# Static, Dynamic, or Adaptive? Comparing Interface Scaffolding Methods in an Intelligent Tutoring System

Jennifer M. Reddig
jreddig3@gatech.edu
Georgia Institute of Technology
Atlanta, GA, USA

Adit Gupta
Drexel University
Philadelphia, PA, USA

Momin Siddiqui
Georgia Institute of Technology
Atlanta, GA, USA

Elias Izmirilian
Georgia Institute of Technology
Atlanta, GA, USA

Christopher J. MacLellan
Georgia Institute of Technology
Atlanta, GA, USA

## Abstract

Intelligent Tutoring Systems (ITS) support enhanced student learning outcomes, but sustained engagement can be a challenge for adult learners. In this work, we present an A/B experiment comparing three scaffolding interfaces in the Apprentice Tutors ITS, deployed as supplemental support across nine sections of College Algebra at a technical college: Full Scaffolding, which presents all fine-grained problem steps; Adaptive Scaffolding, which fades steps based on estimated mastery; and On-Demand Scaffolding, which presents only a final answer field with optional steps. We find that both dynamic conditions improve course outcomes relative to Full Scaffolding, with students assigned to On-Demand earning significantly higher overall grades ($\beta = 9.82$, $p = .018$) and completing substantially more problems ($b = 239.04$, $p = .001$). Causal mediation analysis indicates that approximately 28% of the On-Demand grade advantage can be attributed to increased tutor engagement and more problems solved. These results suggest that for supplemental educational technology in higher education, improving learner retention can be a gateway to improving learning outcomes.

## CCS Concepts

• **Applied computing** → **Interactive learning environments**; • **Information systems** → *Personalization.*

## Keywords

Intelligent Tutoring Systems, Adult Learning, Learning Engineering

## 1 Introduction

Intelligent Tutoring Systems (ITS) have proven to be effective at supporting student learning at scale, particularly in K-12 contexts [15, 27]. However, much of this success hinges on sustained use, with students spending up to 40% of their class time working with Cognitive Tutor [22] or at least two hours each week with ALEKS [19]. While this kind of extended use works well in a K-12 setting where students are in class five days a week, maintaining this level of engagement in a college or university course can be difficult. The target population for this work is non-traditional students [16], who have an average age of 25 and are pursuing education outside the conventional schooling pathway. As a result, they have to manage many responsibilities, including full-time employment and family commitments. These learners need to see value in their supplemental educational technology and practice needs to fit into their sporadic and disjoint free time [21].

Unlike K-12 students, non-traditional learners have considerably more autonomy over how, when, and whether they engage with supplemental tools. When educational technology is optional, adoption is the first barrier to overcome in order to improve learning outcomes. When tool use is optional, adult learners' decisions to engage are shaped by factors like perceived usefulness and whether the platform meets their expectations. These factors have been shown to be predictive of continued usage intention [18]. If a tool feels burdensome or misaligned with the time constraints of a working adult student, it will simply go unused, regardless of its effectiveness. This means that for supplemental educational technology in higher education, improving learner retention is a prerequisite for improving learning outcomes.

Apprentice Tutors is an Intelligent Tutoring System platform that provides supplemental resources for College Algebra classes at the Technical College System of Georgia. While learners who choose to use the Apprentice Tutors tend to perform better on the associated classroom assessments, most learners do not complete one problem and even fewer ever open the tool [9]. Focus groups with students revealed that because the interfaces contain lots of fine-grained steps to solve each problem, students felt overwhelmed [10]. Though the tutor interfaces were created in collaboration with the instructors, students still balked at the length of the problems. This raises the question: can we adjust how the problem solving process is presented to surface key problem-solving steps, while also not overwhelming the learner with scaffolding?

This work aims to answer the following research questions:

(1) What is the impact of progressive disclosure of scaffolding on learner retention?
(2) Does adaptive problem presentation improve learner outcomes compared to static problem presentation?
(3) Do learners achieve better outcomes when they can choose to progressively disclose necessary steps?

To answer these questions, we ran an A/B experiment comparing the impact of the standard static interface that provides full scaffolding to two dynamic interfaces within the Apprentice Tutors ITS. We found that while both dynamic interfaces significantly improved course grades, the effect was driven primarily by the students completing more problems compared to the static interface.

## 2 Background

Intelligent Tutoring Systems have consistently demonstrated strong learning outcomes across a range of educational contexts. ITS produce effect sizes comparable to human tutoring, making them one of the most effective scalable educational interventions available [27]. Across almost all education levels and domains, ITS are effective tools for learning [15]. Despite their demonstrated effectiveness, widespread adoption of ITS remains inconsistent, as learner acceptance and willingness to engage with these systems are not guaranteed [11]. Even when ITS are made available, actual usage frequently falls well below recommended levels [19]. When learners have the autonomy to opt in or out on their own terms, which is typical in adult learning scenarios, usage levels may fall even more.

Non-traditional learners (and adult learners more broadly) present distinct challenges for educational technology design that are not well addressed by systems built for K-12 students. Adult learners tend to be self-directed and motivated by perceived relevance [12]. Unlike traditional students, non-traditional learners face competing demands from work and family that fragment their study time and contribute to higher rates of attrition [3, 5]. Adult learners tend to be problem-centered and oriented toward learning that has immediate relevance and practical application [13]. These characteristics have concrete implications for educational technology design, which must be flexible, low-friction, and respectful of limited time to sustain engagement [21]. These considerations are especially important for supplemental tools since interface complexity and time burden can prevent adoption before any learning occurs [10].

Whether adult learners choose to engage with an educational technology is shaped substantially by their perceptions of the tool. Perceived usefulness and ease of use are among the primary determinants of technology adoption and continued use [7]. Within higher education, these factors consistently emerge as the strongest predictors of whether learners choose to adopt and sustain engagement with educational technology [8]. When engagement is voluntary, these factors have a stronger influence on intention to use, as the external pressure from the institution or instructor is no longer present [25].

Scaffolding is one method of making the learning process more attainable for students by helping them through a task they could not accomplish themselves [28]. In ITS specifically, step-by-step problem decomposition is one of the most common forms of scaffolding, breaking complex tasks into manageable steps with feedback at each stage [26, 27]. Fine-grained scaffolding of steps provides structure for complex, problem-solving tasks by reminding learners of what they have yet to do [20]. However, too much scaffolding can reduce the productive struggle that drives learning, while too little may prevent students from making progress [14]. Scaffolding can also be overwhelming for a complex problem when presenting the student with the detailed sequential process they need to accomplish to complete a task [10]. Fading the scaffolds is one potential solution to this problem, gradually withdrawing support as learners internalize the problem-solving process. Fading scaffolding will help with productive struggle as learners progress, but it does not help with the initial overwhelming presentation. In this work, we use a progressive disclosure method that allows students to reveal additional scaffolded steps as needed, keeping the initial interface minimal while still having supportive structures in place.

Apprentice Tutors is a web-based intelligent tutoring platform built in collaboration with instructors from the Technical College System of Georgia. Tutors on this platform provides immediate correctness feedback, on-demand hints, dynamic problem generation, and adaptive problem sequencing driven by Bayesian Knowledge Tracing [6]. Tutor content was based on the OpenStax Intermediate Algebra textbook [17] and built in collaboration with course instructors. The tutors are entirely supplemental and students are not required to use the tool. While students who engage with the tutors tend to perform better on course assessments, the vast majority of enrolled students never complete even a single problem [9]. Interface complexity has been identified as a key barrier, with students describing the fine-grained step structure as overwhelming and appear to discontinue using the tool for this reason [10]. While Apprentice Tutors provide detailed step-level support to learners, they also fixes the level of scaffolding for every student.

To investigate how presentation of scaffolding influences engagement and learning within the ITS, we built three versions of the initial problem interface, seen in Figure 1. The Full Scaffolding interface presents the problems as they were designed by the instructor, with maximum guidance and all steps visible. The On-Demand interface presents only the final answer initially, but students can request additional step guidance as needed (by clicking the red plus button). Steps are revealed in a sequential manner from the first step, progressively leading students through the problem-solving process. Each step added reminds the learner what the very next step is. The Adaptive interface is dynamically adjusted based on learner proficiency, fading from Full Scaffolding to On-Demand as learners progress, while also providing learners the ability to request additional scaffolding.

## 3 Methodology

The Apprentice Tutors System was deployed during the 2024-2025 school year in TCSG's College Algebra classes. Students were not required to use the tutors; they were available as a supplementary resource and the teacher encouraged their use. The tutors were available in 9 sections of College Algebra. All sections had the same instructor. The tutors were embedded in BlackBoard so students could launch and log in to the system with a single click. Apprentice Tutors has 30 problem types that span 8 math topics, with most of the topics concentrated at the beginning of the semester.

Teachers also had access to a teacher dashboard that let them see all progress and problems solved by each student enrolled in their class. Apprentice Tutors uses Bayesian Knowledge Tracing [6] to track student mastery as students interact with the tutors. This allows the tutors to adapt problem sequencing and initial interface appearance in the following study. To allow the Apprentice Tutors to adapt problem sequencing to student performance and give teachers insight into student progress, the tutors capture all student interactions and correctness.

Before beginning this study, our protocol was reviewed and approved by our institutional review board. Upon first accessing the tutors, students were asked to consent to participate in the study.

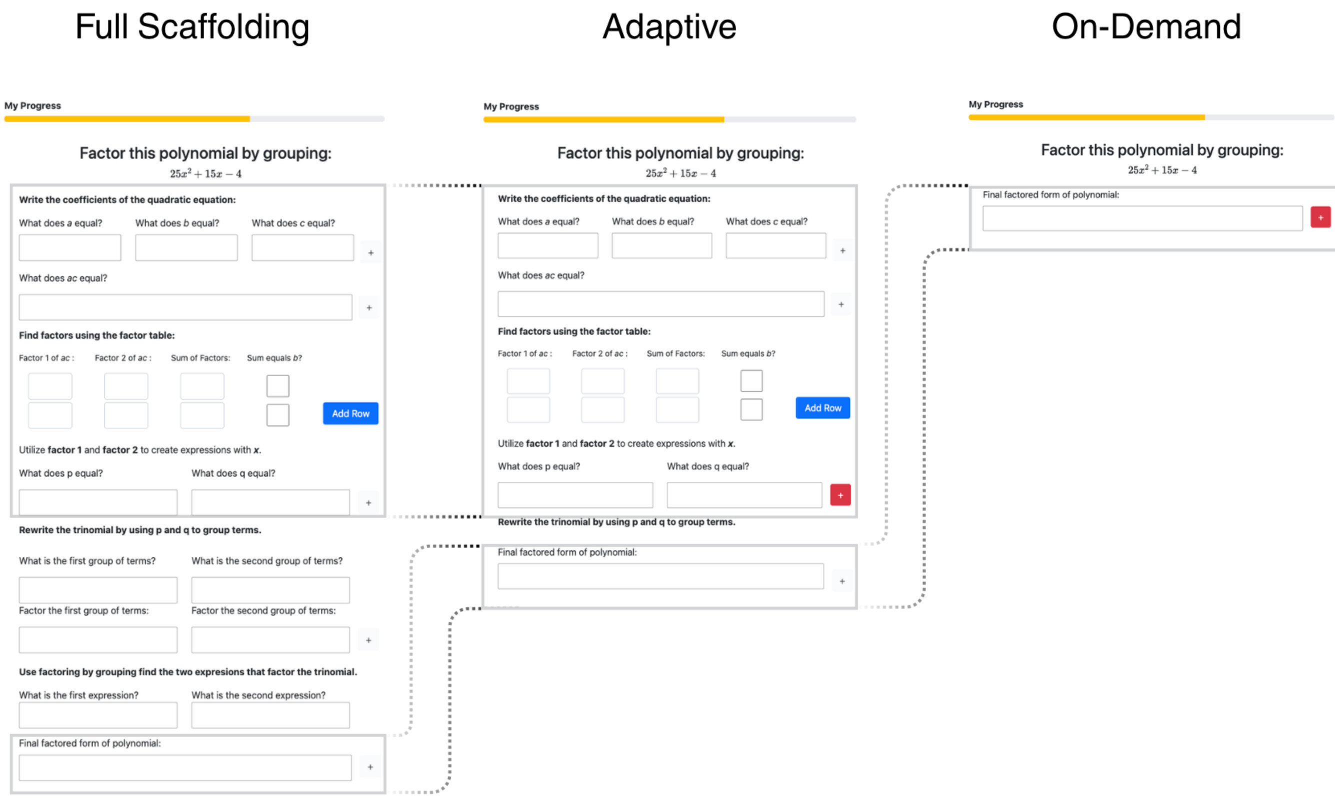


**Figure 1: Example interfaces for the three scaffolding conditions used in the tutoring system. *Full Scaffolding* (left) breaks the polynomial factoring problem into explicit sequential steps. *Adaptive* (center) collapses steps as students progress, but students may add the steps back as needed, revealing steps in a sequential manner. *On-Demand* (right) starts with only a single answer field, but allows students to add additional steps as needed.**

Students who did not consent were removed from data analysis. All students from the targeted course sections consented to participate.

After receiving consent, students were randomly assigned to one of three conditions: Full Scaffolding, Adaptive Scaffolding, or On-Demand Scaffolding. Students retained this condition during the entire semester. Students in the Full Scaffolding condition received the tutors in the default format with all fine-grained steps visible, and students were required to complete all steps to finish a problem. Students in the On-Demand Scaffolding condition saw only the final answer field and had the option to add additional supportive steps by clicking a button. If students added additional steps, the new steps were required for problem completion. Steps were added in the same sequence for every student. Steps were revealed in a sequential manner from the first step, progressively leading students through the problem-solving process. Each step added reminded the learner of the very next step. Students in the Adaptive Scaffolding condition received a custom initial interface based on their estimated problem mastery, the average mastery of all KCs within the problem. As a student approaches full mastery, the initial problem interface shows fewer and fewer steps. If the platform estimated that the student had achieved 25% mastery of the entire problem, the student would see the first 75% of steps, including the final answer field. If the platform estimated that the student had achieved 75% mastery, the student would see the first 25% of steps. As with the On-Demand condition, students always had the option to add more supportive steps as needed. In this way, the Adaptive condition faded scaffolding from Full Scaffolding to a single final answer field (where On-Demand starts) as students progress. The different scaffolding approaches are shown in Figure 1.

When students first connect to Apprentice Tutors, they are assigned an anonymous identifier, which all of their transaction data is linked with. Overall, 110 students participated in the study and used the tutor, where 36 students were assigned the Full Scaffolding condition, 37 were assigned the Adaptive condition, and 36 were assigned the On-Demand condition. There were 276 students in the classes that did not use the tutor. Since student usage was voluntary, we captured and analyzed student usage and performance data to evaluate the impact of our interventions. We also analyzed test scores for students in the targeted classes, and in collaboration with the Technical College staff, we were able to link test scores with the random identifiers of students that used our tutors without breaking anonymity [1].

## 4 Results

We began by analyzing the error rates on problems solved by learners in each condition. To compare across conditions, which have

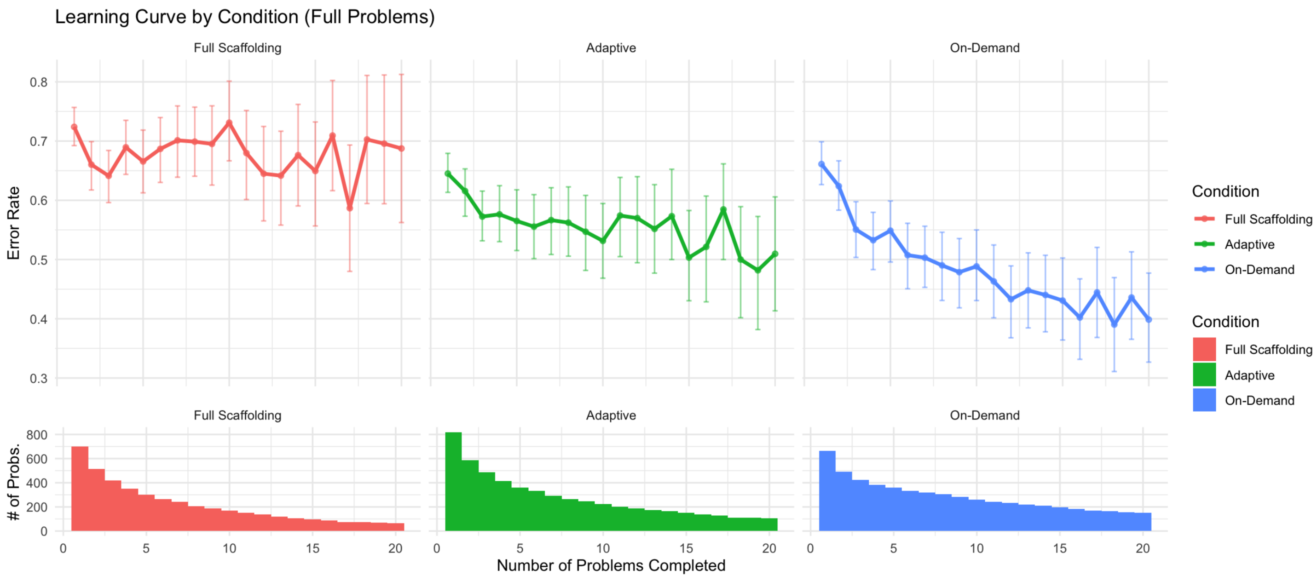


**Figure 2: Learning curves (top) and observation counts (bottom) across the first 20 problem opportunities, separated by scaffolding condition. Error bars represent 95% confidence intervals at each opportunity.**

**Table 1: Mixed-effects logistic regression model predicting whole-problem errors by the interaction between scaffolding condition and opportunity. Both the Adaptive and On-Demand conditions had significantly lower baseline error odds. The significant condition × opportunity interactions indicate that error rates declined more steeply with practice in the Adaptive and On-Demand conditions than in Full Scaffolding, consistent with the learning curves in Figure 2. The model includes random intercepts for students and problem types, and a random slope for opportunity by problem topic.**

| **Predictor** | **OR** | **95% CI** | ***p*** |
|---|---|---|---|
| *Fixed Effects* | | | |
| Intercept | 4.07 | [2.24, 7.40] | **<.001** |
| Condition: Adaptive | 0.56 | [0.32, 0.97] | **.038** |
| Condition: On-Demand | 0.39 | [0.22, 0.68] | **.001** |
| Opportunity | 0.99 | [0.97, 1.01] | .305 |
| Condition: Adaptive × Opportunity | 0.99 | [0.98, 0.99] | **<.001** |
| Condition: On-Demand × Opportunity | 0.98 | [0.98, 0.99] | **<.001** |
| *Random Effects* | | | |
| $\sigma^2$ | 3.29 | | |
| $\tau_{00}$ (students) | 1.11 | | |
| $\tau_{00}$ (problem type) | 4.08 | | |
| $\tau_{11}$ (problem type.Opp.) | <0.01 | | |
| ICC | 0.61 | | |
| Marginal $R^2$ | 0.240 | | |
| Conditional $R^2$ | 0.705 | | |
| $N$ (students) | 110 | | |
| $N$ (problem type) | 89 | | |
| Observations | 27,829 | | |

vastly different granularities and intermediate step checkpoints so KCs directly measured between conditions are different, we analyzed the probability of completing a problem without error. Figure 2 shows learning curves by condition over whole problems. We modeled the effect of the interaction between condition and opportunity on error rate using a generalized linear mixed-effects model, with random intercepts for students ($N = 110$) and problem types ($N = 89$), and a random slope for opportunity by problem type. On the first practice opportunity, students in the Full Scaffolding condition had approximately an 73% error rate (intercept: $\beta = 1.40$, $SE = 0.30$, $z = 4.62$, $p < .001$). That is, the probability of learners completing a problem without error is about 27%. Both intervention conditions showed significantly lower initial error rates (Adaptive: $\beta = -0.58$, $SE = 0.28$, $z = -2.07$, $p = .038$; On-Demand: $\beta = -0.95$,

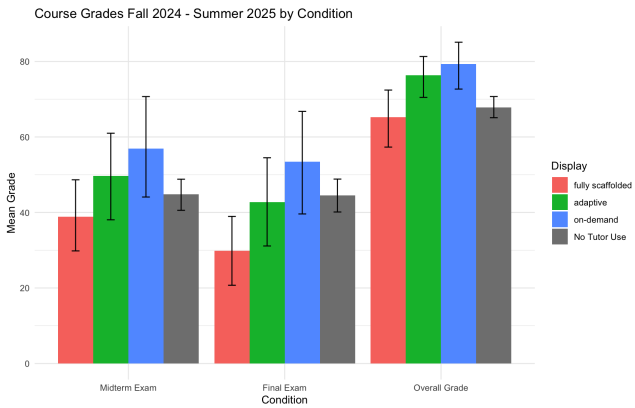


**Figure 3: Mean course grades by scaffolding condition for the Fall 2024 - Summer 2025 academic year. Students in the On-Demand condition scored nearly a full letter grade higher than students in the Full Scaffolding condition. Students in the Adaptive condition scored about 8 percentage points higher, but the difference was not significant. The mean grades of students who did not use the tutors is included for reference. Error bars represent 95% bootstrapped confidence intervals.**

$SE = 0.29$, $z = -3.28$, $p = .001$). Opportunity was not a significant predictor for the Full Scaffolding condition ($\beta = -0.01$, $SE = 0.01$, $z = -1.03$, $p = .305$), suggesting little learning over practice in that condition. However, the Full Scaffolding condition also experienced the largest attrition rate, which may have impacted the results. Both intervention conditions showed significantly steeper error reduction slopes (Adaptive: $\beta = -0.01$, $SE = 0.002$, $z = -4.27$, $p < .001$; On-Demand: $\beta = -0.02$, $SE = 0.002$, $z = -7.11$, $p < .001$). The intervention conditions also had higher retention across opportunities compared to the Full Scaffolding condition.

We found that students in the On-Demand Scaffolding condition had significantly higher final course grades than students in the Full Scaffolding condition, as seen in Figure 3. Students in the Full Scaffolding condition finished the course with 65% on average ($\beta = 62.33$, $SE = 5.3$, $t = 11.77$, $p < .001$), while students in the On-Demand condition finished the course with 79%, on average a full letter grade higher ($\beta = 9.82$, $SE = 4.11$, $t = 2.39$, $p = .018$). Students in the adaptive condition scored around 76%, although the difference was only marginally significant ($\beta = 7.60$, $SE = 4.03$, $t = 1.89$, $p = .062$).

To examine whether engagement with the Apprentice Tutors explained these grade differences, we first examined if the different conditions displayed different in-tutor behavior. Students in the On-Demand condition completed significantly more problems than students in the Full Scaffolding condition ($b = 239.04$, $SE = 69.74$, $t = 3.43$, $p = .001$), while students in the Adaptive condition did not differ significantly from the Full Scaffolding baseline ($\beta = 68.34$, $SE = 68.86$, $t = 0.99$, $p = .323$). These results are shown in Figure 4.

We used a generalized linear mixed-effects model to predict whether a student will complete a problem they start by their assigned condition, with random intercepts for students (N = 110) and

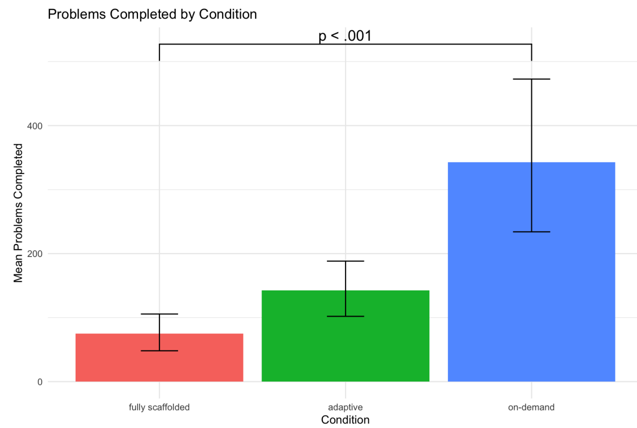


**Figure 4: Mean number of problems completed by scaffolding condition. On-Demand students completed substantially more problems than those in the Full Scaffolding condition, with Adaptive students falling between the two. Error bars represent 95% bootstrapped confidence intervals.**

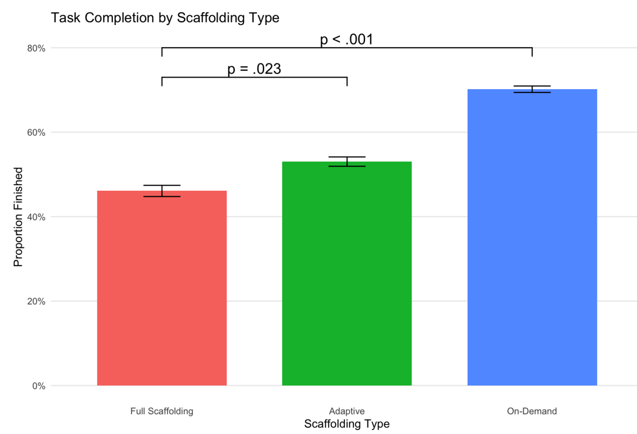


**Figure 5: Predicted probability of problem completion by scaffolding type. Both the Adaptive and On-Demand conditions were associated with significantly higher odds of completion relative to Full Scaffolding. Error bars represent 95% bootstrapped confidence intervals.**

problem types (N = 89). Students in the Full Scaffolding condition had an estimated probability of problem completion of 34.6%. Both the Adaptive and On-Demand conditions were associated with significantly higher odds of completion relative to Full Scaffolding. The Adaptive condition showed approximately 1.90 times the odds of completion ($\beta = 0.64$, $SE = 0.28$, $z = 2.27$, $p = .023$), while the On-Demand condition showed approximately 3.56 times the odds of completion ($\beta = 1.27$, $SE = 0.29$, $z = 4.34$, $p < .001$), as seen in Figure 5. Students in the On-Demand and Adaptive conditions are more likely to finish the problems they start, while students in the Full Scaffolding leave their problems incomplete about 65% of the time.

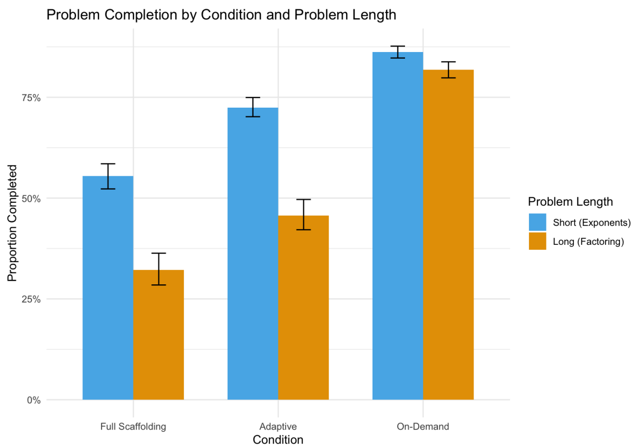


**Figure 6: Predicted probability of problem completion by scaffolding type and problem length. Short problems are more likely to be completed overall, and students in the On-Demand condition are also more likely to finish both types of problems. The greatest difference between problem types is the Full Scaffolding condition and the smallest difference is the On-Demand condition. Error bars represent 95% bootstrapped confidence intervals.**

We examined whether the length and complexity of a problem was an additional factor of whether or not students completed a problem by comparing two tutored topics: exponents ($\sim$ 3 fields in the Full Scaffolding condition) and factoring (between 10 and 20 fields in the Full Scaffolding condition). Because the factoring tutors have more lengthy problems that require many steps to complete, the appearance of the On-Demand interface and the Full Scaffolding interface will be drastically different, while the difference will not be as great with exponents which have few steps to complete. We used a generalized linear mixed-effects model to model the effect of the interaction between Condition and topic on problem completion, with random intercepts for students (N = 84). Exponent problems had a 56% probability of completion in the Full Scaffolding condition, and the Adaptive condition did not differ significantly ($\beta = 0.495, SE = 0.26, z = 1.895, p = 0.058$). Students in the on-demand condition were more likely to complete the problems they start ($\beta = 1.48, SE = 0.27, z = 5.408, p < .001$). Factoring problems were substantially less likely to be completed overall ($\beta = -1.26, SE = 0.14, z = -9.29, p < .001$). The Adaptive condition was not significantly different from the Full Scaffolding baseline ($\beta = 0.297, SE = 0.18, z = 1.62, p = .105$). However, the On-Demand condition was significantly more likely to complete factoring problems ($\beta = 0.45, SE = 0.18, z = 2.44, p = 0.015$). This interaction is shown in Figure 6.

To examine whether engagement with the Apprentice Tutors explained these grade differences, we created a second model with problems completed as a covariate. When accounting for tutor activity, the condition effects were reduced and no longer statistically significant on overall course grades (adaptive: $\beta = 6.62$, $SE = 3.97$, $t = 1.67$, $p = .098$; on-demand: $\beta = 6.74$, $SE = 4.26$, $t = 1.58$, $p = .116$). Instead, problems completed (across conditions) was a significant predictor of final grade ($\beta = 0.01$, $SE = 0.005$, $t = 2.14$, $p = .034$), such that every additional problems completed was associated with approximately a 0.01% percentage point increase in course grade. This suggests that the differences across conditions were driven by differing levels of engagement with the tutoring system. Full results from all three models are shown in Table 2.

Given our A/B experimental design, and the randomization of participants across conditions, our results suggest that assigning students to the on-demand condition does causally translate into an improvement in their overall grade in the course compared to the full scaffolding condition. Our secondary analysis suggests that effect is driven by the number of problems the student completes, with students completing more problems in the on-demand condition. Unfortunately, we cannot experimentally manipulate the number of problems that students choose to do within our A/B framework. To formally test whether problems completed mediated the effect of condition on course grade, we conducted a causal mediation analysis [24]. In alignment with our primary analysis (model 1), the total effect of the On-Demand condition on grade was significant ($b = 9.57$, 95% CI $[1.27, 17.69]$, $p = .026$). Further, our analysis found that the average causal mediation effect (ACME) for the On-Demand condition was significant ($b = 2.85$, 95% CI $[0.21, 6.00]$, $p = .038$), indicating that the indirect path through tutor engagement reliably accounted for a portion of the grade advantage. In contrast, the average direct effect (ADE) was not significant ($b = 6.72$, 95% CI $[-2.03, 14.94]$, $p = .130$), suggesting that completing more problems was the primary mechanism by which On-Demand causally increases students' overall grades. Overall, approximately 28% of the main effect of condition operated through the problems completed mediator as compared to the direct effect of the condition (i.e., proportion mediated). No significant mediation was observed for the Adaptive condition (ACME: $b = 0.85$, 95% CI $[-0.82, 3.22]$, $p = .352$). Full results are shown in Table 3.

## 5 Discussion

In this work, we found that students who received On-Demand step scaffolding completed more problems within the Apprentice Tutors, and as a consequence, earned higher overall course grades. The On-Demand condition also saw improved learning gains, with a faster decline in problem-level error rates than students in the Full Scaffolding condition. Students in the Adaptive condition also received some benefits, lower initial error rates on the first attempted problem and a faster decline in error rates than Full Scaffolding. While the Adaptive condition was more likely to finish a problem, students did not complete significantly more problems or have higher course grades. Though the Full Scaffolding condition appears to have little learning occurring, we hypothesize that it is because students solved few problems, and quickly abandoned the intimidating tool. We see the greatest effect on retention during longer problems, which suggests that keeping the presentation of problems short is vital for maintaining engagement.

Our mediation analysis suggests that completing more problems was a significant factor in the higher course grades. Worked examples and [4] and deliberate practice [2] are some of the most effective ways to acquire mathematical skills. Other interventions that encourage students to complete more problems may see similar

**Table 2: Linear mixed-effects models predicting overall course grade and problems completed by scaffolding condition. The first model estimates the effect of condition on course grade; the second adds problems completed as a covariate to assess whether engagement mediates the condition effect; the third models problems completed as the outcome. The On-Demand condition was associated with significantly higher grades (Model 1) and substantially more problems completed (Model 3) relative to Full Scaffolding. When problems completed was included as a covariate (Model 2), the effect of On-Demand on grade became non-significant, suggesting mediation through engagement. All models include random intercepts for course section.**

| | Grade by Condition | | | Grade by Condition and Problems Completed | | | Problems Completed | | |
|---|---|---|---|---|---|---|---|---|---|
| **Predictor** | **Est.** | **95% CI** | ***p*** | **Est.** | **95% CI** | ***p*** | **Est.** | **95% CI** | ***p*** |
| *Fixed Effects* | | | | | | | | | |
| Intercept | 62.33 | [51.84, 72.83] | **<.001** | 61.67 | [50.68, 72.67] | **<.001** | 51.50 | [−70.45, 173.44] | .404 |
| Condition: Adaptive | 7.60 | [−0.39, 15.58] | .062 | 6.62 | [−1.25, 14.49] | .098 | 68.34 | [−68.11, 204.80] | .323 |
| Condition: On-Demand | 9.82 | [1.68, 17.95] | **.018** | 6.74 | [−1.70, 15.18] | .116 | 239.04 | [100.83, 377.25] | **.001** |
| Problems completed | — | — | — | 0.01 | [0.00, 0.02] | **.034** | — | — | — |
| *Random Effects* | | | | | | | | | |
| $\sigma^2$ | 305.22 | | | 292.99 | | | 90510.33 | | |
| $\tau_{00}$ (Course Section) | 151.59 | | | 176.89 | | | 9924.30 | | |
| ICC | 0.33 | | | 0.38 | | | 0.10 | | |
| Marginal $R^2$ | 0.039 | | | 0.059 | | | 0.092 | | |
| Conditional $R^2$ | 0.358 | | | 0.413 | | | 0.182 | | |
| $N$ (Course Section) | 9 | | | 9 | | | 9 | | |
| Observations | 115 | | | 115 | | | 115 | | |

**Table 3: Causal mediation analysis examining whether problems completed mediates the effect of scaffolding condition on overall course grade, with Full Scaffolding as the reference condition. ACME denotes the average causal mediation effect. ADE denotes the average direct effect. For the On-Demand condition, the indirect effect was significant with problems completed accounting for approximately 28% of the total grade advantage. No significant mediation was found for the Adaptive condition.**

| **Effect** | **Estimate** | **95% CI** | | ***p*** |
|---|---|---|---|---|
| | | Lower | Upper | |
| *On-Demand condition* | | | | |
| ACME (indirect) | 2.85 | 0.21 | 6.00 | **.038*** |
| ADE (direct) | 6.72 | −2.03 | 14.94 | .130 |
| Total effect | 9.57 | 1.27 | 17.69 | **.026*** |
| Prop. mediated | 0.28 | −0.01 | 1.37 | **.064.** |
| *Adaptive condition* | | | | |
| ACME (indirect) | 0.85 | −0.82 | 3.22 | .352 |
| ADE (direct) | 6.55 | −1.43 | 14.58 | .114 |
| Total effect | 7.40 | −0.96 | 15.44 | **.094.** |
| Prop. mediated | 0.09 | −0.55 | 0.74 | .382 |

improvements in course grades. During this time period, Apprentice Tutors served over 200 course sections in mathematics and nursing, with nearly 1000 unique students accessing the tutors. These insights on problem presentation leading to increased engagement and higher course grades have the potential to help many students across disciplines.

The On-Demand intervention caused students to persist in using the tutor, and their overall grades benefited from the additional practice. For this population of adult learners, getting students to practice and engage with technology is a greater challenge than individualizing the learning experience, and the intervention meaningfully improved student retention and adoption. One hypothesis for why we see this effect is that the students' perceived effort is lower for the minimal interface, as well as the estimated time-cost for completing a problem. Longer problems showed the strongest completion rate gains, which may indicate that learners found the minimal interface more approachable and chose to work on the problem until completion.

Another hypothesis is that these adult learners may not require all the fine-grained steps to complete a problem. The Adaptive interface also saw improved in-tutor learning gains, and were more likely to finish a problem. It is possible that for this population and topic, completely tutoring the problem-solving process from the ground up is not needed, and tutoring should focus more on refreshing concepts learned in K-12 and encouraging students to internalize the process. Presenting all fine-grained steps may also increase the cognitive load, rather than offloading the process to the tutor. Since classroom assessments do not scaffold problem-solving in the same way that the Full Scaffolding interfaces do, it is also possible that initial problem/answer format of the On-Demand condition better matched the format of class assessments, better preparing students for the exams (midterm and final).

Though the Adaptive condition displayed similar in-tutor learning gains as the On-Demand condition, we do not see the same improvement in course outcomes. There are several reasons why the Adaptive condition may have under-performed. First, the initial problem interface was no different from the Full Scaffolding interface, so the Adaptive interface did not benefit from any of the On-Demand interface's perceived ease. In addition, all students received the same step collapsing sequence, which collapsed the later steps first. An alternative approach to adaptation could be initially assuming knowledge and adapting by progressively adding more steps instead of removing steps. The adaption was also based on the overall problem mastery, rather than individual KC mastery, so students may have been required to practice steps they had already mastered, but were only given the unmastered steps if they expanded the problem further. Overall mastery drove the amount of expansion, but not which specific steps were revealed. Finally, we set all KCs to use the same initial BKT parameters, rather than learning the KC parameters to those that best fit the data.[1] Perhaps using better fitting parameters would change both the initial appearance of the Adaptive interface and how steps are collapsed between problems. A unique initial mastery probability for each KC would drastically change the initial appearance of the Adaptive condition, and perhaps better align with the prerequisite knowledge of the users.

For supplemental educational technology, motivating continued use should be a priority. While learning technologies should be evaluated in terms of learning outcomes, engagement is a key factor in mediating overall learning. This work linked key engagement metrics to overall learning outcomes through mediation analysis. We increased student retention by lowering the appearance of difficulty and letting the user control how much detail to see. Scaffolding was always available so students were never without additional help, but allowing the student to choose whether they needed additional steps may have encouraged continued use over locking students into a system-prescribed breakdown. Other tutors may want to give students the ability to adjust scaffolding as needed instead of using static, fine-grained interfaces. The Full Scaffolding interface, though fine-grained with all the structure needed to solve the problem, was not an experience that students chose to persist with. Students should receive the level of support they need, but ultimately be encouraged to solve problems independently without additional structure.

Future work should improve the Adaptive condition by personalizing based on granular KCs instead of overall problem mastery. The BKT parameters used should also be set to those that best fit to student data, as not all KCs will have the same initial probability of being known. This experiment also used a linear problem breakdown. Since the Apprentice Tutors use a unique Hierarchical Task Network system for expert models [23], revising the tutors to have a more hierarchical structure would allow us to provide supportive structure at different levels of problem granularity, instead of just next steps. Adapting based on problem depth instead of problem length could also be a way to improve the Adaptive condition. The On-Demand condition could also be improved by revealing additional steps by depth rather than by a sequential task list.

A key tension in scaffolding design is the tradeoff between granularity of support and learner engagement. Fine-grained scaffolding, like the Full Scaffolding condition, provides detailed guidance that reduces the chance of unproductive struggle, but it also increases the perceived length and complexity of a problem, which reduces voluntary engagement. Coarser scaffolding lowers this barrier but risks leaving learners without the support they need to make progress. One method for solving this problem was giving the learner complete control over how many steps they see. However, not all learners may be able to effectively self-regulate their learning. The tutoring system can support learners by adapting the level of granularity to each student. However, the optimal level of granularity likely varies by learner, topic, and prior knowledge. Learner models could let us navigate this tradeoff dynamically. A tutor can infer using learner models when a learner is likely to benefit from additional guidance versus when that guidance is counterproductive or would drive the learner away from the tool. Appropriately modeling students will allow us to find the right level of granularity for each student to keep them engaged while still providing a problem breakdown that supports their skill acquisition.

A major contribution of this work is the mediation analysis that ties the intervention to course grades through tutor engagement. This allows us to say not only that the intervention was successful, but *why* the intervention was successful. Designing A/B experiments that allow for this kind of testing and analysis is important for overall explainability. Identifying through what channels student learning was made possible is important for generalizable insights that apply more broadly to other educational technology. In this experiment, 28% of the main effect from the On-Demand condition operated through the problems completed mediator. There may be additional mediators that could explain more of the overall results, such as time on task, help-seeking behaviors, or session frequency.

Future work should also evaluate the student experience using each interface to see if the perceived effort or ease differs between conditions and is a driving factor behind the sustained use. Because students in the Full Scaffolding condition did not continue to use the tutor, we were not able to evaluate the learning effects of that interface. Future studies should evaluate the impact of each interface through a controlled trial so technology developers can appropriately weigh the learning effects against perceived ease to maximize learning outcomes for all students.

## 6 Conclusion

This study examined whether modifying the way problems are presented within an ITS problem interface could improve engagement and learning outcomes for adult learners in a voluntary-use setting. Students randomly assigned to the On-Demand scaffolding condition completed significantly more problems and earned higher course grades than students in the Full Scaffolding condition. Causal mediation analysis confirmed that this grade advantage was partially explained by increased tutor engagement, suggesting that the interface change influenced learning primarily by sustaining student participation. Students in the Adaptive condition showed improved in-tutor learning gains and higher problem completion rates relative to Full Scaffolding, but did not differ significantly on course outcomes. The mediation analysis showed us how the direct effects of the experiment can be explained, highlighting that similar learning occurred in all conditions, but sustained practice is the primary driver of improved course grades.

Allowing students to control the depth of scaffolding they receive, rather than prescribing a fixed level of support, appears to lower the perceived burden of participation without sacrificing learning. Future work should examine the perceived effort and time cost associated with each interface, evaluate the Adaptive condition with KC-level rather than problem-level mastery estimation, and assess whether a hierarchical depth-based expansion instead of

[1] Data for fitting the KCs was not available until after the experiment was completed.

a linear sequence-based expansion further improves outcomes in both the Adaptive and On-Demand conditions.

## Acknowledgments

The material in this paper is based upon work supported by the National Science Foundation under Grant No. 2247790 and Grant No. 2112532. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Thank you to Trina Lumbard for working with us to build the Apprentice Tutors system and continuing to partner with our research.

## References

[1] 1EdTech. 2026. Learning tools interoperability. https://www.1edtech.org/standards/lti

[2] K Anders Ericsson. 2008. Deliberate practice and acquisition of expert performance: a general overview. *Academic emergency medicine* 15, 11 (2008), 988–994.

[3] Siti Nor Amira Baharudin, Munira Murad, and Nur Hana Hj Mat. 2013. Challenges of adult learners: A case study of full time postgraduates students. *Procedia-social and behavioral sciences* 90 (2013), 772–781.

[4] Christina Areizaga Barbieri, Dana Miller-Cotto, Sarah N Clerjuste, and Kamal Chawla. 2023. A meta-analysis of the worked examples effect on mathematics performance. *Educational Psychology Review* 35, 1 (2023), 11.

[5] John P Bean and Barbara S Metzner. 1985. A conceptual model of nontraditional undergraduate student attrition. *Review of educational Research* 55, 4 (1985), 485–540.

[6] Albert T Corbett and John R Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction* 4, 4 (1994), 253–278.

[7] Fred D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly* 13, 3 (1989), 319–340.

[8] Jiyuan Feng, Bin Yu, Wee Hoe Tan, Zetian Dai, and Zengkun Li. 2025. Key factors influencing educational technology adoption in higher education: A systematic review. *PLOS Digital Health* 4, 4 (2025), e0000764.

[9] Adit Gupta and Christopher MacLellan. 2025. Intelligent tutors beyond K-12: An observational study of adult learner engagement and academic impact. *arXiv preprint arXiv:2502.16613* (2025).

[10] Adit Gupta, Momin Siddiqui, Glen Smith, Jenn Reddig, and Christopher MacLellan. 2024. Intelligent tutors for adult learners: an analysis of needs and challenges. *arXiv preprint arXiv:2412.04477* (2024).

[11] Hanjing Huang, Youjie Chen, and Pei-Luen Patrick Rau. 2022. Exploring acceptance of intelligent tutoring system with pedagogical agent among high school students. *Universal Access in the Information Society* 21, 2 (2022), 381–392.

[12] Malcolm S. Knowles. 1973. *The Adult Learner: A Neglected Species.* Gulf Publishing Company, Houston.

[13] Malcolm Shepherd Knowles. 1984. Andragogy in action. (1984).

[14] Kenneth R Koedinger and Vincent Aleven. 2007. Exploring the assistance dilemma in experiments with cognitive tutors. *Educational psychology review* 19, 3 (2007), 239–264.

[15] Wenting Ma, Olusola O Adesope, John C Nesbit, and Qing Liu. 2014. Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of educational psychology* 106, 4 (2014), 901.

[16] Kris MacDonald. 2018. A review of the literature: The needs of nontraditional students in postsecondary education. *Strategic Enrollment Management Quarterly* 5, 4 (2018), 159–164.

[17] OpenStax. 2026. https://openstax.org/details/books/intermediate-algebra-2e

[18] Zilong Pan and Min Liu. 2022. The Role of Adaptive Scaffolding System in Supporting Middle School Problem-Based Learning Activities. *Journal of Educational Technology Systems* 51, 2 (2022), 117–145.

[19] Andrea Phillips, John F Pane, Rebecca Reumann-Moore, and Oluwatosin Shenbanjo. 2020. Implementing an adaptive intelligent tutoring system as an instructional supplement. *Educational Technology Research and Development* 68, 3 (2020), 1409–1437.

[20] Chris Quintana, Brian J Reiser, Elizabeth A Davis, Joseph Krajcik, Eric Fretz, Ravit Golan Duncan, Eleni Kyza, Daniel Edelson, and Elliot Soloway. 2018. A scaffolding design framework for software to support science inquiry. In *Scaffolding*. Psychology Press, 337–386.

[21] Jennifer Reddig, Glen R. Smith Jr, Sanaz Ahmadzadeh Siyahrood, Wesley G. Morris, Yoojin Bae, Kaitlyn Crutcher, John Kos, Rahul K. Dass, Jinho Kim, Momin Naushad Siddiqui, Daniel Weitekamp, Ploy Thajchayapong, Sandeep Kakar, Alex Endert, Scott Crossley, Min Kyu Kim, Chris Dede, Ashok K. Goel, and Christopher J. MacLellan. 2026. Guidelines for Designing AI Technologies to Support Adult Learning. In *Proceedings of the Designing Interactive Systems Conference (DIS '26)* (Singapore, Singapore). ACM, New York, NY, USA, 1–23. doi:10.1145/3800645.3813102

[22] Steve Ritter, Michael Yudelson, Stephen E Fancsali, and Susan R Berman. 2016. How mastery learning works at scale. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. 71–79.

[23] Momin N Siddiqui, Adit Gupta, Jennifer M Reddig, and Christopher J MacLellan. 2024. HTN-based tutors: A new intelligent tutoring framework based on hierarchical task networks. In *Proceedings of the Eleventh ACM Conference on Learning@ Scale*. 491–495.

[24] Dustin Tingley, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. 2014. Mediation: R package for causal mediation analysis. *Journal of statistical software* 59 (2014), 1–38.

[25] HsingYi Tsai, Deborah Compeau, and Darren Meister. 2017. Voluntary use of information technology: an analysis and synthesis of the literature. *Journal of Information Technology* 32, 2 (2017), 147–162.

[26] Kurt VanLehn. 2006. The behavior of tutoring systems. *International journal of artificial intelligence in education* 16, 3 (2006), 227–265.

[27] Kurt VanLehn. 2011. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational psychologist* 46, 4 (2011), 197–221.

[28] David Wood, Jerome S Bruner, and Gail Ross. 1976. The role of tutoring in problem solving. *Journal of child psychology and psychiatry* 17, 2 (1976), 89–100.