Title

Bridging Declarative, Procedural, and Conditional Metacognitive Knowledge Gap Using Deep Reinforcement Learning

Permalink

Journal

Authors

Abdelshiheed, Mark
Hostetter, John Wesley
Barnes, Tiffany
et al.

Publication Date

Peer reviewed

Bridging Declarative, Procedural, and Conditional Metacognitive Knowledge Gap Using Deep Reinforcement Learning

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

Abstract

In deductive domains, three metacognitive knowledge types in ascending order are declarative, procedural, and conditional learning. This work leverages Deep Reinforcement Learning (*DRL*) in providing *adaptive* metacognitive interventions to bridge the gap between the three knowledge types and prepare students for future learning across Intelligent Tutoring Systems (ITSs). Students received these interventions that taught *how* and *when* to use a backward-chaining (BC) strategy on a logic tutor that supports a default forward-chaining strategy. Six weeks later, we trained students on a probability tutor that only supports BC without interventions. Our results show that on both ITSs, DRL bridged the metacognitive knowledge gap between students and significantly improved their learning performance over their control peers. Furthermore, the DRL policy adapted to the metacognitive development on the logic tutor across declarative, procedural, and conditional students, causing their strategic decisions to be more autonomous.

Keywords: Deep Reinforcement Learning; Preparation for Future Learning; Intelligent Tutoring Systems; Declarative Knowledge; Procedural Knowledge; Conditional Knowledge

Introduction

A demanding required feature of learning is being continuously prepared for future learning (Bransford & Schwartz, 1999). Our incremental knowledge is the evidence that preparation for future learning exists yet is hard to predict and measure (Detterman & Sternberg, 1993). Considerable research has found that one factor that facilitates preparing students for future learning is their metacognitive knowledge (Kua et al., 2021; Cutrer et al., 2017; Hershkovitz et al., 2013).

Three types of metacognitive knowledge in deductive domains are *declarative*, *procedural*, and *conditional* learning (Anderson, 2005; Schraw, 1998; Bloom, 1956). Although substantial research has shown that the three types could be acquired sequentially or simultaneously (Teng, 2020; Yildirim et al., 2001; Schraw, 1998), it was also shown that each learner possesses a *dominant* type of knowledge based on the educational context and learning environment (Kuhn, 2000; Brown, 1987). Moreover, prior work has shown that students' metacognitive knowledge can evolve during learning (Rittle-Johnson et al., 2001; Baker, 1994). Thus, *adaptive* interventions considering such development are needed (Azevedo et al., 2005; Pintrich, 2002; Butler, 1998).

Reinforcement Learning (RL) (Sutton & Barto, 2018) is one of the most effective approaches for *adaptive* support and scaffolding across Intelligent Tutoring Systems (ITSs) (Krueger et al., 2017; Zhou et al., 2019). The deep learning extension of RL, known as deep RL (DRL), has been commonly utilized in pedagogical policy induction across ITSs (Abdelshiheed et al., 2023; Alam et al., 2023; Hostetter et al., 2023a, 2023b; Ju et al., 2021; Sanz-Ausin et al., 2019) due to its higher support of model sophistication. As far as we know, no prior work has leveraged DRL in providing adaptive interventions to bridge the metacognitive knowledge gap and prepare students for future learning across ITSs.

This work builds on our prior work, where students were categorized into declarative, procedural, or conditional learners based on their *dominant* metacognitive knowledge on an ITS. We found that only *conditional* students were prepared for future learning, as they significantly outperformed their declarative and procedural peers across different deductive domains (Abdelshiheed et al., 2021, 2020). Inspired by such findings, this work empirically evaluates the DRL's effectiveness through a classroom study; we leverage DRL to provide *adaptive* metacognitive interventions to bridge the knowledge gap for declarative and procedural students. DRL provided metacognitive interventions that teach students *how* and *when* to use a backward-chaining (BC) strategy on a logic tutor that supports a default forward-chaining (FC) strategy. After six weeks, we trained students on a probability tutor that only supports BC without receiving interventions. Our results showed that DRL indeed sealed the gap, prepared students for future learning, adapted to their metacognitive development, and raised their decision-making autonomy.

Background & Related Work

Declarative, Procedural and Conditional Knowledge

Metacognition indicates cognition about cognition and the ability to control, conceive and regulate knowledge (Livingston, 2003; Roberts & Erdos, 1993; Flavell, 1979). Three types of metacognitive knowledge are declarative, procedural, and conditional (Schraw, 1998; Schraw & Moshman, 1995; Brown, 1987; Jacobs & Paris, 1987; Bloom, 1956).

Declarative knowledge —also described as surface or rote learning (Biggs et al., 1999)— is considered the simplest and lowest level of knowledge, as it involves memorization of facts and data offered in the default settings of a learning environment (Azevedo & Aleven, 2013; De Backer et al., 2012; Krathwohl, 2002; Bloom, 1956). **Procedural** knowledge is a higher form of knowledge that is discerned with the auto-

mated understanding of *how* to use different problem-solving strategies and cognitive skills without conscious attention or reasoning about their rationale (Azevedo & Aleven, 2013; Krathwohl, 2002; Dochy, 1992; Georgeff & Lansky, 1986; Bloom, 1956). **Conditional** knowledge is the highest level of knowledge, as it requires understanding *how*, *when* and *why* to use each strategy and cognitive skill (Kiesewetter et al., 2016; Larkin, 2009; Schraw, 1998).

Considerable research has investigated the significance of acquiring and nurturing each knowledge type (Castro-Schez et al., 2021; Boden et al., 2018; Fossati, 2009) and bridged the gap between them (Teng, 2020; Mbato, 2019; Yildirim et al., 2001). Boden et al. (2018) showed that students' self-efficacy was significantly correlated to declarative rather than procedural knowledge when solving physics problems. Fossati (2009) found that students with high procedural knowledge significantly outperformed their peers on a tutoring system that teaches linked lists. In Castro-Schez et al. (2021), students who had high conditional knowledge significantly surpassed their low peers in learning predictive parsing algorithms on a tutoring system.

Metacognitive Development

Metacognitive development is defined as the shifts in the learning approach and the metacognitive knowledge and skills used by a student (Anderson, 2005; Case & Gunstone, 2002; Rittle-Johnson et al., 2001; Kuhn, 2000; Baker, 1994; Anderson, 1982). We focus on the development across declarative, procedural, and conditional knowledge. Anderson (1982) proposed a framework for developing and acquiring metacognitive skills. He presented two major stages in skill acquisition: the declarative stage, where facts about the domain are interpreted, and the procedural stage, where domain knowledge is incorporated into procedures for performing the skill. Anderson (1982) stated that knowledge compilation occurs when the learner transitions from the declarative to the procedural stage. Later, in Anderson (2005), he argued that procedural knowledge depends upon conditional knowledge, which in turn depends on declarative knowledge. Specifically, Anderson (2005) articulated that the learner's metacognitive knowledge develops from memorizing facts about strategies, then knowing the proper situations to use them, and finally, mastering each strategy.

Rittle-Johnson et al. (2001) claimed that declarative and procedural knowledge develop in an iterative fashion through improved problem representations. They conducted two experiments on fifth- and sixth-graders learning decimal fractions and found that initial declarative knowledge predicted gains in procedural knowledge and vice versa. They showed that correct problem representation mediated the relation between declarative and procedural knowledge.

Reinforcement Learning in ITSs

Reinforcement Learning (RL) is a popular machine learning branch ideal in environments where actions result in numeric rewards without knowing a ground truth (Sutton & Barto,

2018). Due to its aim of maximizing the cumulative reward, RL has been widely used in educational domains due to the flexible implementation of reward functions (Gao et al., 2023, 2022; Ju et al., 2021; Sanz-Ausin et al., 2020). Deep RL (DRL) is a category of algorithms that combine RL algorithms with neural networks; for instance, Deep Q-Network (DQN) algorithm is the neural network extension of the Q-learning algorithm (Mnih et al., 2015). Substantial work has used RL and DRL in inducing pedagogical policies across ITSs (Ju et al., 2021; Sanz-Ausin et al., 2020; Zhou et al., 2019). Zhou et al. (2019) utilized hierarchical RL to improve the learning gain on an ITS and showed that their policy significantly outperformed an expert and a random condition.

Ju et al. (2021) presented a DRL framework that identifies the critical decisions to induce a critical policy on an ITS. They evaluated their critical-DRL framework based on two success criteria: *necessity* and *sufficiency*. The former required offering help in *all* critical states, and the latter required offering help *only* in critical states. Their results showed that the framework fulfilled both criteria. Sanz-Ausin et al. (2020) conducted two consecutive classroom studies where DRL was applied to decide whether the student or tutor should solve the following problem. They found that the DRL policy with simple explanations significantly improved students' learning performance more than an expert policy.

Despite the wide use of RL and DRL on ITSs, the attempts to combine either with metacognitive knowledge have been minimal (Krueger et al., 2017). Krueger et al. (2017) used RL to teach the metacognitive skill of knowing how much to plan ahead (Deciding How to Decide). Their metacognitive RL framework builds on the semi-gradient SARSA algorithm developed to approximate Markov decision processes. They defined a meta Q-function that takes the meta state of the environment and the planning horizon action. They evaluated their framework on two planning tasks, where constrained reward functions were defined such that the rewards could be predicted many steps ahead to facilitate forming a plan.

To sum up, despite much prior work on declarative, procedural, and conditional knowledge, it has yet to investigate the impact of closing the gap between them on preparation for future learning across ITSs. Our work utilizes DRL in providing adaptive metacognitive interventions to bridge the gap across ITSs. We investigate the impact of our interventions on students' metacognitive development and preparation for future learning. In brief, we induce and deploy a DRL policy of metacognitive interventions on a logic tutor and investigate its impact on a subsequent probability tutor.

Logic and Probability Tutors

Logic Tutor It teaches propositional logic proofs by applying valid inference rules such as Modus Ponens through the standard sequence of pre-test, training and post-test. The three phases share the same interface, but training is the *only* one where students can seek and get help. The pre-test has two problems, while the post-test is harder and has six prob-
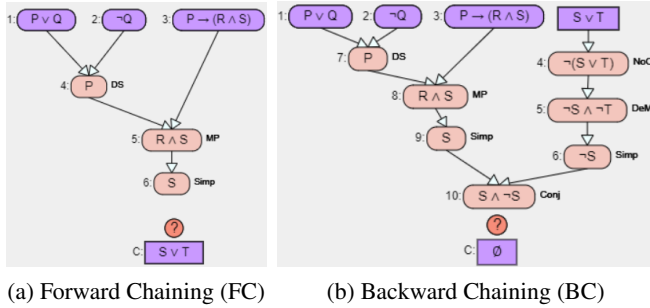
(a) Forward Chaining (FC) (b) Backward Chaining (BC)

Figure 1: Logic Tutor Problem-Solving Strategies

lems; the first two are isomorphic to the pre-test problems. Training consists of five ordered levels with an *incremental degree of difficulty*, and each level consists of four problems. Every problem has a score in the $[0, 100]$ range based on the accuracy, time and solution length.

The *pre-* and *post-test* scores are calculated by averaging their pre- and post-test problem scores. A student can solve any problem throughout the tutor by either a *FC* or a *BC* strategy (Shabrina, Mostafavi, Abdelshiheed, et al., 2023; Shabrina, Mostafavi, Chi, & Barnes, 2023; Shabrina, Mostafavi, Tithi, et al., 2023). Figure 1a shows that for FC, one must derive the conclusion at the bottom from givens at the top, while Figure 1b shows that for *BC*, students need to derive a contradiction from givens and the *negation* of the conclusion. Problems are presented by *default* in FC, but students can switch to BC by clicking a button in the tutor interface.

Figure 2: Training on the Modified Logic Tutor (*DRL*)

Probability Tutor It teaches how to solve probability problems using ten principles, such as the Complement Theorem. The tutor consists of a textbook, pre-test, training, and post-test. Like the logic tutor, training is the only section for students to receive and ask for hints, and the post-test is harder than the pre-test. The textbook introduces the domain principles, while training 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
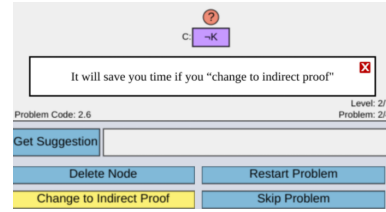

Figure 3: Strategy Switch Nudge

reduced to the givens.

In pre- and post-test, students solve 14 and 20 open-ended problems, where each pre-test problem has an isomorphic post-test one. Answers are graded double-blind by experienced graders using a partial-credit rubric, where grades are based *only* on accuracy in the $[0, 100]$ range. The *pre-* and *post-test* scores are the average grades in their sections.

Methods

Three Metacognitive Interventions As students can choose to switch problem-solving strategies *only* on the logic tutor, our interventions are provided on the logic training. We previously found that **Conditional** students frequently switched *early* (**within the first 30 actions**) to *BC* on the logic tutor, **Procedural** students switched **late** (**after the first 30 actions**), and their **Declarative** peers made **no** switches and used the default strategy (Abdelshiheed et al., 2020). It was also shown that providing metacognitive interventions that presented problems directly in BC or recommended switching to BC —referred to as Nudges— caused *Declarative* and *Procedural* students to catch up with their *Conditional* peers (Abdelshiheed, Hostetter, Shabrina, et al., 2022; Abdelshiheed, Hostetter, Yang, et al., 2022).

This work leverages DRL to provide three metacognitive interventions regardless of the student's metacognitive group: **Nudge (Nud)**, **Present in BC (Prs)**, or **No Intervention (No)**. We trained *Experimental (DRL)* students on the modified tutor shown in Figure 2. Two worked examples on BC were provided for teaching students *how* to use BC, where the tutor showed a step-by-step solution. Since our interventions included the no-intervention option, we intervened in as many problems as possible. Figure 3 shows an example of a nudge, which is prompted after a number of seconds sampled from a probability distribution of prior students' switch behavior (Abdelshiheed et al., 2020). We did not intervene in the last training problem of each level, as it is used to evaluate the improvement on that level.

DRL Policy Induction We used data from four previous studies comprising 867 students (Abdelshiheed, Hostetter, Shabrina, et al., 2022; Abdelshiheed, Hostetter, Yang, et al., 2022; Abdelshiheed et al., 2021, 2020) and performed a $80 - 20$ train-test split. The dataset consisted of a record (**state**, **action**, **reward**) per student per logic training problem. The *state* is the feature vector incorporating 152 features

that capture temporal, accuracy-based and hint-based behaviors. The *action* is Nudge, Present in BC, or No Intervention. The *reward* is the immediate problem score in the logic tutor.

Our objective is to investigate whether DRL works with our metacognitive interventions **rather** than *which* DRL algorithm is better with our interventions. We preferred DRL to RL due to its prevailing success in educational domains (Sanz-Ausin et al., 2020). To select the algorithm, we had to avoid a relatively simple one such as Deep Q-Network (DQN), which overestimates action values (Mnih et al., 2015) and may result in underfitting. Furthermore, we needed to avoid sophisticated DRL algorithms, such as autoencoders and actor-critic approaches, so that DRL does not overshadow the impact of our metacognitive interventions. In other words, a sophisticated DRL algorithm yielding an optimal policy would be acknowledged likely for its sophistication rather than for the metacognitive interventions it provided. Thus, we exploited Double-DQN (DDQN), which solves the overestimation issue in DQN by **decoupling** the action *selection* from *evaluation* in two different neural networks (Van Hasselt et al., 2016). The resulting modified Bellman equation becomes:

$$Q(s,a;\theta) = r + \gamma\, Q(s', argmax_{a'}\, Q(s',a',\theta);\theta^-) \quad (1)$$

where r is the reward; γ is the discount factor; s and s' refer to the current and next states; a and a' denote the current and next actions. DDQN uses the **main** (θ) neural network to *select* the action with the highest Q-value for the next state and then *evaluates* its Q-value using the **target** (θ^-) neural network. After hyperparameter tuning, we picked the model with the lowest mean squared error loss. The deployed policy had two hidden layers with 16 neurons each, 1e-3 learning late, 9e-1 discount factor, 32 batch size, a synchronization frequency of 4 steps between both neural networks (θ and θ^-), and was trained until convergence (≈ 2000 epochs).

Experiment Setup

The experiment took place in an undergraduate Computer Science class in the Fall of 2022 at North Carolina State University. The participants were assigned each tutor as a class assignment and told that completion is required for full credit. We randomly split students into *Experimental (DRL)* and *Control (Ctrl)* conditions, where students were first assigned the logic tutor and then the probability tutor six weeks later. On both tutors, students received the problems in the same order and followed the standard phases described in the Logic and Probability Tutors section. The *only* difference between the two conditions is that *DRL* students received our *adaptive* metacognitive interventions in the **logic training** provided by the DRL policy (Fig. 2), while their *Ctrl* peers received no interventions. On probability, all students received no interventions.

The main challenge in this work was to ensure an even distribution of the metacognitive groups {*Declarative (Decl)*, *Procedural (Proc)*, *Conditional (CDL)*} across the two conditions {*DRL, Ctrl*}. We aimed to investigate whether DRL

would help students with different incoming metacognitive knowledge. Therefore, as per our prior work, we utilized the random forest classifier (RFC) that, based on pre-test performance, predicts the incoming metacognitive group before training on logic and was previously shown to be 96% accurate (Abdelshiheed et al., 2021).

A total of 112 students finished both tutors. We found that our DRL policy provided no interventions for CDL_{DRL} students 94% of the time. Therefore, we combined CDL_{DRL} and CDL_{Ctrl} into *CDL*. After randomly splitting students into conditions and utilizing the RFC for even distribution, we had 22 $Decl_{DRL}$, 24 $Proc_{DRL}$, 22 $Decl_{Ctrl}$, 22 $Proc_{Ctrl}$ and 22 *CDL* students. The RFC was 98% accurate in classifying students who received no interventions —$Decl_{Ctrl}$, $Proc_{Ctrl}$ and *CDL*.

Results

Learning Performance

Table 1: Comparing Groups across Tutors

	Experimental (DRL: $N = 46$)	Control (Ctrl: $N = 44$)	Conditional (CDL: $N = 22$)
	Logic Tutor		
Pre	55.9 (21)	55.8 (21)	58.2 (19)
Iso. Post	**92.1 (5)***	73.4 (17)	83.4 (12)*
Iso. NLG	**0.47 (.1)***	0.16 (.28)	0.35 (.11)*
Post	**87.7 (6)***	69.8 (15)	80.2 (11)*
NLG	**0.45 (.11)***	0.13 (.33)	0.31 (.15)*
	Probability Tutor		
Pre	75.7 (15)	76 (14)	78.6 (14)
Iso. Post	**95.3 (4)***	72.7 (13)	89.1 (7)*
Iso. NLG	**0.4 (.12)***	-0.04 (.18)	0.24 (.15)*
Post	**94.9 (4)***	70.2 (15)	87.7 (8)*
NLG	**0.37 (.14)***	-0.08 (.21)	0.22 (.19)*

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

Experimental vs. Control vs. Conditional   Table 1 shows the groups' performance on both tutors. We display the mean and standard deviation of pre- and post-test scores, isomorphic scores, and the normalized learning gain (*NLG*) defined as ($NLG = \frac{Post - Pre}{\sqrt{100 - Pre}}$), where 100 is the maximum score. We refer to pre- and post-test scores as *Pre* and *Post*, respectively, while the groups are abbreviated as *DRL*[1], *Ctrl* and *CDL*. We performed a Shapiro-Wilk normality test for each metric for each group and found no evidence of non-normality ($p > .05$). A one-way ANOVA using group as factor found no significant difference on *Pre*: $F(2,109) = 0.09, p = .91$ for logic and $F(2,109) = 0.21, p = .81$ for probability. To measure the improvement on isomorphic problems, repeated measures ANOVA tests were conducted (one for each group on each tutor) using {*Pre, Iso. Post*} as factor. On both tutors, we found that *DRL* and *CDL* learned significantly with $p < .0001$, while *Ctrl* did not ($p > .05$).

A one-way ANCOVA[2] using *Pre* as covariate and group

---

[1]Italicized version refers to group; otherwise refers to policy.
[2]General effect size η^2 was reported for conservative results.

as factor found a significant effect on *Post* on both tutors: $F(2, 108) = 38.4, p < .0001, \eta^2 = 0.71$ for logic and $F(2, 108) = 49.6, p < .0001, \eta^2 = 0.79$ for probability. Subsequent Bonferroni-corrected ($\alpha = .05/3$) analyses revealed that *DRL* significantly outperformed both groups, while *CDL* significantly surpassed *Ctrl*; for instance, *DRL* had significantly higher *Post* than *CDL*: $t(66) = 3.6, p < .001$ and $t(66) = 3.8, p < .001$ for logic and probability, respectively. Similar results were found using ANOVA on *NLG*. In brief, these findings on both tutors confirm that $DRL > CDL > Ctrl$.

Groups within Experimental and Control We compared the performance of the metacognitive groups {*Decl, Proc*} across the two conditions {*DRL, Ctrl*} to assess the within- and between-condition impact of the DRL policy. Our results showed the same statistically significant pattern on both tutors for *Post*, *NLG* and their isomorphic versions: $Decl_{DRL}$, $Proc_{DRL} > Decl_{Ctrl}$, $Proc_{Ctrl}$. We display *Pre* and *Post* scores in Figure 4 and report the statistical results for *NLG* in the following paragraph.
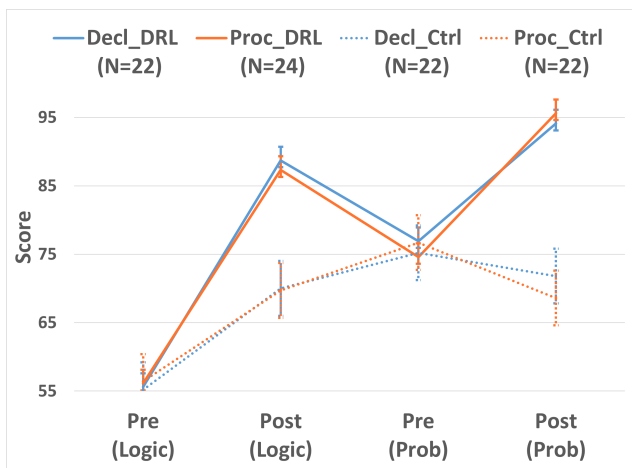

Figure 4: Conditions *Pre* and *Post* Performance across Tutors

A two-way ANOVA using metacognitive group and condition as factors found a significant interaction effect on *NLG* on both tutors: $F(1, 86) = 36.3, p < .0001, \eta^2 = 0.68$ for logic and $F(1, 86) = 45.8, p < .0001, \eta^2 = 0.76$ for probability. Follow-up Bonferroni-adjusted ($\alpha = .05/6$) analyses showed that *DRL* significantly outperformed *Ctrl*. Specifically, on logic, $Decl_{DRL}$ $[0.45(.12)]$ had significantly higher *NLG* than $Decl_{Ctrl}$ $[0.16(.33)]$: $t(42) = 6.2, p < .0001$, and $Proc_{DRL}$ $[0.44(.14)]$ significantly outperformed $Proc_{Ctrl}$ $[0.1(.31)]$: $t(44) = 7.3, p < .0001$. On probability, we found the same patterns, as $Decl_{DRL}$ $[0.34(.13)]$ and $Proc_{DRL}$ $[0.39(.16)]$ surpassed $Decl_{Ctrl}$ $[-0.07(.32)]$ and $Proc_{Ctrl}$ $[-0.1(.25)]$, respectively: $t(42) = 6.5, p < .0001$ and $t(44) = 7.9, p < .0001$.

Policy Decisions and Students' Strategic Behavior

As the learning performance results found no significant difference within the *DRL* group on both tutors, we further investigated the distribution of the DRL policy decisions;

$Decl_{DRL}$ received 94(33%) Nudges, 65(23%) presented in BC and 127(44%) No Intervention, while $Proc_{DRL}$ received 82(26%) Nudges, 74(24%) presented in BC and 156(50%) No Intervention. A chi-square test showed no significant difference in the decisions' distribution between the *DRL* groups: $\chi^2(2, N = 598) = 3.2, p = .2$. Thus, we combined their policy decisions and analyzed the decisions' distribution per logic training level, as shown in Table 2.

A chi-square test showed a significant relationship between the policy decision type and the training level: $\chi^2(8, N = 598) = 81.2, p < .0001$. Post-hoc pairwise chi-square tests with Bonferroni adjustment ($\alpha = .05/10$) showed that the last two levels had significantly more No-Intervention decisions. For instance, the fourth level had more No-Intervention decisions than the third level: $\chi^2(2, N = 276) = 33.4, p < .0001$.

Table 2: Distribution of DRL Policy Decisions across Levels

	L1 (%)	L2 (%)	L3 (%)	L4 (%)	L5 (%)
Nud	**40**	**42**	**37**	17	18
Prs	28	31	31	16	15
No	32	27	32	**67**	**67**

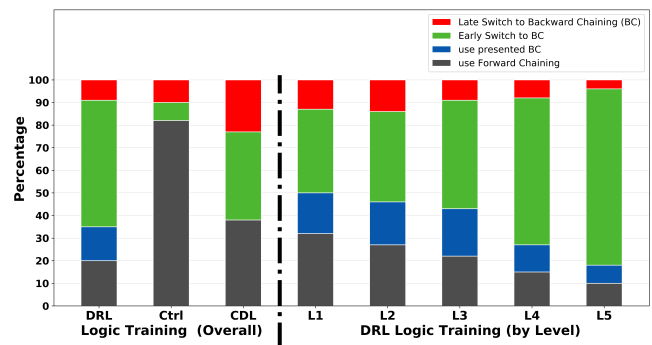In a column (training level), bold is for the highest value.

Figure 5: Students' Strategic Decisions

The students' strategic decisions on logic training are shown in Figure 5 to investigate the impact of the DRL policy on their choices. The first three columns reflect the choices *DRL*, *Ctrl* and *CDL* students made during the entire training, where $Decl_{DRL}$ and $Proc_{DRL}$ decisions had similar distributions and thus were combined. *Early* switches to BC occurred within the first 30 actions, while *late* ones happened after that, as defined in our prior work (Abdelshiheed, Maniktala, et al., 2022; Abdelshiheed et al., 2020). The *Ctrl* and *CDL* students did not have problems presented in BC and hence had one less choice. Bonferroni-corrected ($\alpha = .05/3$) chi-square tests showed that $DRL > CDL > Ctrl$ in their early-switch choices[3]. For example, *DRL* switched early significantly more than *CDL*: $\chi^2(2, N = 794) = 54.9, p < .0001$. The last

---

[3] 'use presented BC' was excluded from pairwise comparisons.

five columns in Figure 5 display the *DRL* students' decisions per training level. Pairwise Bonferroni-adjusted ($\alpha = .05/10$) chi-square tests revealed that *DRL* early-switch choices significantly increased in the last two levels. For example, *DRL* students switched early in the fourth level significantly more than the third one: $\chi^2(3, N = 276) = 25.1, p < .0001$.

In essence, while the DRL policy intervened significantly less in the last two training levels (Table 2), Figure 5 shows that *DRL* students made significantly more early switches in the same two levels. This suggests that *DRL* students became more **autonomous** as training proceeded. Additionally, *DRL* students switched early more frequently than *Ctrl* and *CDL*.

Policy Adaptation for Metacognitive Development

One of our objectives was to investigate whether the DRL policy adapted to the student's metacognitive knowledge as it changes through logic training. We leveraged association rule mining to observe frequent patterns by the DRL policy. Specifically, for a student, each two consecutive policy decisions were encoded into a transaction represented as $\{a_t, c_t, a_{t+1}\}$, where $a_t, a_{t+1} \in \{Nud, Prs, No\}$ and refer to the current and next policy decisions, respectively; $c_t \in \{Agree, Disagree\}$ and indicates whether the student *complied* (agreed/disagreed) with the current policy decision, where agreement and disagreement are defined as:

- *Nudge (Nud)*: early switches to BC represent agreement, while late switches to BC or using FC means disagreement.

- *Present in BC (Prs)*: using BC and FC denote agreement and disagreement, respectively.

- *No Intervention (No)*: agreement is to use FC, while using BC reflects disagreement.

We focused on rules in the form of $\{a_t, c_t\} \Rightarrow a_{t+1}$ to extract meaningful association rules for the next policy decision based on the student's compliance with the current decision. Since c_t has two possible values and a_t, a_{t+1} each has three possible values, there are 18 possible rules. The *support* and *confidence* of each rule were traditionally computed as:

$$\text{Sup}(\{a_t, c_t\} \Rightarrow a_{t+1}) = \frac{count(\{a_t, c_t, a_{t+1}\})}{Total}$$
$$\text{Conf}(\{a_t, c_t\} \Rightarrow a_{t+1}) = \frac{count(\{a_t, c_t, a_{t+1}\})}{count(\{a_t, c_t\})}$$

where $Total = 598$ [46 *DRL* students * 13 training decisions]. Table 3 lists the top six rules for the DRL policy sorted by their support in descending order. The rules learned from $Decl_{DRL}$ and $Proc_{DRL}$ students were similar and thus were combined. The remaining rules were excluded due to their significantly low support and confidence values.

Interpreting Association Rules Table 3 reveals unique perspectives of the DRL policy. In essence, the first **three** rules suggest that the policy treated those who knew in advance *how* and *when* to use BC as *Conditional* students

Table 3: Top Association Rules for DRL policy

Rank	Rule	Supp (%)	Conf (%)
1	$\{No, Disagree\} \Rightarrow No$	23	76
2	$\{Nud, Agree\} \Rightarrow No$	12	61
3	$\{Prs, Agree\} \Rightarrow No$	10	58
4	$\{No, Agree\} \Rightarrow Nud$	10	60
5	$\{Nud, Disagree\} \Rightarrow Prs$	7	69
6	$\{Prs, Disagree\} \Rightarrow Nud$	5	66

by *avoiding* interventions in such situations. The last **two** rules reflect swapping the interventions once a student disagrees with their utility. This finding suggests that students' metacognitive knowledge changes during training and confirms our prior results that metacognitive interventions have different effects on *Declarative* and *Procedural* students (Abdelshiheed, Hostetter, Yang, et al., 2022). The **fourth** rule shows that the DRL policy preferred recommending rather than imposing BC for students who previously used FC.

Discussions & Conclusions

Bridging the Gap We showed that our DRL policy caused students with low incoming metacognitive knowledge (declarative and procedural) to outperform their conditional peers, who had the highest knowledge and received no interventions. In other words, DRL bridged the metacognitive knowledge gap between students on a logic tutor, where the interventions were provided, and on a subsequent probability tutor, where students received no interventions.

Preparation for Future Learning Experimental declarative and procedural students received DRL-based interventions on logic and surpassed their no-intervention control peers on logic and probability. This suggests that DRL prepared students for future learning, as they outperformed control students on probability based on logic interventions.

Autonomy and Metacognitive Development The DRL policy adapted to the back-and-forth metacognitive development between declarative, procedural, and conditional students. Specifically, the association mining rules analyses showed that the DRL policy changed its interventions to adapt to the dynamic metacognitive knowledge of students. Hence, students became more autonomous and made effective strategic decisions, even when DRL intervened significantly less.

Limitations and Future Work There are at least two caveats in our work. First, splitting students into experimental and control resulted in relatively small sample sizes. Second, the probability tutor supported only one strategy, which restricted our intervention ability to logic. Future work involves implementing forward chaining on the probability tutor, comparing multiple DRL algorithms for our interventions, and comparing the DRL interventions against a stronger control that receives random interventions.

Acknowledgments

References

Abdelshiheed, M., Hostetter, J. W., Barnes, T., & Chi, M. (2023). Leveraging deep reinforcement learning for metacognitive interventions across intelligent tutoring systems. In *Proceedings of the 24th international conference on artificial intelligence in education.*

Abdelshiheed, M., Hostetter, J. W., Shabrina, P., Barnes, T., & Chi, M. (2022). The power of nudging: Exploring three interventions for metacognitive skills instruction across intelligent tutoring systems. In *Proceedings of the 44th annual conference of the cognitive science society* (pp. 541–548).

Abdelshiheed, M., Hostetter, J. W., Yang, X., Barnes, T., & Chi, M. (2022). Mixing backward- with forward-chaining for metacognitive skill acquisition and transfer. In *Proceedings of the 23rd international conference on artificial intelligence in education* (pp. 546–552).

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). US Army Combat Capabilities Development Command–Soldier Center.

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* (pp. 2547–2553).

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* (pp. 945–951).

Alam, N., Mostafavi, B., Chi, M., & Barnes, T. (2023). Exploring the effect of autoencoder based feature learning for a deep reinforcement learning policy for providing proactive help. In *Proceedings of the 24th international conference on artificial intelligence in education.*

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological review*, *89*(4), 369.

Anderson, J. R. (2005). *Cognitive psychology and its implications*. Macmillan.

Azevedo, R., & Aleven, V. (2013). *International handbook of metacognition and learning technologies* (Vol. 26). Springer.

Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A. (2005). Adaptive human scaffolding facilitates adolescents' self-regulated learning with hypermedia. *Instructional science*, *33*, 381–412.

Baker, L. (1994). Fostering metacognitive development. *Advances in child development and behavior*, *25*, 201–239.

Biggs, J., Tang, C., & Kennedy, G. (1999). Teaching for quality learning at universities. *Open University, Buckingham.*

Bloom, B. S. (1956). Taxonomy of educational objectives: The classification of educational goals. *Handbook; Cognitive domain*, *1*.

Boden, K., Kuo, E., Nokes-Malach, T., Wallace, T., & Menekse, M. (2018). What is the role of motivation in procedural and conceptual physics learning? an examination of self-efficacy and achievement goals. *Physics Education Research Conference*, 60–63.

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

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. *Metacognition, motivation, and understanding*, 65–116.

Butler, D. L. (1998). The strategic content learning approach to promoting self-regulated learning: A report of three studies. *Journal of educational psychology*, *90*(4), 682.

Case, J., & Gunstone, R. (2002). Metacognitive development as a shift in approach to learning: an in-depth study. *Studies in Higher education*, *27*(4), 459–470.

Castro-Schez, J. J., Glez-Morcillo, C., Albusac, J., & Vallejo, D. (2021). An intelligent tutoring system for supporting active learning: A case study on predictive parsing learning. *Information Sciences*, *544*, 446–468.

Cutrer, W. B., Miller, B., Pusic, M. V., Mejicano, G., Mangrulkar, R. S., Gruppen, L. D., . . . Moore Jr, D. E. (2017). Fostering the development of master adaptive learners: a conceptual model to guide skill acquisition in medical education. *Academic medicine*, *92*(1), 70–75.

De Backer, L., Van Keer, H., & Valcke, M. (2012). Exploring the potential impact of reciprocal peer tutoring on higher education students' metacognitive knowledge and regulation. *Instructional science*, *40*, 559–588.

Detterman, D. K., & Sternberg, R. J. (1993). *Transfer on trial: Intelligence, cognition, and instruction.* Ablex Publishing.

Dochy, F. J. R. C. (1992). *Assessment of prior knowledge as a determinant for future learning*. Open Universiteit.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, *34*(10), 906.

Fossati, D. (2009). *Automatic modeling of procedural knowledge and feedback generation in a computer science tutoring system*. University of Illinois at Chicago.

Gao, G., Gao, Q., Yang, X., Pajic, M., & Chi, M. (2022). A reinforcement learning-informed pattern mining frame-

work for multivariate time series classification. In *Proceedings of the 31st international joint conference on artificial intelligence.*

Gao, G., Ju, S., Ausin, M. S., & Chi, M. (2023). Hope: Human-centric off-policy evaluation for e-learning and healthcare. In *Proceedings of the 22nd international conference on autonomous agents and multiagent systems.*

Georgeff, M. P., & Lansky, A. L. (1986). Procedural knowledge. *Proceedings of the IEEE*, *74*(10), 1383–1398.

Hershkovitz, A., Baker, R., Gowda, S. M., & Corbett, A. T. (2013). Predicting future learning better using quantitative analysis of moment-by-moment learning. In *Educational data mining.*

Hostetter, J. W., Abdelshiheed, M., Barnes, T., & Chi, M. (2023a). Leveraging fuzzy logic towards more explainable reinforcement learning-induced pedagogical policies on intelligent tutoring systems. In *2023 IEEE international conference on fuzzy systems.*

Hostetter, J. W., Abdelshiheed, M., Barnes, T., & Chi, M. (2023b). A self-organizing neuro-fuzzy q-network: Systematic design with offline hybrid learning. In *Proceedings of the 22nd international conference on autonomous agents and multiagent systems.*

Jacobs, J. E., & Paris, S. G. (1987). Children's metacognition about reading: Issues in definition, measurement, and instruction. *Educational psychologist*, *22*(3-4), 255–278.

Ju, S., Zhou, G., Abdelshiheed, M., Barnes, T., & Chi, M. (2021). Evaluating critical reinforcement learning framework in the field. In *Proceedings of the 22nd international conference on artificial intelligence in education.* (pp. 215–227).

Kiesewetter, J., Ebersbach, R., Tsalas, N., Holzer, M., Schmidmaier, R., & Fischer, M. R. (2016). Knowledge is not enough to solve the problems–the role of diagnostic knowledge in clinical reasoning activities. *BMC medical education*, *16*(1), 1–8.

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into practice*, *41*(4), 212–218.

Krueger, P. M., Lieder, F., & Griffiths, T. (2017). Enhancing metacognitive reinforcement learning using reward structures and feedback. In *Cogsci.*

Kua, J., Lim, W.-S., Teo, W., & Edwards, R. A. (2021). A scoping review of adaptive expertise in education. *Medical Teacher*, *43*(3), 347–355.

Kuhn, D. (2000). Metacognitive development. *Current directions in psychological science*, *9*(5), 178–181.

Larkin, S. (2009). *Metacognition in young children*. Routledge.

Livingston, J. A. (2003). *Metacognition: An overview.* ERIC.

Mbato, C. L. (2019). Indonesian efl learners' critical thinking in reading: Bridging the gap between declarative, procedural and conditional knowledge. *Humaniora*, *31*(1), 92.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... others (2015). Human-level control through deep reinforcement learning. *nature*, *518*(7540), 529–533.

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into practice*, *41*(4), 219–225.

Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W. (2001). Developing conceptual understanding and procedural skill in mathematics: An iterative process. *Journal of educational psychology*, *93*(2), 346.

Roberts, M. J., & Erdos, G. (1993). Strategy selection and metacognition. *Educational psychology*, *13*(3-4), 259–266.

Sanz-Ausin, M., Azizsoltani, H., Barnes, T., & Chi, M. (2019). Leveraging deep reinforcement learning for pedagogical policy induction in an intelligent tutoring system. *International Educational Data Mining Society.*

Sanz-Ausin, M., Maniktala, M., Barnes, T., & Chi, M. (2020). Exploring the impact of simple explanations and agency on batch deep reinforcement learning induced pedagogical policies. In *AIED* (pp. 472–485).

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional science*, *26*(1-2), 113–125.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational psychology review*, *7*, 351–371.

Shabrina, P., Mostafavi, B., Abdelshiheed, M., Chi, M., & Barnes, T. (2023). Investigating the impact of backward strategy learning in a logic tutor: Aiding subgoal learning towards improved problem solving. *International Journal of Artificial Intelligence in Education.*

Shabrina, P., Mostafavi, B., Chi, M., & Barnes, T. (2023). Impact of learning a subgoal-directed problem-solving strategy in an intelligent logic tutor. In *Proceedings of the 24th international conference on artificial intelligence in education.*

Shabrina, P., Mostafavi, B., Tithi, S. D., Chi, M., & Barnes, T. (2023). Learning problem decomposition-recomposition with data-driven chunky parsons problem within an intelligent logic tutor. In *Proceedings of the 16th international conference on educational data mining.*

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction.* MIT press.

Teng, F. (2020). The role of metacognitive knowledge and regulation in mediating university efl learners' writing performance. *Innovation in Language Learning and Teaching*, *14*(5), 436–450.

Van Hasselt, H., et al. (2016). Deep reinforcement learning with double q-learning. In *AAAI* (Vol. 30).

Yildirim, Z., Ozden, M. Y., & Aksu, M. (2001). Comparison of hypermedia learning and traditional instruction on knowledge acquisition and retention. *The Journal of educational research*, *94*(4), 207–214.

Zhou, G., Azizsoltani, H., Ausin-Sanz, M., Barnes, T., & Chi, M. (2019). Hierarchical reinforcement learning for pedagogical policy induction. In *Proceedings of the 20th international conference on artificial intelligence in education* (pp. 544–556).