) their goal based on their preference. The algorithm followed a tiered logic: students who underachieved their goal by a margin of 33% or more received encouragement to reduce their goal or reflect on challenges; students who met their goal were praised and encouraged to sustain it; and students who overachieved by more than 33% on average were invited to increase their goal. Goal recommendations were based on the midway point between students' past goal and their historical performance (e.g., three if students chose two but achieved four skills per week) and rounded to whole numbers. This rather simple algorithm was based on the fact that established recommendation algorithms typically require data from nine measurement points or more (Adams et al., 2017), which our design did not allow for. Yet, it followed a similar logic: take a goal close to students' historical average (and round it up as needed to encourage growth). Further, the explicit contrast between students' initial goal and actual performance was designed to improve their goal calibration (Hadwin & Webster, 2013). Students in the static goal condition did not receive a midway report to isolate the effect of goal achievement feedback.

An illustrative excerpt of the paper-based goal-setting and recommendation process is shown in Figure 5.3. Students were always given the option to accept, reject, or modify the recommended goal, thus maintaining autonomy.



You have been knocking your goal out of the park! Do you want to ramp up your goal? This feedback will help you become better at setting and achieving your goals. ***Please fill the parts in blue below.***

Past goal: 20 minutes of IXL practice per week

Your achievement (11/13–11/26): 30 minutes per week (150% of goal)

Updated goal recommendation: 25 minutes of IXL practice per week

I, the student, respond:

- ☐ YES, I AGREE to my updated goal
- ☑ NO, I DO NOT AGREE to my updated goal. Instead, I will practice 28 minutes per week
- ☐ NO, I DO NOT AGREE to my updated goal. Instead, I will keep my old goal

Figure 5.3: Hypothetical illustration of student interaction with an updated goal recommendation based on historical performance averages. Students retained full control over whether to adopt, reject, or modify the suggestion.

Goal Monitoring and Feedback. During the intervention, tutors used a centralized dashboard (Figure 5.4) to monitor all students' progress toward their weekly goals. In this particular study, using an early prototype of the dashboard, tutors were only able to see whether students completed their goal in the past week or not. Tutors were trained to reference this goal achievement outcome and encourage students as-needed (Figure 5.4). Tutors reinforced goal achievement with verbal praise and brief motivational interactions. For students who failed to meet their goals, tutors provided scaffolded guidance based on students' self-reported reasons, drawing on strategies from prior research on human tutoring (D. Thomas et al., 2023). This support was the same in both experimental conditions. Tutors saw only goal-completion status on a dashboard and were not told students' experimental assignment.

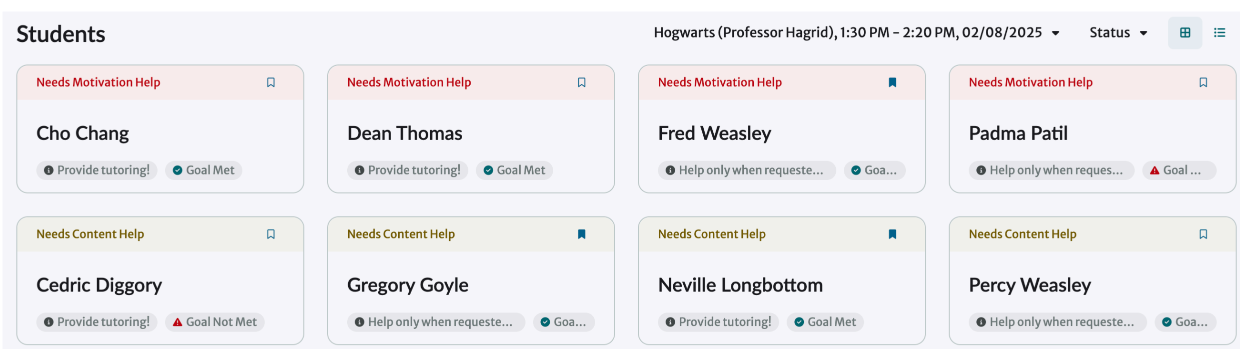


Figure 5.4: Example student data dashboard as seen by remote tutors. The dashboard enabled tutors to monitor weekly goal completion and tailor support accordingly.

5.2.5 Data Analysis Methods

Weekly goal achievement and IXL practice data were logged in a centralized spreadsheet maintained by the tutor program and research team. For each student, we recorded the condition, the number of skills practiced and mastered, the minutes spent in IXL, and whether the weekly goal was achieved. These logs were based on IXL-generated performance reports and aggregated using custom Python scripts.

We used the IXL "skill proficient" flag to count the number of skills completed each week. This lenient threshold (typically marked at 80% proficiency in IXL) was chosen based on observations and teacher input during Week 1 of the intervention. Many students treated this level as the point at which they considered a skill "done." To avoid undermining motivation and ensure alignment with classroom interpretations of progress, we adopted this operationalization for all skill-based outcome variables.

To examine the effects of adaptive versus static goal setting on student behavior and performance, we conducted a series of mixed-effects models aligned with each research question. We used both logistic and linear mixed-effects modeling approaches, accounting for the nested structure of the data (weekly observations nested within students). All models were implemented in `R` using the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015), and model assumptions (e.g., homoscedasticity) were checked using standard diagnostic procedures.

RQ1a: Do adaptive goals improve goal achievement compared to teacher-assigned goals? To estimate whether adaptive, student-set goals increased students' likelihood of achieving weekly goals, we modeled the binary outcome of goal completion using a mixed-effects logistic regression. Fixed effects included goal condition (adaptive vs. static), segment (pre- vs. post-adjustment), and their interaction. A random intercept for each student was included to account for repeated measures. The interaction term allowed us to test whether adaptive goals became more effective over time, particularly after the midpoint goal calibration.

RQ1b: Does goal achievement predict future success (momentum effects)? To test whether goal achievement in one week predicted higher rates of success in the subsequent week, we grouped the data by goal condition and prior-week reward status, then calculated the proportion of students who achieved their goal in the following week. We first tested whether prior-week and next-week goal achievement were independent using a χ^2 test of independence. Rejection of independence indicates association between the two weeks (i.e., a momentum effect). Condition-specific carryover was then compared descriptively using success rates stratified by goal condition and prior-week achievement. For each group, for descriptive plots, we also computed binomial 95% confidence intervals using the Agresti-Coull method.

RQ2: Do adaptive goals improve time spent practicing and skill mastery? To assess whether goal condition predicted future changes in practice time and skill mastery, we employed mixed-effects interrupted time-series models (McDowall et al., 2019). The unit of analysis was the student-week, with repeated measures for each participant across a 3-week baseline period and an 8-week intervention phase.

Model specification. For each outcome, we included: (a) a fixed effect for the overall time trend across the study (*Week Number*), (b) binary indicators for whether the *Adaptive Goals* or *Static Goal* intervention was active in a given week, and (c) post-intervention time variables counting weeks since the start of each goal condition (*Weeks Since Adaptive Goal Started*, *Weeks Since Static Goal Started*).

This structure allowed estimation of the baseline slope before any intervention, the immediate level effect of each goal condition when active, and the change in slope for each goal condition relative to baseline (i.e., interrupted time series trends). All time variables were centered such that Week 1 of the 8-week intervention was coded as $t = 1$, and baseline weeks were coded as $t = -2, -1, 0$. The general interrupted time-series model for outcome Y_{ij} (student i in week j) was:

$$
\begin{aligned}
Y_{ij} = \beta_0 &+ \beta_1(\text{Week Number}_j) \\
&+ \beta_2(\text{Adaptive Goal Active}_j) \\
&+ \beta_3(\text{Static Goal Active}_j) \\
&+ \beta_4(\text{Weeks Since Adaptive Goal Started}_j) \\
&+ \beta_5(\text{Weeks Since Static Goal Started}_j) \\
&+ u_{0i}^{(\text{student})} + u_{0t(i)}^{(\text{teacher})} + u_{0s(i)}^{(\text{school})} + u_{0g(i)}^{(\text{grade})} \\
&+ u_{1i}^{(\text{student})}(\text{Week Number}_j) + \varepsilon_{ij},
\end{aligned}
\tag{5.1}
$$

where β_0 is the baseline intercept, β_1 is the overall time trend, β_2–β_3 are level differences when each goal condition is active (representing the main intervention effect), and β_4–β_5 are slope changes when each goal condition is active. Random intercepts $u_{0i}^{(\text{student})}$, $u_{0t(i)}^{(\text{teacher})}$, $u_{0s(i)}^{(\text{school})}$, and $u_{0g(i)}^{(\text{grade})}$ capture clustering at the student, teacher, school, and grade levels, respectively, where $t(i)$, $s(i)$, and $g(i)$ denote the teacher, school, and grade associated with student i. A random slope $u_{1i}^{(\text{student})}$ for *Week Number* accounts for within-student variation in time trends, following recommendations to assume the highest possible complexity in random effects structure in hierarchical linear models (Barr, Levy, Scheepers, & Tily, 2013). Residual error is represented by ε_{ij}. [1]

This model was (1) estimated for the skill proficiency outcome (counting the number of skills per week), modeled using a mixed-effects Poisson regression suited for count data, and (2) for the weekly time spent practicing using the $\log(1 + x)$ transformation to accommodate zero values and skew arising from the time distribution being non-negative. All models were fit in `lme4` in R (Bates et al., 2015). Parameter estimates were reported with 95% confidence intervals. Fixed-effect significance was assessed using Wald z-tests (generalized models) or t-tests (linear models).

RQ3: Do students' baseline practice time without goals and goal achievement during the intervention correlate with their goal intervention benefits? In RQ3, we investigate whether goal-setting benefits are heterogeneous and predictable from students' prior effort and in-intervention goal completion. We therefore estimated each student's responsiveness to goal-setting (how much their practice time and skill proficiency changed when goal supports were

[1] We did not estimate a separate double-interrupted time-series for the midpoint goal adjustment, as adding an additional intervention indicator and time trend rendered the model non-identifiable for our sample.

active) and then examined whether this responsiveness was associated with baseline effort and goal achievement rates during the intervention. We first derived a student-level estimate of the two intervention effects (adaptive and static goals) from the interrupted time-series models described above. For each outcome (weekly skill proficiency and weekly minutes practiced), we refit the models with added random slopes for the goal condition indicators at the student level, so that the magnitude of the goal-setting main effects (β_2 and β_3 in Equation 1) could vary across students.

From each model, we extracted the student-specific random slopes for the adaptive and static skills goal condition and added them to the corresponding fixed effects to obtain the total goal-setting effect for that student and model. For the skill proficiency model (Poisson), these estimates were expressed as log-rate ratios (i.e., incidence rate ratios). To form a single composite metric of the goal-setting effects per student across outcomes, we combined the skill proficiency and practice time effects (for each model, leading to a total of four effects) by averaging them. The resulting composite score thus reflects each student's overall responsiveness to goal setting (be it adaptive or static) across both outcomes. The rationale behind averaging multiple student-level intervention effects per student is that more measurements are expected to yield a more reliable estimate. With that said, we also separately explored an average of student-level goal intervention effects of adaptive and static goals, to observe any difference by goal type.

To answer RQ3, we then used this composite as the dependent variable in separate analyses, entering prior effort (mean weekly minutes practiced during the baseline period before any goal setting occurred) and goal achievement (% of goals completed during the interventions) as predictors, including their interaction. In an exploratory analysis, we separately fit two additional models predicting the student-level adaptive or static goal effect, respectively. All moderation models were fit using standard multiple linear regression with students as the unit of analysis.

Throughout all analyses, model assumptions were checked through residual diagnostics. These diagnostics included Q–Q plots to assess residual normality, residuals-versus-fitted plots to check homoscedasticity and linearity, scale–location plots to detect heteroscedasticity, and residuals-versus-leverage plots (with Cook's distance) to identify influential observations.

5.3 Results

5.3.1 RQ1a: Do Adaptive, Self-Set Goals Improve Goal Achievement over Static, Teacher-Set Goals?

Across both study phases, students achieved a higher percentage of goals in the adaptive goals condition (29–36%) than in the static goals condition (20–21%). In addition, students in the adaptive goals condition showed an increase in goal achievement from Phase 1 to Phase 2 (29% to 36%), consistent with the provision of a goal adjustment opportunity based on performance data (see Table 5.1). Answering RQ1a, these patterns suggest that the adaptive, student-set goal condition (which included the calibration opportunity) produced higher rates of goal achievement.

Based on the mixed-effects logistic regression (see Data Analysis Methods Section under RQ1), these differences were statistically significant. Students in the *Adaptive Skills* condition had higher odds (as signified by odds ratios, or OR, of greater than 1) of achieving their weekly goals than those in the *Static Minutes* condition (OR $= 1.85$, $p < .001$). In addition, a signif-

Table 5.1: Goal achievement rates (%) by goal-setting condition across study phases. Students were more likely to reach goals when setting them themselves with data feedback (adaptive goals condition) compared to teacher-set goals (static goals condition).

Goal Type	Phase 1	Phase 2
Static Minutes	19.7% [15.6%, 23.8%]	20.8% [16.6%, 25.0%]
Adaptive Skills	29.4% [24.6%, 34.1%]	36.4% [31.4%, 41.4%]

icant interaction between goal condition and phase (OR = 1.36, $p < .001$) indicated that the advantage of adaptive, student-set goals increased from pre- to post-adjustment, aligning with the introduction of data-driven goal calibration. Descriptively, students met self-set goals more often than teacher-assigned goals in Phase 1 (29.4% vs. 19.7%, a 9.7-percentage-point difference) and in Phase 2 (36.4% vs. 20.8%, a 15.6-percentage-point difference).

As a manipulation check, we confirmed that these improvements in goal achievement coincided with students, on average, adjusting their goal upward when overachieving, and adjusting their goal downward when underachieving (see Figure 5.5). Students who were not given a recommendation retained their goal. Therefore, any changes from the first to second phase ("Initial Goal" to "Adjusted Final Goal" in Figure 5.5) among students without a recommendation can be attributed to students who were absent during Phase 1 and therefore first set their goal in Phase 2.

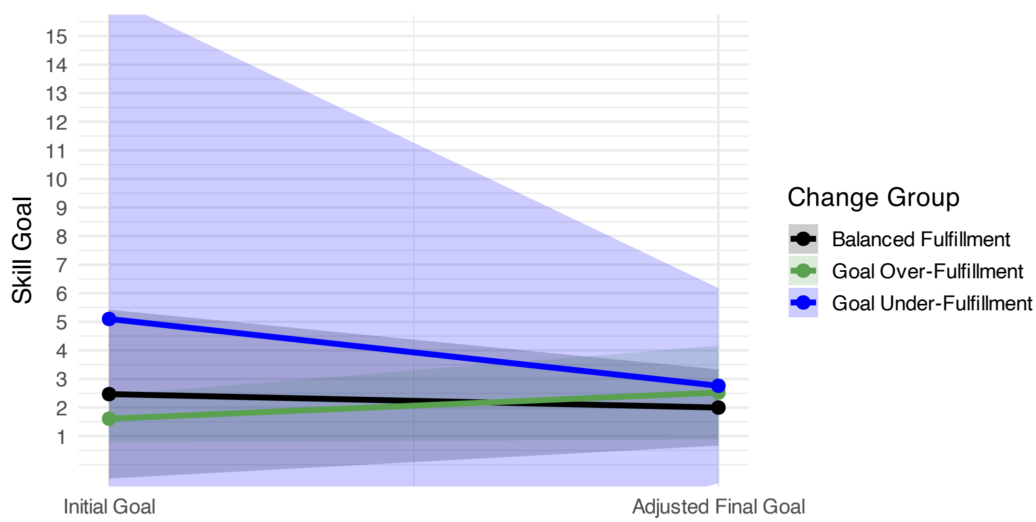


Figure 5.5: Students' self-set goals in the adaptive goals condition before and after data-driven feedback based on the type of recommendation, which was linked to performance. The shaded region represents $\pm$ 1 *SD*.

5.3.2 RQ1b: Do Adaptive, Self-Set Goals Enhance Goal Achievement Momentum?

We examined whether achieving a goal in one week predicted higher rates of success in the subsequent week, testing for a momentum effect in goal achievement. Descriptively, Figure 5.6 shows that students who received a reward in the previous week were more likely to meet their goal in the following week, consistent with a momentum or reinforcement effect. This pattern held for both goal types, but the increase was more pronounced for students in the *skills goal* condition. Students with *Adaptive Skills* goals who earned a reward in the prior week achieved their goal the following week 57% of the time, compared to 26% after not earning a reward (31 percentage points; factor 2.23). For *Static Minutes* goals, the increase was smaller, from 18% to 36% (about 18 points; factor 2.05).

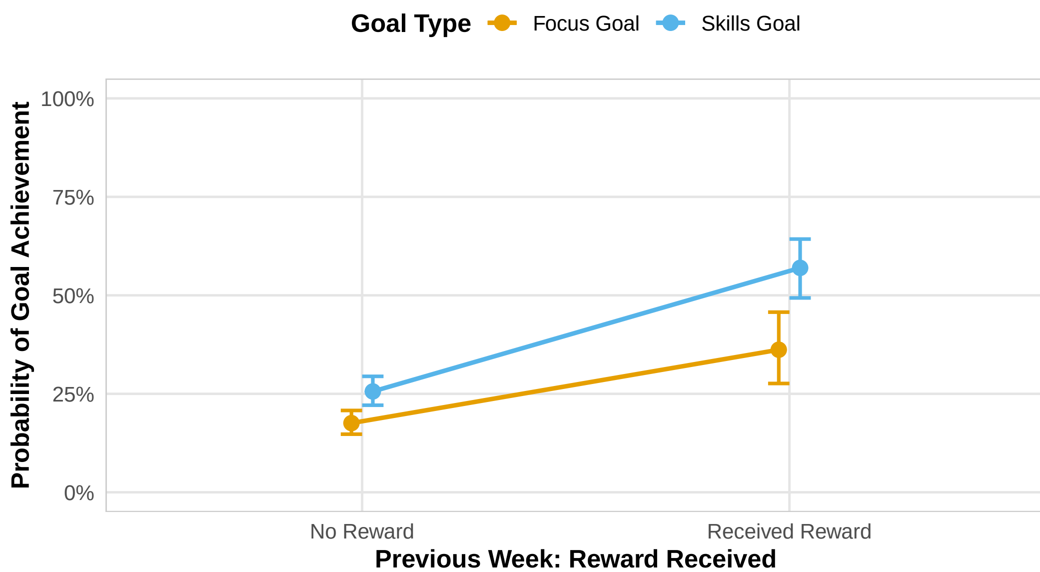


Figure 5.6: Probability of meeting goal again after meeting goal or not in the previous week (momentum) across conditions.

Prior-week and next-week goal achievement were not independent ($\chi^2(1, N = 379) = 7.49$, $p = .006$), indicating that achievement in one week was associated with achievement in the next (a week-to-week momentum effect). Descriptively, momentum was stronger under *Adaptive Skills* goals (57% vs. 26% following prior-week success vs. failure; 31 percentage points) than under *Static Minutes* goals (36% vs. 18%; 18 percentage points), consistent with the pattern in Figure 5.6.

5.3.3 RQ2: Do Adaptive, Self-Set Goals Improve Practice Time and Skill Proficiency?

Reporting descriptive differences first, the *No Goal* baseline condition showed the lowest practice time ($M = 23.6$ min) and proficiency ($M = 1.03$ skills), with an average of 22.8 minutes required per skill. Relative to baseline, the *Static Minutes* condition was associated with a 15.7% increase in practice time and a 34.0% increase in skills gained, while the *Adaptive Skills* condition showed a 21.2% increase in practice time and a 41.7% increase in skills gained. Both goal-setting conditions also showed greater efficiency (fewer minutes per skill). The *Adaptive Skills* condition had the highest proficiency ($M = 1.46$) and the lowest minutes per skill (19.6), slightly outperforming the *Static Minutes* condition ($M = 1.38$, 19.9 min/skill; see Table 5.2). Next, we turn to whether these differences are significant based on interrupted time-series models.

Table 5.2: Descriptive statistics for practice time, proficiency, and efficiency by goal condition. Values are presented as $M \pm SD$.

Goal Condition	Practice (min)	Proficient	Min./Skill
No Goal	23.6 ± 30.1	1.03 ± 2.09	22.8
Static Minutes	27.3 ± 23.5	1.38 ± 2.10	19.9
Adaptive Skills	28.6 ± 24.7	1.46 ± 2.18	19.6

Skills proficient per week. Using the mixed-effects interrupted time-series model specified above, students demonstrated a positive overall week-to-week trend in skills proficient (Incidence Rate Ratio representing changes in skill frequency, IRR $= 1.05$, 95% CI [1.02, 1.08], $p < .001$), corresponding to a 5% weekly increase. Weeks under the self-set *skill* goal condition were associated with a 24% higher proficiency rate relative to non-intervention weeks in terms of skills proficient per week (IRR $= 1.24$, 95% CI [1.03, 1.50], $p = .023$). In contrast, the teacher-set *minutes* goal condition showed no reliable difference in proficiency (IRR $= 1.08$, 95% CI [0.90, 1.31], $p = .416$). Post-start slopes for both goal conditions (i.e., static and adaptive) were not statistically different from zero (adaptive goals: IRR $= 0.96$, $p = .141$; static goals: IRR $= 0.98$, $p = .505$), indicating no changes in trend. Random effects showed that most variance in skill proficiency per week was attributable to differences between students (ICC = 0.758), with smaller contributions from teachers (ICC = 0.053) and schools (ICC = 0.018). Grade-level differences were negligible beyond these factors (ICC ≈ 0).

Minutes practiced per week. In the linear mixed-effects model on $\log(1 + \text{minutes})$, students showed an overall significant week-to-week increase (estimate $= 0.05$ on the log scale, 95% CI [0.02, 0.09], $p = .003$). Weeks under the self-set adaptive *skill* goal condition were associated with a significant increase in minutes practiced (estimate $= 0.52$, 95% CI [0.26, 0.79], $p < .001$). Weeks under the *minutes* goal were also associated with a significant increase (estimate $= 0.37$, 95% CI [0.10, 0.63], $p = .007$). Post-start slopes for skills and minutes goals were not statistically different from zero (adaptive skill goals: -0.07, $p = .143$; static minutes goals: -0.04, $p = .358$). Variance partitioning indicated modest clustering at the student level (ICC $= .109$), with smaller components for school (ICC $= .054$) and teacher (ICC $= .042$). Grade-level variance was too small (i.e., near-zero) to be reliably estimated.

Direct comparison of intervention coefficients. Although adaptive, self-set *skill* goals showed larger point estimates than teacher-set *minutes* goals across models, the difference between a significant and a non-significant coefficient is not, by itself, necessarily statistically significant. We therefore tested the null hypothesis that the two intervention coefficients are equal ($\beta_{\text{skills}} - \beta_{\text{minutes}} = 0$) using planned Wald contrasts with single-step multiplicity adjustment.

In the skills–per–week model (Poisson), the skills-vs-minutes contrast was 0.138 (SE $= 0.110$), $z = 1.257$, $p = .209$; on the incidence-rate scale this corresponds to a ratio of IRRs of $e^{0.138} = 1.15$ with a 95% CI [0.93, 1.42], which trended toward an adaptive goal benefit but yielded no significant difference between the two goal conditions in their effect on proficiency. In the minutes–per–week model (linear), the contrast was -0.19 minutes (SE $= 2.93$), $z = -0.065$, $p = .948$, with a 95% CI [-5.94, 5.56] minutes, likewise showing no significant difference between the intervention effects. Taken together, while adaptive *skill* goals significantly increased proficiency and both goal conditions increased time-on-task compared to students not having goals, we did not find evidence that the adaptive *skill* goal effect was statistically larger than the *minutes* goal effect.

5.3.4 RQ3: Do individual differences in prior effort and achievement explain differences in goal benefits?

Descriptively, Figure 5.7 summarizes the student-level composite goal effect across all outcomes (skills vs. minutes practiced) and goal condition (static vs. adaptive). It displays the distribution of the composite intervention effect, with a dashed line at zero. Most students (84%) exhibit positive goal intervention effects, while 30 of 182 students (16%) fall below zero, indicating a negative composite response to goal setting with rewards. To study intervention differences by prior effort, students were also stratified into low, mid, and high prior-effort groups using standardized pre-intervention effort, with cutoffs set at one-half of a standard deviation below and above the mean. This classification separates meaningfully lower- and higher-effort students while retaining sufficient sample sizes within each group. When breaking out students by prior effort, the distribution shifts systematically: students with low prior effort exhibit the highest share of positive composite effects, followed by those with mid prior effort, while students with high prior effort show the lowest proportion of positive effects (at about 50% benefiting) and a more concentrated mass near zero.

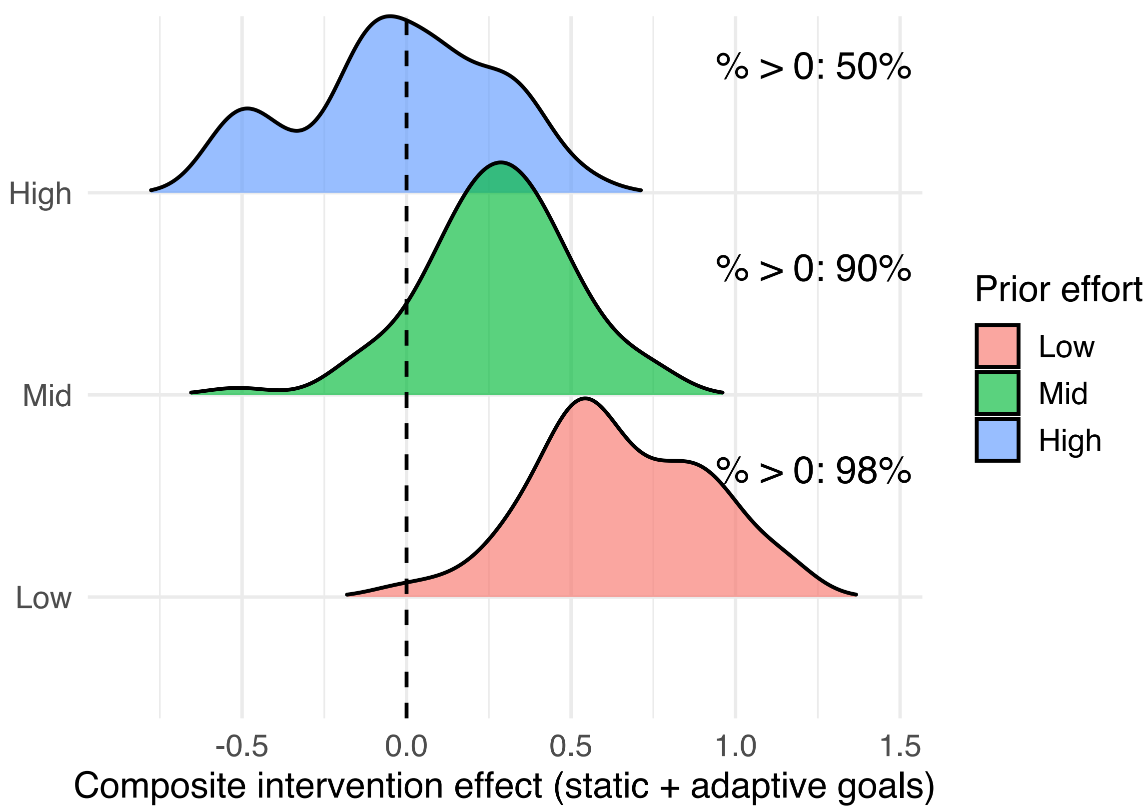


Figure 5.7: Histograms of the *raw* composite intervention effect (dashed line at 0; panel annotations report the share of students with a positive intervention effect composite > 0).

Across all models ($N = 182$ students for which intervention effects were estimable), *prior effort during baseline* was a strong, negative predictor of subsequent goal-setting benefits: students who practiced more minutes before any goal-setting showed smaller estimated intervention gains (combined composite: $b = -0.77$, 95% CI $[-0.87, -0.68]$, $p < .001$; static-goal effect: $b = -0.75$, 95% CI $[-0.85, -0.64]$, $p < .001$; adaptive-goal effect: $b = -0.79$, 95% CI $[-0.89, -0.69]$, $p < .001$).

Goal achievement during the intervention was positively associated with the composite score of student-level intervention benefit, but its strength varied by goal condition. For the combined composite outcome, higher goal achievement predicted larger benefits ($b = 0.14$, 95% CI $[0.04, 0.24]$, $p = .007$). Disaggregating by goal condition, the association was significant for the *static goal* effect ($b = 0.19$, 95% CI $[0.08, 0.29]$, $p < .001$) and positive but not statistically reliable for the *adaptive goal* effect ($b = 0.09$, 95% CI $[-0.01, 0.18]$, $p = .074$).

Across all models, there was no significant interaction between prior effort and goal achievement when predicting the student-level intervention effect (all $p > .416$). Another potential confounder for the intervention effect could be goal difficulty, as highly challenging goals are sometimes known to be very motivating (Locke & Latham, 2002). However, Pearson's product–moment correlation between goal difficulty (expressed in the average, student-level number of skills per week set as their target) and the composite student-level intervention effect was non-significant, $r = -0.06$, 95% CI $[-0.21, 0.09]$, $p = .458$, indicating no evidence that goal difficulty was associated with overall effectiveness of goal setting with rewards. Model fit was moderate to strong: $R^2_{\text{adj}} = .568$ for the combined composite, .527 for the static-goal effect, and .594 for the adaptive-goal effect.

Taken together, students who more consistently *met their weekly goals* derived greater overall benefit from goal setting, while students with *higher baseline practice effort* tended to show smaller incremental intervention benefits.

5.4 Discussion

This study advances research on motivational adaptivity in educational technology by examining when and for whom *autonomy-supportive, data-driven goal calibration* is associated with engagement and learningwithin a bundled classroom intervention. In an 11-week randomized crossover field experiment in a tutoring program, we compared teacher-assigned, static goals to student-set goals paired with a single feedback report that recommended performance-contingent goal adjustments in the middle of the adaptive goal intervention. Weekly rewards were held constant across conditionsin their contingency structure.

Three contributions emerge. First, adaptive, student-set goals yield higher goal attainment than static goals, and this advantage grows after the calibration opportunity. Simple performance feedback improves goal realism and sustains success. Second, goal achievement carries over week to week, with descriptively stronger carryover under adaptive goals. Third, effects are heterogeneous: students with lower baseline practice time benefit most and higher-practice students show smaller incremental gains.

5.4.1 Benefits of Goal Adjustment and Selection for Goal Achievement

We found that self-set goals supported by performance-based recommendations produced higher rates of weekly goal completion than static, teacher-assigned goals. This pattern provides a direct answer to RQ1a. This advantage became more pronounced following the midpoint calibration, accentuating the potential of timely, data-driven feedback to help students align their goals with actual performance. Students tended to adjust their goals downward after underachievement and upward after overachievement, consistent with our recommendations and with meta-analytic evidence on post-success and post-failure goal adjustment (Theobald, Lin, Sakaki, Murayama, & Brod, 2025). In that meta-analysis, self-efficacy emerged as a key mediator of upward adjustment after success, suggesting that assessing self-efficacy in the final study may provide further explanatory insight. While it is worth noting that teacher-assigned goals were challenging and only met by about 20% of students, they were not unrealistic and were aligned with in-class practice targets of roughly 30–40 minutes per session.

The momentum analysis (RQ1b) further revealed that meeting a goal in one week more than doubled the likelihood of meeting it again the following week. This pattern aligns with theories of positive reinforcement, in which goal achievement and self-efficacy can mutually reinforce one another to sustain or improve performance (Wäschle et al., 2014). Students who more frequently met their goals also derived larger benefits from the intervention in terms of their weekly practice minutes and the number of proficient math skills. This correlation suggests that earlier and more frequent opportunities for goal adjustment and selection could strengthen these effects.

Finally, these findings raise important questions about the temporal dynamics of adaptive goal support. While the present results are consistent with the idea that performance-based feedback can help students calibrate goals and sustain engagement, it remains an open question whether the observed benefits would persist, attenuate, or stabilize over longer periods as students become accustomed to the goal-setting routine. Improvements in goal achievement need not imply continual week-over-week gains (unrealistic given session time and ceiling effects) and may reflect convergence toward a stable, maintainable level of effort and performance. From this perspective, future work should examine not only whether more frequent, data-driven calibration opportunities enhance short-term outcomes, but also whether they help maintain stable engagement and autonomy over time once novelty effects diminish. Future research should address these questions by enabling the refinement and longitudinal evaluation of automated, adaptive goal-setting models that balance effectiveness with student autonomy, a known moderator of the relationship between extrinsic rewards and performance (Cerasoli et al., 2014; Patall et al., 2008).

5.4.2 Low-Effort Students Benefited More from Goal Support

Across both conditions, goal setting was associated with increased practice time; however, only adaptive goals produced a statistically significant improvement in skill proficiency relative to the no-goal baseline, and proficiency was the metric on which adaptive students set weekly targets. While the adaptive condition showed descriptively larger gains than the static condition, the difference between interventions was not statistically significant (RQ2). At the same time, moderation analyses revealed that students with higher prior baseline practice time derived significantly smaller benefits from either intervention (RQ3). In contrast, students with lower baseline practice time showed the largest relative gains. Although this pattern was robust across the RQ3 analyses, the strength of the negative association with baseline practice should be interpreted cautiously because baseline levels and estimated intervention gains are not fully independent, potentially introducing regression-to-the-mean effects. Overall, 84% of students demonstrated some improvement in weekly time spent and weekly skill proficiency during the goal support interventions, whereas the remaining 16%, typically already exerting high effort without formal goals, showed no improvement. Notably, answering the other part of RQ3, students with higher rates of goal attainment derived stronger intervention benefits, too.

These findings contribute novel evidence to the literature on adaptive, autonomy-supportive goal setting in three ways. First, on a theoretical level, they highlight the importance of initial baseline practice time as a moderator of intervention effectiveness, offering an empirical basis for tailoring goal-setting supports to student profiles. Conceptually, this pattern is consistent with a diminishing-returns account, common in research on scaffolding (Kalyuga, 2009). When effort regulation is already high, additional goal structure and incentives provide little marginal scaffolding, whereas for low-effort students, they can supply the initial structure needed to initiate and maintain persistence. Second, on a practical level, they extend prior work by showing that data-informed goal-setting activities and recommendations can meaningfully benefit students in authentic classroom contexts. Third, they suggest that the greatest returns from adaptive goal setting with rewards may come from engaging students who might otherwise remain under-involved in technology-enhanced practice due to low intrinsic motivation, providing evidence for hypotheses from prior research framing extrinsic motivators as an entry point for engaging the academically unmotivated (Asher & Harackiewicz, 2024; Hidi & Harackiewicz, 2000). Collectively, these insights highlight the potential of adaptive goal-setting systems to reduce effort-related student differences in learning, thereby enabling more broadly effective access to the benefits of personalized, technology-enhanced instruction.

5.4.3 Intrinsic Motivation as a Potential Moderator of Intervention Differences

Next to a ceiling effect (given that classrooms have limited instructional time), another possible explanation for the finding that high-prior-effort students benefited less from the goal-setting intervention is that they may have entered the study with higher intrinsic motivation (or more internalized reasons for practicing), whereas students who benefited most may have started with lower intrinsic motivation. Under this interpretation, performance-contingent rewards can have different motivational consequences depending on students' starting points. For students already practicing frequently for self-endorsed reasons, tying rewards to goal attainment may shift attention from the inherent value of practice to the external payoff, reducing perceived autonomy and producing a crowding-out effect (smaller marginal gains or attenuated engagement) (Cerasoli et al., 2014; Deci et al., 1999; Patall et al., 2008). For students with initially low intrinsic motivation, extrinsic motivators can function as a short-term "on-ramp" that increases participation; repeated exposure and successful experiences may then create opportunities for interest and persistence to develop (Asher & Harackiewicz, 2024; Hidi & Harackiewicz, 2000; Priniski et al., 2018). Importantly, prior meta-analytic work suggests that such undermining effects are contingent on perceived autonomy: incentives tied to performance are most likely to displace intrinsic motivation when choice is limited, whereas perceived choice can enhance both performance and motivation

(Cerasoli et al., 2014; Patall et al., 2008). Consistent with this boundary condition, the present intervention sought to preserve autonomy by allowing students to set and revise their own goals, though future work should directly measure perceived autonomy and intrinsic motivation to test this mechanism more explicitly.

5.4.4 Explanatory Mechanisms and Novelty: Why Did Adaptive Calibration Help, and For Whom?

Beyond establishing that adaptive, student-set goals increased goal attainment, our findings contribute new evidence about the *mechanisms* through which motivational supports in educational technology can change learning behavior over time. In particular, the midpoint calibration opportunity appears to have increased students' experienced rate of success (i.e., attainable goals), which in turn strengthened week-to-week persistence (goal-achievement momentum), with the largest benefits concentrated among initially low-effort students. This connects autonomy-supportive goal design to a temporally unfolding reinforcement dynamic that is rarely measured in adaptive learning technologies and is not fully specified by prior theory.

A first explanation is that adaptive goal calibration may have increased *perceived competence* by making targets more attainable without removing challenge. Self-determination theory links autonomy and competence support to improved motivation, and recent meta-analytic evidence suggests these supports are effective in educational contexts (Howard, Bureau, Guay, Chong, & Ryan, 2021; Wang, Wang, Wang, Wind, & Gill, 2024). In our setting, calibration feedback explicitly contrasted students' intended goals with their realized performance and offered adjustment suggestions; this may have helped students update goal difficulty into a range that produced more frequent mastery experiences.

Second, our momentum results suggest that autonomy-supportive goal structures may amplify the downstream motivational impact of success. Achieving a goal increased the likelihood of achieving the next goal in both conditions, but the carryover was descriptively stronger under adaptive, self-set goals. One interpretation consistent with self-determination theory is that success may translate more strongly into future autonomous motivation when students experience the goal as self-endorsed rather than imposed (Howard et al., 2021). This would add on top of the known motivational benefits of regular academic achievement events (Wäschle et al., 2014). In this view, autonomy not only raises effort levels; it may increase the *return of success* on subsequent persistence.

Third, heterogeneity by baseline effort suggests an important boundary condition: goal supports with rewards appear to function primarily as an *effort-regulation scaffold* for students who would otherwise practice little. This pattern aligns with expectancy–value frameworks, which posit that externally structured motivational supports (e.g., performance goals or incentives) are most effective for students with low initial interest or engagement, while offering diminishing marginal returns for already highly motivated learners (Harackiewicz, Smith, & Priniski, 2016; Hidi & Harackiewicz, 2000). In this view, adaptive goal setting reduces effort disparities by supplying structure and reinforcement where self-regulation is weakest, rather than uniformly boosting engagement. At the same time, reduced benefits for some high-effort students may reflect countervailing motivational processes: externally framed goals and rewards can be experienced as evaluative or threatening, increasing performance-avoidance tendencies and fear of failure when success is uncertain (Covington & Müeller, 2001). Finally, domain-specific self-efficacy offers a complementary explanation for the strong gains observed among lower-effort students. Self-efficacy is shaped by mastery experiences and performance feedback and predicts persistence and goal pursuit; in our study, the adaptive condition's explicit feedback and opportunity for goal revision may have helped students with lower initial confidence translate early successes into stronger perceived capability, thereby reinforcing week-to-week persistence. Together, these mechanisms motivate future work that directly measures students' perceived autonomy, competence, performance-avoidance orientations, and self-efficacy to produce more precise explanatory accounts of who benefits most from adaptive goal calibration (Bureau, Howard, Chong, & Guay, 2022; Saks, 2024).

5.4.5 Limitations and Future Work

Several design features limit causal interpretation of the findings. The adaptive and static conditions differed simultaneously in goal source (student-set versus teacher-assigned), goal metric (skills mastered versus minutes practiced), and midpoint performance feedback (present only in the adaptive condition). The study therefore compares two intervention packages as opposed to isolating the effect of adaptive calibration alone. In addition, the crossover design introduced the possibility of carryover effects, as students may have transferred habits, reward expectations, or calibrated performance expectations from one phase to the next. Although the use of distinct goal metrics was intended to reduce student confusion and calibration transfer across phases, motivational and behavioral carryover cannot be ruled out. Future factorial designs will help disentangle the mechanisms underlying the observed effects.

A second limitation concerns unobserved heterogeneity in student motivation and the contextual scope of the intervention. Factors such as intrinsic motivation, perceived control, failure avoidance, and goal orientation were not directly measured, limiting our ability to identify which students are most responsive to different forms of goal setting, goal choice, and reward tracking. Future adaptive interventions could incorporate these constructs to support psychometrically informed personalization of goal recommendations and incentive structures. Pairing such interventions with *implementation intentions*—concrete if–then plans linking situational cues to goal-directed actions—may further strengthen performance while supporting autonomous goal pursuit (Brickell & Chatzisarantis, 2007; Gollwitzer, 1999; Gollwitzer & Sheeran, 2006). Moreover, the study was conducted exclusively within the IXL mathematics environment. While the weekly summary metrics used to generate adaptive recommendations (e.g., time spent and skills mastered) are common across many educational technologies, future research should examine whether adaptive goal-setting frameworks generalize to other learning domains, instructional settings, and digital practice platforms.

Finally, the relatively short intervention periods (four weeks per condition) may have limited the ability to observe longer-term changes in self-regulation, goal calibration, and the transfer of goal-setting strategies beyond the immediate tutoring context. Longer implementations spanning an academic year could clarify whether adaptive goal setting produces durable improvements in effort regulation and learning outcomes or whether effects diminish as novelty and extrinsic rewards fade. Relatedly, the adaptive condition permitted a wider range of student-selected goal difficulty despite an initial anchoring recommendation, creating additional variation in challenge level that was not experimentally controlled. Some students may therefore have selected goals that were substantially easier or harder than intended. Future work could leverage digital goal-setting supports to provide autonomy-preserving guardrails that encourage appropriately challenging targets while maintaining meaningful choice (Borchers, Peng, et al., 2025). More systematic co-design with teachers may also improve instructional alignment, feasibility of implementation, and students' perceptions of goal-setting supports as empowering rather than controlling.

5.5 Conclusion

The present findings provide empirical support for the study's proposed account of how adaptive goal calibration can enhance learning. Students who set their own adaptive goals, supported by data-driven calibration, achieved weekly goals at higher rates than students assigned static, teacher-set goals. Adaptive goals were also associated with larger learning gains, though the difference in learning outcomes between goal conditions was not statistically reliable. Relative to the no-goal baseline, both goal conditions increased practice time, whereas only the adaptive goal condition showed a reliable improvement in skill proficiency. Goal achievement was itself a meaningful predictor of the benefits students derived from goal support, highlighting mechanisms that help students set appropriately challenging targets and experience repeated success.

Taken together, these results point toward a dual benefit of adaptive goal setting in technology-enhanced learning: (1) immediate improvements in productivity, reflected in increased practice time and higher proficiency gains, and (2) enhanced goal calibration, as students adjusted their targets in response to performance-based feedback and were subsequently more likely to achieve them. These benefits were particularly pronounced for students with lower baseline practice time, suggesting that adaptive, autonomy-supportive goal setting can help close participation gaps in engagement and in-platform achievement for students who initially practiced the least. The findings also reveal that students with high prior baseline practice benefited less from the intervention, raising questions about the interplay between intrinsic motivation, autonomy, and the perceived value of extrinsic incentives. Addressing this heterogeneity will require deeper investigation into motivational profiles, goal orientations, and perceptions of control.

From a practical standpoint, the study demonstrates that data-driven goal-setting processes, when paired with timely recommendations, can be integrated into classroom tutoring environments without altering core instructional routines. Prior research on goal setting in education, in both classroom-based interventions and educational technology, has most commonly relied on fixed dosage targets (assigned minutes, assignments, or problem counts), often determined by teachers or vendors and typically without ongoing performance-based goal revision or sustained extrinsic rewards. In contrast, the present study evaluates student-selected performance goals combined with an explicit mid-intervention feedback process that contrasts intended goals with realized performance and provides concrete adjustment recommendations. The intervention was implemented within an existing tutoring program (adaptive practice software, synchronous remote tutoring, in-person coordination by teachers and research staff), not a custom-built research system. The practical contribution lies in providing field-based evidence that structured, data-informed goal calibration changes goal attainment and persistence relative to the static goal formats that dominate prior classroom practice.

As theoretical contributions, these results refine prior work on self-regulated learning and self-determination theory by providing field-based evidence on how autonomy in goal selection and revision relates to student behavior under performance-contingent rewards. Holding rewards constant across conditionsin contingency structure, goal attainment and week-to-week persistence differed depending on whether goals were teacher-assigned and static or student-set and adjusted in response to performance feedback. The results point to effort regulation and goal calibration as instructional targets distinct from knowledge-state adaptivity, manipulable through data-driven analytics in technology-enhanced learning environments.

Overall, the most consequential feature of goal support in technology-supported practice may be whether students experience repeated, attainable successes that carry forward from week to week. Goal calibration based on recent performance increased goal attainment and persistence in the studied tutoring context, with the largest gains among students who initially practiced the least. These forms of adaptivity have the potential to inform how persistence differences in individualized practice in adaptive STEM learning systems might be addressed.

Acknowledgments

This work was made possible with the support of the Learning Engineering Virtual Institute. The opinions, findings and conclusions expressed in this material are those of the authors. This research was funded by the Institute of Education Sciences (IES) of the U.S. Department of Education (Award #R305A220386).

Chapter 6

Adaptive Goal Support at Scale

Summary Statement in Relation to the Thesis

This chapter extends the thesis by demonstrating that adaptive goal support can be implemented at scale within a hybrid human-AI program and widely used adaptive learning systems while preserving its core behavioral mechanisms and benefits. The findings show that tutor-mediated, technology-enhanced goal support remains associated with meaningful increases in student practice under real-world conditions, even in the presence of heterogeneous implementation and limited instructional time. At the same time, the results refine the theoretical framework of the thesis by identifying an implementation tradeoff in which time devoted to behavioral scaffolding directly competes with time available for mathematics practice.

Beyond overall effects, the chapter provides insight into the mechanisms and boundary conditions of goal support. Process-level features, particularly implementation intentions, were associated with increased practice, with benefits concentrated among lower-effort students, while higher-effort students did not experience additional gains. In contrast, baseline motivational measures showed limited explanatory power for variation in outcomes, suggesting that responsiveness to goal support is shaped more by structured processes and implementation context than by stable individual traits. Relatedly, students' implementation-intention responses suggest that many viewed goal attainment as a problem of managing the learning environment. The most commonly selected strategies involved reducing distractions or negotiating quieter working conditions, such as wearing headphones, moving to a quiet spot, or asking peers for quiet time. This pattern indicates that, when students translated goals into concrete plans, they often focused less on changing their motivation or study habits broadly and more on controlling the immediate classroom conditions that interfered with practice. Taken together, these findings highlight that the effectiveness of adaptive goal support depends not only on its ability to influence student behavior, but also on how it is targeted and integrated into classroom environments that are constrained by limited instructional time.

Taken together, this chapter contributes to the thesis by providing evidence that effort-aware adaptive systems can be both effective and scalable, while clarifying the conditions under which behavioral scaffolds are most beneficial and the practical constraints that shape their impact in authentic classroom settings.

6.1 Introduction

Scaling the data-driven goal support intervention developed in earlier chapters requires reducing the level of human researcher facilitation needed to set, monitor, and adjust goals. In prior studies, these routines were conducted by an on-site researcher. Such an approach is not sustainable in typical instructional settings, where teachers and tutors must balance multiple responsibilities. The present study therefore examines whether core elements of the intervention can be embedded within existing tutoring workflows using infrastructure that automates key aspects of tracking and feedback, with minimal researcher involvement. This deployment occurs within the same human-AI hybrid tutoring context as prior studies of this thesis.

Hybrid human-AI tutoring is a suitable context for testing scalable goal support as it includes synergy between human and AI tutors that cover motivational and cognitive aspects of learning (Lin et al., 2023; D. R. Thomas et al., 2024). In this context, the central design question for adaptive goal support is whether a tutor-mediated goal support routine can operate effectively when standardized (to scale) across sites that differ in platform usage, tutoring cadence, and implementation constraints (Borchers, Houk, et al., 2025). Earlier chapters of this thesis argue that weekly, personalized practice targets, presented within tutors' existing workflows, can increase observable engagement and indirectly support learning by providing structure, feedback, and opportunities for adjustment. The present chapter evaluates this theory of change under scaled conditions. Further, it tests explanatory accounts of *how* goal support works (from the standpoint of goal setting and adjustment processes) and *why*, based on motivational variables that explain differences in intervention outcomes.

During the Fall 2025 term, the hybrid tutoring program implemented a common goal-setting routine with a program-wide launch on October 20, 2025. Across six partner schools (with four schools including a no-goal baseline similar to previous studies in this thesis), students logged weekly practice on two primary mathematics platforms, tutors used a shared dashboard displaying individual effort and progress targets, and a subset of students completed surveys at the beginning and end of the term. This setting provides the first opportunity in the thesis to observe the intervention operating at scale, without intensive researcher facilitation and with technology-mediated goal review with achievement-contingent rewards integrated into routine hybrid tutoring practice.

6.1.1 Study goals

Two study goals organize the empirical work in this chapter, and map onto this dissertation's broader contributions.

Study Goal 1 (Generalization of the intervention effect). The primary goal is to test whether the intervention effect observed in prior chapters, namely that a tutor-facilitated weekly goal-setting workflow is associated with increases in visible practice, replicates under scaled implementation. This includes replication across multiple educational technologies, under tutor-mediated rather than researcher-mediated goal setting and adjustment, and in the presence of heterogeneous implementation fidelity across sites driven by teachers rewarding goal achievement, as well as a frequent, data-driven goal-adjustment routine delivered through an online tutoring platform.

Study Goal 2 (The goal-setting *process* and who benefits). The second goal is to characterize the goal-setting process under scaled implementation and to relate process indicators to student-reported motivation. Specifically, this chapter examines how students select goals when provided with a data-driven goal recommendation through their tutor, and whether motivational constructs including intrinsic motivation, achievement goal orientations, and self-efficacy are associated with individual-level goal-related gains in practice. Finally, it tests the impact of additional goal scaffolding, namely implementation intentions, on student outcomes. This goal complements the question of replication by addressing for whom the intervention is most effective and through what processes, thereby foregrounding the mechanisms proposed in the thesis (namely goal feedback, recommendation, and student-driven adaptation; see Chapter 2).

6.1.2 Research questions

Three research questions operationalize the study goals and distinguish between intervention effects, process variation, and individual differences.

RQ1. Is the introduction of goal support associated with changes in weekly practice behavior, measured as minutes practiced and skills completed?

RQ2. Does prompting for implementation intentions, as an optional additional component of goal support, enhance student effort benefits during goal support?

RQ3. How do motivational factors (i.e., intrinsic motivation, self-efficacy, and goal orientation) relate to variation in practice gains through goal support?

6.1.3 Hypotheses

Three hypotheses correspond directly to RQ1–RQ3.

H1. Students exhibit higher weekly minutes practiced and skills completed following the introduction of goal support.

H2. Access to implementation-intention scaffolding is associated with higher levels of practice during goal support.

H3. Motivational and self-regulatory factors, including intrinsic motivation, self-efficacy, and achievement goal orientations, will moderate the effects of adaptive goal support. In particular, stronger effects are expected for students with more mastery-oriented goals and weaker performance-avoidance tendencies.

6.1.4 Novelty and significance

This chapter makes three primary contributions that extend the theory of adaptive goal support developed in Chapter 2 into a realistic, scaled middle school math implementation context.

First, it provides a field test of the goal-setting workflow under scaled conditions with limited researcher supervision. In prior studies in this thesis, goal setting, monitoring, and adjustment were facilitated by a researcher. In contrast, the present study embeds these processes within a hybrid human-AI tutoring system in which tutors mediate goal support, teachers distribute rewards, and technology supports weekly feedback and tracking. The intervention operates across multiple mathematics platforms and heterogeneous classroom contexts. This contribution is important because it evaluates whether the core principles outlined in Chapter 2, namely adaptive feedback (Roll et al., 2014), performance-calibrated goal recommendations (Adams et al., 2017; Locke & Latham, 2019), and learner autonomy in goal selection (Deci & Ryan, 1985; Ryan & Deci, 2000), can be sustained outside of controlled research settings.

Second, the chapter links log-based measures of student practice to student-reported motivational constructs across two survey waves in order to examine the mechanisms underlying the intervention. Building on the theory of change in Chapter 2, the analyses test whether changes in math self-efficacy are associated with patterns of practice consistent with mastery experiences. This is theoretically important because prior work on adaptive educational systems has focused primarily on in-task support such as hints and feedback, with limited attention to effort regulation and goal-directed behavior (Aleven et al., 2016; Azevedo et al., 2022). By connecting behavioral data to motivational change, this chapter provides evidence relevant to explaining why goal support affects practice and why effects differ across students. In addition, this contribution clarifies the role of externally prompted goal support within frameworks such as self-determination theory, where autonomy-supportive structures can still facilitate engagement in the presence of extrinsic rewards (Cerasoli et al., 2014; Koestner et al., 2008).

Third, the chapter characterizes the goal-setting process under realistic implementation constraints, including limited tutor time, variation in classroom practices, and differences in implementation fidelity. In particular, it examines process features such as goal adjustment and the use of implementation intentions (Gollwitzer, 1999). This contributes to the literature by extending the thesis focus from whether goal support works to how it works in practice. It also provides an empirical account of process-level scaffolding in goal pursuit, which is less frequently studied than outcome-level effects in educational technology. These findings identify where the intervention signal is robust and where it is sensitive to contextual variation, thereby defining concrete targets for future work on adaptive goal support and effort-aware learning systems.

6.2 Study Preparation and Goal Support Tool Design

Scaling adaptive goal support from a researcher-facilitated model to routine implementation introduces several important considerations. In this context, participating teachers assume responsibility for distributing goal-based rewards, while hybrid tutors set and review goals with students through an application that automates core components of tracking and feedback.

6.2.1 Teacher Reward Distribution Based on Goal Achievement Reports

A pilot study was conducted at a single participating school in Spring 2025 to inform the design of a teacher-facing goal report. The primary objective was to translate findings from prior chapters into a practical report that enables teachers to reward students with fruit snacks (as opposed to researchers doing so on-site). The resulting report summarized weekly student effort, goal attainment, and progress in a format designed for ease of use during tutoring sessions.

Researchers were present on-site during the initial week of goal setting to support implementation and introduce the reward structure. During this period, students set goals using the paper-based goal contract from the previous chapter (setting one goal for minutes and one for skills) and were introduced to the use of fruit snacks as a small, tangible reward for goal attainment. For the subsequent three weeks, teachers independently managed reward distribution without on-site researcher facilitation. Notably, this school had not participated in prior goal-setting interventions described in earlier chapters, and therefore represents a newly onboarded implementation context.

To support implementation, weekly reports were distributed to teachers via email in the form of a spreadsheet generated by a member of the research team. Based on this interim evaluation, the report design included columns such as teacher, grade, period, student name, goal targets (minutes and skills), weekly practice metrics (minutes practiced, skills mastered), and an indicator of reward eligibility (i.e., "Fruit Snack Eligible"). In response to teacher feedback, the report was refined to improve usability after the pilot concluded: students eligible for rewards were highlighted in green, while others were marked in red, and rows were sorted accordingly to streamline reward distribution during class time.

Analyses of IXL math activity, using an interrupted time series approach consistent with prior chapters, indicated patterns closely mirroring earlier findings. The sample consisted of one school (199 students, 9 teachers) participating in the hybrid tutoring twice per week during one class period of mathematics instruction. Results indicated a 22% increase in time-on-task and a substantial increase in skill mastery (approximately 65%), suggesting not only greater engagement but also higher efficiency in practice (i.e., more skills mastered per unit of time). These findings replicate prior evidence of increased effort alongside disproportionately larger gains in skill acquisition. Detailed statistical analyses are not reported here for brevity and will be presented in future work.

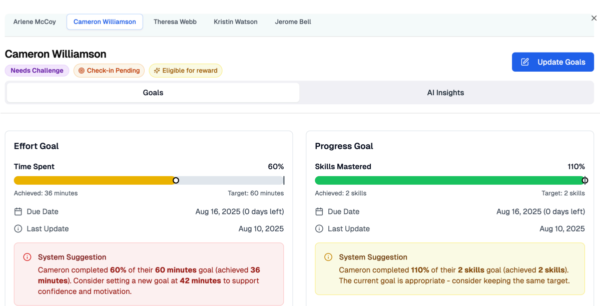
A teacher debrief was conducted at the end of the semester with all participating teachers, the lead researcher, and a representative from the tutoring program's partnerships team. Based on informal notes from this discussion, teachers reported that students generally responded positively to the reward system. Additionally, some teachers (n = 2) reported that students were particularly motivated when shown real-time progress toward their goals based on IXL Math dashboard displays (i.e., live updates of time-on-task and skills mastered). This observation informed a subsequent design decision to integrate more visible, student-facing progress tracking into the hybrid tutoring platform (see next subsection), enabling regular feedback on goal-relevant metrics.

6.2.2 Goal Support Platform

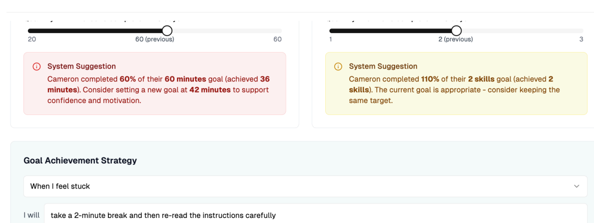
Prototyping design sessions with five middle school students were used as a basis for the design of the platform (see details in Borchers, Peng, et al. (2025)). In this study, students interacted with early dashboard mockups under simulated conditions of under- and overachievement to elicit their responses to system-generated goal recommendations. Findings indicated that students valued adaptive recommendations grounded in their prior performance, but strongly preferred retaining final control over their goals. At the same time, students expressed appreciation for guardrails that prevented excessively large or small adjustments, particularly as a means of avoiding overwork or disengagement.

These findings directly informed the platform's design in two ways. First, goal recommendations were implemented as advisory rather than prescriptive, requiring explicit student acceptance or modification. Second, constraints were introduced on allowable deviations from recommended goals to ensure that student-defined targets remained within a reasonable and achievable range, thereby operationalizing the calibration principles outlined in Chapter 2. In practice, goal bounds were determined based on prior discussions with teachers or administrators. In the absence of site-specific preferences, default constraints were applied. For effort goals, values were bounded between 10 and 50 minutes per week, reflecting typical classroom instructional time. For progress goals on IXL and Khan Academy, targets were constrained between 1 and 5 skills per week, based on approximately the 80th percentile of historical performance distributions. For MobyMax, the upper bound was set to 10 skills per week, reflecting higher rates of skill completion observed on that platform relative to IXL.

The resulting platform (Figure 6.1) automatically records goal-setting events and stores them in a relational database, enabling persistent tracking of both time-based and skill-based goals. On a weekly basis, data pipelines ingest exported usage logs from partner educational technology platforms (e.g., IXL, Khan Academy, MobyMax), which are then matched to individual student goals. These processes update goal progress metrics and compute goal attainment indicators used for both student-facing feedback and teacher-facing reward reports. Tutors would log into the platform to share their screen and review these progress metrics with their student before updating goals (Figure 6.1a). Goals would always be updated such that all students had an effort (minutes, problems completed per week) and progress goal (skills mastered per week) at all times. Effort and progress goals were always updated together. At each adjustment, students had the choice of choosing minutes or problems as their effort goal in MobyMax. For the other ed-tech platforms, students would always have a minutes per week and a skills per week goal. In cases where students did not set goals during tutoring sessions, a default goal is assigned based on edtech-level recommendations to ensure that all students receive a minimum level of structure and feedback (set to 30 minutes per week and 2 skills per week).



(a) Adaptive goal calibration dashboard with current progress and recommended adjustment.



(b) Implementation intentions prompt for specifying an "if-then" plan which will be accompanied by a personal goal rationale.

Figure 6.1: Hybrid tutoring goal support interface designs for the adaptive goal-setting intervention. The left panel shows the goal calibration dashboard; the right panel displays the goal setting and implementation intentions interface.

The platform further incorporates adaptive goal recommendations based on recent student performance. Recommendations are computed as the midpoint between the current goal and observed performance rounded to the nearest integer, effectively implementing an adjustment rule that moves goals toward achievable levels. This formulation aligns with students' expressed preference for guardrails during calibration and supports early-stage stabilization of goal difficulty. Consistent with findings from the prototyping study, all recommendations are accompanied by simple, interpretable feedback and are presented alongside visualizations of progress, including past weeks' effort and skill mastery trajectories. These design choices aim to support reflection in both pre-actional and post-actional phases of goal pursuit, while maintaining transparency about the rationales for recommendations.

When setting goals (see Figure 6.1b), tutors shared their screen and prompted students to express their goal preference in response to the system's recommendation, following the design described in Borchers, Peng, et al. (2025). Tutors then adjusted the goal using a slider with built-in guardrails. For 50% of students, this process was followed by a prompt to set implementation intentions (see Section 6.2.4).

6.2.3 Tutor Training and Implementation Support

Implementation fidelity was supported through a combination of structured tutor training, ongoing coordination mechanisms, and real-time support channels with the hybrid tutoring team. First, all tutors completed a standardized training module as part of their onboarding for the semester. The module introduced the rationale for goal support, emphasizing its role in increasing student engagement and learning, and provided step-by-step guidance on how to integrate goal-setting routines into regular tutoring sessions. Tutors were trained to (a) provide brief, actionable goal feedback, (b) facilitate student decision-making when system-recommended goal adjustments were presented, and (c) follow a consistent workflow embedded within the PLUS app and goal-support dashboard. This workflow included reviewing student progress, discussing recommendations, and recording updated goals on behalf of the student during screen-sharing while maintaining student ownership of the decision. Tutors were also instructed to support two types of goals—effort-based (e.g., minutes practiced) and progress-based (e.g., skills mastered)—and to keep goal-related interactions brief (ideally under one minute) to preserve instructional time for mathematics support.

The training further emphasized key implementation principles aligned with the intervention design. These included prioritizing student autonomy in goal setting, using system recommendations as guidance rather than prescriptions, and maintaining a balance between goal support and core tutoring responsibilities (i.e., prioritizing immediate math support first when multiple students are seeking help simultaneously). Tutors were also introduced to optional scaffolds such as goal achievement strategies (i.e., implementation intentions), which could be used to support students in translating goals into concrete action plans when appropriate.

Second, weekly debriefs were conducted with lead tutors, who oversee tutoring sessions across sites. These meetings served as a forum to clarify questions about goal support procedures, share implementation challenges, and reinforce best practices. The lead researcher occasionally participated in these meetings, especially in the first few weeks of implementation. These debriefs also allowed the research and program teams to monitor emerging issues in real time and iteratively refine guidance to tutors based on observed patterns in the field.

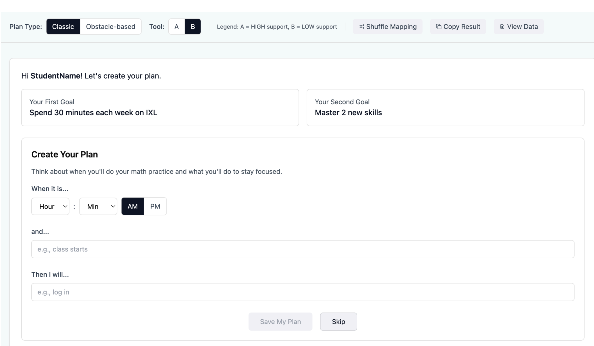
Third, a dedicated Slack channel was maintained to provide real-time implementation support. Tutors and lead tutors could escalate questions related to goal-setting workflows, dashboard functionality, or classroom coordination. These inquiries were addressed promptly by members of the tutoring, technical, or research teams, ensuring that implementation issues could be resolved during active tutoring sessions.

6.2.4 Goal Implementation Intention Design

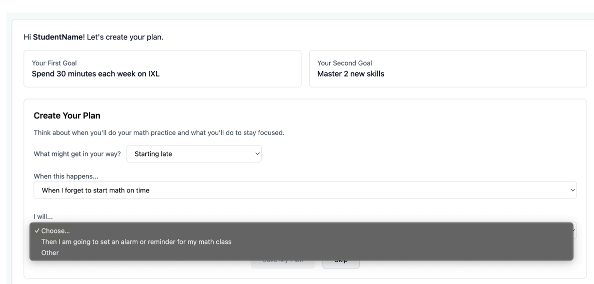
A key research design consideration was the randomized inclusion of implementation intentions as an additional scaffold for goal-directed behavior. Implementation intentions, typically structured as “if-then” plans, are intended to support the translation of abstract goals into concrete, actionable behaviors (Gollwitzer, 1999; Gollwitzer & Sheeran, 2006). This is especially so in contexts where students must initiate and sustain effort independently.

A pilot conducted at one participating school - the site of the first intervention study in Chapter 4 - informed the design of this feature. As part of regular hybrid tutoring program delivery in Spring 2025 (as part of the study described in Chapter 5), students were already pursuing weekly goals with fruit snack rewards. During one full school day, an on-site research team member conducted short interviews with approximately 20 students across seven class periods in brief, opportunistic interactions embedded within tutoring sessions. In these interactions, students were prompted to reflect on potential obstacles to goal attainment and were guided through the process of forming implementation intentions using early design mockups (Figure 6.2).

To explore alternative scaffolding approaches, the researcher rotated across four conditions varying along two design dimensions: (1) the anchor of the implementation intention (time-based vs. situation-based) and (2) the level of support (free-form entry vs. structured pre-selection). Time-based anchors prompted students to specify when they would work toward their goals (e.g., “after school at 4pm”; see Figure 6.2a), whereas situation-based anchors focused on contextual triggers (e.g., “when I finish my homework” or “when I feel distracted”). Support levels varied between open-ended text entry and interfaces that provided selectable options for common situations or time slots (see Figure 6.2b). At the end, students were prompted to compare whether they would find the respective alternative anchor design (time vs. obstacle-based) more helpful. Design decisions regarding the level of support were based on the quality and fluency with which students were able to complete the setting of implementation intentions.



(a) Implementation intentions design exemplifying time-based anchoring with free-form entry.



(b) Implementation intentions design exemplifying situation-based anchoring with pre-selection.

Figure 6.2: Implementation intention scaffolding design used in a pilot implementation to decide about the anchor of the intention (time-based vs. situation-based) and level of support (free-form vs. pre-selection).

Following the pilot, the lead researcher and a member of the hybrid tutoring design team reviewed structured observation notes to derive design implications. A primary finding was that many students struggled to independently generate complete and meaningful if-then statements, particularly the "if" component involving identification of realistic obstacles or triggering situations independently. Students often defaulted to vague or non-actionable conditions (e.g., "when I have time") or omitted the conditional component altogether. This is important because implementation intentions require an easily recognizable recognition trigger to be recalled from memory and acted upon. Therefore, the decision was made to add scaffolding of concrete situations to which students could relate their implementation intentions.

In response, the final design incorporated structured scaffolding for the "if" component by providing a curated set of common situations derived from pilot responses (e.g., "when I arrive to class," "when I am distracted by talking to my neighbor," "when I feel distracted by YouTube"). This reduced the cognitive load associated with generating implementation intentions while maintaining some degree of personalization through selection and optional refinement, as students were able to indicate a free-form entry in an "Other" category. Second, situation-based anchoring was prioritized over time-based formulations. Observations suggested that situational cues were more intuitive for students and better aligned with classroom contexts, where the start and end of practice are largely determined by schedules and teacher-imposed structures. As a result, students were unlikely to independently track clock time while practicing.

6.3 Methods

6.3.1 Sample

The present chapter uses two overlapping analytic samples, corresponding to two rollout patterns in Fall 2025. Students were included in the sample if their legal guardians did not opt them out of participation in the hybrid tutoring program. The **full sample** comprises all partner middle-grades schools that participated in the PLUS hybrid tutoring program with adaptive goal support under the program-wide implementation calendar. Concretely, this includes six partner middle schools (School A through School F). In the analysis repository used for this thesis, the full weekly panel contains 4,736 student-week rows from 1,417 distinct students. This sample supports program-wide goal support process analyses and analyses that link logs to fall and winter surveys wherever roster matching permits.

The **interrupted time-series subsample** is restricted to the four schools in which goal support began only after an initial period of hybrid tutoring without logged goal-setting targets in Fall 2025, yielding interpretable within-student contrast between a no-goal phase and a with-goal phase on the same calendar. These schools are School B, School C, School D, and School F. Analyses that require this contrast - including the school-level descriptives in Table 6.1, the interrupted time series in Table 6.2, and the implementation-intention models in Table 6.3 - are estimated on this subsample. As reported in those tables, the ITS models use 2,513 student-week observations from 522 students, with practice aggregated across two dominant mathematics platforms (IXL and MobyMax) and 16 unique classroom teachers.

6.3.2 Materials

Mathematics practice platforms (IXL, Khan Academy, and MobyMax). Students completed independent mathematics practice on adaptive educational technologies, chiefly **IXL Math**, **Khan Academy**, and **MobyMax**. These systems deliver standards-aligned item or exercise practice, supply skill- or topic-level progress summaries, and log time on task. The products differ in how they label mastery, completion, and proficiency, which statistical analyses adjust for by adding an ed-tech specific intercept in regression models. For this study, weekly exports from each vendor were harmonized into a common panel of *minutes practiced*, *problems completed*, and *skills* counts so that goal targets (effort and progress) could be reviewed on a single tutor dashboard regardless of which product a student used most often. Using multiple partner platforms extends prior thesis work that relied primarily on a single technology and is central to testing whether adaptive goal support generalizes beyond one vendor's platform and analytics.

Student surveys (fall and winter). Brief motivational surveys were administered online early in the hybrid tutoring term (fall) and again toward the end of the term (winter) during the final week of tutoring. Surveys were timed such that they removed approximately equal instructional time from both the pre-intervention and intervention phases, thereby minimizing potential bias in estimating the main intervention effect.

Survey measures were selected to align with the mechanism-focused priorities outlined in Chapter 2. **Intrinsic motivation** was measured using the intrinsic value subscale of the Motivation Questionnaire (MMQ) for secondary school students, consisting of three items on a 5-point Likert scale (Fiorella et al., 2021). **Perceived autonomy** was captured using the Perceived Choice subscale of the Intrinsic Motivation Inventory (IMI), comprising seven items on a 7-point Likert scale (Self-Determination Theory, 2022). **Achievement goal orientations** were measured using an adapted version of the Achievement Goal Questionnaire–Revised (AGQ-R), modified for middle school mathematics contexts to assess mastery and performance orientations as well as approach and avoidance dimensions. This scale included four constructs with three items each, using a 5-point Likert scale (Elliot & Murayama, 2008). Intrinsic motivation, goal orientation, and self-efficacy were assessed twice. All other constructs were only assessed once at the beginning of the semester.

In addition, **math self-efficacy** was measured using a single confidence item suitable for repeated classroom administration, and **college expectancy** was assessed with a single item capturing perceived likelihood of completing college. All Likert-type responses were numerically coded for analysis. Fall responses were matched to roster identifiers to enable linkage with weekly behavioral log data. Students provided assent for survey participation in accordance with the approved institutional review board protocol.

6.3.3 Procedure

At the beginning of the school year, meetings were conducted with school administrators and, when available, teacher representatives at each participating site to coordinate implementation logistics. These meetings established tutoring schedules, clarified the integration of goal support into classroom routines, and aligned expectations regarding teacher-led reward distribution for goal attainment.

Hybrid tutoring sessions took place as part of regular mathematics instruction or designated intervention periods, typically occurring one or two times per week. During each session, students joined virtual tutoring rooms (on the Zoom or Pencil video conferencing software) and were placed into small breakout groups with assigned college tutors. Tutors were instructed to begin each session with a brief check-in to establish contact with all students and take attendance.

Goal support was embedded within these ongoing tutoring interactions. After the initial check-in, tutors revisited students individually during the session to review their goal progress using the dashboard. When prompted by the system, tutors facilitated goal-setting or goal-adjustment conversations by sharing their screen, discussing current performance relative to goals, and supporting students in selecting updated targets. These interactions were designed to be brief (typically under one minute per student) to minimize disruption to core mathematics instruction while ensuring regular exposure to goal feedback and adjustment opportunities.

For a subset of students (randomized at 50% chance), implementation intention scaffolding was incorporated into this workflow throughout the Fall semester (see Section 6.2.4). In these cases, tutors additionally guided students through selecting or refining a goal achievement strategy following goal setting. This step was integrated directly into the same interaction as goal setting.

The timing of goal support followed a program-wide rollout (October 20, 2025) when goal-setting routines were introduced and maintained consistently for the remainder of the term. Student surveys were administered online during tutoring sessions at two time points: early in the term (fall) and during the final week of tutoring (winter). Surveys were completed under tutor supervision and screen-sharing to ensure participation while minimizing disruption to instruction.

Throughout the study period, teachers reviewed weekly goal attainment reports (sent once per week via email at 7:00 AM local time) and distributed small rewards (e.g., fruit snacks) to students who met their goals. Reports were sent and goal week cycles were started on the first week day with hybrid tutoring per school. Teachers were also provided with a one-page handout summarizing when and from what email address they would receive fruit snack reports at the beginning of the semester. Fruit snacks were delivered to schools via mail 1-2 weeks ahead of the intervention start.

6.3.4 Analysis Methods

The primary unit of analysis is a **weekly student-week summary**: each row summarizes one student's practice for a calendar week aligned to the program's export schedule. Observations are statistically nested within students and within classrooms, teachers, schools, and practice platforms (IXL, Khan Academy, MobyMax) which models adjust for via random intercepts unless otherwise specified. Primary outcomes are **minutes practiced** and **skills completed** (since those are the two variables available across all ed-techs). Where noted in the results, these two outcomes are modeled on a **log scale** to temper right skew (since both variables cannot be lower than zero) for inferential models.

RQ1: interrupted time series for goal support. To estimate the main effect of introducing goal support (RQ1), interrupted time series specifications were fit on the delayed-start subsample that contains both a pre-period without goal support and a during-period with goal support. Predictors include an overall linear calendar time trend (week index), an indicator for the post-launch segment of the term, representing a differential linear trend after introduction of the intervention, and a goal-support indicator contrasting weeks with active goal setting and feedback against weeks without it, representing the main intervention effect (and test of H1).

To account for weeks in which students recorded no activity, models include an indicator for no activity (i.e., zero minutes practiced). This adjustment addresses a potential confound whereby zero-activity weeks may reflect a mixture of factors such as student absence, late program entry, differences in roster management, or intermittent platform use, as opposed to true differences in effort. Including this control helps separate changes in participation (whether students engage at all) from changes in effort conditional on engagement, yielding more interpretable estimates of goal support on observed practice behavior. We included this adjustment because zero-activity rates were substantially higher in the pre-intervention period than during the intervention (approximately 72% vs. 53% overall, with variation across schools), which could otherwise inflate estimated intervention effects by conflating increased participation with increased effort.

RQ2: goal-setting process and implementation intentions. To examine goal-setting processes (RQ2), models relate **post-period** practice (minutes and skills) to **pre-period baseline practice**, a binary indicator of **implementation-intention exposure** (whether the student could articulate or record an achievement strategy in the goal-support tool; corresponding to the test of H2), and their interaction effect. The interaction tests the prediction that the effectiveness of implementation intentions depends on baseline weekly effort.

RQ3: Motivation and student-level intervention benefit (G). For analyses linking logs to surveys, intervention benefit is defined as a student-level composite intervention effect G, following the student-specific intervention-effect approach used in Chapter 5. To estimate G, the interrupted time-series models for weekly minutes and weekly skills are refit with student-level random slopes for the goal-support indicator, allowing the intervention effect to vary across students. For each outcome, the student-specific goal-support effect is computed by adding the student's random slope to the fixed effect for goal support. These student-level effects are then standardized (z-scored), and the standardized minutes and skills effects are averaged to form G. This averaging is warranted by a moderate-to-high correlation between these two effects ($r = .486$). Higher values indicate larger estimated intervention benefits.

Associations between motivational measures and G are estimated using Pearson correlations and ordinary least squares regression. For students with both fall and winter surveys, within-year change scores (winter minus fall on shared items) are used in exploratory regressions of G on changes in intrinsic motivation, mastery approach, math self-efficacy, and college expectancy. All survey-based models use listwise deletion, resulting in smaller analytic samples.

6.4 Results

6.4.1 RQ1: Main Goal Support Intervention Effect

Table 6.1 reports average weekly minutes practiced and skills completed by school, comparing periods without goal support (0) and with goal support (1).

Table 6.1: Average weekly practice by school and goal support condition after removing student-weeks with no activity. The Student-weeks column counts analytic rows in the ITS subsample (not unique students); models in Table 6.2 use 522 unique students across these schools.

	Sample		Minutes		Skills	
School	Student-weeks	Teachers	No Goals (0)	With Goals (1)	No Goals (0)	With Goals (1)
School B	373	6	9.71	23.90	1.40	4.65
School C	109	2	18.60	26.70	0.76	1.08
School D	113	1	36.50	35.60	2.12	1.37
School F	147	2	11.80	20.60	1.91	2.19

Table 6.2 reports interrupted time series models predicting weekly minutes practiced and skills completed, controlling for zero-activity weeks. The introduction of goal support is associated with increases in both outcomes. For minutes practiced, the estimated coefficient is 0.25 (95% CI [0.17, 0.33], $p < .001$), corresponding to an approximate 28% increase in weekly minutes practiced. For skills completed, the estimated coefficient is 0.42 (95% CI [0.34, 0.50], $p < .001$), corresponding to an approximate 52% increase in weekly skills completed. These results support H1. These percentage increases are computed by exponentiating the log-scale coefficients (i.e., $e^{\beta} - 1$), reflecting multiplicative changes in the original outcome scale.

At the same time, the introduction of goal support counteracted a negative time trend. Weekly practice declines over the course of the term, with coefficients of -0.07 for minutes and -0.11 for skills (both $p < .001$). The post-launch period is associated with a positive linear trend shift relative to this trend. The estimated coefficients are 0.07 for minutes ($p = .005$) and 0.07 for skills ($p = .017$), indicating a modest upward adjustment in practice following the introduction of goal support. Taken together, these results indicate that goal support is associated with substantially higher levels of student practice in both minutes and skills, even after accounting for zero-activity weeks and an overall declining baseline trend over time.

In a post-implementation discussion with partner schools, one concern raised—particularly regarding MobyMax—was that some students may have selected easier math fact skills to meet their goals. This raises the possibility that observed increases in skills completed could partially reflect changes in task difficulty rather than true improvements in learning. To assess whether skill improvement estimates were driven by platform-specific effects, we estimated a model allowing the effect of goal support to vary by educational technology (i.e., adding a random slope for ed-tech). Model comparison did not support this additional complexity: the model with random slopes showed worse fit (AIC: 4140.68 vs. 4136.68; BIC: 4210.63 vs. 4194.97), suggesting that allowing the intervention effect to vary by platform does not improve explanatory power. These results provide no evidence that skill gains are disproportionately driven by a single platform or by platform-specific behavior such as completing easier skills.

Table 6.2: Interrupted time series models predicting weekly practice (absence-controlled)

	Minutes (log)	Skills (log)
Intercept	3.02***	0.84*
	[2.91, 3.13]	[0.12, 1.56]
Week (time)	-0.07***	-0.11***
	[-0.09, -0.06]	[-0.13, -0.10]
Post-launch period	0.07**	0.07*
	[0.02, 0.11]	[0.01, 0.12]
Goal support	0.25***	0.42***
	[0.17, 0.33]	[0.34, 0.50]
No Activity	-2.85***	-0.91***
	[-2.90, -2.80]	[-0.96, -0.86]
R^2 (marginal)	0.877	0.570

Note: 95% confidence intervals in brackets. * $p < .05$, ** $p < .01$, *** $p < .001$.

6.4.2 RQ2: Impact of Implementation Intentions

6.4.2.1 Quantitative Impact on Student Effort and Skill Mastery

Table 6.3 reports models predicting post-period minutes and skills as a function of baseline practice, the ability to articulate an achievement strategy, and their interaction.

For minutes practiced, baseline effort is strongly associated with post-period effort ($\beta = 0.71$, 95% CI [0.62, 0.80], $p < .001$). Students who can articulate an achievement strategy exhibit higher post-period minutes ($\beta = 0.23$, 95% CI [0.05, 0.41], $p = .014$). This effect is moderated by baseline effort, with a negative interaction ($\beta = -0.23$, 95% CI [-0.38, -0.08], $p = .003$), indicating that the association between implementation intentions and practice is larger for students with lower baseline minutes.

For skills completed, baseline performance is again positively associated with post-period outcomes ($\beta = 0.27$, 95% CI [0.18, 0.36], $p < .001$). Students who can articulate an achievement strategy complete more skills on average ($\beta = 0.16$, 95% CI [0.01, 0.30], $p = .038$). The interaction with baseline skills is not statistically significant ($\beta = -0.04$, $p = .540$), indicating that this association does not vary meaningfully across baseline levels.

These results indicate that implementation intentions are associated with higher levels of practice, with stronger effects for lower-effort students in the case of minutes practiced. Together, findings lend partial support to H2.

6.4.2.2 Qualitative Analysis of Implementation Intentions

A total of 762 strategy responses were analyzed using inductive thematic coding. To account for interface effects, responses were separated into system-provided dropdown selections ($n = 520$) and self-reported free-text responses ($n = 242$). Dropdown strategies reflect constrained choices from predefined options, whereas free-text responses capture spontaneously generated strategies.

Dropdown responses clustered into several distinct strategy categories. The most common category involved managing environmental distractions, including "wear headphones or move to a quiet spot" ($n = 125$, 22.4%) and "ask my friends to give me quiet time" ($n = 115$, 20.6%). To-

Table 6.3: Implementation intentions models predicting post-period practice

	Minutes (post)	Skills (post)
Intercept	-0.13**	-0.32***
	[-0.22, -0.04]	[-0.43, -0.21]
Baseline practice	0.71***	0.27***
	[0.62, 0.80]	[0.18, 0.36]
Can set strategy	0.23*	0.16*
	[0.05, 0.41]	[0.01, 0.30]
Baseline × Strategy	-0.23**	-0.04
	[-0.38, -0.08]	[-0.17, 0.09]
Observations	494	494
Schools	4	4
Platforms	2	2
R^2 (marginal)	0.493	0.153

Note: 95% confidence intervals in brackets. * $p < .05$, ** $p < .01$, *** $p < .001$.

gether, these two strategies accounted for 43.0% of all structured selections. A second frequently selected category was support-seeking, including "ask my tutor for help" ($n = 99$, 17.7%) and "ask for a hint or help (from the system or tutor)" ($n = 28$, 5.0%), which together accounted for 22.7% of selections. The remaining responses reflected self-management strategies. Attention-regulation strategies included "take a short break to focus" ($n = 58$, 10.4%) and "take a short break and come back with a fresh mind" ($n = 25$, 4.5%). Task-initiation and time-management strategies included "get my laptop and start working first thing in the class" ($n = 43$, 7.7%), "set a short timer (10 minutes) and just start" ($n = 22$, 3.9%), and "set an alarm or reminder for my math class" ($n = 19$, 3.4%). Less frequently selected strategies focused on reducing digital distractions by "close all tabs except my learning related ones" ($n = 13$, 2.3%) and on motivation regulation through "remind myself why finishing math will help me (grades, goals)" ($n = 12$, 2.1%). Overall, responses were concentrated in strategies that modified the learning environment or leveraged external support, while internally managed attention, planning, and motivation strategies were selected less often.

A comparable pattern emerged in the thematic analysis of self-reported responses ($n = 242$), although the distribution of strategies was more heterogeneous than in the structured dropdown selections. As with the dropdown responses, the most common themes involved managing the learning environment and seeking support. Environment-control strategies accounted for approximately 25–30% of responses and included statements such as "move somewhere quieter," while help-seeking strategies represented approximately 20–25% of responses and included responses such as "ask the teacher for clarification." Social-regulation strategies, such as requesting space from peers or adjusting social interactions to support concentration, accounted for approximately 15–20% of responses. Smaller proportions reflected attention-management and recovery strategies, including responses such as "take a short break and come back later." In contrast to the structured selections, self-reported responses also included a wider range of individualized self-management strategies (though such strategies continued to be rare). Examples included task-initiation approaches (e.g., "start with the easiest problem first") and planning or time-management strategies (e.g., "set a timer to stay focused").

6.4.3 RQ3: Intervention Benefits and Motivational Measures

Intervention benefit was indexed by the composite standardized practice shift G, which captures student-level variation in response to the goal-support intervention. G was defined as the mean of within-platform z-scored shifts in weekly minutes and weekly coalesced skills between the pre- and post-goal-support periods. Higher values indicate students who exhibited larger practice gains than their peers on the same dominant practice platform.

Associations with pre-intervention (fall) motivation. Among students with a linked earliest fall survey and valid G (n = 250; restricted to those with both pre- and post-launch weekly observations), we first examined the association between intrinsic motivation and G, given its central theoretical relevance in the presence of extrinsic incentives. In a univariate regression, the standardized coefficient was positive and statistically significant ($\beta = 0.18$, 95% CI [0.07, 0.29], $p = .002$). In a multivariate model adjusting for failure avoidance, mastery approach, and math self-efficacy, the association between intrinsic motivation and intervention benefit remained positive but did not reach conventional levels of statistical significance ($\beta = 0.14$, 95% CI [-0.03, 0.32], $p = .101$). This attenuation is consistent with substantial overlap among the motivational constructs, particularly between intrinsic motivation and math self-efficacy ($r = .71$ in the regression sample). None of the other fall motivational predictors were significant (failure avoidance: $p = .434$; mastery approach: $p = .370$; math self-efficacy: $p = .545$). Overall explanatory power was low ($R^2 = .049$, adjusted $R^2 = .033$). Reliability was acceptable for the three-item intrinsic motivation scale (Cronbach's $\alpha = .848$) but more modest for the composite outcome measure G. Because G was defined as the average of within-platform standardized shifts in weekly minutes and weekly skills, its reliability was constrained by the correlation between these two components. Spearman–Brown corrected split-half reliability for G, based on repeated random halving of the weekly panel data and recomputation of the composite, was approximately .59. Thus, the weak multivariate effects may reflect both measurement noise in the outcome and overlap among the predictor constructs. Taken together, the results provide some evidence for a positive bivariate association between intrinsic motivation and intervention benefit, but this relationship was modest and not robust to multivariate adjustment. Overall, the findings provide only weak support for H3.

Association between intervention benefit and changes in motivation. Among students with both an earliest fall match and a winter match on **shared survey items**, complete data on G, and complete winter-minus-fall change scores on the three scales entered into the change-score model ($n = 65$), within-year deltas were defined as **Winter** $-$ **Fall** on (i) intrinsic motivation (three-item composite), (ii) mastery approach (single item), and (iii) math self-efficacy (single item)—the same constructs carried in both administrations. The simple Pearson correlation between G and change in math self-efficacy was $r = 0.26$ (two-sided $p = .039$). An exploratory ordinary least squares model regressed G simultaneously on Δintrinsic motivation, Δmastery approach, and Δmath self-efficacy. Only the coefficient on Δmath self-efficacy was positive and statistically distinguishable from zero at $\alpha = .05$ (estimate ≈ 0.26, two-sided $p = .018$); Δintrinsic motivation (estimate ≈ -0.06, $p \approx .69$) and Δmastery approach (estimate ≈ -0.11, $p \approx .34$) were not significant in that specification.

6.4.4 Exploratory Analysis of Goal Revision Behavior

Selected goals closely tracked recommended targets, with substantial variance explained in both categories. The relationship was stronger for progress goals ($R^2 = 0.75$) than for effort goals ($R^2 = 0.61$), indicating higher alignment with recommendations in progress-based goals. In both categories, selected goals increase with recommended targets, though less than proportionally, based on simple linear regressions predicting the chosen goal target based on the recommended target. For effort goals, the slope of 0.653 indicates that students raise their selected goals by about 0.65 units for each additional unit recommended (due to guardrails for the maximum value), while the positive intercept (10.5) suggests a substantial baseline level of effort even when recommendations are low due to guardrails to the minimum value. For progress goals, the slope is similar (0.679), indicating comparable responsiveness, with a similar, non-zero intercept (0.894). Overall, students partially adjust toward recommendations, with stronger alignment (higher $R^2 = 0.75$) in progress than in effort goals ($R^2 = 0.61$).

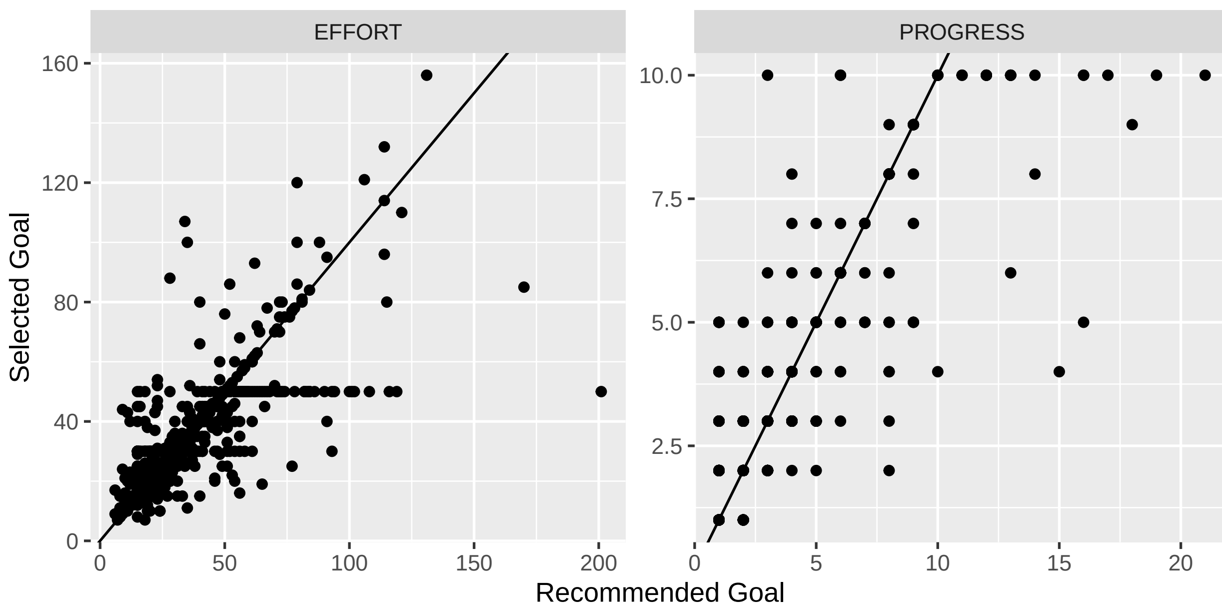


Figure 6.3: Recommended vs. selected goal by category (effort vs. progress) including estimated linear regression trend line

We hypothesized that students with higher intrinsic motivation would set more goals and engage more with tutoring. However, this was not supported by the data. A Poisson regression predicting the number of goals from intrinsic motivation showed no significant relationship (IRR = 1.02, 95% CI [0.94, 1.10], $p = .696$). The model explained virtually no variance (Nagelkerke $R^2 = 0.001$), indicating that intrinsic motivation was not meaningfully associated with goal-setting behavior in this sample.

We estimated a linear mixed-effects model predicting selected goals from prior recommended targets and student motivational characteristics, including intrinsic motivation and multiple achievement goal orientations, with a random intercept for student to account for repeated observations. Results indicated that selected goals were strongly associated with prior recommended targets ($\beta = 1.01$, $p < .001$), suggesting near one-to-one alignment. Among motivational predictors, only performance-approach orientation was significantly related to higher selected goals ($\beta = 0.05$, $p = .007$), while intrinsic motivation and other orientations were not significant ($p > .20$). The model explained a large proportion of variance overall (marginal $R^2 = 0.878$, conditional $R^2 = 0.884$), largely driven by the strong effect of prior recommended targets. Multicollinearity diagnostics indicated no concerns (all VIFs < 2), suggesting that predictor estimates are stable and not distorted by correlations among covariates.

6.5 Discussion

This chapter extends the thesis by demonstrating that adaptive goal support can be implemented at scale within a hybrid human–AI tutoring system while preserving its core theoretical mechanisms. Building on prior chapters, the central objective was to test whether a tutor-mediated, technology-supported goal-setting routine generalizes beyond researcher-facilitated settings, and to examine the processes and learner characteristics that shape its effectiveness. The novelty lies in embedding adaptive goal support into real-world instructional workflows with minimal researcher involvement, while simultaneously linking behavioral outcomes to motivational and process-level mechanisms.

6.5.1 Summary of results

This study provides evidence that adaptive goal support can be implemented at scale within a hybrid human–AI tutoring context and remains associated with meaningful increases in student practice. The introduction of goal support was linked to substantial gains in both minutes practiced and skills completed, even in the presence of an overall declining time trend. Students' selected goals closely tracked system recommendations, indicating that the platform's calibration mechanisms effectively guided goal-setting behavior.

Motivational factors played a more limited role than expected. Intrinsic motivation showed a modest positive association with intervention gains, which is notable given the presence of extrinsic rewards and prior literature suggesting that such incentives can, under some conditions, crowd out intrinsic motivation (Bénabou & Tirole, 2003). Instead, results suggest that intrinsic motivation enhanced the effect of goal support. Additional analyses ruled out a simple selection effect whereby more intrinsically motivated students engaged more frequently in goal-setting interactions. One potential explanation is that intrinsically motivated students are more responsive to externally provided goal structures. In this account, intrinsic motivation does not simply predict baseline engagement, but increases the likelihood that students act on feedback and follow through on calibrated targets. This suggests that intrinsic motivation moderates responsiveness to goal support, extending prior work that has largely treated intrinsic and extrinsic influences as additive by showing that intrinsic motivation can shape how effectively students translate externally structured goals into sustained effort (Cerasoli et al., 2014).

In contrast, process features of the intervention were more strongly associated with outcomes. In particular, the use of implementation intentions was linked to higher levels of practice, with effects concentrated among lower-effort students. For students with higher baseline effort, implementation intentions did not produce additional gains. One explanation for this pattern is the tradeoff imposed by limited instructional time. Implementation intentions require time within tutoring sessions, reducing time available for direct mathematics practice. For lower-effort students, this cost is offset by gains in goal-directed behavior, as the scaffold helps initiate and sustain effort that would otherwise not occur (Gollwitzer, 1999; Gollwitzer & Sheeran, 2006). For higher-effort students, however, who are already consistently engaging, the same time investment yields little additional benefit and may even displace productive practice. This highlights an important design implication: behavioral scaffolds are not universally beneficial, but are most effective when targeted to students who lack established effort routines. More broadly, the results introduce a constraint often overlooked in goal-setting research, namely, that scaffolding competes with core instructional time, suggesting that the effectiveness of such supports depends not only on their behavioral impact, but also on their opportunity cost in real classroom settings.

The qualitative implementation-intention responses further clarify what kind of self-regulatory support students appeared to need. Students most often selected strategies focused on managing the immediate learning environment, followed by help-seeking from tutors, teachers, or the software, whereas purely internal self-management strategies (e.g., taking breaks) were less common. This pattern suggests that students perceived obstacles to persistence in their online math practice as primarily environmental or external. One interpretation is developmental, namely that middle-school students may still be developing the regulatory capacity needed to independently manage distraction, initiate work, and persist through difficulty. Another interpretation is contextual, namely because tutors, teachers, and software supports were highly available in this hybrid setting, students may have reasonably treated help-seeking as a central goal-attainment strategy. In either case, implementation intentions may have been beneficial because they helped students actively regulate environmental factors such as distractions and access to support. This interpretation is consistent with the stronger effects observed among lower-effort students, who may have entered the intervention with fewer established routines for managing their learning environment or initiating practice independently. For these students, implementation intentions may have complemented the broader goal-support intervention by making regulatory behaviors more explicit and actionable, whereas higher-effort students may have already possessed effective strategies for managing environmental or help-seeking challenges.

Exploratory analyses further indicate that increases in math self-efficacy are positively associated with intervention benefit. While causal direction cannot be established due to limited measurement points, this pattern is consistent with prior work suggesting a reciprocal relationship between performance and self-efficacy, potentially mediated by reductions in procrastination (Wäschle et al., 2014). Future work could examine this mechanism more directly, for example by analyzing delayed start behavior in response to goal support (Gurung et al., 2025).

Overall, the findings support the generalizability of adaptive goal support, highlight the importance of structured goal-setting processes, and suggest that behavioral scaffolds may be more influential than baseline motivational traits in driving student engagement. They further point to the role of externally supported regulation and process-level scaffolding as key components of effort-aware adaptivity in educational technology, particularly under realistic classroom constraints.

6.5.2 Theoretical and practical implications for the thesis

This chapter advances the thesis by showing that adaptive goal support remains effective under realistic implementation constraints, while also revealing important boundary conditions. A central finding is that goal-support mechanisms operate within a constrained instructional environment, where time spent on scaffolding competes directly with time available for practice. This extends the theoretical framing of adaptive goal support by highlighting that effectiveness depends not only on behavioral impact, but also on the opportunity cost of implementation.

This tradeoff is most evident in the implementation intention results: scaffolding improved outcomes for lower-effort students but did not benefit higher-effort students, for whom the additional time may displace productive work. This suggests that goal-support features should be targeted rather than uniformly applied, with more intensive scaffolding reserved for students who are less likely to initiate or sustain effort independently.

A second implication concerns the design of goal recommendations. In the present implementation, recommendations were based primarily on recent behavior, and selected goals closely tracked these recommendations. This is surprising, as prior work shows that students want to retain final control over their goals (Borchers, Peng, et al., 2025). Despite this, their selected goals largely follow system recommendations, indicating a tension between preference for control and reliance on guidance. This raises the question of whether preserving autonomy is required for effectiveness, or whether similar outcomes would emerge under more prescriptive recommendation schemes. Practically, this makes recommendation construction a central design component of adaptive goal support, as small changes in how recommendations are generated can directly affect student behavior. An open question is whether incorporating additional information—such as anticipated task difficulty or the consistency of prior effort—would improve calibration. While such extensions have been proposed in recent work on effort forecasting (Qiu, Thomas, Guo, Aleven, & Borchers, 2026), it is unclear whether increased model complexity would translate into better outcomes or is necessary in practice.

Motivational measures showed limited explanatory power for variation in intervention benefit. Most constructs were not significantly associated with G, and overall model fit was low, indicating that baseline self-reports account for little variance in outcomes. While some heterogeneity in responsiveness is present by construction of the student-level effects, the current analyses do not identify clear, actionable predictors of who benefits more. This points instead to the role of implementation context. Variation across schools highlights the importance of implementation fidelity: one participating school did not show descriptive improvements following the introduction of goal support, in contrast to the overall pattern of results. Given that this same school showed positive effects in prior implementations, this difference is unlikely to reflect structural limitations of the intervention, and more likely reflects differences in how consistently goal-support routines and reward structures were enacted. Informal observations indicate variability in teacher follow-through compared to other participating schools, particularly in reward distribution, which may attenuate intervention effects. Together, this suggests that variation in outcomes is driven less by measured student characteristics and more by differences in how the intervention is enacted in practice. In this setting, consistent implementation of goal-support routines and reward structures appears to be a prerequisite for observing effects, implying that fidelity constraints may dominate individual differences in determining intervention impact.

6.5.3 Limitations

This study should be interpreted in light of several limitations. Motivational measures are based on two survey waves, limiting the ability to capture fine-grained changes over time and introducing potential measurement noise, particularly for single-item constructs. Behavioral outcomes are derived from platform log data, which capture observable practice but not off-platform activity or variation in task difficulty. Finally, implementation fidelity likely varied across sites, including differences in tutor practices and teacher reward distribution, which likely contributed to heterogeneity in observed effects. These factors reflect the realities of deploying interventions at scale and should be considered when interpreting the magnitude and generalizability of the results. An additional limitation is that the study did not assess students' perceptions of the usefulness of different implementation-intention strategies or compare the effectiveness of specific strategy categories. Future work could combine qualitative strategy data with student ratings and behavioral outcomes to identify which forms of environmental management, help-seeking, or self-management support are most effective for different learners.

Chapter 7

General Discussion

A central problem motivating this dissertation is that adaptive educational technologies have concentrated primarily on optimizing instruction and metacognitive support during learning activities while devoting comparatively little attention to the conditions that determine whether sustained practice occurs at all. Intelligent tutoring systems can now personalize problem difficulty, pacing, feedback, and sequencing with considerable sophistication (Aleven, McLaughlin, Glenn, & Koedinger, 2025). Even so, the effectiveness of these systems still depends on students repeatedly engaging in effortful practice across days and weeks. In authentic classroom settings, the benefits of adaptive instruction are constrained not only by the quality of cognitive personalization but also by students' willingness and ability to persist in consistent practice. This dissertation argues that educational AI systems should adapt not only to learners' knowledge states but also to the patterns of effort regulation that shape participation before, between, and after instructional interactions. The dissertation studies, accordingly, outline how such *effort-aware* adaptivity in classroom AI might be designed.

7.1 Summary of Results

The studies presented across this dissertation examine adaptive goal support as one possible form of effort-aware adaptivity. The interventions investigated here attempted to regulate practice behavior itself by shaping when students practiced, how consistently they returned to practice, and whether goals remained appropriately calibrated over time. Goal support was therefore conceptualized as part of a broader feedback process involving target setting, progress monitoring, accountability, adjustment, and repeated opportunities for attainable success. Across the dissertation, these components were examined through a sequence of field studies that moved from early feasibility questions to large-scale deployment within authentic hybrid tutoring environments. The thesis first investigated whether low-friction accountability structures could increase participation in supplemental mathematics practice. It then evaluated whether goal contracts, contingent rewards, and tutor monitoring could improve both practice and proficiency outcomes. Later studies examined whether adaptive calibration produced different effects than static goal assignment and whether these patterns generalized across schools, tutors, and instructional platforms under operational conditions.

Chapter 3 established the feasibility of low-friction accountability structures through caregiver-supported goal contracts. The findings demonstrated that motivational coordination around practice could be implemented much more successfully than instructional tutoring provided directly by caregivers. These results suggested that scalable support for student effort may depend less on expanding instructional labor within households and more on creating lightweight structures that encourage accountability and persistence.

Building on this foundation, Chapter 4 examined whether goal contracts combined with contingent rewards and lightweight tutor monitoring could improve student outcomes within hybrid tutoring environments. The intervention produced meaningful increases in both weekly practice and mathematics performance, demonstrating that goal support could operate effectively within realistic instructional settings without requiring intensive teacher oversight. Improvements in skill outcomes exceeded corresponding increases in practice time, which raised the possibility that the intervention influenced not only the quantity of practice but also the quality and productivity of student engagement.

Chapter 5 then investigated whether adaptive calibration produced different outcomes than static goal assignment. Although both forms of goal support increased practice time relative to no-goal conditions, only the adaptive goal condition showed a statistically reliable improvement in proficiency relative to the no-goal baseline. Although the direct comparison between the two goal-support conditions was not significant, adaptive goals were associated with substantially higher rates of goal attainment. The results also revealed strong heterogeneity by baseline effort. Students with initially low levels of participation benefited most from adaptive support, whereas students who already practiced consistently often experienced limited gains. These findings suggested that adaptive goal support functions less as a universal motivational enhancement and more as an effort-regulation scaffold that becomes most valuable when self-regulatory processes are weakest. The chapter also introduced evidence for a momentum process in which successful goal attainment increased the likelihood of future success, with descriptively stronger carryover under adaptive calibration than under externally fixed targets.

Chapter 6 examined whether these patterns generalized under authentic implementation conditions across multiple schools, instructional platforms, and tutoring contexts. The deployment replicated earlier improvements in practice and proficiency outcomes while also revealing the importance of implementation fidelity, session constraints, and contextual variation across sites. Researcher-led facilitation was replaced with tutor-mediated dashboards and teacher-managed reward systems, allowing the intervention to operate within routine program structures instead of tightly controlled experimental conditions. Cross-site variation demonstrated that the effectiveness of goal support depended heavily on how consistently accountability structures, progress visibility, and rewards were enacted in practice. These findings positioned implementation as a central explanatory factor within the dissertation's broader theory of change.

The progression across studies reframes adaptive goal support as a problem of orchestration involving analytics, accountability, calibration, and classroom implementation within the practical constraints of limited instructional time. The remainder of this chapter synthesizes these findings, revises the dissertation's theory of change, and situates effort-aware adaptivity within broader theories of self-regulation, motivation, scaffolding, and human-AI coordination in classroom learning environments.

7.1.1 Toward an Effort-Aware Model of Classroom Resources

The empirical results across the dissertation suggest that adaptive goal support influenced student learning through multiple pathways related to effort regulation and classroom time use. The interventions consistently increased mathematics practice and proficiency outcomes across studies, yet the mechanisms underlying these changes require further interpretation. One possible explanation is that the interventions increased the total amount of time students were willing to spend practicing math during instructional periods with limited available time, for example, by initiating effort earlier. A second possibility is that students used existing instructional time more productively once effort became structured through goals, feedback, and accountability. A third possibility is that effort scaffolding itself competed with instructional time and therefore produced benefits only when the value of the scaffold exceeded its opportunity cost. Together, these findings point toward a broader theoretical model in which classroom learning depends not only on instructional quality, but also on how effort and time are allocated within constrained educational environments. In the following, each of these interpretations is explored in more detail to derive a process model that can (a) generate hypotheses for future work and (b) guide better future effort-aware adaptivity in classrooms using educational technologies.

Use of classroom time. One interpretation supported by the dissertation findings is that many students do not fully utilize the instructional time already available to them (assuming, as in the context of this dissertation, that the primary usage of educational technologies occurs during in-class seatwork). Students may begin work slowly, disengage before sessions end, or spend substantial portions of class time idle despite nominal participation. Emerging evidence on classroom "coasting" behavior suggests that learners frequently conserve effort even during allocated learning periods (Gurung et al., 2026). Several results from this dissertation align with that interpretation. In particular, larger intervention gains appeared among students with the lowest prior levels of class time usage, especially in Chapter 5. A conservative interpretation of these findings is that low-effort students simply had greater unused instructional capacity available. Students who already practiced consistently may have approached practical ceilings imposed by fixed tutoring periods, which limited the possibility of further gains. This explanation is also consistent with the final deployment study, which found relatively weak evidence that stable motivational traits such as intrinsic motivation accounted for heterogeneity in intervention effects. Taken together, baseline effort predicted intervention responsiveness more consistently than dispositional motivational measures, suggesting that existing patterns of participation may have constrained improvement most.

Increase in per-minute effort. At the same time, the results also suggest that the intervention influenced how students used time that had already been allocated for practice. All three quasi-experimental intervention studies produced disproportionately large improvements in skill mastery relative to corresponding increases in logged practice time. This pattern implies that effort cannot be understood solely through duration measures such as minutes on task. Students may vary substantially in how intensely, efficiently, or consistently they engage during the same amount of time. Prior work has shown that behavioral indicators such as problems completed can predict mathematics achievement more strongly than raw time-on-task measures (Ritter, Joshi, Fancsali, & Nixon, 2013). Time itself is also an imperfect proxy for activity because logged minutes often fail to distinguish active engagement from inactivity or distraction (Kovanovic, Gašević, Dawson, Joksimovic, & Baker, 2015). The dissertation findings are consistent with a view of effort as a latent behavioral construct that includes persistence, sustained attention, and productive engagement during practice opportunities (e.g., via self-explanation). Emerging work suggests that these latent dimensions can be inferred from process data and may predict both learning outcomes and future participation in terms of the number of minutes practiced, skills mastered, but also student learning rates (Borchers, Zhang, Yang, Nagashima, & Domingue, 2026; Qiu et al., 2026). The strong skill gains observed across several interventions therefore may reflect improvements in productive effort per minute in addition to increases in total practice time.

Effort scaffolding costs and benefits. A third theoretical implication concerns the cost of scaffolding itself within constrained instructional environments. The final deployment study demonstrated that additional effort supports did not uniformly improve outcomes across students. Implementation intentions produced benefits primarily for students with low prior effort and in some cases appeared to reduce practice opportunities for students who already engaged consistently. These findings suggest that effort scaffolding competes directly with mathematics practice for limited classroom time. In short tutoring sessions, even brief goal-setting routines consume instructional minutes that could otherwise be devoted to academic work. The effectiveness of a scaffold therefore depends not only on whether it improves regulation, but also on whether the improvement justifies the time required to implement it. This interpretation positions instructional time as a central constraint within effort-aware educational systems.

These findings motivate a broader theoretical model centered on the balance between under-scaffolding and over-scaffolding of self-regulation goals in classroom learning environments. Students with weak self-regulation may require substantial external structure to initiate and sustain practice, whereas highly engaged students may experience diminishing returns when additional scaffolds consume time without meaningfully improving effort. The interventions examined in this dissertation adapted primarily to prior levels of effort through historical platform data and goal calibration. They did not dynamically adapt to evolving classroom constraints, instructional pacing, or opportunity costs associated with scaffolding during live instruction. Future effort-aware systems may therefore require forms of adaptivity that account not only for prior participation patterns, but also for how instructional time is being used moment by moment within classroom environments. All three theoretical components are summarized in Figure 7.1.

7.1.2 Interpreting Mechanisms in Adaptive Goal Support

A discussion-worthy limitation across the dissertation studies is that the interventions bundled together multiple components of adaptive goal support rather than systematically isolating them experimentally. Although Chapter 5 partially isolated adaptive goal feedback and calibration effects, and Chapter 6 experimentally varied implementation intentions, most interventions combined several elements simultaneously, including goal targets, progress feedback, adaptive calibration, accountability structures, implementation intentions, and contingent rewards. As a result, the present findings cannot definitively determine which individual components were necessary or sufficient for producing the observed improvements in practice time and proficiency outcomes. The dissertation therefore provides stronger evidence for the effectiveness of the broader intervention bundle, paving the way for ablation studies in future work.

At the same time, the studies and the revised effort-aware adaptivity model presented in Figure 7.1 still provide important clues about how adaptive goal support may operate in classroom settings. Differences between static and adaptive goal conditions, variation in implementation

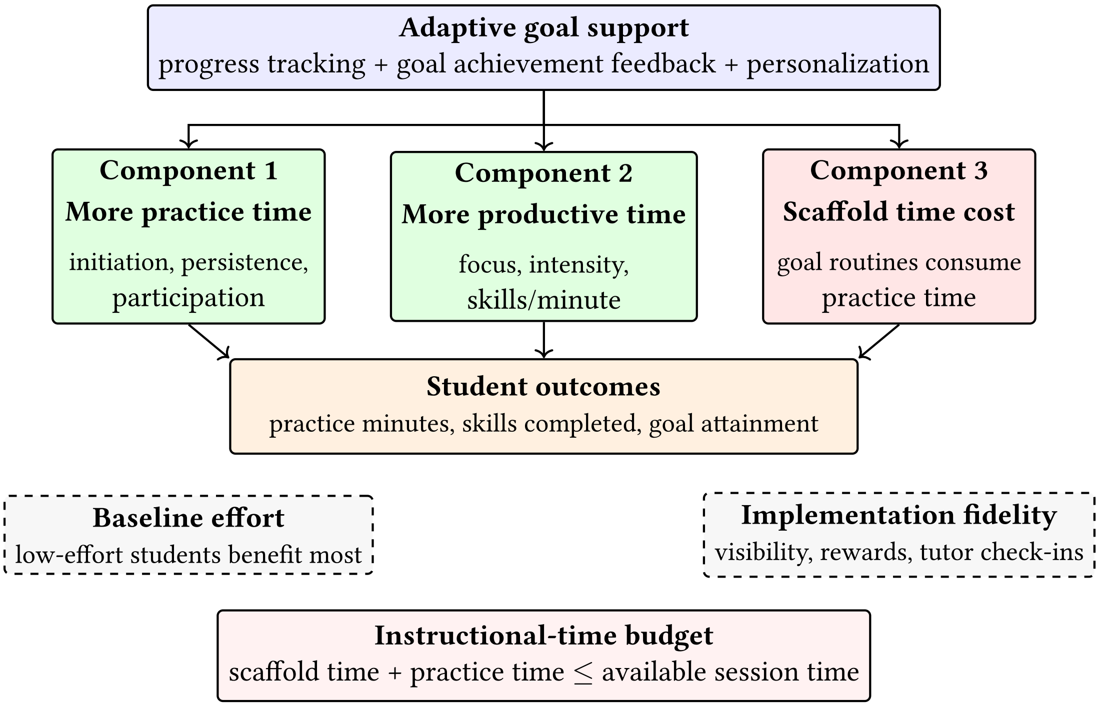


Figure 7.1: Final three-component theory of change for effort-aware adaptive goal support. Adaptive goal support may improve student outcomes by increasing practice time, increasing productive effort per minute, and balancing these gains against the opportunity cost of scaffolding within a fixed instructional-time budget.

fidelity across sites, and qualitative reports from tutors all suggest that some components likely contributed more strongly than others under particular classroom conditions. The revised theoretical model in Figure 7.1 provides a provisional process account for organizing these observations and generating future hypotheses.

In particular, the findings motivate more careful theorizing about which intervention components primarily influence whether students initiate practice at all, which components influence the intensity and productivity of practice once engagement has begun, and which components justify their instructional-time cost within constrained classroom environments. Some forms of support may function primarily as effort-initiation scaffolds, whereas others may improve focus, persistence, or productive engagement during practice. Still other components may derive their effectiveness less from direct motivational effects and more from increasing visibility, coordination, accountability, or social reinforcement during implementation. The following sections synthesize empirical findings, tutor observations, and exploratory analyses across the dissertation to develop provisional mechanistic interpretations of adaptive goal support. These interpretations should not be understood as definitive causal claims. Instead, they are intended to generate more precise hypotheses and stronger experimental designs for future work on effort-aware educational AI systems, as discussed next.

7.1.2.1 Hypothesized Mechanisms and Directions for Future Research

One notable pattern across the dissertation studies was that adaptive goal support produced the largest gains among students with initially low levels of effort and participation. Students who already practiced consistently often experienced smaller improvements. One possible explanation is that these students approached practical ceilings in available practice time, leaving limited room for additional gains. Another possibility is that highly engaged students required qualitatively different forms of support than those emphasized in the intervention bundle. Externally structured goals, rewards, and accountability systems may become less effective when students already possess stable engagement habits or internally endorsed reasons for participation. In some cases, highly visible performance structures may even shift attention away from meaningful engagement and toward external evaluation or reward attainment (Cerasoli et al., 2014; Covington & Müeller, 2001; Deci et al., 1999; Patall et al., 2008). These findings point toward the importance of differentiating between students who require support for effort initiation (e.g., through lacking regulation of environmental factors such as in-class distractions, as revealed by qualitative analysis of implementation intentions) and students who require support for sustaining, organizing, or deepening already existing engagement patterns.

The findings also raise questions about the role of attainable success experiences in maintaining persistence over time. Across several studies, students who achieved weekly goals were substantially more likely to achieve future goals, with descriptively stronger carryover under adaptive calibration conditions in which targets could be adjusted in response to prior performance. One interpretation is that adaptive calibration increased the frequency of attainable mastery experiences by moving goals closer to students' actual performance ranges, especially since student compliance with goal recommendations was generally strong. This process may have strengthened perceived competence and increased willingness to continue engaging with practice activities in subsequent weeks. Earlier analyses in Chapter 5 suggested that calibration improved goal attainment without requiring proportionally large increases in total practice time. The intervention may therefore have influenced not only how much students practiced, but also how appropriately goals matched students' evolving capabilities, which in turn predicted greater goal achievement rate and intervention benefit. Although goal compliance was high, students were able to or were prompted to independently set goals, such that their success experiences may carry greater motivational weight when students perceive goals as self-endorsed instead of externally imposed (Howard et al., 2021; Koestner et al., 2008).

The implementation-intention findings provide a related perspective on how effort regulation may operate in classroom environments. The intervention appeared to help students translate abstract weekly goals into concrete actions linked to specific contexts and routines, particularly among students with lower baseline effort. Qualitative responses revealed that students frequently focused on environmental and social regulation strategies, including finding quieter workspaces, reducing distractions, wearing headphones, or asking tutors and peers for assistance. These patterns suggest that effort regulation in classroom settings depends heavily on the management of attention, environment, and social coordination. Students did not simply describe needing more motivation. Many instead described needing help structuring the conditions under which productive work could occur. This observation has important implications for future effort-aware systems. Adaptive educational technologies may benefit from supporting contextual regulation processes alongside numeric goal calibration and progress tracking. Features related to distraction management, help-seeking, environmental planning, and social accountability may therefore represent important extensions of existing goal-support systems.

The dissertation findings ultimately support a provisional process account in which adaptive goal support operates through a recurring cycle of target setting, progress visibility, accountability, calibrated success, rewards, and renewed participation. The intervention appears most effective when baseline effort is low and especially when implementation support is provided to students with low prior effort. The studies also suggest that intervention effects emerge through multiple pathways simultaneously. Some gains likely resulted from increases in total practice time, especially among students who previously underutilized available instructional opportunities. Other gains may have resulted from improvements in the quality and intensity of engagement during already allocated practice time. Future research should therefore move beyond asking whether adaptive goal support improves outcomes in general. More important questions concern which components influence practice initiation, which improve productive engagement during practice, and which forms of support justify their instructional-time cost within constrained educational environments. These distinctions are likely to become increasingly important as educational AI systems attempt to personalize support for effort regulation in addition to knowledge acquisition.

7.1.3 Adaptive Goal Support in Relation to Prior Theories of Motivation and Goal Setting

The findings across this dissertation align most directly with classical goal-setting theory, particularly accounts emphasizing the motivational effects of specific, proximal goals paired with performance feedback (Locke & Latham, 2002). Across studies, students who participated in adaptive goal support consistently increased mathematics practice relative to comparison conditions, and these improvements were accompanied by corresponding gains in proficiency outcomes. The results therefore provide additional evidence that structured goals can organize attention, sustain effort, and increase persistence in educational settings. However, it should be noted that intervention benefits may also be attributable to the prospect of goal achievement rewards more than the intrinsic value of set goals themselves, subject to future work. At the same time, the dissertation extends prior goal-setting work by situating goals within a broader adaptive feedback process involving calibration, accountability, implementation support, and repeated adjustment over time. In this sense, the interventions examined here were not static goal manipulations alone. They functioned as ongoing coordination systems intended to regulate practice behavior across weeks of classroom participation.

The dissertation findings also connect closely to self-determination theory, especially perspectives emphasizing competence and autonomy support (Ryan & Deci, 2000). Adaptive calibration appeared to increase rates of goal attainment by moving targets closer to students' actual performance ranges. This process likely increased the frequency of attainable success experiences, which may have strengthened perceived competence and encouraged continued engagement. Momentum effects observed across studies further support this interpretation. Students who successfully achieved goals in one week became substantially more likely to achieve subsequent goals, particularly under adaptive self-set conditions. The findings therefore suggest that calibration may matter not only because it improves prediction accuracy, but also because it shapes how students interpret success and failure over time. At the same time, the generally high rates of recommendation compliance complicate simple distinctions between externally imposed and autonomous goals. Many students accepted recommended goals while still perceiving themselves as active participants in the goal-setting process (as supported by design research findings on the goal setting mechanism studied in this research; see (Borchers, Peng, et al., 2025)). This pattern suggests that autonomy in classroom AI systems may depend less on unrestricted independence and more on whether students experience recommendations as reasonable, attainable, and open to negotiation.

The observed heterogeneity in intervention effects also connects to scaffolding and expertise-reversal perspectives (Kalyuga, 2009). Students with initially low levels of effort consistently benefited most from adaptive support, whereas students who already practiced consistently often showed weaker effects. This pattern suggests that motivational scaffolds may follow boundary conditions similar to cognitive scaffolds. Supports that are highly beneficial for learners with weak self-regulatory routines may become redundant or inefficient for students who already possess stable engagement habits. The implementation-intention findings reinforce this interpretation. Additional scaffolds improved outcomes primarily among lower-effort students and occasionally appeared to consume valuable instructional time for students who already engaged productively. The dissertation therefore extends expertise-reversal arguments beyond cognitive instruction into the domain of effort regulation and classroom participation.

Finally, the dissertation contributes to emerging perspectives on human-AI complementarity in educational systems (Holstein et al., 2020). Across studies, adaptive analytics alone were insufficient to produce consistent intervention effects. Outcomes depended on tutor implementation and teacher reward follow-through (as evidenced by a lack of intervention benefit in the school with one teacher exhibiting inconsistent reward practices). The progression across studies moved from researcher-led facilitation toward increasingly distributed systems involving dashboards, tutors, teachers, and platform analytics operating together within authentic classroom environments. These findings position effort-aware educational AI as a coordination problem involving both technological and human infrastructure. The central challenge is embedding those recommendations within classroom routines that sustain visibility, accountability, and follow-through over time. A related direction for future research is envisioning the goal support structure studied here without human tutors but instead orchestrated by teachers.

7.2 Future Directions for Effort-Aware Educational AI

The findings across this dissertation point toward a broader research agenda centered on effort-aware adaptivity as a complementary lens to existing forms of instructional personalization. Although the interventions examined here demonstrated that adaptive goal support can improve practice and proficiency outcomes under authentic classroom conditions, the studies also revealed important limitations in current models of effort regulation, measurement, and implementation. Future work should therefore focus not only on refining intervention components, but also on developing more precise theories and computational representations of how effort evolves over time within constrained educational environments.

One important direction concerns the development of more accurate predictive models for goal calibration and intervention targeting. The recommendation systems examined in this dissertation relied primarily on prior practice behavior and relatively simple historical trends. Future systems may benefit from richer predictive models that incorporate trajectories of goal attainment, recent failure streaks, engagement volatility, help-seeking behavior, idle time, assignment completion patterns, and contextual classroom variables (Qiu et al., 2026). Such models could support more dynamic forms of calibration that adjust not only target difficulty, goal guardrails (e.g., min and max values) but also scaffold intensity, feedback cadence, and accountability structures in response to evolving student behavior. More sophisticated prediction systems may also help distinguish students who require effort initiation support from students who would benefit more from organizational, environmental, or planning-oriented scaffolds.

A related challenge concerns the measurement of effort itself. Across much of the educational technology literature, effort is still operationalized primarily through coarse indicators such as minutes on task or total activity counts. These measures might fail to capture substantial variation in engagement quality, persistence, attention, and productive participation (Kovanovic et al., 2015). Future effort-aware systems will likely require richer latent models of engagement derived from process data, including effort initiation delay (Gurung et al., 2026, 2025) and student response times when faced with difficult tasks (Borchers, Zhang, et al., 2026). Productive effort may ultimately prove more important than duration alone for understanding both learning gains and future participation patterns. The development of better latent effort measures and their correlation with long-term learning outcomes (e.g., test scores) therefore represents a central direction for future research in effort-aware AI.

Another promising direction concerns the use of generative AI agents to deliver adaptive goal support at scale. The interventions studied in this dissertation relied on tutors and teachers to facilitate goal setting, monitor progress, provide accountability, and encourage continued participation. While human support likely contributed to intervention effectiveness, it also introduced implementation variability across tutors, classrooms, and schools. Recent work suggests that conversational AI systems may be capable of supporting several components of self-regulated learning, particularly by helping students identify learning resources, enact learning strategies, monitor progress, and engage in self-reflection through adaptive dialogue and personalized feedback (Chiu, 2024; Guan, Raković, Chen, & Gašević, 2025). Notably, however, a recent systematic review found that educational chatbots have provided comparatively limited support for goal setting, planning, and adaptation despite their central role in self-regulated learning (Guan et al., 2025). This gap suggests a promising opportunity for generative AI systems to extend adaptive goal-support interventions by providing scalable assistance with data-driven goal calibration, progress monitoring, and implementation planning. Such systems could potentially provide more scalable and consistent goal-support experiences while reducing the personnel demands associated with human-mediated implementations. At the same time, it remains unclear whether generative AI can replicate the social accountability and relational support that human tutors provide. Future research should therefore investigate which aspects of adaptive goal support can be effectively automated through generative AI and which continue to benefit from human involvement. More broadly, the most effective future systems may involve human-AI partnerships in which generative agents provide scalable monitoring and personalization while educators contribute social accountability and contextual understanding.

Future work should also investigate how motivational states and expectations evolve through repeated cycles of goal success and failure. The present studies suggest that attainable success experiences may strengthen persistence over time, with descriptively stronger effects under adaptive calibration conditions. However, the mechanisms underlying these momentum effects remain poorly understood. More temporally sensitive research designs (with more frequent motivational survey assessments) could examine how expectancy judgments, perceived competence, frustration, or confidence shift following specific goal-achievement events. Adaptive systems may eventually benefit from estimating motivational states continuously through behavioral traces instead of relying exclusively on infrequent self-report surveys. This direction aligns with broader efforts in learning analytics to infer latent psychological processes from interaction data while minimizing measurement burden in classroom settings (Borchers, Deininger, & Pardos, 2026).

7.3 Contributions and Significance

This dissertation contributes to research on educational AI, motivation, and self-regulated learning by positioning effort regulation as a meaningful target of adaptive educational support. Existing adaptive learning systems have concentrated primarily on modeling knowledge states and optimizing instructional support during active learning interactions. The studies presented here argue that these systems must also account for whether students engage in sustained practice at all, how consistently they participate across time, and how productively they use available instructional opportunities. Across multiple field studies, the dissertation demonstrates that adaptive goal support can improve both mathematics practice and proficiency outcomes within authentic hybrid tutoring environments, particularly for students with initially low levels of participation. The findings additionally extend existing theories of scaffolding and expertise reversal into the domain of effort regulation by showing that motivational supports exhibit important boundary conditions tied to baseline engagement and instructional context. More broadly, the dissertation advances a theoretical model in which classroom learning depends on the interaction between adaptive support and constrained instructional time.

The dissertation also contributes a scalable design and implementation framework for adaptive goal support in real classroom settings. The interventions integrated analytics dashboards, tutors, teachers, rewards, and multi-platform practice data into a coordinated workflow capable of operating across schools and instructional environments. The findings suggest that accountability and coordination structures can produce meaningful improvements in participation without requiring intensive instructional involvement from caregivers or teachers. Several practical design principles additionally emerged across studies, including the importance of guardrailed goal recommendations, student-visible progress tracking, environmental and implementation-intention scaffolds, and targeted scaffold allocation for students with low prior effort.

These contributions matter because they address a central limitation facing many adaptive learning systems at scale: students cannot benefit from personalized instruction when they disengage from practice opportunities altogether. The dissertation therefore reframes human-AI educational systems as coordination infrastructures that support persistence, participation, and productive engagement in addition to cognitive learning. This perspective has particular importance for students with historically low levels of practice participation, who demonstrated the largest gains from adaptive goal support across studies. Future educational AI systems may ultimately need to adapt simultaneously to what students know, how they learn, and whether they remain willing and able to engage in sustained effort over time.

7.4 Limitations

Several limitations qualify the conclusions that can be drawn from this dissertation. Most importantly, the studies provide stronger evidence for the effectiveness of adaptive goal support as an implemented bundle than for the causal contribution of each individual component. Goal targets, feedback, calibration, accountability, implementation support, and rewards were often combined in authentic program settings, which limits claims about necessary and sufficient mechanisms. In addition, the evidence comes from field implementations in which platform logs, classroom routines, tutor practices, teacher reward follow-through, and student attendance all shaped what the intervention became in practice. These conditions strengthen ecological validity, but they also introduce variation that is difficult to fully control. The outcome measures also relied heavily on platform-generated indicators such as minutes practiced and skills completed, which are useful for studying participation at scale but cannot fully capture long-term, off-platform learning (e.g., measured in state test scores). Finally, the observation windows were limited relative to the broader developmental goals of self-regulated learning. The studies show that adaptive goal support can improve practice and proficiency outcomes in the short and medium term, but they do not yet determine whether students internalize calibration routines, sustain effort after rewards are removed, or transfer these regulatory skills to other learning contexts.

7.5 Conclusion

This dissertation investigated adaptive goal support as a mechanism for effort-aware educational AI within authentic middle school mathematics hybrid tutoring environments. Across multiple field studies, the findings demonstrated that personalized goals, calibrated feedback, accountability structures, and contingent rewards can increase both mathematics practice and proficiency outcomes under realistic classroom conditions. The strongest and most consistent benefits emerged among students with initially low levels of participation, suggesting that adaptive goal support functions most effectively as an effort-regulation scaffold for students who have not yet developed stable routines for sustained engagement. The studies also showed that intervention effectiveness depended on implementation conditions, including reward consistency, and the practical constraints imposed by limited instructional time.

Beyond the specific interventions examined here, the dissertation advances a broader argument about the future direction of educational AI. Existing adaptive learning systems have focused predominantly on modeling cognition during active learning interactions. The findings presented throughout this dissertation suggest that these systems must also account for the behavioral and organizational processes that determine whether meaningful engagement occurs consistently over time. Students cannot benefit from personalized instruction if they disengage from practice opportunities, use instructional time inefficiently, or fail to sustain participation across weeks of learning. Effort regulation therefore represents an important and underdeveloped dimension of educational adaptivity. The dissertation proposes that future systems should adapt to patterns of participation, persistence, calibration, and productive engagement across instructional contexts.

The dissertation additionally contributes a theoretical perspective in which classroom learning depends on the interaction between adaptive support and constrained instructional resources.

The findings suggest that effort-aware interventions can improve outcomes through multiple pathways, including increasing the amount of time students spend practicing and improving the productivity of engagement during already allocated instructional time. At the same time, the studies showed that scaffolds themselves consume valuable classroom resources. Goal-setting routines, implementation planning, and accountability conversations require time that could otherwise be devoted to academic work. Effective effort-aware systems must therefore balance the benefits of scaffolding against the instructional costs required to implement it. This tension emerged repeatedly across the dissertation and motivated the proposed model of classroom effort regulation centered on practice initiation, productive engagement, and scaffold opportunity costs.

The findings also position effort-aware educational AI as a problem of coordination involving both technological and human infrastructure. Coordination among analytics, researchers, teachers, and tutors produced goal support benefits that early caregiver pilot studies, which showed limited participation across households, could not sustain at scale (Chapter 3). The progression across the dissertation moved from tightly researcher-supported implementations toward increasingly distributed systems operating across schools and tutoring sites. Future educational AI systems without dedicated human tutoring and researcher support will require stronger integration into the practical realities of classroom instruction and human support systems.

References

Adams, M. A., Hurley, J. C., Todd, M., Bhuiyan, N., Jarrett, C. L., Tucker, W. J., ... Angadi, S. S. (2017). Adaptive goal setting and financial incentives: a 2× 2 factorial randomized controlled trial to increase adults' physical activity. *BMC public health*, *17*, 1–16.

Aleven, V., Mclaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, *16*(2), 101–128.

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2016). Instruction based on adaptive learning technologies. *Handbook of research on learning and instruction*, *2*, 522–560.

Aleven, V., McLaughlin, E. A., Glenn, R. A., & Koedinger, K. R. (2025). Instruction based on adaptive learning technologies. In R. E. Mayer, P. A. Alexander, & L. Fiorella (Eds.), *Handbook of research on learning and instruction* (3rd ed., pp. 477–518). New York, NY: Routledge.

Alwahbi, A. (2020). The use of contingency contracting in educational settings: A review of the literature. *Educational Research and Reviews*, *15*(6), 327–335.

Arroyo, I., Muldner, K., Schultz, S. E., Burleson, W., Wixon, N., & Woolf, B. P. (2016). Addressing affective states with empathy and growth mindset. In *Umap (extended proceedings)*.

Asher, M. W., & Harackiewicz, J. M. (2024). Using choice and utility value to promote interest: Stimulating situational interest in a lesson and fostering the development of interest in statistics. *Journal of Educational Psychology*, *117*(4), 647–662. doi: 10.1037/edu0000921

Aurino, E., Tsinigo, E., & Wolf, S. (2022). *Nudges to improve learning and gender parity: preliminary findings on supporting parent-child educational engagement during covid-19 using mobile phones* (Tech. Rep.). EdTech Hub.

Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., ... others (2022). Lessons learned and future directions of metatutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology*, *13*, 813632.

Azevedo, R., Witherspoon, A. M., Chauncey, A., Burkett, C., & Fike, A. (2009). Metatutor: A metacognitive tool for enhancing self-regulated learning. In *Aaai fall symposium: Cognitive and metacognitive educational systems* (pp. 14–19).

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research*, *19*(2), 185–224.

Baker, R. S. (2007). Modeling and understanding students' off-task behavior in intelligent tutoring systems. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 1059–1068).

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of memory and language*, *68*(3), 255–278.

Bashkov, B. M. (2021). Assessing the impact of ixl math over three years: A quasi-experimental study. *ESSA Research Report. Online Submission.*

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of statistical software*, *67*, 1–48.

Bear, G. G., Slaughter, J. C., Mantz, L. S., & Farley-Ripple, E. (2017). Rewards, praise, and punitive consequences: Relations with intrinsic and extrinsic motivation. *Teaching and Teacher Education*, *65*, 10–20.

Bénabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *The review of economic studies*, *70*(3), 489–520.

Bolker, B. M. (2015). Linear and generalized linear mixed models. *Ecological statistics: contemporary theory and application*, 309–333.

Borchers, C., Deininger, H., & Pardos, Z. A. (2026). Toward trait-aware learning analytics. In *Proceedings of the lak26: 16th international learning analytics and knowledge conference* (pp. 250–260).

Borchers, C., Houk, A., Aleven, V., & Koedinger, K. R. (2025). Engagement and learning benefits of goal setting with rewards in human-ai tutoring. In *Proceedings of the 26th international conference on artificial intelligence in education.*

Borchers, C., Nguyen, H. T., Carvalho, P. F., Koedinger, K. R., & Aleven, V. (2025a). Goal setting engages more caregivers in online math homework than instructional support. In *European conference on technology enhanced learning* (pp. 61–75). Cham: Springer Nature Switzerland.

Borchers, C., Nguyen, H. T., Carvalho, P. F., Koedinger, K. R., & Aleven, V. (2025b). Involving parents in tutoring systems to increase content confidence: A design probe study. In *Proceedings of the 26th international conference on artificial intelligence in education (aied 2025).* Palermo, Italy.

Borchers, C., Ooge, J., Peng, C., & Aleven, V. (2025). How learner control and explainable learning analytics about skill mastery shape student desires to finish and avoid loss in tutored practice. In *Proceedings of the 15th international learning analytics and knowledge conference (LAK '25)* (pp. 810–816). Dublin, Ireland: Association for Computing Machinery. doi: 10.1145/3706468.3706484

Borchers, C., Peng, C., Lyu, Q., Carvalho, P. F., Koedinger, K. R., & Aleven, V. (2025). Student perceptions of adaptive goal setting recommendations: A design prototyping study. In *Proceedings of the 26th international conference on artificial intelligence in education (AIED '25)* (pp. 244–251). Cham: Springer.

Borchers, C., Zhang, L., Yang, K., Nagashima, T., & Domingue, B. W. (2026). Understanding student effort using response-time propensities during problem solving. In *Proceedings of the thirteenth ACM conference on learning @ scale (L@S '26).* Seoul, Republic of Korea: Association for Computing Machinery. doi: 10.1145/3774398.3811624

Brickell, T. A., & Chatzisarantis, N. L. (2007). Using self-determination theory to examine the motivational correlates and predictive utility of spontaneous exercise implementation intentions. *Psychology of Sport and exercise*, *8*(5), 758–770.

Bronfenbrenner, U. (2000). Ecological systems theory. In A. E. Kazdin (Ed.), *Encyclopedia of psychology* (Vol. 3, pp. 129–133). Washington, DC: American Psychological Association.

Bull, S. (2020). There are open learner models about! *IEEE Transactions on Learning Technologies*, *13*(2), 425–448.

Bureau, J. S., Howard, J. L., Chong, J. X., & Guay, F. (2022). Pathways to student motivation: A meta-analysis of antecedents of autonomous and controlled motivations. *Review of educational research*, *92*(1), 46–72.

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic motivation and extrinsic incentives jointly predict performance: a 40-year meta-analysis. *Psychological bulletin*, *140*(4), 980.

Chang, D. H., Lin, M. P.-C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using ai chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, *15*(17), 12921.

Chiu, T. K. (2024). A classification tool to foster self-regulated learning with generative artificial intelligence by applying self-determination theory: A case of chatgpt. *Educational technology research and development*, *72*(4), 2401–2416.

Cooper, J. O., Heron, T. E., & Heward, W. L. (2007). *Applied behavior analysis* (2nd ed.). Upper Saddle River, NJ: Pearson/Merrill-Prentice Hall.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, *13*(1), 3–21.

Covington, M. V., & Müeller, K. J. (2001). Intrinsic versus extrinsic motivation: An approach/avoidance reformulation. *Educational psychology review*, *13*(2), 157–176.

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological bulletin*, *125*(6), 627.

Deci, E. L., & Ryan, R. M. (1985). The general causality orientations scale: Self-determination in personality. *Journal of research in personality*, *19*(2), 109–134.

DeMink-Carthew, J., Olofson, M. W., LeGeros, L., Netcoh, S., & Hennessey, S. (2017). An analysis of approaches to goal setting in middle grades personalized learning environments. *RMLE Online*, *40*(10), 1–11.

Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. a meta-analysis on intervention studies at primary and secondary school level. *Metacognition and learning*, *3*(3), 231–264.

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior*, *52*, 338–348.

Elliot, A. J., & Murayama, K. (2008). On the measurement of achievement goals: critique, illustration, and application. *Journal of educational psychology*, *100*(3), 613.

Feng, M., Huang, C., & Collins, K. (2023). Promising long term effects of assistments online math homework support. In *International conference on artificial intelligence in education* (pp. 212–217).

Fiorella, L., Yoon, S. Y., Atit, K., Power, J. R., Panther, G., Sorby, S., ... Veurink, N. (2021). Validation of the mathematics motivation questionnaire (MMQ) for secondary school students. *International Journal of STEM Education*, *8*(1), 52. doi: 10.1186/s40594-021-00307-0

Gollwitzer, P. M. (1999). Implementation intentions: strong effects of simple plans. *American psychologist*, *54*(7), 493.

Gollwitzer, P. M. (2012). Mindset theory of action phases. *Handbook of theories of social psychology*, *1*, 526–545.

Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. *Advances in experimental social psychology*, *38*, 69–119.

Gonzalez-DeHass, A. R., & Willems, P. P. (2024). Middle-school students and digital homework: The evolving role of family engagement. *Middle School Journal*, *55*(4), 25–34.

Greene, B. A., DeBacker, T. K., Ravindran, B., & Krows, A. J. (1999). Goals, values, and beliefs as predictors of achievement and effort in high school mathematics classes. *Sex roles*, *40*(5), 421–458.

Guan, R., Raković, M., Chen, G., & Gašević, D. (2025). How educational chatbots support self-regulated learning? a systematic review of the literature. *Education and Information Technologies*, *30*(4), 4493–4518.

Gurung, A., Gutterman, J., Thomas, D. R., Feng, M., Aleven, V., & Koedinger, K. R. (2026). Coasting through class: Learning opportunity loss from practice avoidance during individual seatwork. *arXiv preprint arXiv:2604.25014*.

Gurung, A., Lin, J., Huang, Z., Borchers, C., Baker, R. S., Aleven, V., & Koedinger, K. R. (2025). Starting seatwork earlier as a valid measure of student engagement. *arXiv preprint arXiv:2505.13341*.

Hadwin, A. F., & Webster, E. A. (2013). Calibration in goal setting: Examining the nature of judgments of confidence. *Learning and instruction*, *24*, 37–47.

Hammer, D., & Berland, L. K. (2014). Confusing claims for data: A critique of common practices for presenting qualitative research on learning. *Journal of the Learning Sciences*, *23*(1), 37–46.

Harackiewicz, J. M., Smith, J. L., & Priniski, S. J. (2016). Interest matters: The importance of promoting interest in education. *Policy insights from the behavioral and brain sciences*, *3*(2), 220–227.

Hasan, M. A., Noor, N. F. M., Rahman, S. S. B. A., & Rahman, M. M. (2020). The transition from intelligent to affective tutoring system: a review and open issues. *IEEE Access*, *8*, 204612–204638.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of educational research*, *70*(2), 151–179.

Holstein, K., Aleven, V., & Rummel, N. (2020). A conceptual framework for human–ai hybrid adaptivity in education. In *Artificial intelligence in education: 21st international conference, aied 2020, ifrane, morocco, july 6–10, 2020, proceedings, part i 21* (pp. 240–254).

Hoover-Dempsey, K. V., & Sandler, H. M. (1997). Why do parents become involved in their children's education? *Review of educational research*, *67*(1), 3–42.

Howard, J. L., Bureau, J., Guay, F., Chong, J. X., & Ryan, R. M. (2021). Student motivation and associated outcomes: A meta-analysis from self-determination theory. *Perspectives on Psychological Science*, *16*(6), 1300–1323.

Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., … others (2003). Technology probes: inspiring design for and with families. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 17–24).

Jeynes, W. H. (2007). The relationship between parental involvement and urban secondary school student academic achievement: A meta-analysis. *Urban education*, *42*(1), 82–110.

Jin, Q., Borchers, C., Fancsali, S., & Aleven, V. (2025). Who to help? a time-slice analysis of K-12 teachers' decisions in classes with AI-supported tutoring. In *Proceedings of the 18th international conference on educational data mining (EDM 2025)* (pp. 640–645). Palermo, Italy.

Kahle, A. L., & Kelley, M. L. (1994). Children's homework problems: A comparison of goal setting and parent training. *Behavior therapy*, *25*(2), 275–290.

Kalyuga, S. (2009). The expertise reversal effect. In *Managing cognitive load in adaptive multimedia learning* (pp. 58–80). IGI Global Scientific Publishing.

Kamenetz, A. (2019). It's a smartphone life: More than half of us children now have one. *National Public Radio*, *31*.

Karumbaiah, S., Borchers, C., Falhs, A.-C., Holstein, K., Rummel, N., & Aleven, V. (2023). Teacher noticing and student learning in human-AI partnered classrooms: a multimodal analysis. In *Proceedings of the 17th international conference of the learning sciences (ICLS 2023)* (pp. 1042–1045).

Karumbaiah, S., Borchers, C., Shou, T., Falhs, A.-C., Liu, P., Nagashima, T., … Aleven, V. (2023). A spatiotemporal analysis of teacher practices in supporting student learning and engagement in an ai-enabled classroom. In *International conference on artificial intelligence in education* (pp. 450–462).

Koedinger, K. R., Baker, R. S., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the edm community: The pslc datashop. *Handbook of educational data mining*, *43*, 43–56.

Koedinger, K. R., Carvalho, P. F., Liu, R., & McLaughlin, E. A. (2023). An astonishing regularity in student learning rate. *Proceedings of the National Academy of Sciences*, *120*(13), e2221311120.

Koedinger, K. R., Kim, J., Jia, J. Z., McLaughlin, E. A., & Bier, N. L. (2015). Learning is not a spectator sport: Doing is better than watching for learning from a mooc. In *Proceedings of the second (2015) acm learning@scale* (pp. 111–120).

Koestner, R. (2008). Reaching one's personal goals: A motivational perspective focused on autonomy. *Canadian Psychology/Psychologie Canadienne*, *49*(1), 60.

Koestner, R., Otis, N., Powers, T. A., Pelletier, L., & Gagnon, H. (2008). Autonomous motivation, controlled motivation, and goal progress. *Journal of personality*, *76*(5), 1201–1230.

Kovanovic, V., Gašević, D., Dawson, S., Joksimovic, S., & Baker, R. (2015). Does time-on-task estimation matter? implications on validity of learning analytics findings. *Journal of Learning Analytics*, *2*(3), 81–110.

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

Legaspi, R., Xu, W., Konishi, T., Wada, S., & Ishikawa, Y. (2022). Multidimensional analysis of sense of agency during goal pursuit. In *Proceedings of the 30th acm conference on user modeling, adaptation and personalization* (pp. 34–47).

Li, X., Fan, Y., Li, T., Raković, M., Singh, S., van der Graaf, J., … Gašević, D. (2025). The FLoRA engine: Using analytics to measure and facilitate learners' own regulation activities. *Journal of Learning Analytics*, *12*(1), 391–413. doi: 10.18608/jla.2025.8349

Lin, J., Thomas, D. R., Han, Z., Tan, W., Nguyen, N. D., Gupta, S., … Koedinger, K. R. (2023). Personalized learning squared (plus): Doubling math learning through ai-assisted tutoring.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American psychologist*, *57*(9), 705.

Locke, E. A., & Latham, G. P. (2019). The development of goal setting theory: A half century retrospective. *Motivation Science*, *5*(2), 93.

Long, Y., & Aleven, V. (2017). Enhancing learning outcomes through self-regulated learning support with an open learner model. *User Modeling and User-Adapted Interaction*, *27*, 55–

88.

Long, Y., Holstein, K., & Aleven, V. (2018). What exactly do students learn when they practice equation solving? refining knowledge components with the additive factors model. In *Proceedings of the 8th international conference on learning analytics and knowledge* (pp. 399–408).

Matthews, C. D.-M. (2025). *Middle school mathematics teachers' use of diverse mathematics educational technology tools: Ixl, i-ready, and/or mathia* (Unpublished doctoral dissertation). Walden University.

McDowall, D., McCleary, R., & Bartos, B. J. (2019). *Interrupted time series analysis.* Oxford University Press.

Mrazek, A. J., Mrazek, M. D., Ortega, J. R., Ji, R. R., Karimi, S. S., Brown, C. S., ... others (2021). Teenagers' smartphone use during homework: an analysis of beliefs and behaviors around digital multitasking. *Education Sciences*, *11*(11), 713.

Muis, K. R., Winne, P. H., & Ranellucci, J. (2016). The role of calibration bias and performance feedback in achievement goal regulation. *International Education Research*, *4*(1), 14–36.

Newland, L. A. (2015). Family well-being, parenting, and child well-being: Pathways to healthy adjustment. *Clinical psychologist*, *19*(1), 3–14.

Nguyen, H. T., Borchers, C., Xia, M., & Aleven, V. (2024). Designing tools for caregiver involvement in intelligent tutoring systems for middle school mathematics. In *Proceedings of the 18th international conference of the learning sciences-icls 2024, pp. 865-872.*

Patall, E. A., Cooper, H., & Robinson, J. C. (2008). The effects of choice on intrinsic motivation and related outcomes: a meta-analysis of research findings. *Psychological bulletin*, *134*(2), 270.

Peacock, G. G., Ervin, R. A., Daly, E. J., & Merrell, K. W. (2009). *Practical handbook of school psychology: Effective practices for the 21st century.* Guilford Press.

Pelemo, P.-E. K. (2022). *Parental involvement: Investigating the roles, barriers, and strategies of effective parent-teacher relationships in middle school math* (Unpublished doctoral dissertation). Southeastern University.

Peng, C., Borchers, C., & Aleven, V. (2024). Designing homework support tools for middle school mathematics using intelligent tutoring systems. In *Proceedings of the 18th international conference of the learning sciences-icls 2024, pp. 1730-1733.*

Priniski, S. J., Hecht, C. A., & Harackiewicz, J. M. (2018). Making learning personally meaningful: A new framework for relevance research. *The Journal of Experimental Education*, *86*(1), 11–29.

Qiu, E. S., Thomas, D. R., Guo, B., Aleven, V., & Borchers, C. (2026). From heuristics to analytics: Forecasting effort and progress in online learning. In *Proceedings of the 19th international conference on educational data mining (edm 2026).*

Ritter, S., Joshi, A., Fancsali, S., & Nixon, T. (2013). Predicting standardized test scores from cognitive tutor interactions. In *Educational data mining 2013.*

Roll, I., Wiese, E. S., Long, Y., Aleven, V., & Koedinger, K. R. (2014). Tutoring self-and co-regulation with intelligent tutoring systems to help students acquire better learning skills. *Design recommendations for intelligent tutoring systems*, *2*, 169–182.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, *25*(1), 54–67.

Saks, K. (2024). The effect of self-efficacy and self-set grade goals on academic outcomes. *Frontiers* *in Psychology*, *15*, 1324007.

Schweizer, A., Niedlich, S., Adamczyk, J., & Bormann, I. (2017). Approaching trust and control in parental relationships with educational institutions. *Studia paedagogica*, *22*(2), 97–115.

Self-Determination Theory. (2022). Intrinsic motivation inventory (imi) – complete [Computer software manual]. PDF document. Retrieved from `https://selfdeterminationtheory.org/wp-content/uploads/2022/02/IMI_Complete.pdf`

Shogren, K. A., Mazzotti, V. L., Hicks, T. A., Raley, S. K., Gerasimova, D., Pace, J. R., ... others (2024). The goal setting challenge app: Impact on transition goal attainment outcomes of students with disabilities. *Career Development and Transition for Exceptional Individuals*, *47*(1), 4–18.

Simpson, M. A., Norberg, K. A., & Fancsali, S. E. (2024). Replicating an "astonishing regularity in student learning rate". In *Proceedings of the 17th international conference on educational data mining* (pp. 420–425).

Theobald, M., Lin, W. M., Sakaki, M., Murayama, K., & Brod, G. (2025). How (much) do people revise their goals after success and failure? the triple-a model of goal revision. *Educational Psychologist*, 1–21.

Thomas, D., Yang, X., Gupta, S., Adeniran, A., Mclaughlin, E., & Koedinger, K. (2023). When the tutor becomes the student: Design and evaluation of efficient scenario-based lessons for tutors. In *Lak23: 13th international learning analytics and knowledge conference* (pp. 250–261).

Thomas, D. R., Lin, J., Gatz, E., Gurung, A., Gupta, S., Norberg, K., ... others (2024). Improving student learning with hybrid human-ai tutoring: A three-study quasi-experimental investigation. In *Proceedings of the 14th learning analytics and knowledge conference* (pp. 404–415).

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.

van Leeuwen, A., Knoop-van Campen, C. A., Molenaar, I., & Rummel, N. (2021). How teacher characteristics relate to how teachers use dashboards: Results from two case studies in k-12. *Journal of Learning Analytics*, *8*(2), 6–21.

Venugopalan, D., Yan, Z., Borchers, C., Lin, J., & Aleven, V. (2025). Combining large language models with tutoring system intelligence: A case study in caregiver homework support. In *Proceedings of the 15th international learning analytics and knowledge conference* (pp. 373–383).

Wang, Y., Wang, H., Wang, S., Wind, S. A., & Gill, C. (2024). A systematic review and meta-analysis of self-determination-theory-based interventions in the education context. *Learning and Motivation*, *87*, 102015.

Wäschle, K., Allgaier, A., Lachner, A., Fink, S., & Nückles, M. (2014). Procrastination and self-efficacy: Tracing vicious and virtuous circles in self-regulated learning. *Learning and instruction*, *29*, 103–114.

Xu, Z., Zhao, Y., Zhang, B., Liew, J., & Kogut, A. (2023). A meta-analysis of the efficacy of self-regulated learning interventions on academic achievement in online and blended environments in k-12 and higher education. *Behaviour & Information Technology*, *42*(16), 2911–2931.

Yang, J., Guo, H., Yin, Y., Bai, J., Wang, B., Liu, J., ... Li, Z. (2024). m3p: Towards multimodal multilingual translation with multimodal prompt. *arXiv preprint arXiv:2403.17556*.