UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
The Power of Nudging: Exploring Three Interventions for Metacognitive Skills Instruction across Intelligent Tutoring Systems

Permalink
https://escholarship.org/uc/item/8qh9z65n

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors
Abdelshiheed, Mark
Hostetter, John Wesley
Shabrina, Preya
et al.

Publication Date
2022

Peer reviewed

eScholarship.org
Powered by the California Digital Library
University of California

The Power of Nudging: Exploring Three Interventions for Metacognitive Skills Instruction across Intelligent Tutoring Systems

Mark Abdelshiheed, John Wesley Hostetter, Preya Shabrina, Tiffany Barnes, and Min Chi
Department of Computer Science
North Carolina State University
Raleigh, NC 27695
{mnabdels, jwhostet, pshabri, tmbarnes, mchi}@ncsu.edu

Abstract

Deductive domains are typical of many cognitive skills in that no single problem-solving strategy is always optimal for solving all problems. It was shown that students who know *how* and *when* to use each strategy (*StrTime*) outperformed those who know neither and stick to the default strategy (*Default*). In this work, students were trained on a logic tutor that supports a *default* forward-chaining and a backward-chaining (BC) strategy, then a probability tutor that only supports BC. We investigated three types of interventions on teaching the *Default* students *how* and *when* to use which strategy on the logic tutor: *Example*, *Nudge* and *Presented*. Meanwhile, *StrTime* students received no interventions. Overall, our results show that *Nudge* outperformed their *Default* peers and caught up with *StrTime* on both tutors.

Keywords: metacognitive skills instruction; worked examples; tutoring nudges; strategy instruction

Introduction

Deductive task domains are those in which a solution requires an argument, proof, or derivation; each step is the outcome of applying a domain principle, operator, or rule. Deductive domains such as geometry, logic and probability are standard components of STEM fields. Two common problem-solving strategies in such domains are forward-chaining (FC) and backward-chaining (BC) (Russell & Norvig, 2020). In FC, the reasoning proceeds from the given propositions toward the target goal, whereas BC is goal-driven in that it works backward from a goal state to a given state. Early studies show that experts often use a mixture of FC and BC strategies, and more importantly, they often use past experience, heuristics, and many other kinds of knowledge to determine their strategies (Priest & Lindsay, 1992). Our prior work showed that students who know *which* problem-solving strategies to use *when*, referred to as *StrTime*, consistently learn across different deductive domains, as they possess the necessary *metacognitive skills*, unlike their peers who follow the default strategy, known as *Default* (Abdelshiheed et al., 2020).

It has been believed that metacognitive skills are essential for academic achievements (de Boer et al., 2018; Erskine, 2010; Zimmerman, 1990), and teaching such skills impacts learning outcomes (Zepeda et al., 2015; M. Chi & VanLehn, 2010) as well as strategy use (Lee & Oxford, 2008; Chambres et al., 2002; Roberts & Erdos, 1993). STEM domains often demand the use of various problem-solving strategies, and some prior research has categorized knowing *how* and *when* to use each strategy as two metacognitive skills (Winne & Azevedo, 2014; Cardelle-Elawar, 1992), referred to as strategy- and time-awareness, respectively.

Prior work has shown the positive impact of strategy awareness on preparing students for future learning (Belenky & Nokes-Malach, 2012; Abdelshiheed et al., 2021) and time awareness on planning skills (Winne & Azevedo, 2014; Fazio et al., 2016). Thus various attempts were made to teach students the two metacognitive skills, such as teaching the strategy by example (Likourezos & Kalyuga, 2017; Glogger-Frey et al., 2015), prompting nudges to use the strategy (Richey et al., 2015; Belenky & Nokes, 2009) and explicitly presenting it (Fellman et al., 2020; Spörer et al., 2009).

Our work directly compares three types of interventions on teaching *Default* students *how* and *when* to use which strategy on the logic tutor in this ascending order of instructional support: *Example*, *Nudge* and *Presented*. All interventions provided BC worked examples. The main difference is that *Nudge* prompted students to switch to BC in problems proper to do so, while for *Presented*, those problems were presented in BC by default. Our primary research question is: Which of the three types of interventions would make *Default* students catch up with their *StrTime* peers?

Our study involved two intelligent tutoring systems (ITSs) (Vanlehn, 2006): logic and probability. Students were first assigned to a logic tutor that supports FC and BC strategies, with FC being the default, then to a probability tutor six weeks later that supports BC only. During the logic instruction, *Default* students were split into four conditions: three intervention groups —*Example*, *Nudge* and *Presented*— and a *Control* group without any intervention. On the other hand, we believe that *StrTime* students already have the two metacognitive skills and thus are considered the gold standard and received no intervention. All students went through the same probability tutor and were asked to decide whether they wanted to solve the following problem on their own (problem-solving (PS)), the tutor to present it as a worked example (WE), or to solve it collaboratively with the tutor in the form of a faded worked example (FWE). Overall, our results show that *Nudge* students outperformed their other *Default* peers and caught up with *StrTime* on both tutors. Additionally, *Nudge*'s strategy behavior on the logic tutor was similar to *StrTime* as both knew how and when to use BC. Surprisingly, *Nudge* chose significantly more *PS* on the probability tutor, and *StrTime* chose significantly less FWE.

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*. ©2022 The Author(s). This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

Related Work

Teaching by Example, Nudging and Presenting

Substantial work has explored many approaches for teaching strategies and highlighted their tradeoffs. We focus on the possible combinations of three approaches: teaching a strategy by example (Likourezos & Kalyuga, 2017; Glogger-Frey et al., 2015), prompting nudges to use a strategy (Richey et al., 2015; Zepeda et al., 2015; Belenky & Nokes, 2009) and directly presenting it (Fellman et al., 2020; Spörer et al., 2009; M. Chi & VanLehn, 2007; Schwartz & Martin, 2004).

Glogger-Frey et al. (2015) found that students receiving worked examples of journal extracts reviews outperformed their peers, who had to come up with the reviews, on post-test performance. However, Likourezos and Kalyuga (2017) reported no significant difference between students who received fully-guided worked examples, partially-guided ones and unguided assistance on post-test geometry tasks.

Zepeda et al. (2015) showed that students who received tutoring nudges and worked examples performed better on a physics test and a novel self-guided activity than their peers who received no such instruction. Conversely, Richey et al. (2015) found no significant difference between students who were instructed to study the worked examples and their peers, who received the same examples with tutoring nudges, on near, intermediate and far transfer electric circuit tasks.

Spörer et al. (2009) found that students who were explicitly instructed on comprehensive reading strategies surpassed their peers, who were taught by the instructors' text interactions, on a transfer task and follow-up test. On the other hand, Fellman et al. (2020) found no significant difference between students who were presented explicit strategy instruction to practice the single-digit n-back task and their peers who practiced without such instruction, as both groups showed emerging transfer to untrained variants of the same task.

Metacognitive Skills Instruction

Metacognitive skills regulate one's awareness and control of their cognition (Chambres et al., 2002; Roberts & Erdos, 1993). Many studies have demonstrated the significance of metacognitive skills instruction on academic performance (de Boer et al., 2018; Erskine, 2010), learning outcomes (Zepeda et al., 2015; M. Chi & VanLehn, 2010, 2008) and regulating strategy use (Schraw & Gutierrez, 2015).

Schraw and Gutierrez (2015) argue that metacognitive skill instruction involves feeling what is known and not known about a task, as this allows learners to gather information efficiently, adapt to changes in task requirements, and develop strategies to overcome the task. They state that such instruction should further compare strategies according to their feasibility and familiarity from the learner's perspective.

Belenky and Nokes (2009) showed that students who were prompted with metacognitive nudges, which reflect on the current problem-solving processes, outperformed their peers who received problem-focused nudges, which focus on the current goal, on a permutation transfer task. M. Chi and VanLehn (2010) found that teaching students principle-emphasis skills closed the gap between high and low learners, not only in the domain where they were taught (probability) but also in a second domain where they were not taught (physics).

Strategy- and Time-Awareness

Strategy- and time-awareness have been regarded as metacognitive skills as they respectively address *how* and *when* to use a problem-solving strategy (de Boer et al., 2018; Winne & Azevedo, 2014; Lee & Oxford, 2008; Cardelle-Elawar, 1992). Researchers have emphasized the role of strategy awareness in learning a foreign language (Teng, 2020; Lee & Oxford, 2008) and preparation for future learning (Abdelshiheed et al., 2021; Belenky & Nokes-Malach, 2012; Chamot, 1998), and the impact of time awareness on planning skills and academic performance (de Boer et al., 2018; Fazio et al., 2016; Winne & Azevedo, 2014).

Lee and Oxford (2008) studied the role of strategy awareness in teaching English to Korean students; specifically, students aware of various learning strategies employed these strategies more frequently than their peers. In Abdelshiheed et al. (2021), we found that students who knew two problem-solving strategies were the best learners in two independent domains. Belenky and Nokes-Malach (2012) showed that students who had a higher aim to master presented materials and strategies outperformed their peers on a transfer task.

In Fazio et al. (2016), students who knew when to use each strategy to pick the largest fraction magnitude had high mathematical proficiency. Their peers who did not know when to apply each strategy failed to choose the correct alternative when offered choices. de Boer et al. (2018) showed that students who knew when and why to use a given strategy exhibit long-term metacognitive knowledge that improves their academic performance. de Boer et al. emphasized that knowing *when* and *why* has the same importance as knowing *how* when it comes to strategy choice in multi-strategy domains.

To sum up, much of the prior work has highlighted the importance of metacognitive skill instruction and teaching strategy- and time-awareness. Many approaches for teaching strategies have been investigated, such as teaching by example, prompting nudges, and direct presentation. However, as far as we know, no agreement has been found on the most effective combination of these approaches, and no work has compared these approaches in intelligent tutoring systems. This work compares three ways to teach a backward-chaining (BC) strategy on two intelligent tutoring systems: logic and probability. First, by examples alone ($Example$), then by examples and nudges to switch to BC ($Nudge$), and finally, by examples and directly presenting BC ($Presented$).

Methods

Participants

They are Computer Science undergraduates at North Carolina State University. Each tutor is a class assignment whose completion is required for full credit, and students are told that

Table 1: Tutors' Assignment and Completion Counts

	Logic		Probability	
	Assigned	Completed	Assigned	Completed
Control	23	21	19	**17**
Example	23	20	20	**19**
Nudge	22	21	21	**20**
Presented	20	17	16	**15**
StrTime	49	45	41	**40**
	$\chi^2(4, N = 261) = 0.09$, $p = .99$		$\chi^2(4, N = 228) = 0.05$, $p = .99$	

Only students who completed Logic were assigned to Probability.

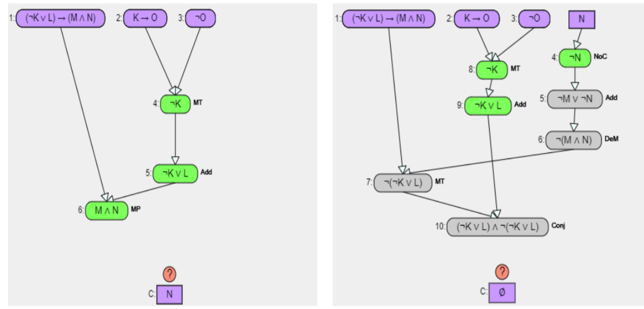
grades are based on effort, not performance. The main challenge in this work is that the student's metacognitive label —*Default* or *StrTime*— can be calculated only at the end of logic training, but the label is needed at its beginning to determine the intervention possibility. Specifically, *StrTime* students frequently follow the desired behavior of switching *early* (within the first 30 actions) to *BC*, while *Default* students make no switches and stick to *FC* (Abdelshiheed et al., 2022, 2020). Such switch behaviors are recorded at the end of the logic training, and hence, can not be calculated before training. Therefore, as per Abdelshiheed et al. (2021), we utilize the random forest classifier (RFC) that, based on pre-test performance, predicts the metacognitive label before training on logic and was previously shown to be 96% accurate.

Among 230 students assigned to the logic tutor, 137 were classified by the RFC into 88 *Default* and 49 *StrTime*[1]. *Default* students were randomly split into four conditions: a control —*Control*— and three experimental —*Example*, *Nudge* and *Presented*. Table 1 shows the assigned and completed counts on both tutors for *Default* (top four rows) and *StrTime* (fifth row). The last column is for students who finished both tutors since we excluded dropout logic students from the probability assignment. Hence, only the last column students were included in our analyses resulting in 17 *Control*, 19 *Example*, 20 *Nudge*, 15 *Presented* and 40 *StrTime*. As shown in Table 1, a chi-square test found no significant difference between the groups' completion rates on both tutors. The RFC was 97% accurate in classifying students who received no interventions —*Control* and *StrTime*.

Two Tutors and Our Interventions

Logic Tutor and Our Interventions The logic tutor teaches propositional logic proofs by applying valid inference rules such as Modus Ponens and Constructive Dilemma. It consists of five ordered levels with an *incremental degree of difficulty*, and each level consists of four problems. A student can solve any problem by either a **FC** or **BC** strategy. Figure 1a shows that in *FC*, one must derive the conclusion at the bottom from givens at the top, while Figure 1b shows that in *BC*, students derive a contradiction from givens and the *negation* of the conclusion. Problems are presented by *default* in FC with the ability to switch to BC by clicking the yellow button in Figure 2. The logic tutor was adjusted, as shown in Figure 3, to accommodate the following interventions for *Default* students:

[1]The remaining students were excluded from further analyses, as their label is irrelevant to this work



(a) Forward Chaining (FC) (b) Backward Chaining (BC)

Figure 1: Logic Tutor Problem-Solving Strategies

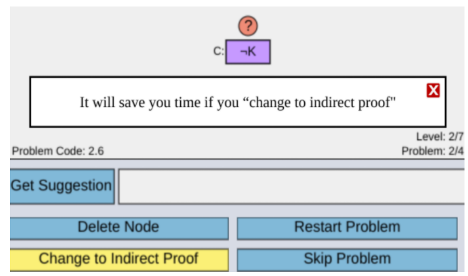


Figure 2: Prompted Strategy Switch Nudge

- *No Intervention*: students are assigned to the original tutor.
- *Example*: two WEs on BC are provided.
- *Nudge*: same as *Example*, and nudges (shown in Figure 2) are prompted to switch to BC in some problems.
- *Presented*: same as *Example*, and students are presented some problems in BC by default.

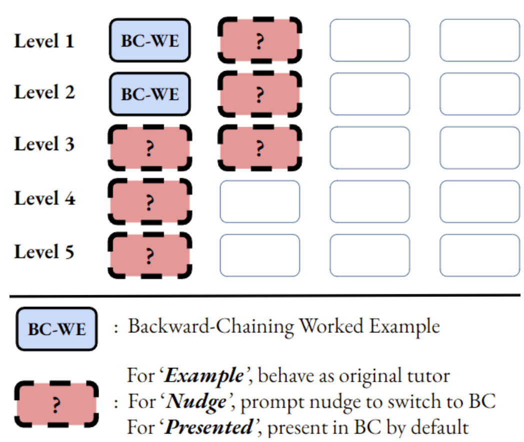


Figure 3: Training on the Adjusted Logic Tutor

In Figure 3, it is crucial to note that: 1) *white* problems behave the same as the original tutor, 2) *red* problems are selected based on the historical strategy switches in our data, and 3) nudges are prompted after a number of seconds sampled from a probability distribution of prior students' switch behavior.

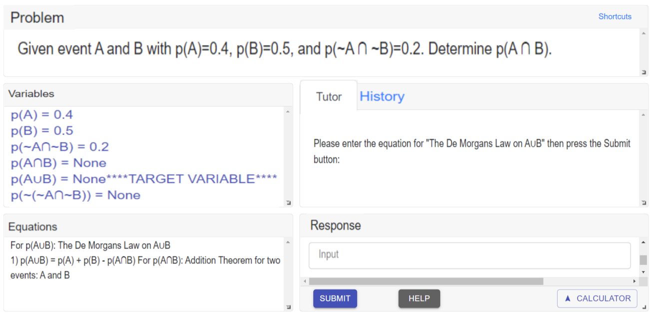


Figure 4: Probability Tutor Interface

Probability tutor: It teaches how to solve probability problems using ten principles, such as the Complement Theorem and De Morgan's Law, as shown in Figure 4. It consists of 12 problems, each of which can *only* be solved by BC as it requires deriving an answer by *writing and solving equations* until the target is ultimately reduced to the givens. A problem can be *PS*, *WE* or *FWE*. *PS* requires students to solve alone, *WE* involves a step-by-step solution from the tutor, and *FWE* demands student and tutor collaboration.

Table 2: Overview of the Study Procedure

Logic	Pre-test (2 problems)
	Training (20 problems):
	Control : No Intervention
	Example, *Nudge*, *Presented* : Intervention (Fig. 3)
	StrTime : No Intervention
	Post-test (6 problems, including 2 isomorphic)
	Six weeks later
Prob.	Textbook
	Pre-test (14 problems)
	Training (12 problems):
	On ten problems, students choose $PS/WE/FWE$
	Post-test (20 problems, including 14 isomorphic)

Procedure

Table 2 summarizes our procedure. During the logic instruction, students went through the standard sequence of pre-test, training and post-test. The first two post-test problems are isomorphic to the two pre-test problems. The *only* difference occurred during training on logic, as shown in Table 2.

Six weeks later, students were trained on the probability tutor following the standard procedure: textbook, pre-test, training, and post-test. In the textbook, they studied the domain principles; In pre- and post-test, students solved 14 and 20 open-ended problems that required them to derive an answer by writing and solving one or more equations. Each pretest problem has a corresponding isomorphic post-test problem. For the training section, shown in Figure 4, students went through 12 problems and selected the type on ten of them; two problems were fixed as *PS*. For *FWE* problems, each step was randomly decided to determine whether the student or tutor should solve it. Note that on both tutors, the post-test is *much more challenging* than the pre-test, and the problem order is the same for all students.

Grading criteria

On logic, a problem score is a function of time, accuracy, and solution length. The *pre-* and *post-test* scores are calculated by averaging the pre- and post-test problem scores. On probability, students' answers are graded by experienced graders in a double-blind manner using a partial-credit rubric, and grades are based *only* on accuracy. The *pre-* and *post-test* scores are the average grades in their respective sections. On both tutors, test scores are in the range of $[0, 100]$.

Results

Learning Performance

Table 3: Comparing Groups across Tutors

	Condition				
	Control ($N=17$)	*Example* ($N=19$)	*Nudge* ($N=20$)	*Presented* ($N=15$)	*StrTime* ($N=40$)
	Logic Tutor				
Pre	59.1(19)	56.9(25)	60.5(13)	60.4(15)	60(18)
Iso-Post	65.4(8)	69.7(7)	**89.8**(**5**)*	83.4(4)*	85.3(6)*
Iso-NLG	0.04(.24)	0.09(.3)	**0.4**(**.13**)*	0.34(.14)*	0.35(.19)*
Post	59.9(9)	65.5(8)	**86.1**(**5**)*	80(5)*	81.7(6)*
NLG	-0.05(.3)	0.05(.37)	**0.39**(**.15**)*	0.29(.16)*	0.3(.23)*
Time	5.5(7)	4.8(4)	5.3(4)	6.2(6)	4.6(7)
	Probability Tutor				
Pre	79.4(12)	74.5(17)	77(14)	74.1(14)	76(15)
Iso-Post	73.1(22)	77(14)	**94.2**(**6**)*	85.8(17)	92.6(13)*
Iso-NLG	-0.06(.39)	0.03(.28)	**0.32**(**.19**)*	0.16(.22)	0.28(.2)*
Post	70.3(20)	73.6(16)	**91.9**(**5**)*	83.5(20)	89.3(11)*
NLG	-0.09(.36)	-0.04(.35)	**0.27**(**.24**)*	0.13(.23)	0.26(.17)*
Time	4.3(6)	3.9(4)	4.2(5)	3.5(4)	4.4(5)

In a row, bold is for the highest value, and asterisk means significance over no asterisks.

Table 3 compares the groups' performance across the two tutors showing the mean and standard deviation of pre- and post-test scores, isomorphic scores, training time in hours, and the learning outcome in terms of the normalized learning gain (NLG) defined as ($NLG = \frac{Post-Pre}{\sqrt{100-Pre}}$), where 100 is the maximum test score. We refer to pre-test, post-test and NLG scores as *Pre*, *Post* and *NLG*, respectively. A one-way ANOVA using condition as factor found no significant difference on *Pre*: $F(3,67) = 0.14, p = .93$ for logic, and $F(3,67) = 0.49, p = .69$ for probability. Similarly, no significant difference was found in the training time on both tutors. In order to measure the students' improvement on isomorphic problems, several repeated measures ANOVA were conducted (one for each group on each tutor) using {*Pre*, *Iso-Post*} as factor. Results showed that *Nudge* and *StrTime* learned significantly with $p < 0.0001$ on both tutors, *Presented* learned significantly with $p = 0.0001$ on logic and $p = 0.02$ on probability. *Example* and *Control* did not perform significantly higher on *Iso-Post* than *Pre* on both tutors. These findings verify the RFC's accuracy, as *StrTime* learned significantly on both tutors, while *Control* did not, despite both groups receiving no interventions.

Comparing Conditions A comparison between the four conditions in Table 3 was essential to assess the performance of *Default* students. On the logic tutor, a one-way ANCOVA[2] using condition as factor and *Pre* as covariate found a significant difference on *Post*: $F(3,66) = 59.7, p < .0001, \eta^2 = 0.69$. Follow-up post-hoc analyses with Bonferroni[3] adjustment[4] revealed that *Nudge* and *Presented* significantly outperformed *Example* ($t(37) = 5.9, p < .0001$ and $t(32) = 5.2, p < .0001$) as well as *Control* ($t(35) = 7.8, p < .0001$ and $t(30) = 6.3, p < .0001$). No significant difference was found between *Nudge* and *Presented*, or between *Example* and *Control*. Similar patterns were observed on *NLG* using *ANOVA*. These findings show that *Nudge*, *Presented* > *Example*, *Control*.

On the probability tutor, a one-way ANCOVA using condition as factor and *Pre* as covariate reported a significant difference on *Post*: $F(3,66) = 14.5, p < .0001, \eta^2 = 0.31$. Subsequent Bonferroni-corrected analyses showed that *Nudge* significantly outperformed *Presented* ($t(33) = 3.6, p = .001$), *Example* ($t(37) = 5.6, p < .0001$) and *Control* ($t(35) = 6.2, p < .0001$); meanwhile, *Presented* significantly surpassed *Example* and *Control* ($t(32) = 3.1, p = .004$ and $t(30) = 3.4, p = .002$). No significant difference was found between between *Example* and *Control*. Similar patterns were found using *ANOVA* on *NLG*. In short, these results show that *Nudge* > *Presented* > *Example*, *Control*.

In essence, *Nudge* students were the best on both tutors, followed by *Presented*, who learned less on probability. Surprisingly, *Example* learned no different from *Control* on both tutors, which signifies the additional instructional support that *Nudge* and *Presented* were given on logic.

Comparing with *StrTime* To determine whether any condition caught up with *StrTime* students, post-hoc pairwise analyses were conducted on logic and probability *Post* using Bonferroni correction. On logic, results revealed that *Nudge* and *Presented* caught up with *StrTime* as no significant difference was found between their *Post* and that of *StrTime* ($t(58) = 0.9, p = .37$ and $t(53) = 0.3, p = .77$). On the other hand, *StrTime* significantly outperformed *Example* ($t(57) = 5.4, p < .0001$) and *Control* ($t(55) = 6.7, p < .0001$). Similar results were found on *NLG*.

On the probability tutor, only *Nudge* caught up with *StrTime* as no significant difference was found on *Post* ($t(58) = 0.2, p = .84$). Meanwhile, *StrTime* significantly surpassed *Presented* ($t(53) = 3.1, p = .003$), *Example* ($t(57) = 5.1, p < .0001$) and *Control* ($t(55) = 5.7, p < .0001$). Similar patterns were observed on *NLG*.

In brief, *Nudge* and *Presented* caught up with *StrTime* in the presence of our interventions on logic. Only *Nudge* caught up with *StrTime* on probability without such interventions. Lastly, *Example* and *Control* performed significantly worse than *StrTime* on both tutors.

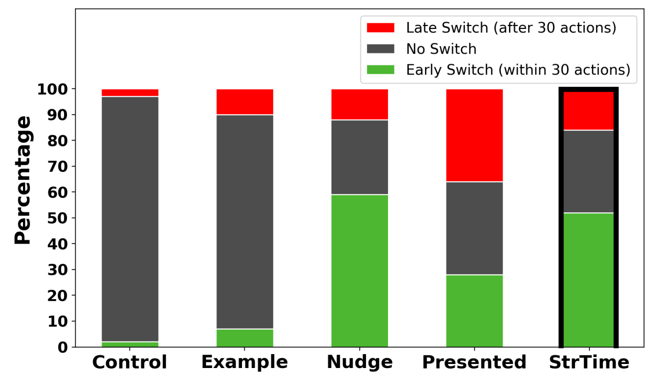


Figure 5: Strategy Switch Behavior on Logic

Strategy Switch on Logic

The strategy switch behavior on the logic tutor (from FC into BC) is displayed in Figure 5 to investigate the impact of our intervention on students' strategy choices. Decisions are combined across the training and post-test sections, as no significant difference was found in their distribution between the two sections. Additionally, *StrTime* is highlighted in bold as the gold standard.

A one-way ANOVA using condition as factor showed a significant difference in the frequency of *early* switches: $F(3,67) = 6.7, p < .001, \eta^2 = 0.23$. Moreover, a chi-square test showed a significant relationship between the switch type and student group[5]: $\chi^2(8, N = 2664) = 934.3, p < .0001$. Post-hoc pairwise chi-square tests with Bonferroni adjustment showed that for *early* switches: *Nudge*, *StrTime* > *Presented* > *Example*, *Control*. For instance, *Nudge* and *StrTime* made early switches significantly more than *Presented*: $\chi^2(2, N = 840) = 100.2, p < .0001$ and $\chi^2(2, N = 1320) = 84.2, p < .0001$, respectively. No significant difference was found between *Nudge* and *StrTime*, or between *Example* and *Control*.

[2]General effect size η^2 was reported for conservative results

[3]Bonferroni was chosen for more conservative results

[4]($\alpha = .05/10$) throughout the results section

[5][111 students] * [20 training - 2 WE + 6 post] = 2664 decisions

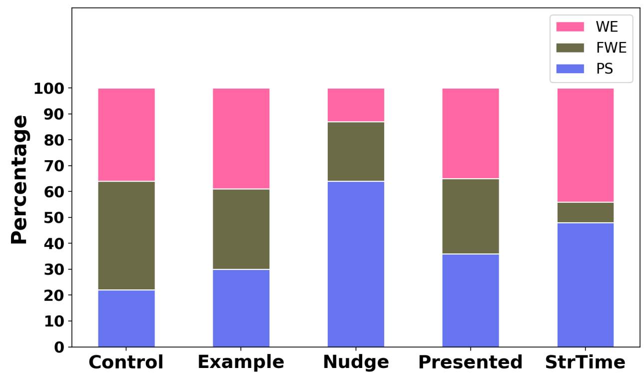


Figure 6: Problem-level Decisions on Probability

Student Decision on Probability

Table 3 showed that our interventions' impact on logic might also extend to probability. Therefore, in Figure 6, we investigate the problem-level decisions in the probability training section as students chose them. Step-level decisions were not considered since the tutor randomly chose them. It is important to note that for each student group, there was no significant correlation between any problem-level decision type shown in Figure 6 and any performance metric in Table 3.

A chi-square test found a significant relationship between the problem-level decision type and student group[6]: $\chi^2(8, N = 1110) = 162.1, p < .0001$. Follow-up pairwise chi-square tests with Bonferroni correction showed that for *PS*: *Nudge* > *Presented*, *Example*, *Control*; for *FWE*: *StrTime* < *Nudge*, *Presented*, *Example*, *Control*. For instance, *Nudge* chose *PS* significantly more than *Presented*: $\chi^2(2, N = 350) = 32.9, p < .0001$, while *StrTime* chose *FWE* significantly less than *Nudge*: $\chi^2(2, N = 600) = 67.5, p < .0001$. No significant difference was found between any pair of *Presented*, *Example* and *Control* on any decision type. In short, *Nudge* and *StrTime* made decisions different from each other and their peers, while *Presented*, *Example* and *Control* made similar decisions.

Discussions & Conclusions

We showed that to teach students *how* and *when* to use a strategy, using worked examples alone may not be very effective, as *Example* did not significantly outperform *Control*. However, students learned better when we reinforced examples by prompting BC nudges or presenting problems in BC by default, as *Nudge* and *Presented* significantly surpassed *Example* and *Control*. Additionally, providing nudges was even more beneficial as *Nudge* continued to outperform *Presented* on probability significantly.

Catching up with StrTime While *Nudge* and *Presented* caught up with *StrTime* on logic, only *Nudge* caught up with *StrTime* on their logic early-switch behavior and probability learning performance. This finding suggests that the *Nudge* students are prepared for future learning (Bransford & Schwartz, 1999) as they performed well on probability based on interventions they received on logic.

Relation to ICAP Framework We believe that our results show that the effectiveness of the Interactive, Constructive, Active and Passive (ICAP) framework (M. T. Chi & Wylie, 2014; M. T. Chi, 2009) can be extended to teaching students metacognitive skills. Precisely, *Control* encountered passive learning as they received no interventions, while *Example* received an active treatment as students were required to go through the examples and proceed to the next steps. *Presented* can be seen as a constructive intervention since students were asked to generate solutions in a strategy presented to them beyond the default one. Finally, *Nudge* received an interactive intervention where the tutor offered nudges to switch strategies, but the actual switch had to come from students. Our findings are consistent with ICAP in that interactive learning activities achieve the highest learning outcomes, as is the case with *Nudge* students on the two tutors.

Relation to Nudge Theory The nudge theory (Thaler & Sunstein, 2008) states that nudges have an essential role in behavioral sciences (Simon & Tagliabue, 2018) and influence individuals' social and cognitive behavior (Smith et al., 2013; Goldstein et al., 2008). Our results suggest that the impact of this theory is evident in teaching *Default* students the BC strategy on a logic tutor. Precisely, the strategy behavior of *Nudge* students changed after receiving prompted nudges to use BC, resulting in the best performance on both tutors.

Students' Choices and Personalities The evaluation of students' choices on probability revealed that *StrTime* students preferred minimal collaboration with the tutor; they chose *WE* or *PS* likely to save time or show effort, respectively. On the other hand, *Nudge* students chose *PS* significantly more than their intervention and *Control* peers, likely to demonstrate their acquired BC knowledge. At the end of probability training, students were provided the ten-item personality inventory[7] (TIPI) (Gosling, Rentfrow, & Swann Jr, 2003), which showed that *Nudge* and *StrTime* identified themselves as *critical* and *quarrelsome* significantly more than their peers.

Limitations and Future Work There are at least two caveats in our study. First, our study focused on different interventions for *Default* students, and hence, the conditions ended up with relatively small sample sizes. Second, the logic tutor offered a strategy by default, and the probability tutor supported only one strategy. A more convincing testbed would be having the tutors support both strategies, where students will be asked to choose the default strategy on each problem. The future work includes combining nudges and presentation into one intervention, implementing FC on the probability tutor, and providing explanations in the nudges on why *BC* is helpful.

[6] [111 students] * [10 choices on training] = 1110 decisions

[7] This was not stated earlier for not being our main scope

Acknowledgments

This research was supported by the NSF Grants: MetaDash: A Teacher Dashboard Informed by Real-Time Multichannel Self-Regulated Learning Data (1660878), Integrated Data-driven Technologies for Individualized Instruction in STEM Learning Environments (1726550), Generalizing Data-Driven Technologies to Improve Individualized STEM Instruction by Intelligent Tutors (2013502) and CAREER: Improving Adaptive Decision Making in Interactive Learning Environments (1651909).

References

Abdelshiheed, M., Maniktala, M., Barnes, T., & Chi, M. (2022). Assessing competency using metacognition and motivation: The role of time-awareness in preparation for future learning. In *Design recommendations for intelligent tutoring systems* (Vol. 9, pp. 121–131).

Abdelshiheed, M., Maniktala, M., Ju, S., Jain, A., Barnes, T., & Chi, M. (2021). Preparing unprepared students for future learning. In *Proceedings of the 43rd annual conference of the cognitive science society* (Vol. 43).

Abdelshiheed, M., Zhou, G., Maniktala, M., Barnes, T., & Chi, M. (2020). Metacognition and motivation: The role of time-awareness in preparation for future learning. In *Proceedings of the 42nd annual conference of the cognitive science society* (Vol. 42).

Belenky, D. M., & Nokes, T. J. (2009). Examining the role of manipulatives and metacognition on engagement, learning, and transfer. *The Journal of Problem Solving*, *2*(2), 6.

Belenky, D. M., & Nokes-Malach, T. J. (2012). Motivation and transfer: The role of mastery-approach goals in preparation for future learning. *Journal of the Learning Sciences*, *21*(3), 399–432.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. *Review of research in education*, *24*(1), 61–100.

Cardelle-Elawar, M. (1992). Effects of teaching metacognitive skills to students with low mathematics ability. *Teaching and teacher education*, *8*(2), 109–121.

Chambres, P., et al. (2002). *Metacognition: Process, function, and use*. Kluwer Academic Publishers.

Chamot, A. (1998). Teaching learning language strategies to language students. *Language and Linguistics*.

Chi, M., & VanLehn, K. (2007). The impact of explicit strategy instruction on problem-solving behaviors across intelligent tutoring systems. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 29).

Chi, M., & VanLehn, K. (2008). Eliminating the gap between the high and low students through meta-cognitive strategy instruction. In *Intelligent tutoring systems* (pp. 603–613).

Chi, M., & VanLehn, K. (2010). Meta-cognitive strategy instruction in intelligent tutoring systems: How, when, and why. *Educational Technology & Society*, *13*(1), 25–39.

Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science*, *1*(1), 73–105.

Chi, M. T., & Wylie, R. (2014). The icap framework: Linking cognitive engagement to active learning outcomes. *Educational psychologist*, *49*(4), 219–243.

de Boer, H., et al. (2018). Long-term effects of metacognitive strategy instruction on student academic performance: A meta-analysis. *Educational Research Review*, *24*, 98–115.

Erskine, D. L. (2010). *Effect of prompted reflection and metacognitive skill instruction on university freshmen's use of metacognition*. Brigham Young University.

Fazio, L. K., et al. (2016). Strategy use and strategy choice in fraction magnitude comparison. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(1), 1.

Fellman, D., Jylkkä, J., Waris, O., Soveri, A., Ritakallio, L., Haga, S., … Laine, M. (2020). The role of strategy use in working memory training outcomes. *Journal of Memory and Language*, *110*, 104064.

Glogger-Frey, I., et al. (2015). Inventing a solution and studying a worked solution prepare differently for learning from direct instruction. *Learning and Instruction*, *39*, 72–87.

Goldstein, D. G., et al. (2008). Nudge your customers toward better choices. *Harvard Business Review*, *86*(12), 99–105.

Gosling, S. D., Rentfrow, P. J., & Swann Jr, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in personality*, *37*(6), 504–528.

Lee, K. R., & Oxford, R. (2008). Understanding efl learners' strategy use and strategy awareness. *The Asian EFL Journal Quarterly March 2008 Volume 10, Issue*, *10*(1), 7–32.

Likourezos, V., & Kalyuga, S. (2017). Instruction-first and problem-solving-first approaches: alternative pathways to learning complex tasks. *Instructional Science*, *45*, 195–219.

Priest, A., & Lindsay, R. (1992). New light on novice—expert differences in physics problem solving. *British journal of Psychology*, *83*(3), 389–405.

Richey, J. E., et al. (2015). Transfer effects of prompted and self-reported analogical comparison and self-explanation. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 37).

Roberts, M. J., & Erdos, G. (1993). Strategy selection and metacognition. *Educational Psychology*, *13*, 259–266.

Russell, S. J., & Norvig, P. (2020). *Artificial intelligence: a modern approach* (4th ed.). Pearson.

Schraw, G., & Gutierrez, A. P. (2015). Metacognitive strategy instruction that highlights the role of monitoring and control processes. In *Metacognition: Fundaments, applications, and trends* (pp. 3–16). Springer.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and instruction*, *22*(2), 129–184.

Simon, C., & Tagliabue, M. (2018). Feeding the behavioral revolution: Contributions of behavior analysis to nudging and vice versa. *Journal of Behavioral Economics for Policy*, *2*(1), 91–97.

Smith, N. C., et al. (2013). Choice without awareness: Ethical and policy implications of defaults. *Journal of Public Policy & Marketing*, *32*(2), 159–172.

Spörer, N., et al. (2009). Improving students' reading comprehension skills: Effects of strategy instruction and reciprocal teaching. *Learning and instruction*, *19*(3), 272–286.

Teng, F. (2020). The benefits of metacognitive reading strategy awareness instruction for young learners of english as a second language. *Literacy*, *54*(1), 29–39.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness.* HeinOnline.

Vanlehn, K. (2006). The behavior of tutoring systems. *International journal of artificial intelligence in education*, *16*(3), 227–265.

Winne, P. H., & Azevedo, R. (2014). Metacognition. In R. K. Sawyer (Ed.), *The cambridge handbook of the learning sciences* (2nd ed.). Cambridge University Press.

Zepeda, C. D., et al. (2015). Direct instruction of metacognition benefits adolescent science learning, transfer, and motivation: An in vivo study. *Journal of Educational Psychology*, *107*(4), 954.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational psychologist*, *25*(1), 3–17.