Will They Try Again? A Large-Scale RCT on Scaffolds that Support Persistence in an Intelligent Tutoring System

Michael W. Asher
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
masher@andrew.cmu.edu

Yumou Wei
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
yumouw@andrew.cmu.edu

Adam Daniel Reynolds
Siyavula Foundation
Johannesburg, South Africa
adam@siyavula.com

Amy Ogan
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
aeo@andrew.cmu.edu

Paulo F. Carvalho
Human-Computer Interaction Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania, USA
pcarvalh@cs.cmu.edu

Abstract

Persistence after failure is critical for learning—but when students make mistakes in intelligent tutoring systems, they often choose not to try again. How can digital platforms encourage students to persist at these moments? We conducted a randomized controlled trial in an intelligent tutoring system for math and science, involving 164,532 students (Grades 8-12) who completed 17 million practice problems. We tested two scalable interventions: a brief persuasive prompt encouraging students to try again, and a visual default nudge that highlighted the retry option. Both interventions increased persistence after failure, and when combined, their effects were additive—suggesting they operate through distinct psychological mechanisms. The nudge had a much larger immediate effect, but the prompt showed proportionally greater spillover to untreated problems. These findings advance theories of persuasive design, demonstrating that implicit, interface-level nudges and explicit motivational prompts can be combined to avoid redundancy while amplifying impact.

CCS Concepts

• **Applied computing** → **E-learning**; **Interactive learning environments**; **Computer-assisted instruction**.

Keywords

Student engagement, Nudge interventions, Persuasive prompts

ACM Reference Format:
Michael W. Asher, Yumou Wei, Adam Daniel Reynolds, Amy Ogan, and Paulo F. Carvalho. 2026. Will They Try Again? A Large-Scale RCT on Scaffolds that Support Persistence in an Intelligent Tutoring System. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3772318.3791885

ACM ISBN 979-8-4007-2278-3/26/04
https://doi.org/10.1145/3772318.3791885

1 Introduction

Intelligent tutoring systems have become a core component of instruction in many classrooms, offering students self-paced practice, immediate feedback, and on-demand support [1, 23, 30, 60]. In these systems, the immediate feedback is critical for learning: real-time corrections allow students to identify misconceptions and learn from mistakes [22, 24, 40]. However, it comes with a tradeoff: students disengage at higher rates after making mistakes in digital environments—precisely when persistence is most beneficial [3, 14, 35]. A critical challenge, therefore, is to design intelligent tutoring systems that promote persistence after failure.

While many intelligent tutoring systems personalize content delivery, they rarely capitalize on their adaptive potential to maintain engagement over time. Interface features often remain static and one-size-fits-all, even at moments when students are most at risk of disengaging. As a result, the potential for these platforms to foster persistence through psychologically-informed design remains largely untapped.

In this paper, we examine two psychologically informed, adaptive interface features for promoting persistence after failure in a digital learning platform: default-based nudges, which subtly shape behavior by making retrying more visually salient, and persuasive prompts, which explicitly encourage students to try again. Although nudges and prompts are both applications of persuasive design—the intentional use of interactive systems to influence user attitudes and behaviors without coercion or deception [28, 47]—they operate through distinct psychological mechanisms. Whereas nudges reshape the structure of a choice by leveraging automatic, perception-driven processes, prompts influence behavior through explicit messaging that engages deliberate evaluation.

One might expect that combining two persuasive design features would amplify their effectiveness. However, research in social psychology suggests that simultaneous interventions operating through similar psychological mechanisms may produce redundant effects [32, 49]. Given that nudges and prompts both aim to promote persistence but potentially engage different processing pathways—automatic versus deliberative—a critical question about persuasive design emerges: can interventions that target different persuasive pathways operate independently? And more practically, can combining these interventions increase the rate that students retry problems, beyond what either achieves alone?

To answer these questions, we collaborated with Siyavula, a South African non-profit that develops educational technology, to conduct and analyze a large-scale randomized controlled trial. This study involved 164,532 students who completed approximately 17 million mathematics and science problems during the 2024 academic year.

This work makes three primary contributions. First, we present a large-scale, causal investigation about whether implicit (nudges) versus explicit (persuasive prompts) approaches to persuasive design can combine to increase student persistence after failure in a real-world, digital learning environment.

Second, by directly comparing these implicit and explicit strategies, we test how distinct elements of persuasive design influence behavior when they are combined. Whereas combining interventions with similar mechanisms can lead to redundancy and diminishing returns [32, 49], we examine whether pairing strategies with distinct psychological foundations can produce additive or even synergistic effects.

Third, we move beyond immediate effects to examine the longer-term consequences of these interventions, assessing whether they generalize to future problems, fade over time, and ultimately improve downstream learning outcomes.

We pursue these goals through minimal but psychologically informed technology-based interventions that target the critical moment after failure—an approach inspired by research on wise interventions showing that small, well-timed changes can produce substantial effects [33, 61, 64] while remaining readily scalable across educational platforms.

2 Related Work

2.1 Persistence and Productive Struggle

Persistence, the willingness to sustain effort in the face of difficulty, is a cornerstone of effective learning. Although learners can sometimes acquire knowledge through carefully scaffolded "errorless learning," a large body of research demonstrates that more durable and transferable understanding emerges when learners actively grapple with challenging material [44]. Decades of work on topics like the testing effect, deliberate practice, and active learning converge on the conclusion that students learn best when they are given opportunities to test their understanding, make mistakes, and refine their performance through practice and feedback [18, 27, 29, 52].

Research on productive failure illustrates this dynamic. Kapur [38] found that groups of students who were asked to collaboratively solve complex and unfamiliar, "ill-structured" problems achieved stronger conceptual understanding than peers who were given simpler, better-scaffolded versions of the same material . Although the ill-structured problems led to high rates of initial failure, the process of struggling with multiple potential solutions prepared students to recognize underlying structure when it was later introduced.

Other lines of research highlight similar benefits of struggle before instruction. Studies on pre-questioning and practice-centered instruction suggest that the challenge of upfront practice questions can pique curiosity and direct attention, making learners more likely to retain key ideas from subsequent lectures or readings [20, 41, 48]. For example, Carpenter and colleagues [19] found that when participants completed a set of short-answer practice questions before watching a history lecture, this helped them remember key details. Asher et al. [6] showed that learners who solved problems and received elaborated feedback—without any prior instruction—improved their memories and generalization abilities more efficiently than those who watched a lecture on the same content. Follow-up studies indicate that such practice enhances metacognitive calibration, helping learners identify gaps in their knowledge [4].

A complementary perspective comes from mastery learning, which assumes that—with sufficient time, feedback, and support—most students can achieve mastery of nearly any skill [12, 13]. This approach relies on an instructional structure that requires learners to demonstrate proficiency in prerequisite skills before advancing to new ones [13], a process that depends on repeated opportunities to practice and correct errors. However, prior research shows that students in digital environments are especially likely to disengage after making mistakes—precisely when persistence would be most beneficial [3, 14, 35]. If a learner routinely skips ahead after an error, the promise of mastery learning is undermined. Digital platforms, with their ability to deliver unlimited practice and immediate feedback, are well-positioned to support mastery at scale—but only if students are willing to use these opportunities.

In summary, struggling with difficult problems—whether through ill-structured tasks, prequestions, or practice with feedback— activates prior knowledge, improves metacognitive calibration, and sensitizes learners to the critical features of subsequent instruction. Persistence in the face of initial failure therefore is not merely about endurance; it is a process that prepares learners to integrate new information, encode it more deeply, and apply it flexibly to new situations.

2.2 Persuasive Design Through Choice Architecture and Nudges

Fogg's Behavior Model [28] provides a helpful framework for understanding persuasive design—how to intentionally shape user behavior in digital environments. According to the model, three elements must converge for a target behavior to occur: the user must be motivated to perform the behavior, able to do so, and the behavior must be triggered by the environment. Triggers can vary. Sometimes, when a user is already motivated and able, a simple "signal" can prompt them to act. If a user lacks ability, a "facilitator" can make the behavior easier. If a user lacks motivation, a "spark" can provide it.

One powerful way to facilitate behavior is through "choice architecture"—the way options are presented and structured in an environment [56]. Choice architecture is effective because people rarely weigh all available information carefully when making decisions. Instead, when faced with complex choices, they frequently apply mental shortcuts to contextual cues [37, 58]. For instance, people often exhibit a status quo bias, assuming that whatever option is presented as the default must be the recommended or best choice [43, 53]. To capitalize on this, a persuasive designer can set the desired behavior as the default option, leveraging status quo bias to make the behavior easier.

Interventions like this have been called "nudges" by behavioral economists [56]. By reducing cognitive effort or highlighting certain options, nudges can steer behavior while preserving individual autonomy to choose.

Digital environments are particularly well-suited for implementing nudges because designers have precise control over the choice architecture, allowing them to carefully engineer how decisions are framed [17, 62]. Digital platforms also create new possibilities for how nudges can be implemented. For example, while defaults traditionally involve pre-selecting an option [36, 55], in digital platforms designers can convey the principle of defaults—presenting one option as the standard path—through visual salience. Visual features such as button size, color, contrast, or position can highlight a particular option, subtly guiding user attention and shaping choice [34]. Digital designers frequently use this principle to distinguish between primary and secondary actions—for example, by presenting a bright "Save" button next to a muted "Cancel."

Digital nudges have raised ethical concerns when used to manipulate users against their interests—for instance, exploitative "dark patterns" in cookie consent designs that prioritize data collection over user privacy [31, 42, 59]. Berdichevsky and Neuenschwander [9] distinguish ethical persuasion, which respects user autonomy and benefits users, from unethical persuasion, which deceives or manipulates users to act against their best interests. When nudges meet these criteria of user benefit and respect for autonomy, they can support beneficial outcomes. For example, changing the default setting on campus copiers to double-sided printing reduced paper use by 15% [26], and setting renewable energy as the default for electricity contracts increased green energy purchases nearly tenfold [25].

Educational contexts offer a similarly clear case for ethical persuasive design: when platforms guide students toward persistence and deeper engagement, the interests of designers and learners align. Moreover, in educational technology where many platforms fully control problem selection and remove student agency entirely [39, 60], well-designed nudges can actually increase autonomy by introducing meaningful choices—for instance, a highlighted "Try Again" button after failure communicates a recommended path while preserving student choice about how to proceed.

While the effects of default nudges are well established in consumer domains, often in the context of online privacy decisions [8, 10, 59], their application in education remains underexplored. This gap presents both a theoretical and practical opportunity to examine whether small adjustments to interface design can guide learners toward deeper engagement at moments when persistence is at risk.

2.3 Persuasive Prompts

Unlike nudges, which facilitate behavior by reducing friction with choice architecture, persuasive prompts work as motivators—explicitly encouraging learners to persist after failure by supplying reasons, offering support, or reframing the meaning of trying again. In Fogg's Behavior Model [28], when users have the ability to perform a behavior but lack motivation, a "spark" can provide the necessary impetus to act. Persuasive prompts serve this function: rather than altering the structure of the choice itself, they aim to increase students' willingness to retry by making re-engagement feel more worthwhile, supported, or aligned with their goals.

The Persuasive Systems Design (PSD) model [47] extends Fogg's framework by identifying specific design principles organized into four categories: primary task support, dialogue support, system credibility, and social support. Persuasive prompts draw primarily on principles of dialogue support—features that provide feedback to guide users toward target behaviors—including praise, reminders, and suggestions to motivate continued engagement. They may also leverage social support through appeals to peer behavior and system credibility principles such as authority and expertise.

Consider, for example, this message:

> "Why not try again? Our questions have worked solutions to help you learn. Work through the solution, then try the question again! Many students perform better on their second try–you can do it!

Although it is brief, this message operationalizes several PSD principles simultaneously. It employs dialogue support through suggestion (recommending use of worked solutions), personalization ("to help you learn") and praise ("you can do it!"). It invokes system credibility through authority cues (positioning the platform as an effective learning tool). It also leverages social support via social comparison (referencing peer performance). These design choices also align with broader principles from the social psychology of persuasion [46]. The inclusion of a rationale supports central-route processing [50], where individuals are persuaded by the content of a message rather than by superficial features, and the reference to peer behavior serves as social proof (signaling that a behavior is common or normative) [21].

Prior work supports the effectiveness of brief, persuasive messages in digital learning contexts. Yeomans and Reich [65] found that short prompts encouraging students to reflect on goals and strategies improved online course completion, particularly when messages emphasized autonomy and capability. Dozens of studies show that prompting learners to periodically explain things to themselves (e.g., as they complete readings or math problems) can improve learning outcomes [2, 11, 51]. Growth mindset interventions have demonstrated that prompting students to reframe mistakes as opportunities to learn can increase their willingness to persist through difficult tasks [16, 63].

Taken together, this body of work suggests that persuasive prompts can increase the likelihood that learners persist through failure by directly addressing motivation—supplying rationales, providing affective support, and encouraging re-engagement. Whereas nudges operate through implicit, perception-driven mechanisms that facilitate behavior, persuasive prompts engage deliberate reasoning to spark motivation. Given their distinct psychological mechanisms, an important question is whether these intervention approaches can be complementary when used together—each addressing a different element in Fogg's behavioral framework [28]—or whether they might overlap or interfere.

3 The Present Research: Combining Persuasive Prompts with Default Nudges

Because persuasive prompts and default nudges rely on distinct psychological mechanisms, dual-process models of persuasion [50] offer a useful lens for thinking about their combined effects. Visual salience functions peripherally: it captures attention quickly, reduces cognitive effort, and implicitly signals endorsement. Persuasive prompts, by contrast, engage central processing by providing reasons and explicitly encouraging persistence. In behavioral economics, this contrast has been described as the difference between fast, effortless "System 1" responses and slower, more effortful "System 2" thinking [37].

But what happens when these automatic and deliberative interventions are deployed together? One possibility is redundancy: if either intervention is already sufficient to guide behavior, their joint use may offer little additional benefit. For example, a salient visual cue and a motivational message might each independently increase the likelihood that a student retries a problem, but combining them may not produce further gains.

A second possibility is additivity: because the interventions act on different cognitive routes, their effects may accumulate. In this case, learners might be most likely to persist when their attention is captured by a salient cue and they simultaneously receive a rationale that encourages effort.

A synergistic result may also occur if many students lack both the motivation and ability to repeat problems: in this case a nudge may lower friction (making the behavior easier) but students may still avoid repeated problems if they do not see the value. This is the prediction that emerges from Fogg's Behavioral Model of persuasive design, as well as expectancy-value theories of motivation, which suggest that students need both ability and motivation to persist [5, 45, 57].

A final possibility is interference: rather than boosting each other, the prompt and nudge could inadvertently undermine each other. For instance, if the presence of both creates cognitive overload, the combined experience might feel confusing or overwhelming, leading learners to ignore the interventions altogether. Although we expect that this possibility is unlikely because the nudge should be subtle and processed unconsciously, it is important to monitor for potential interference.

Understanding whether default nudges and persuasive prompts are redundant, additive, synergistic, or interfering would advance theory by clarifying how central and peripheral pathways jointly shape student decision-making. It would also inform the persuasive design of digital learning platforms that aim to support persistence using both interface-level cues and motivational messaging.

We investigate these possibilities through a randomized controlled trial involving 164,532 students, conducted in partnership with Siyavula, a South African educational technology provider. The interventions were deployed across Siyavula's mathematics and science platforms (Grades 8–12) during the 2024 academic year, taking place in real classrooms with authentic coursework, occurring across millions of learning interactions. To our knowledge, this represents one of the largest experimental studies of educational technology conducted to date.

Our study was guided by five research questions (RQs):

- **RQ1**: Can visual default nudges and/or persuasive prompts encourage students to repeat challenging problems after failure in an online learning environment?
- **RQ2**: Are the effects of these interventions additive, redundant, synergistic, or interfering?
- **RQ3**: Do the effects of the interventions generalize to problems where the nudges are not shown (i.e., are there spillover effects)?
- **RQ4**: Do the interventions lose effectiveness with repeated exposures over time?
- **RQ5**: Do the interventions, individually or in combination, have downstream effects on student learning?

We hypothesized that both interventions would increase persistence following failure (RQ1). However, we did not advance a directional hypothesis regarding RQ2, as theoretical arguments support the possibility of redundancy, additivity, synergy, or interference. For RQ3, we expected that visual nudges would have little to no spillover effect because their influence is likely non-conscious and context-bound. In contrast, we predicted that persuasive prompts could yield spillover effects by shaping students' conscious beliefs and learning strategies—encouraging persistence even in the absence of a prompt.

Regarding RQ4, we anticipated different patterns of fade-out across the two interventions. Because visual nudges operate through low-effort mechanisms, we expected minimal fatigue. Persuasive prompts, however, might become less effective over time if students experience motivational depletion or cognitive fatigue from repeated effort regulation.

Finally, for RQ5, we hypothesized that persistence following failure—encouraged by either intervention—would enhance learning outcomes by increasing exposure to corrective feedback and practice opportunities aligned with the relevant learning objectives.

4 Methods

Siyavula is a South African non-profit that develops educational technology to support math and science learning. Its core product is an intelligent tutoring system (ITS) that serves secondary school students in Grades 8 through 12 for math and Grades 10 through 12 for science. For these subjects, Siyavula provides curriculum-aligned exercises that are automatically scored and accompanied by worked solutions. The present research took place in the context of the Siyavula ITS and these curricula. The research was covered by Siyavula's terms of service, in compliance with South Africa's Protection of Personal Information Act (POPIA). Direct consent for this research was not collected from participants by the authors. The relevant IRB at Carnegie Mellon University deemed analysis of this data to be exempt research with non-identifiable participants.

4.1 System Design

Within Siyavula's ITS, practice problems are organized into chapters by general topic (e.g., factorization of algebraic expressions), and further grouped into narrower sections with more specific learning objectives (e.g., identifying common factors or factorizing quadratic expressions). Within sections, each practice exercise is based on a template that allows surface features of the problem and solution (e.g., numbers and names) to be automatically generated. This design enables students to repeatedly practice "isomorphic" versions of the same problem, which are nearly identical in structure but vary in surface details. Figures 1A and 1B show two isomorphic versions of an algebra problem with the learning objective "factorizing quadratic expressions."

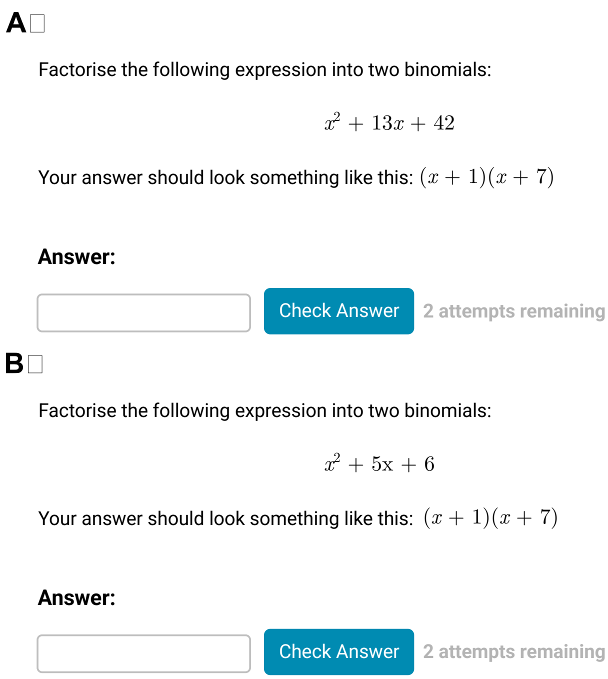

Figure 1: Isomorphic versions of a Siyavula factorization problem

After completing each practice exercise, students receive feedback about the correctness of their answer as well as a worked solution that explains how the correct answer was obtained. Students are then presented with two options (Figure 2): advance to the next exercise (typically a related problem) or repeat an isomorphic version of the problem they just completed.

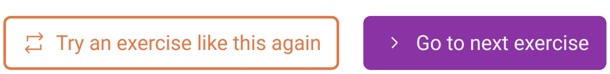

Figure 2: The standard options shown to students after each practice exercise

4.2 The Randomized Controlled Trial

To examine how visual default nudges and persuasive prompts could affect students' persistence after failure (and their subsequent mastery of the associated objectives), Siyavula conducted a randomized controlled trial (RCT) with new users in its mathematics and science curricula during the 2024 school year.

The experiment used a between-subjects, 2 x 2 factorial design to test the independent and combined effects of visual default nudges and persuasive prompts. Upon sign-up, all participants were randomly assigned to one of four experimental conditions: (1) a control group that received no intervention, (2) a visual default nudge-only group, (3) a persuasive prompt-only group, and (4) a combined "prompt + default" group that received both interventions. Students remained in their assigned condition for the duration of the study.

4.2.1 Design of the Visual Default Nudge and Persuasive Prompt. The default nudge was implemented as a visual design cue in the choice students saw after each practice exercise. When students were eligible, the button labeled "Try an exercise like this again" was highlighted in bright orange to draw attention and implicitly recommend this option. Otherwise, students saw the version of the buttons shown in Figure 2, in which "Go to next exercise" was highlighted in purple as the default.

The persuasive prompt was implemented as a text box that appeared at the bottom of the screen (above the "try again" and "next exercise" buttons) after an incorrect attempt, encouraging students to examine the solution and try again. To reduce repetitiveness, the prompt's exact wording varied, but it consistently included three components: a reminder, a rationale, and encouragement. Figure 3 illustrates the visual nudge and provides an example of the persuasive prompt.

4.2.2 Eligibility for the Interventions. Students became eligible for an intervention when they made a *fully incorrect* first attempt on a practice exercise (i.e., a templated item with randomly generated values). For exercises with multiple questions on the same page, students were only eligible when all questions were answered incorrectly. Eligibility was determined at the level of the individual practice session, and students could re-enter eligibility if they revisited the same templated item in a new session (e.g., after logging out or returning to the main dashboard and then re-entering a section). Overall, students made 2,398,813 eligibility-qualifying mistakes during the study.

For the students who were assigned to conditions with visual default nudges, the nudges were shown consistently whenever eligibility criteria were met. In contrast, prompts were shown probabilistically to reduce fatigue from excessive pop-up messages. On a learner's first day on the platform, prompts appeared on 60% of eligible exercises. This probability declined by 10 percentage points each day until reaching 20% on day 5, and then it dropped to a stable floor of 5% from day 6 onward. A subset of students (the control group) did not receive nudges or prompts when they were otherwise eligible.

4.3 Measures

4.3.1 Repeated problems. To examine whether students chose to repeat problems, we analyzed time-stamped log files of all responses

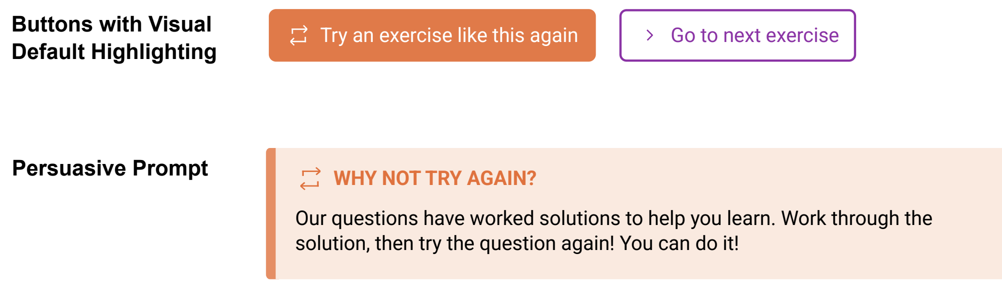

Figure 3: Intervention visuals for the nudge and prompt.

submitted between January 1 and October 5, 2024. We defined a "repeated" attempt as one in which a student chose to solve another problem generated from the same template as the problem they had just completed. In other words, if two consecutive responses came from isomorphic problems—different in surface details but identical in structure—the second was coded as a repeat. For example, if a student first solved Question A in Figure 1 and then immediately attempted Question B, the second attempt would be considered a repeated problem.

Because the Siyavula platform's practice algorithm is designed to avoid presenting the same template twice in a row unless a student actively clicks the "Try Again" button, repeated attempts provide a clear behavioral signal of persistence: they reflect a deliberate decision to revisit the same type of problem. During the study, students submitted 16,999,381 responses to 4,499 distinct problem templates, including 14,199,246 first attempts and 2,800,135 repeated attempts.

4.3.2 Mastery of Associated Practice Problems. To examine the downstream effects of the interventions on learning, we examined whether students eventually solved at least one isomorphic problem (generated from the same template as the original exercise) correctly. Overall, after making 2,398,813 initial errors that qualified students for an intervention, students eventually solved 60% of the associated problems correctly.

4.4 Participants

To determine the sample for the study, we began with all 164,697 students in Grades 8 - 12 who enrolled in a Siyavula math or science course, completed at least one grade-level practice problem, and were assigned to an experimental condition between January 1st 2024 (when the study began) and October 5th 2024 (when the Siyavula team concluded the study and shared their data). While the majority of students in the dataset submitted a modest number of responses during this study period (the median was 37 problems and 99% submitted fewer than 1,470 responses), a small number submitted implausibly large totals (e.g., one account submitted more than 70,000 problems), suggesting that an account may have been shared. To address these extreme cases, we applied a conservative statistical cutoff at the 99.9th percentile of the distribution of practice problems completed. This threshold corresponded to 3,819 problems submitted by a student. Students above this cutoff (N = 165) were excluded from analyses, leaving a final analytic sample of 164,532 students: 41,156 in the control group, 41,178 in the prompt group, 40,976 in the nudge group, and 41,222 who received both interventions.

Among students in the final sample, 94% (N = 154,690) were enrolled in Siyavula's mathematics courses and 35% (N = 57,172) were enrolled in science courses (47,330 students, or 29%, were enrolled in both). Overall, the largest proportion of students was in Grade 10 (28%), followed by Grades 8 (27%), 9 (19%), 11 (15%), and 12 (11%). Other demographic data (e.g., gender, race) were not collected.

5 Results

To address research questions 1-5, we estimated two types of mixed-effects models in R (version 4.5; [54]) using the glmmTMB package [15]. For continuous outcomes, we fit linear mixed-effects models with Gaussian errors and an identity link. For dichotomous outcomes, we fit generalized linear mixed-effects models with binomial errors and a logit link (i.e., logistic mixed-effects models). All models included three between-subjects predictors: a contrast-coded *Default Nudge vs. No Nudge* variable (Nudge = .5, No Nudge = -.5), a contrast-coded *Persuasive Prompt vs. No Prompt* variable (Prompt = .5, No Prompt = -.5), and their interaction.

To address research questions 2-5, we also included additional moderators in two- and three-way interactions with the intervention contrasts (e.g., an "exposure number" variable to test whether intervention effects changed over time). These models are described in more detail in their respective sections of the Results. To account for non-independence due to the nested structure of the data (because students completed multiple problems each), all models included a by-student random intercept, along with by-student random slopes for any predictors that varied within subjects (e.g., exposure), following the recommendations of Barr [7].

Project data are openly available at https://osf.io/ftzgh, along with code to reproduce all analyses.

5.1 RQ1 and RQ2: Do the interventions work? Are the effects additive or redundant?

To test the independent and combined effects of the two interventions on the primary outcome of the study (whether students chose to repeat math problems after failure), we started with the

2.4 million problem attempts in the dataset where students made a mistake that qualified them for an intervention (getting a problem in a practice section fully incorrect for the first time) and we used our primary regression model to test the likelihood that students in each condition chose to repeat the problem. Figure 4 shows the results of this analysis.

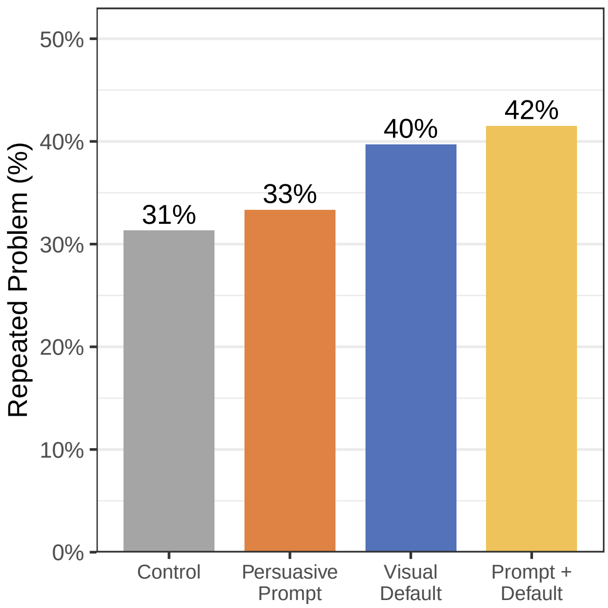

Figure 4: The interventions additively promoted persistence after failure

Overall, both interventions increased the rate at which students persisted after failure. In the control condition, participants chose to repeat a failed problem 31% of the time. The persuasive prompt increased this rate by 2 percentage points (OR = 1.10, z = 10.19, $p < .001$), while the nudge had a larger effect, increasing persistence rates by approximately 9 percentage points (OR = 1.65, z = 52.27, $p < .001$). This difference may reflect the interventions' delivery schedules: the prompt was shown probabilistically—on 60% of eligible problems on day 1, declining to a 5% floor by day 6—whereas the nudge was shown on all eligible problems. We return to this possibility in our analysis of how effectiveness changed with repeated exposures.

The analysis also revealed no interaction between the two interventions (OR = 1.00, z = -0.17, p = .864), suggesting that each influenced persistence independently.

5.2 RQ3: Do intervention effects generalize beyond eligible problems?

To test whether the interventions had any spillover effects—that is, whether they influenced students' behavior even when a prompt or nudge was no longer active—we examined a second category of errors: practice exercises in which students were partially correct. This analysis was possible because students were only eligible to receive an intervention after fully incorrect responses (i.e., when all components of a multi-question exercise were answered incorrectly). In contrast, with partially correct responses students were ineligible for any prompt or nudge. There were 1,123,523 partially correct responses that occurred after a student previously received an intervention.

If either intervention had a lasting influence on how students responded to failure, we would expect students in the corresponding conditions to be more likely to reattempt partially incorrect items, even though no intervention was shown. Conversely, if the effects of the interventions were entirely contingent on their presence at the point of failure, then treatment differences observed for fully incorrect responses (Figure 4) should vanish when students made only partial errors. Figure 5 plots the rate at which students chose to repeat problems following a partially correct response.

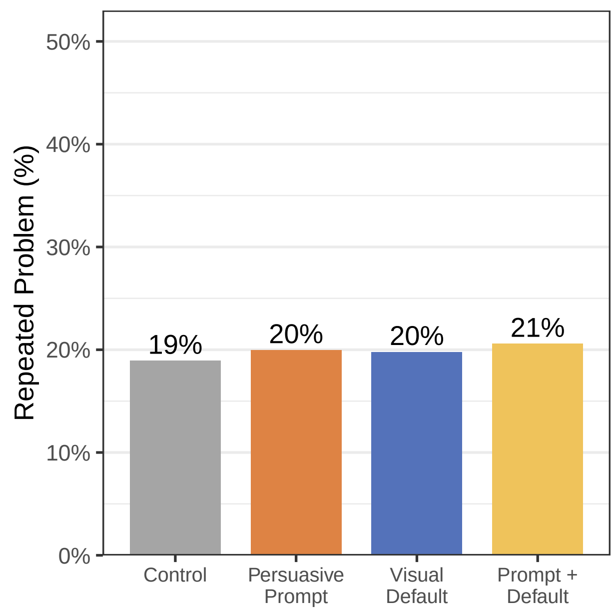

Figure 5: Limited spillover of intervention effects to non-intervention problems

Across all conditions, students were less likely to repeat problems after partial errors than after complete failure. For example, students in the control group reattempted 19% of partially incorrect problems, compared to 31% of fully incorrect ones. Treatment effects were also substantially attenuated, shrinking to approximately one percentage point for each intervention individually and two points when both were combined. Although these effects were statistically significant (for the prompt: OR = 1.03, z = 2.64, p = .008; for the nudge: OR = 1.08, z = 5.97, $p < .001$), their magnitude was extremely small.

Nonetheless, the pattern of attenuation is informative. The absence of the visual default nudge was associated with a 90% decline in repeat rates relative to control (from 9 points when the nudge was present to 1 point when it was removed). For the persuasive prompt, the reduction was smaller—about 50%—with the treatment-control difference dropping from 2 points (on fully incorrect problems) to 1 point (on partially incorrect problems).

This asymmetry suggests that the prompt's influence may be somewhat less dependent on its continued presence than that of the nudge, which appeared to have more transient, context-bound effects. Again, the interaction between the prompt and the nudge was non-significant (OR = 1.02, z = 0.65, p = .513).

5.3 RQ4: Do visual nudges and prompts lose effectiveness with repeated exposure?

We hypothesized that the effectiveness of nudges and prompts might diminish over time, and we considered two distinct mechanisms that could contribute to such a fade-out. First, both interventions might lose their impact as students accumulate exposures. With repeated use, the interventions may become less novel or engaging, students may grow tired of retrying similar problems, or students may become frustrated after retrying some problems but continuing to struggle. This form of decline would reflect a general waning of responsiveness due to habituation or fatigue.

Second, we considered whether the prompt's declining effectiveness might stem from the intervention's design. To minimize disruption, the prompt was delivered with decreasing frequency: it was shown 60% of the time on a learner's first day, decreasing each day until it reached a 5% floor on day six, after which the probability remained constant. In contrast, the visual nudge was shown consistently whenever students met the eligibility criteria. This difference in delivery raises the possibility that potential reductions in prompt effectiveness may reflect differences in exposure, rather than a loss of learner receptiveness.

To test these possibilities, we estimated a mixed-effects model including both *within-day* and *between-day* exposure terms. The within-day variable indexed how many problems a student had already attempted on a given day, capturing the effects of repeated exposures within a single session. The between-day variable tracked how many different days a student had logged in and practiced, reflecting longer-term engagement. Interaction terms between each intervention and both exposure variables were included to assess whether treatment effects declined with time. Because both exposure variables varied within individuals, we included by-subject random slopes for each. To aid convergence, we removed correlations among random effects.

This approach allowed us to distinguish between two types of fade-out: one reflecting a general decline in effectiveness over time (due to fatigue or habituation), and one specific to the decreasing frequency of prompt delivery. If intervention effects decline because students become less responsive over time, we would expect diminishing treatment effects both within and across days. If the decline is instead a function of delivery frequency, then we would expect only between-day changes for the prompt, and those changes should level off beginning on day six, once delivery probability plateaus. To test for this leveling-off, we also fit a secondary model using only responses submitted after day five (N = 1,276,128).

5.3.1 Between-Day Changes: Prompt Effectiveness Declines as Delivery Decreases. Figure 6 shows the predicted probability of students reattempting an eligible problem over time in each condition. Predictions for the first five days (plotted with dotted lines) come from the full model fit on all data. This model showed that overall persistence rates declined gradually with each additional day (OR = 0.97, z = -35.22, p < .001), and that the visual nudges slowed the day-to-day decline very slightly (OR = 1.01, z = 6.18, p < .001). In contrast, the effectiveness of the prompt decreased modestly over time (OR = 0.996, z = -2.93, p = .003).

Predictions for days six and beyond (plotted solid lines) come from the secondary model fit only to data from days six and higher. In this interval—after the prompt delivery probability had stabilized—the visual nudges continued to grow very slightly more effective each day (OR = 1.003, z = 2.06, p = .039), and critically, the prompts no longer decreased in their effectiveness with each day (OR = 1.00, z = 1.03, p = .304).

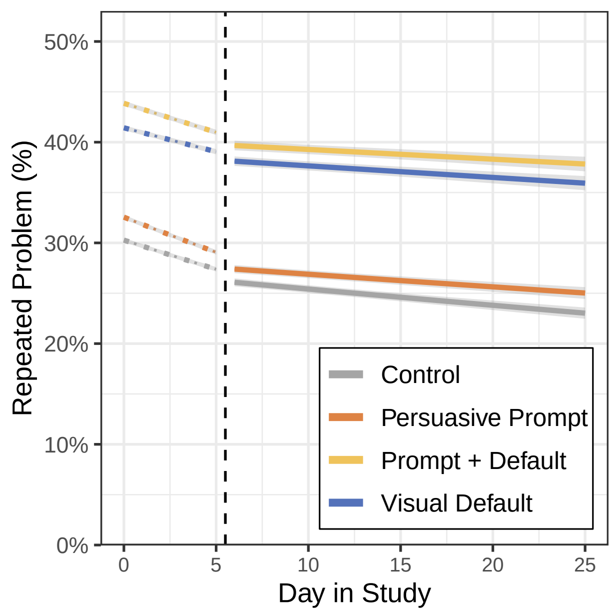

Figure 6: Between-day changes in intervention effectiveness

5.3.2 Within-Day Changes: Prompt Effects Remain Stable; Slight Decline in Nudge Effects. Figure 7 shows the predicted probability of repeating an eligible problem as students encountered multiple failures within a single day. As with the between-day pattern, persistence declined over successive errors (OR = 0.94, z = -69.86, p < .001). However, the effectiveness of the interventions remained largely stable. Relative to control, the effectiveness of the visual nudges declined very slightly with repeated within-day exposures (OR = .997, z = -1.89, p = .059), and the prompt showed no evidence of decline (OR = 1.00, z = 1.07, p = .283).

Together, these results suggest that the interventions were generally resilient to fatigue and habituation. Although the nudge showed a small decline in effectiveness within sessions, it actually became slightly more effective across days. The prompt's effectiveness also declined modestly across days, but this change appears to be driven by reduced delivery frequency rather than decreased student responsiveness. Once delivery stabilized after day six, the prompt's effectiveness did as well—suggesting that persuasive messages may retain their impact over time when delivered with consistent probability.

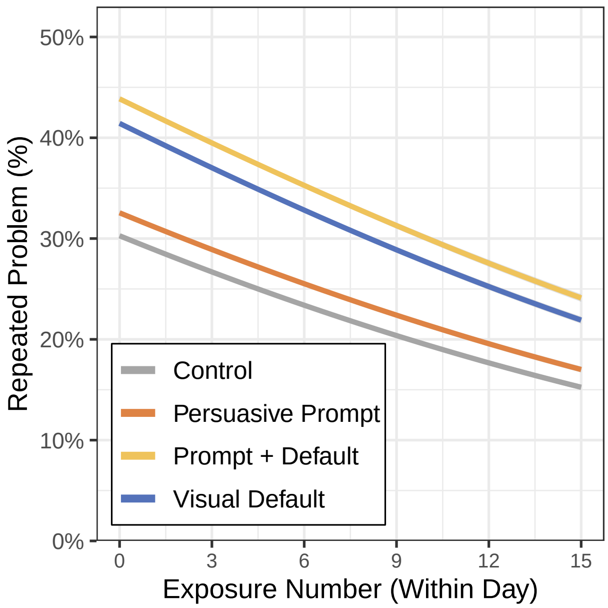

Figure 7: Within-day changes in intervention effectiveness

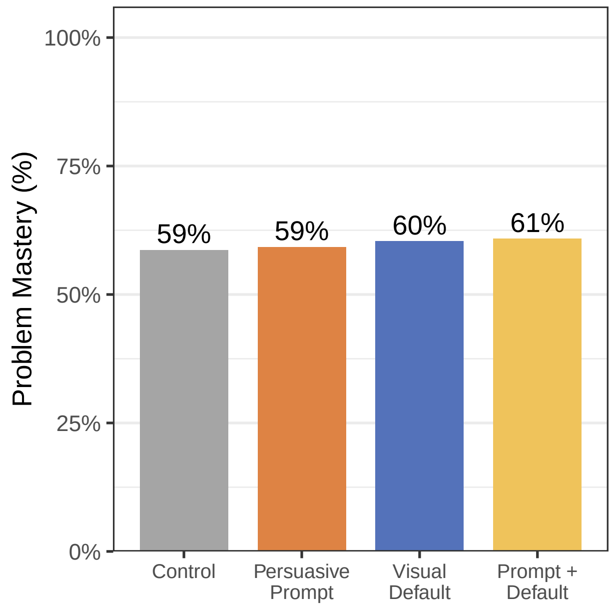

Figure 8: Interventions had statistically significant but very small downstream effects on learning

5.4 RQ5: Do the Interventions Have Downstream Effects on Learning?

Finally, to investigate whether the persuasive prompts and visual nudges had downstream effects on learning, we examined whether students in each condition went on to eventually demonstrate mastery of a problem after getting it completely incorrect (and qualifying for an intervention, if eligible). Importantly, although Siyavula's practice system does not serve the same templated problem twice in immediate succession unless a student chooses to try again, it will often return to that same problem later—particularly when it is one of the few remaining problems required for a student to master a topic. This means that even learners who did not immediately click "Try Again" after an error still had the opportunity to revisit and solve that problem later. Figure 8 displays the results of this analysis.

Overall, the effects of both the prompt (OR = 1.02, $z = 3.02$, $p = .003$) and this nudge (OR = 1.11, $z = 14.44$, $p < .001$) were statistically significant, but their interaction was not (OR = 1.02, $z = 0.60$, $p = .551$), suggesting very small and additive benefits of the two interventions on learning. In the control condition, students made 467,551 initial errors on templated problems, and ultimately went on to solve 58.7% of these problems correctly. Alone, persuasive prompts increased this "solve" rate to 59.2%, the visual default nudges increased it to 60.3%, and the combination of prompts and nudges increased it to 60.8%.

6 Discussion

This study examined whether two lightweight interface design interventions—a visual default nudge and a brief persuasive prompt—could promote persistence after failure in a large-scale digital learning environment. With a randomized controlled trial that included 164,532 students (and analysis of behavior following nearly 2.4 million errors), we tested whether these interventions increased students' likelihood of repeating challenging problems after failure (RQ1), and whether their effects were additive, redundant, synergistic, or interfering when combined (RQ2). We also asked whether their influence generalized beyond directly targeted problems to affect behavior more broadly (RQ3), whether their effectiveness diminished with repeated exposures (RQ4), and whether encouraging persistence ultimately improved students' learning outcomes (RQ5).

Beyond documenting behavioral effects, our findings advance theoretical understanding of persuasive design in interactive systems—specifically, how implicit and explicit persuasive strategies operate independently and in combination. In the sections that follow, we discuss what our results reveal about the mechanisms underlying persuasive interaction design, articulating generalizable design principles for supporting persistence after failure.

6.1 Independent Mechanisms and Implications for Persuasive Design

Both interventions improved persistence after failure (RQ1), and critically, their effects were fully additive (RQ2). The visual default nudge increased reattempt rates by 9 percentage points, the prompt increased them by 2 percentage points, and the combination increased them by 11 percentage points—with no statistical interaction. This pattern of independent, non-overlapping effects provides strong evidence that implicit nudges and explicit prompts operate through distinct psychological mechanisms.

This finding has important implications for theories of persuasive design. Fogg's Behavior Model [28] posits that three elements must converge for behavior to occur: motivation, ability, and a trigger. According to this model, when users lack ability, "facilitators" make behavior easier; when they lack motivation, "sparks" provide it; and when both are present, simple "signals" suffice. Our nudge was designed to be a facilitator—reducing friction by making the retry option visually salient and perceptually easier to select. Our prompt was designed as a spark—providing encouragement and a rationale to boost motivation.

The fact that the nudge succeeded equally well without the motivational prompt raises questions about whether motivation is always necessary for behavior change. The nudge was highly effective on its own, suggesting that implicit strategies operating below conscious awareness may influence behavior without the need for deliberate motivation. By making the target behavior perceptually dominant and cognitively effortless, the nudge may have enabled students to persist "automatically"—before conscious motivation could become relevant.

This challenges strict interpretations of the Behavior Model that treat motivation as a necessary precondition for action. If students can be guided to retry simply by making that option visually salient, then perhaps ability (or reduced friction) can sometimes substitute for motivation rather than merely complement it.

At the same time, the prompt increased persistence even when the nudge was absent, indicating that motivation can drive behavior independently. Students who read the encouraging message consciously decided to retry, despite the visual interface providing no special guidance toward that option. This demonstrates that explicit motivational appeals can succeed on their own when the target action is sufficiently accessible.

The independence of these effects is particularly noteworthy when compared to prior work on combining explicit interventions. When researchers layer multiple message-based strategies—such as combining growth mindset interventions with belonging interventions—they often observe diminishing returns [32, 49]. This redundancy likely occurs because multiple explicit interventions engage overlapping psychological processes, leading to cognitive overload, motivational fatigue, or redundant messaging.

Our results suggest a key design principle: pairing implicit interface-level changes with explicit motivational messaging may circumvent the redundancy problem that plagues combinations of explicit interventions. Because implicit nudges target automatic, perception-driven processes while explicit prompts engage deliberate evaluation, they may be able to work in parallel without interference. For interaction designers, this implies that layering persuasive strategies across cognitive systems—rather than stacking multiple strategies within the same system—may be a more effective approach to behavior change.

This principle extends beyond education. Any interactive system seeking to influence user behavior might benefit from combining interface-level design choices (visual salience, default settings), with targeted messaging (reminders, encouragement, information). Our findings suggest these dual-layered approaches are likely to produce additive rather than redundant effects, maximizing persuasive impact without overwhelming users.

6.2 Comparing Implicit and Explicit Strategies

By directly comparing a nudge and a prompt within the same experimental design, we can assess not only whether they work, but how they differ in their effectiveness. Four patterns emerge that have implications for interaction design.

First, there were stark differences in the effect sizes of the two manipulations. The visual nudge produced substantially larger immediate effects than the prompt—9 percentage points versus 2 (RQ1). This disparity suggests that when designers seek to influence split-second decisions in interactive systems, low-friction interface design may exert stronger influence than motivational messaging. By subtly shifting attention and reducing decision effort, the nudge likely made retrying feel easier and more automatic. In contrast, the prompt required students to read and interpret a message; while it offered encouragement and a rationale, its effects were smaller—possibly because it relied on students' cognitive engagement in the moment.

Second, there were small but theoretically informative differences in spillover of the two manipulations. While the nudge was more powerful in-the-moment, the prompt showed proportionally greater spillover to non-targeted contexts (RQ3). When students were exposed to prompts on some problems, they became slightly more likely to persist on other problems where no prompt was shown—maintaining about half the original effect size. In contrast, the nudge's influence appeared tightly bound to the visual context in which it was delivered: when the highlighted button was absent, its effect size shrank to less than 15% of its original magnitude.

This pattern reveals an important trade-off: implicit nudges offer stronger immediate effects but remain context-bound, while explicit prompts produce weaker in-the-moment effects but may reshape users' broader beliefs and behaviors. For interaction designers, this suggests that the choice between implicit and explicit strategies should depend on design goals. If the objective is to maximize beneficial behaviors at specific decision points (e.g., ensuring users enable privacy settings), implicit nudges are likely more effective. However, if the goal is to foster more generalizable behavior change that extends beyond the immediate interaction (e.g., encouraging healthier habits, productive learning strategies, or sustained engagement), our results suggest that explicit prompts may offer greater long-term value despite their smaller immediate effects.

Third, both interventions maintained their effectiveness across sessions (RQ4). The nudge remained stable or even slightly increased in effectiveness over time, and the prompt's initial decline was fully explained by its decreasing delivery rate. Once prompt delivery stabilized, so did its effect. These results suggest that simple interventions can remain effective even with repeated use. In fact, maintaining consistent delivery is key to sustaining their impact

Finally, students who received either intervention were slightly more likely to eventually solve problems they initially got wrong—approximately 0.6 percentage points for the prompt and 1.6 points for the nudge (RQ5). While these gains are modest, they align with theories of mastery learning and productive struggle, which emphasize that learning depends not just on exposure to content but on learners' willingness to persist through errors and engage with corrective feedback [13, 38].

The modest effect sizes also reflect an important boundary condition. In the Siyavula system, even when students chose not to repeat a failed problem, they typically advanced to a closely related follow-up item. As such, the cost of "disengagement" was low, and repeating the original problem offered only marginal additional practice aligned with the learning objective. From a theoretical perspective, it is notable that repeating the same problem provided any benefit above and beyond practicing a related one—suggesting that error correction on the specific item may confer unique cognitive advantages. However, in interactive systems where the alternative to persistence is genuine disengagement—such as self-paced online courses, productivity tools, or applications where users can abandon tasks entirely—the impact of persistence-supporting interventions on outcomes could be substantially larger. The value of interventions encouraging persistence should scale with the cost of disengagement.

6.3 Limitations and next steps

Several limitations suggest directions for future work. First, the persuasive prompt was delivered probabilistically (60% of eligible moments on Day 1, declining to 5% by Day 6) to prevent user fatigue. While this design was intentional, it may have limited the prompt's impact. On Day 1, when exposure was highest, the prompt's effect reached nearly 3 percentage points—suggesting more consistent delivery might produce stronger effects. Future work should empirically determine optimal delivery frequencies, potentially using adaptive algorithms that adjust based on individual user responses.

The prompt was also deliberately brief and minimally disruptive for scalability. However, "wise" interventions that prompt deeper reflection or meaning-making have produced more substantial and generalizable effects [61, 64]. Future work could explore hybrid approaches combining frequent lightweight prompts with occasional reflective activities—for example, showing brief encouragement after most errors while periodically prompting users to reflect on what they're learning from mistakes.

Second, while we have argued for the generalizability of our findings, our empirical work was conducted entirely within mathematics and science learning. Testing these design principles in other domains (e.g., coding environments or fitness applications) would strengthen claims about generalizability and reveal potential boundary conditions. The relative effectiveness of implicit versus explicit strategies might differ when persistence requires different forms of effort.

Third, our analysis focused on average effects of the prompt and nudge across all classroom contexts, mathematics and science subjects, and problem types. In reality, these implicit and explicit strategies likely vary in their effectiveness across different contexts. For instance, completing an isomorphic problem after failure may be more beneficial for certain types of problems—such as those requiring procedural practice—than others that involve deeper conceptual reasoning. Similarly, the interventions may have different effects in mathematics versus science, where errors reflect different underlying challenges (e.g., computational mistakes versus misconceptions about scientific phenomena). The interventions may also vary in their effectiveness for different groups of students, based on the background knowledge, prior experience, and motivation-related beliefs that students have. For example, there may be a motivational "sweet spot" for a nudge, if they are unnecessary for highly motivated students (who persist in the absence of intervention) and ineffective for particularly unmotivated students (who may need a motivational "spark" rather than a facilitating nudge). Or perhaps because nudges operate implicitly, students' baseline levels of motivation will be less relevant. By understanding heterogeneous treatment effects across problem types, subjects, and student characteristics, researchers could enable adaptive, personalized systems that deploy the right intervention type for each user and context. This is an important future direction for research.

7 Conclusion

Across these findings, a consistent picture emerges: implicit and explicit persuasive strategies operate through distinct mechanisms, produce different behavioral signatures, and can be effectively combined to support persistence after failure. The visual nudge and persuasive prompt each increased student persistence, with fully additive effects that challenge theories requiring both motivation and ability for behavior change. Their contrasting profiles—nudges offering larger immediate but context-bound effects, prompts providing smaller but more generalizable influence—suggest design trade-offs that extend beyond educational technology to any interactive system where users encounter setbacks. While effect sizes on learning outcomes were modest, they were theoretically informative, revealing that persistence after failure can be shaped through interface design at critical moments of potential disengagement. Together, these results advance our understanding of how persuasive design elements operate in interactive systems and provide actionable guidance for designers seeking to support user persistence through thoughtful, ethically grounded interface choices.

Acknowledgments

This research was supported in part by National Science Foundation grant #2301130 to PFC, as well as a Carnegie Mellon University Seed Grant for Generative AI and Education Research to PFC and MWA, and a Mastercard Foundation Grant to AO. We thank our collaborators at Siyavula for their contributions to this project, which include designing and implementing this study and sharing data. During the preparation of this work the authors used Claude Sonnet 4.5 to copyedit the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Vincent Aleven, Elizabeth A McLaughlin, R Amos Glenn, and Kenneth R Koedinger. 2016. Instruction based on adaptive learning technologies. *Handbook of research on learning and instruction* 2 (2016), 522–560.

[2] Vincent A.W.M.M. Aleven and Kenneth R. Koedinger. 2002. An effective metacognitive strategy: learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science* 26, 2 (March 2002), 147–179. doi:10.1207/s15516709cog2602_1

[3] Michaela Arztmann, Jessica Lizeth Domínguez Alfaro, Lisette Hornstra, Jacqueline Wong, Johan Jeuring, and Liesbeth Kester. 2025. Game over? Investigating students' working memory, situational interest, and behavioral patterns as predictors of dropout in an educational game. *International Journal of Child-Computer Interaction* 43 (March 2025), 100721. doi:10.1016/j.ijcci.2025.100721

[4] Michael W. Asher and Paulo F. Carvalho. 2026. Conditions for effective learning without upfront instruction: How practice with feedback supports memory,

generalization, motivation, and metacognition. *Educational Psychology Review* 38, 1 (Dec. 2026), 12. doi:10.1007/s10648-025-10103-6
[5] Michael W. Asher, Cameron A. Hecht, Judith M. Harackiewicz, John J. Curtin, Cora Parrisius, and Benjamin Nagengast. 2025. Why elusive expectancy × value interactions may be critical for theory and intervention: A simulated power analysis. *Motivation Science* (April 2025). doi:10.1037/mot0000394
[6] Michael W. Asher, Faria Sana, Kenneth R. Koedinger, and Paulo F. Carvalho. 2025. Practice with feedback versus lecture: Consequences for learning, efficiency, and motivation. *Journal of Applied Research in Memory and Cognition* (Jan. 2025). doi:10.1037/mac0000205
[7] Dale J. Barr, Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68, 3 (April 2013), 255–278. doi:10.1016/j.jml.2012.11.001
[8] Jan M. Bauer, Regitze Bergstrøm, and Rune Foss-Madsen. 2021. Are you sure, you want a cookie? The effects of choice architecture on users' decisions about sharing private online data. *Computers in Human Behavior* 120 (July 2021), 106729. doi:10.1016/j.chb.2021.106729
[9] Daniel Berdichevsky and Erik Neuenschwander. 1999. Toward an ethics of persuasive technology. *Commun. ACM* 42, 5 (May 1999), 51–58. doi:10.1145/301353.301410
[10] Carlos Bermejo Fernandez, Dimitris Chatzopoulos, Dimitrios Papadopoulos, and Pan Hui. 2021. This website uses nudging: MTurk workers' behaviour on cookie consent notices. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (Oct. 2021), 1–22. doi:10.1145/3476087
[11] Kiran Bisra, Qing Liu, John C. Nesbit, Farimah Salimi, and Philip H. Winne. 2018. Inducing self-explanation: a meta-analysis. *Educational Psychology Review* 30, 3 (Sept. 2018), 703–725. doi:10.1007/s10648-018-9434-x
[12] James H Block and Robert B Burns. 1976. Mastery learning. *Review of research in education* 4 (1976), 3–49. Publisher: JSTOR.
[13] Benjamin S Bloom. 1968. Learning for mastery. Instruction and curriculum. Regional education laboratory for the Carolinas and Virginia, topical papers and reprints, number 1. *Evaluation comment* 1, 2 (1968), n2. Publisher: ERIC.
[14] Anthony F. Botelho, Ashvini Varatharaj, Eric G. Van Inwegen, and Neil T. Heffernan. 2019. Refusing to try: Characterizing early stopout on student assignments. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*. ACM, Tempe AZ USA, 391–400. doi:10.1145/3303772.3303806
[15] Mollie Brooks, Ben Bolker, Kasper Kristensen, Martin Maechler, Arni Magnusson, Hans Skaug, Anders Nielsen, Casper Berg, and Koen Van Bentham. 2017. glmmTMB: Generalized linear mixed models using template model builder. doi:10.32614/CRAN.package.glmmTMB Institution: Comprehensive R Archive Network Pages: 1.1.12.
[16] Jeni L. Burnette, Joseph Billingsley, George C. Banks, Laura E. Knouse, Crystal L. Hoyt, Jeffrey M. Pollack, and Stefanie Simon. 2023. A systematic review and meta-analysis of growth mindset interventions: For whom, how, and why might such interventions work? *Psychological Bulletin* 149, 3-4 (March 2023), 174–205. doi:10.1037/bul0000368
[17] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 ways to nudge: A review of technology mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–15. doi:10.1145/3290605.3300733
[18] Shana K. Carpenter, Steven C. Pan, and Andrew C. Butler. 2022. The science of effective learning with spacing and retrieval practice. *Nature Reviews Psychology* 1, 9 (Aug. 2022), 496–511. doi:10.1038/s44159-022-00089-1
[19] Shana K. Carpenter and Alexander R. Toftness. 2017. The effect of prequestions on learning from video presentations. *Journal of Applied Research in Memory and Cognition* 6, 1 (March 2017), 104–109. doi:10.1016/j.jarmac.2016.07.014
[20] Paulo F. Carvalho, Michael W. Asher, Faria Sana, and Kenneth R. Koedinger. 2024. Skip the reading assignment: Effective and efficient learning with only practice and feedback. doi:10.31219/osf.io/9y2tb
[21] Robert B. Cialdini. 2001. *Influence: science and practice* (4th ed ed.). Pearson Higher Education, Harlow.
[22] Albert T. Corbett and John R. Anderson. 2001. Locus of feedback control in computer-based tutoring: impact on learning rate, achievement and attitudes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Seattle Washington USA, 245–252. doi:10.1145/365024.365111
[23] Albert T Corbett, Kenneth R Koedinger, and John R Anderson. 1997. Intelligent tutoring systems. In *Handbook of human-computer interaction*. Elsevier, 849–874.
[24] Nada Dabbagh, Randall Bass, MJ Bishop, Anthony G Picciano, Jennifer Sparrow, Sarah Costelloe, Kristen Cummings, Brian Freeman, Michael Frye, and Allan Porowski. 2019. *Using technology to support postsecondary student learning: A practice guide for college and university administrators, advisors, and faculty.* Technical Report. Institute of Education Sciences, What Works Clearinghouse, National Center for Education Evaluation and Regional Assistance. Publisher: ERIC.
[25] Felix Ebeling and Sebastian Lotz. 2015. Domestic uptake of green energy promoted by opt-out tariffs. *Nature Climate Change* 5, 9 (Sept. 2015), 868–871. doi:10.1038/nclimate2681
[26] Johan Egebark and Mathias Ekström. 2016. Can indifference make the world greener? *Journal of Environmental Economics and Management* 76 (March 2016), 1–13. doi:10.1016/j.jeem.2015.11.004
[27] K. Anders Ericsson, Ralf T. Krampe, and Clemens Tesch-Römer. 1993. The role of deliberate practice in the acquisition of expert performance. *Psychological Review* 100, 3 (July 1993), 363–406. doi:10.1037/0033-295X.100.3.363
[28] Bj Fogg. 2009. A behavior model for persuasive design. In *Proceedings of the 4th International Conference on Persuasive Technology*. ACM, Claremont California USA, 1–7. doi:10.1145/1541948.1541999
[29] Scott Freeman, Sarah L. Eddy, Miles McDonough, Michelle K. Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences* 111, 23 (June 2014), 8410–8415. doi:10.1073/pnas.1319030111
[30] Arthur C. Graesser, Mark W. Conley, and Andrew Olney. 2012. Intelligent tutoring systems. In *APA educational psychology handbook, Vol 3: Application to learning and teaching.*, Karen R. Harris, Steve Graham, Tim Urdan, Adriana G. Bus, Sonya Major, and H. Lee Swanson (Eds.). American Psychological Association, Washington, 451–473. doi:10.1037/13275-018
[31] Colin M. Gray, Yubo Kou, Bryan Battles, Joseph Hoggatt, and Austin L. Toombs. 2018. The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–14. doi:10.1145/3173574.3174108
[32] Judith M. Harackiewicz, Elizabeth A. Canning, Yoi Tibbetts, Stacy J. Priniski, and Janet S. Hyde. 2016. Closing achievement gaps with a utility-value intervention: Disentangling race and social class. *Journal of Personality and Social Psychology* 111, 5 (2016), 745–765. doi:10.1037/pspp0000075 ISBN: 0022-3514.
[33] Judith M. Harackiewicz and Stacy J. Priniski. 2018. Improving student outcomes in higher education: The science of targeted intervention. *Annual Review of Psychology* 69, 1 (2018), annurev–psych–122216–011725. doi:10.1146/annurev-psych-122216-011725
[34] Laurent Itti and Christof Koch. 2001. Computational modelling of visual attention. *Nature Reviews Neuroscience* 2, 3 (March 2001), 194–203. doi:10.1038/35058500
[35] Annie Johansson, Alexander Olof Savi, Han L. J. Van Der Maas, and Abe Dirk Hofman. 2024. Three strikes and who is out? Individual differences in error-induced quitting. doi:10.31234/osf.io/6tduv
[36] Eric J. Johnson and Daniel Goldstein. 2003. Do defaults save lives? *Science* 302, 5649 (Nov. 2003), 1338–1339. doi:10.1126/science.1091721
[37] Daniel Kahneman. 2011. *Thinking, fast and slow* (1st ed ed.). Farrar, Straus and Giroux, New York.
[38] Manu Kapur. 2008. Productive failure. *Cognition and Instruction* 26, 3 (July 2008), 379–424. doi:10.1080/07370000802212669
[39] Kenneth R Koedinger, John R Anderson, William H Hadley, and Mary A Mark. 1997. Intelligent tutoring goes to school in the big city. (1997), 165322 Bytes. doi:10.1184/R1/6470153.V1 Artwork Size: 165322 Bytes Publisher: [object Object].
[40] James A. Kulik and Chen-Lin C. Kulik. 1988. Timing of Feedback and Verbal Learning. *Review of Educational Research* 58, 1 (1988), 79. doi:10.2307/1170349
[41] Jeri L. Little and Elizabeth Ligon Bjork. 2016. Multiple-choice pretesting potentiates learning of related information. *Memory & Cognition* 44, 7 (Oct. 2016), 1085–1101. doi:10.3758/s13421-016-0621-z
[42] Arunesh Mathur, Gunes Acar, Michael J. Friedman, Eli Lucherini, Jonathan Mayer, Marshini Chetty, and Arvind Narayanan. 2019. Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–32. doi:10.1145/3359183
[43] Craig R.M. McKenzie, Michael J. Liersch, and Stacey R. Finkelstein. 2006. Recommendations Implicit in Policy Defaults. *Psychological Science* 17, 5 (May 2006), 414–420. doi:10.1111/j.1467-9280.2006.01721.x
[44] Janet Metcalfe. 2017. Learning from errors. *Annual Review of Psychology* 68, 1 (Jan. 2017), 465–489. doi:10.1146/annurev-psych-010416-044022
[45] Benjamin Nagengast, Herbert W. Marsh, L. Francesca Scalas, Man K. Xu, Kit Tai Hau, and Ulrich Trautwein. 2011. Who took the "×" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological Science* 22, 8 (2011), 1058–1066. doi:10.1177/0956797611415540 ISBN: 1467-9280 (Electronic)$\backslash$n0956-7976 (Linking).
[46] Harri Oinas-Kukkonen and Marja Harjumaa. 2008. Towards Deeper Understanding of Persuasion in Software and Information Systems. In *First International Conference on Advances in Computer-Human Interaction*. IEEE, Sainte Luce, Martinique, 200–205. doi:10.1109/ACHI.2008.31
[47] Harri Oinas-Kukkonen and Marja Harjumaa. 2018. Persuasive systems design: Key issues, process model and system features. In *Routledge handbook of policy design*, Michael Howlett and Mukherjee Ishani (Eds.). Routledge, 87–105.
[48] Steven C. Pan and Shana K. Carpenter. 2023. Prequestioning and pretesting effects: a review of empirical research, theoretical perspectives, and implications for educational practice. *Educational Psychology Review* 35, 4 (Dec. 2023), 97. doi:10.1007/s10648-023-09814-5
[49] David Paunesku, Gregory M. Walton, Carissa Romero, Eric N. Smith, David S. Yeager, and Carol S. Dweck. 2015. Mind-set interventions are a scalable treatment

for academic underachievement. *Psychological Science* 26, 6 (2015), 784–793. doi:10.1177/0956797615571017

[50] Richard E. Petty and John T. Cacioppo. 1986. The elaboration likelihood model of persuasion. In *Advances in Experimental Social Psychology*. Vol. 19. Elsevier, 123–205. doi:10.1016/S0065-2601(08)60214-2

[51] Adam Reynolds. 2023. *Investigating automated prompts to guide at-risk learners in an online practice platform.* Ph. D. Dissertation. University of Johannesburg. https://hdl.handle.net/10210/511990

[52] Henry L. Roediger and Jeffrey D. Karpicke. 2006. Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science* 17, 3 (March 2006), 249–255. doi:10.1111/j.1467-9280.2006.01693.x

[53] William Samuelson and Richard Zeckhauser. 1988. Status quo bias in decision making. *Journal of Risk and Uncertainty* 1, 1 (1988), 7–59. Publisher: Springer.

[54] R Core Team. 2025. R: A Language and Environment for Statistical Computing. https://www.R-project.org/

[55] Richard H. Thaler and Shlomo Benartzi. 2004. Save More Tomorrow™: Using behavioral economics to increase employee saving. *Journal of Political Economy* 112, S1 (Feb. 2004), S164–S187. doi:10.1086/380085

[56] Richard H. Thaler and Cass R. Sunstein. 2009. *Nudge: Improving decisions about health, wealth, and happiness* (rev. and expanded ed ed.). Penguin Books, New York.

[57] Ulrich Trautwein, Herbert W. Marsh, Benjamin Nagengast, Oliver Lüdtke, Gabriel Nagy, and Kathrin Jonkmann. 2012. Probing for the multiplicative term in modern expectancy–value theory: A latent interaction modeling study. *Journal of Educational Psychology* 104, 3 (2012), 763–777. doi:10.1037/a0027470 ISBN: 0022-0663.

[58] Amos Tversky and Daniel Kahneman. 1974. Judgment under uncertainty: Heuristics and biases. *Science* 185, 4157 (Sept. 1974), 1124–1131. doi:10.1126/science.185.4157.1124

[59] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)informed consent: Studying GDPR consent notices in the field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. ACM, London United Kingdom, 973–990. doi:10.1145/3319535.3354212

[60] Kurt VanLehn. 2006. The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education* 16, 3 (Aug. 2006), 227–265. doi:10.3233/IRG-2006-16(3)02

[61] Gregory M. Walton and Timothy D. Wilson. 2018. Wise interventions: Psychological remedies for social and personal problems. *Psychological Review* 125, 5 (Oct. 2018), 617–655. doi:10.1037/rev0000115

[62] Markus Weinmann, Christoph Schneider, and Jan Vom Brocke. 2016. Digital nudging. *Business & Information Systems Engineering* 58, 6 (Dec. 2016), 433–436. doi:10.1007/s12599-016-0453-1

[63] David S. Yeager, Paul Hanselman, Gregory M. Walton, Jared S. Murray, Robert Crosnoe, Chandra Muller, Elizabeth Tipton, Barbara Schneider, Christopher S. Hulleman, Cintia P. Hinojosa, David Paunesku, Carissa Romero, Kate Flint, Alice Roberts, Jill Trott, Ronaldo Iachan, Jenny Buontempo, Sophia M. Yang, Carlos M. Carvalho, P. Richard Hahn, Maithreyi Gopalan, Pratik Mhatre, Ronald Ferguson, Angela Lee Duckworth, and Carol S. Dweck. 2019. A national experiment reveals where a growth mindset improves achievement. *Nature* 573 (2019), 364–369. doi:10.1038/s41586-019-1466-y

[64] David S. Yeager and Gregory M. Walton. 2011. Social-psychological interventions in education: they're not magic. *Review of Educational Research* 81, 2 (June 2011), 267–301. doi:10.3102/0034654311405999

[65] Michael Yeomans and Justin Reich. 2017. Planning prompts increase and forecast course completion in massive open online courses. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. ACM, Vancouver British Columbia Canada, 464–473. doi:10.1145/3027385.3027416