利用深度强化学习弥合陈述性、程序性和条件性元认知知识差距

Mark Abdelshiheed、John Wesley Hostetter、Tiffany Barnes 和 Min Chi北卡罗来纳州立大学计算机科学系，北卡罗来纳州罗利市，邮编 27695{mnabdels、jwhostet、tmbarnes、mchi}@ncsu.edu

arXiv:2304.11739v1 [cs.CY] 23 Apr 2023

摘要

在演绎领域中，按层次递增排列的三种元认知知识类型是陈述性、程序性和条件性学习。本研究利用深度强化学习（深度强化学习）提供自适应元认知干预，以弥合这三种知识类型之间的差距，并帮助学生为在不同智能辅导系统（智能辅导系统）中的未来学习做好准备。学生接受了这些干预，学习如何以及何时在逻辑辅导系统中使用反向链式推理（反向链式）策略；该系统默认支持前向链推理策略。六周后，我们在没有干预的情况下，让学生使用仅支持反向链式推理的概率辅导系统进行学习。结果显示，在这两种智能辅导系统中，深度强化学习都弥合了学生之间的元认知知识差距，并显著提升了学生相较于对照组同伴的学习表现。此外，深度强化学习策略能够根据学生在逻辑辅导系统中的元认知发展进行调整，适用于陈述型、程序型和条件型学生，从而使他们的策略决策更加自主。

关键词： 深度强化学习；为未来学习做准备；智能辅导系统；陈述性知识；程序性知识；条件性知识

引言

学习的一项重要且必不可少的特征，是持续为未来学习做好准备（Bransford & Schwartz，1999年）。我们逐步积累的知识证明了为未来学习做准备这一过程确实存在，但它难以预测和衡量（Detterman & Sternberg，1993年）。大量研究发现，有助于学生为未来学习做准备的一个因素是他们的元认知知识（Kua 等人，2021年；Cutrer 等人，2017年；Hershkovitz 等人，2013年）。

在演绎领域中，元认知知识有三种类型：陈述性、程序性和条件性 学习（Anderson，2005年；Schraw，1998年；Bloom，1956年）。尽管大量研究表明，这三种类型可以按顺序或同时习得（Teng，2020年；Yildirim 等人，2001年；Schraw，1998年），研究也表明，根据教育情境和学习环境的不同，每位学习者都具有一种主导知识类型（Kuhn，2000年；Brown，1987年）。此外，既有研究表明，学生的元认知知识会在学习过程中不断发展（Rittle-Johnson 等人，2001年；Baker，1994年）。因此，需要采取适应性干预措施，以考虑这种发展（Azevedo 等人，2005年；Pintrich，2002年；Butler，1998年）。

强化学习（RL）（Sutton 和 Barto，2018年）是智能辅导系统（ITSs）中提供自适应 支持和脚手架支持的最有效方法之一（Krueger 等人，2017年；Zhou 等人，2019年）。深度学习的扩展形式深度强化学习（DRL），由于能够支持更复杂的模型，已普遍用于智能辅导系统中的教学策略归纳（Abdelshiheed 等人，2023年；Alam 等人，2023年；Hostetter 等人，2023年（a）、2023年（b）；Ju 等人，2021年；Sanz-Ausin 等人，2019年）。据我们所知，此前尚无研究利用深度强化学习提供自适应干预，以弥合元认知知识差距，并帮助学生为未来学习做好准备。

这项工作建立在我们先前的研究之上。在先前的研究中，我们根据学生在智能辅导系统（ITS）上占主导地位的元认知知识，将其分为陈述性、程序性或条件性学习者。我们发现，只有条件性学生为未来学习做好了准备，因为在不同的演绎领域中，他们的表现显著优于陈述性和程序性同伴（Abdelshiheed等，2021年，2020年）。受这些发现启发，本研究通过课堂研究实证评估深度强化学习的有效性；我们利用深度强化学习为陈述性和程序性学生提供自适应元认知干预，以弥合知识差距。深度强化学习提供元认知干预，教授学生如何以及何时在支持默认正向链式推理策略的逻辑辅导系统上使用反向链式推理（反向链式）策略。六周后，我们在学生未接受干预的情况下，使用仅支持反向链式推理的概率辅导系统对其进行训练。结果表明，深度强化学习确实弥合了差距，使学生为未来学习做好了准备，适应了他们的元认知发展，并提升了他们的决策自主性。

背景与相关工作

陈述性、程序性和条件性知识

元认知指对认知的认知，以及控制、构想和调节知识的能力（Livingston，2003年；Roberts 和 Erdos，1993年；Flavell，1979年）。元认知知识包括陈述性、程序性和条件性三种类型（Schraw，1998年；Schraw 和 Moshman，1995年；Brown，1987年；Jacobs 和 Paris，1987年；Bloom，1956年）。

陈述性 知识也被称为表层学习或机械记忆学习（Biggs 等人，1999年），被认为是最简单、最低层次的知识，因为它涉及对学习环境默认设置中提供的事实和数据进行记忆（Azevedo 和 Aleven，2013年；De Backer 等人，2012年；Krathwohl，2002年；Bloom，1956年）。**程序性**知识是一种更高层次的知识，其特点是能够自动

理解如何使用不同的问题解决策略和认知技能，而无需有意识地关注或推理这些策略和技能的依据（Azevedo 和 Aleven，2013年；Krathwohl，2002年；Dochy，1992年；Georgeff 和 Lansky，1986年；Bloom，1956年）。**条件性**知识是最高层次的知识，因为它要求理解如何、何时 以及为何使用每种策略和认知技能（Kiesewetter 等人，2016年；Larkin，2009年；Schraw，1998年）。

大量研究探讨了习得和培养每种知识类型的重要性（Castro-Schez 等人，2021年；Boden 等人，2018年；Fossati，2009），并弥合了这些知识类型之间的差距（Teng，2020年；Mbato，2019年；Yildirim 等人，2001年）。Boden 等人（2018年）发现，在解决物理问题时，学生的自我效能与陈述性知识的相关性显著高于与程序性知识的相关性。Fossati（2009）发现，在教授链表的辅导系统中，程序性知识较强的学生表现显著优于同伴。Castro-Schez 等人（2021年）的研究发现，在辅导系统中学习预测分析算法时，条件性知识较强的学生表现显著优于条件性知识较弱的学生。

元认知发展

元认知发展被定义为学生在学习方式以及所运用的元认知知识和技能方面发生的变化（Anderson，2005年；Case 和 Gunstone，2002年；Rittle-Johnson 等人，2001年；Kuhn，2000年；Baker，1994年；Anderson，1982年）。我们关注陈述性知识、程序性知识和条件性知识的发展。Anderson（1982年）提出了一个用于发展和习得元认知技能的框架。他提出技能习得包含两个主要阶段：陈述阶段，即理解有关领域的事实；程序阶段，即将领域知识融入执行技能的程序中。Anderson（1982年）指出，学习者从陈述阶段过渡到程序阶段时，会发生知识编译。后来，Anderson（2005年）提出，程序性知识取决于条件性知识，而条件性知识又取决于陈述性知识。具体而言，Anderson（2005年）阐述了学习者的元认知知识如何逐步发展：先记住有关策略的事实，再了解适合运用这些策略的情境，最后掌握每种策略。

Rittle-Johnson 等人（2001年）声称，陈述性知识和程序性知识通过不断改进问题表征，以迭代方式发展。他们对学习小数分数的五、六年级学生进行了两项实验，发现初始陈述性知识能够预测程序性知识的增长，反之亦然。他们表明，正确的问题表征在陈述性知识与程序性知识之间的关系中发挥了中介作用。

智能辅导系统中的强化学习

强化学习（RL）是机器学习中一个广受欢迎的分支，适用于动作会产生数值奖励、但不知道真实值的环境（Sutton 和 Barto，2018年）。由于强化学习旨在最大化累积奖励，且奖励函数的实现方式具有灵活性，因此强化学习已被广泛应用于教育领域（Gao 等人，2023年、2022年；Ju 等人，2021年；Sanz-Ausin 等，2020年）。深度强化学习是将强化学习算法与神经网络相结合的一类算法；例如，深度 Q 网络（DQN）算法是 Q 学习算法在神经网络方面的扩展（Mnih 等，2015）。大量研究已将强化学习和深度强化学习应用于各类智能辅导系统中的教学策略生成（Ju 等人，2021年；Sanz-Ausin 等人，2020年；Zhou 等人，2019年）。Zhou 等人（2019年）利用分层强化学习提高智能辅导系统中的学习增益，并表明其策略的表现显著优于专家策略和随机条件。

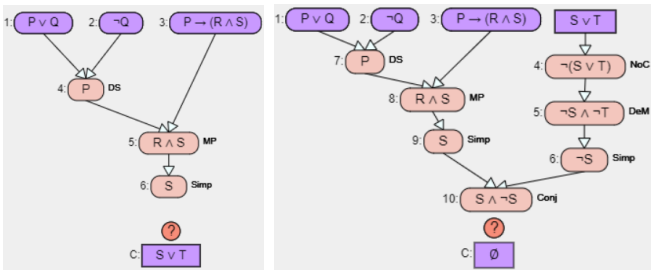
Ju 等人（2021年）提出了一个深度强化学习框架，用于识别关键决策，从而在智能辅导系统中形成关键策略。他们依据两个成功标准评估了其关键决策深度强化学习框架：必要性和充分性。前者要求在所有关键状态下提供帮助，后者则要求仅在关键状态下提供帮助。他们的结果表明，该框架满足了这两个标准。Sanz-Ausin 等人（2020年）开展了两项连续的课堂研究，将深度强化学习应用于决定应由学生还是辅导教师解决下一道问题。他们发现，与专家策略相比，配有简单解释的深度强化学习策略能够显著提升学生的学习表现。

尽管强化学习和深度强化学习已广泛应用于智能辅导系统，但将二者中的任何一种与元认知知识相结合的尝试仍然很少（Krueger 等人，2017年）。Krueger 等人（2017年）使用强化学习教授“知道应提前规划多远”这一元认知技能（决定如何决策）。他们的元认知强化学习框架建立在用于近似马尔可夫决策过程的半梯度SARSA算法之上。他们定义了一个元Q函数，该函数接收环境的元状态和规划时域动作。在两项规划任务中，他们对该框架进行了评估；在这些任务中，他们定义了受约束的奖励函数，使得奖励可以提前许多步进行预测，从而促进计划的形成。

总而言之，尽管此前已有大量关于陈述性知识、程序性知识和条件性知识的研究，但跨智能辅导系统的研究尚未探讨弥合它们之间的知识差距对为未来学习做准备所产生的影响。我们的工作利用深度强化学习，在智能辅导系统中提供自适应元认知干预，以弥合这一差距。我们考察了这些干预对学生元认知发展和为未来学习做准备的影响。简而言之，我们在逻辑辅导系统中诱导并部署元认知干预的深度强化学习策略，并研究其对后续概率辅导系统的影响。

逻辑与概率辅导工具

逻辑辅导系统 通过应用有效的推理规则（如肯定前件式），按照前测、训练和后测的标准顺序教授命题逻辑证明。这三个阶段共享相同的界面，但训练是学生可以寻求并获得帮助的唯一阶段。前测包含两个问题，而后测难度更高，包含六个问-



(a) 前向链推理（前向链推理） (b) 后向链推理（反向链式）

图 1：逻辑辅导系统的问题解决策略

题；其中前两个问题与前测问题同构。训练包括五个按顺序排列、具有递增难度的级别，每个级别包含四个问题。每个问题都有一个 $[0, 100]$ 范围内的分数，该分数取决于准确性、时间和解答长度。

前测- 和后测 分数通过分别对前测和后测中的题目得分取平均值计算得出。学生可以在整个辅导过程中通过前向链推理（ FC ）或反向链式 （ BC ）策略解答任意题目（ Shabrina, Mostafavi, Abdelshiheed, 等，2023；Shabrina, Mostafavi, Chi, 和 Barnes，2023；Shabrina, Mostafavi, Tithi, 等，2023）。图1a显示，对于FC，必须根据顶部给出的已知条件推导出底部的结论；而图1b显示，对于反向链式，学生需要根据已知条件和结论的否定 推导出矛盾。在默认 情况下，题目采用FC方式呈现，但学生可以通过点击辅导系统界面中的按钮切换到BC方式。

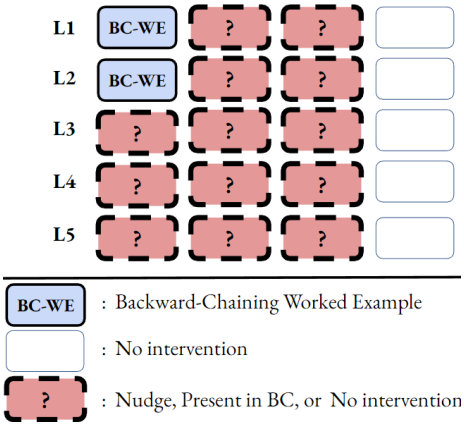


图2：在修改版逻辑辅导系统（深度强化学习）上进行训练

概率辅导系统 它教授如何运用十条原理（例如补集定理）来解概率题。该辅导系统由教科书、前测、训练和后测组成。与逻辑辅导系统一样，训练是学生唯一可以获取提示并请求提示的部分，而后测比前测更难。教科书介绍了领域原理，训练则包含 12 道题，每道题只能 通过反向链式 方法求解，因为这需要列出并求解方程 ，直到最终将目标

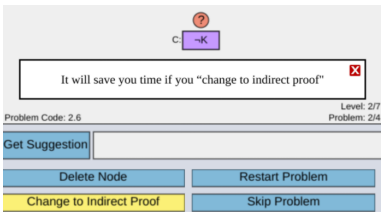


图3：策略切换提示

化简为已知条件。

在前测和后测中，学生分别解答14道和20道开放式问题，其中每道前测题都有一道与之同构的后测题。评分由经验丰富的评分者采用部分得分评分标准进行双盲评定，评分仅 依据 $[0, 100]$ 范围内的准确性。前测-和后测 分数为各部分的平均得分。

方法

三种元认知干预 由于学生只能在逻辑辅导系统上选择切换问题解决策略，因此我们的干预在逻辑训练中提供。我们此前发现，条件性 学生常在逻辑辅导系统中较早切换到反向链式 （**前30个动作内**），程序性 学生则较晚 切换（**前30个动作之后**），而他们的陈述性 同伴**不会** 切换，并使用默认策略（Abdelshiheed 等人，2020年）。研究还表明，提供将问题直接以反向链式方式呈现或建议切换到反向链式方式的元认知干预——称为提示——会使陈述性 和程序性 学生赶上条件性 同伴（Abdelshiheed、Hostetter、Shabrina等，2022年；Abdelshiheed、Hostetter、Yang等，2022年）。

本研究利用深度强化学习，无论学生属于哪个元认知组，均提供三种元认知干预：提示（***Nud***）、呈现于***BC***（***Prs***）或无干预（***No***）。我们让实验组（深度强化学习）学生使用图2所示的改良导师系统进行训练。我们提供了两个关于反向链式的演示例题，以教学生如何使用反向链式；导师系统会展示逐步解答。由于我们的干预包含无干预选项，因此我们尽可能在更多题目中实施干预。图3展示了一个提示示例：提示会在经过若干秒后触发，等待时长根据先前学生切换行为的概率分布进行抽样（Abdelshiheed 等，2020年）。每个级别的最后一道训练题不实施干预，因为该题用于评估学生在该级别上的进步。

深度强化学习策略诱导 我们使用了来自四项既有研究的数据，共涉及867名学生（Abdelshiheed、Hostetter、Shabrina等，2022年；Abdelshiheed、Hostetter、Yang等，2022年；Abdelshiheed等，2021年、2020年），并进行了 80 – 20 训练测试划分。数据集由每名学生针对每道逻辑训练题的一条记录（**状态**、**动作**、**奖励**）组成。状态 是包含152个特征的特征向量

这些特征反映了时间性、基于准确性以及基于提示的行为。动作包括提示、呈现于BC或无干预。奖励 是逻辑辅导系统中题目的即时得分。

我们的目标是探究深度强化学习能否与我们的元认知干预配合使用，**而不是** 探究哪种深度强化学习 算法更适合我们的干预。由于强化学习在教育领域已取得显著成效（Sanz-Ausin 等，2020），我们选择了深度强化学习。为了选择算法，我们必须避免使用相对简单的算法，例如深度 Q 网络（DQN），它会高估动作值（Mnih 等，2015），并可能导致欠拟合。此外，我们还需要避免使用自编码器和行动者-评论家方法等复杂的深度强化学习算法，以免深度强化学习掩盖元认知干预的影响。换言之，如果一种复杂的深度强化学习算法得出了最优策略，人们很可能会将其归功于算法本身的复杂性，而非其所提供的元认知干预。因此，我们采用了双重 DQN（DDQN），它通过在两个不同的神经网络中**将** 动作选择与评估 分离，解决了 DQN 中的过高估计问题（ Van Hasselt 等，2016 ）。由此得到的修正贝尔曼方程为:

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

其中r 表示奖励； γ 表示折扣因子； s 和s' 分别表示当前状态和下一状态； a 和a' 分别表示当前动作和下一动作。DDQN 使用**主** （θ）神经网络为下一状态选择 Q 值最高的动作，然后使用**目标** （θ^{-}）神经网络评估该动作的 Q 值。经过超参数调优后，我们选择了均方误差损失最低的模型。部署的策略包含两个隐藏层，每层有 16 个神经元，学习率为 1e-3，折扣因子为 9e-1，批量大小为 32，两个神经网络（ θ 和 θ^{-}）之间每 4 步同步一次，并训练至收敛（≈ 2000 个训练周期）。

实验设置

实验于 2022 年秋季在北卡罗来纳州立大学的一门本科计算机科学课程中进行。参与者被分配使用每个辅导系统作为课程作业，并被告知完成作业是获得全部学分的必要条件。我们将学生随机分配至实验组（深度强化学习）和对照组（对照组）条件：学生先被分配使用逻辑辅导系统，六周后再使用概率辅导系统。在两个辅导系统中，学生均按照相同的顺序接收题目，并遵循“逻辑与概率辅导工具”部分所述的标准阶段。两种条件之间 唯一 的区别在于，深度强化学习 学生在**逻辑训练** 中接受由深度强化学习策略提供的自适应 元认知干预（图 2），而他们的对照组同伴不接受任何干预。在概率辅导系统中，所有学生均不接受任何干预。

本研究面临的主要挑战，是确保两种条件 {陈述性 *(Decl)*、程序性 *(Proc)*、条件性 *(CDL)*} 下的元认知组 {深度强化学习、对照组} 分布均衡。我们的目标是探究深度强化学习是否有助于具备不同初始元认知知识的学生。因此，按照我们此前的研究，我们采用了随机森林分类器（RFC），该分类器根据前测表现，在逻辑训练前预测学生的初始元认知组；此前研究显示其准确率达到 96%（Abdelshiheed 等，2021年）。

共有112名学生完成了两种辅导系统。我们发现，深度强化学习策略在94%的时间里不会对CDL_{DRL} 学生进行干预。因此，我们将CDL_{DRL} 和CDL_{Ctrl} 合并为CDL。将学生随机分配到各条件，并使用RFC确保均匀分布后，我们有22名$Decl_{DRL}$学生、24名$Proc_{DRL}$学生、22名$Decl_{Ctrl}$学生、22名$Proc_{Ctrl}$ 学生，以及22名CDL学生。RFC对未接受干预的学生进行分类时，准确率达到98%——$Decl_{Ctrl}$、$Proc_{Ctrl}$ 和CDL。

结果

学习表现

表1：各辅导系统之间的组别比较

	实验组 (深度强化学习: $N = 46$)	对照组 (对照组: $N = 44$)	条件性 (CDL: $N = 22$)
逻辑辅导系统			
Pre	55.9(21)	55.8(21)	58.2(19)
Iso. 后置	**92.1**(**5**)*	73.4(17)	83.4(12)*
Iso. 自然语言生成	**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)*
概率辅导系统			
Pre	75.7(15)	76(14)	78。6(14)
Iso. 后置	**95.3**(**4**)*	72。7（13）	89.1(7)*
Iso. 自然语言生成	**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)*

连续出现时，粗体表示最高值，星号表示相对于 n 具有显著性 o 个星号。

实验组 vs. 对照组 vs. 条件性 表1展示了各组在两种辅导系统上的表现。我们列出了前测和后测分数、同构问题分数的均值和标准差，以及标准化学习增益（NLG），其定义为 (NLG $= \frac{\text{Post}-\text{Pre}}{\sqrt{100}-\text{Pre}}$)，其中 100 为最高分。我们分别将前测和后测分数称为 Pre 和 Post，各组则缩写为 DRL[1]，对照组和 CDL。我们对每组的每项指标进行了 Shapiro-Wilk正态性检验，没有发现数据不服从正态分布的证据（$p > .05$）。以组别为因子的单因素方差分析显示，逻辑的前测分数无显著差异：$F(2, 109) = 0.09, p = .91$；概率的前测分数也无显著差异：$F(2, 109) = 0.21, p = .81$。为衡量同构问题上的提升，我们进行了重复测量方差分析（每个组在每种辅导系统上各进行一次），以 {Pre、Iso. 后置} 为因子。在两种辅导系统上，DRL 和 CDL 均取得了显著学习增益（$p < .0001$），而对照组没有（$p > .05$）。

单因素协方差分析[2] 以前测 为协变量、组别

[1]斜体版本指的是组别；否则指的是策略。[2]报告了总体效应量 η^2，以获得保守的结果。

为因子，发现其对两种后置 辅导系统均有显著影响：逻辑方面$F(2,108) = 38.4, p < .0001, \eta^2 = 0.71$，概率方面$F(2,108) = 49.6, p < .0001, \eta^2 = 0.79$。随后进行的Bonferroni校正（$\alpha = .05/3$）分析显示，深度强化学习 显著优于另外两组，而*CDL*显著优于对照组；例如，逻辑和概率方面深度强化学习 的后置 得分均显著高于*CDL*：分别为$t(66) = 3.6, p < .001$ 和$t(66) = 3.8, p < .001$。对*NLG*进行方差分析也得到了类似结果。简言之，两种辅导系统上的这些发现证实，深度强化学习 >*CDL* >对照组。

实验组和对照组中的组别我们比较了两个条件下元认知组 {陈述性、程序性}的表现 {深度强化学习、对照组}，以评估深度强化学习策略在条件内和条件间的影响。结果显示，在两个辅导系统中，后置、*NLG* 及其同构版本均呈现相同的统计显著模式：陈述性$_{DRL}$、程序性$_{DRL}$ > 陈述性$_{Ctrl}$、程序性$_{Ctrl}$。图4展示了前测和后置得分，下一段报告了*NLG* 的统计结果。

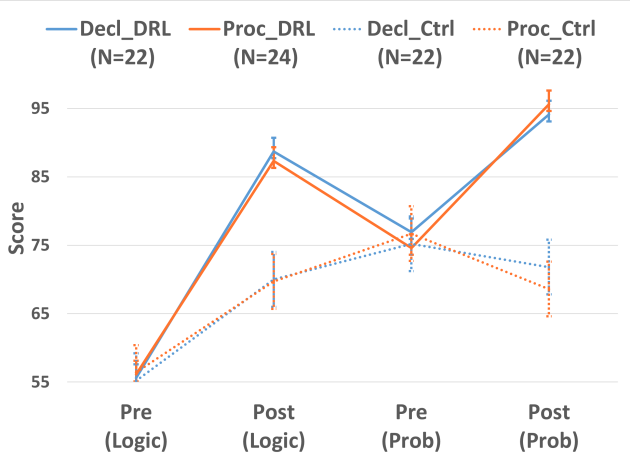


图4：两个辅导系统中各条件前测和后置的表现

以元认知组和条件为因素进行的双因素方差分析发现，两个辅导系统中的NLG均存在显著交互效应：逻辑方面为 $F(1, 86) = 36.3$, $p < .0001, \eta^2 = 0.68$；概率方面为 $F(1, 86) = 45.8$, $p < .0001, \eta^2 = 0.76$。后续的Bonferroni调整（$\alpha = .05/6$）分析显示，深度强化学习组的表现显著优于对照组。具体而言，在逻辑方面，Decl$_{DRL}$ [0.45(.12)]的NLG显著高于Decl$_{Ctrl}$ [0.16(.33)]：$t(42) = 6.2$, $p < .0001$；Proc$_{DRL}$[0.44(.14)]的表现显著优于Proc$_{Ctrl}$ [0.1(.31)]：$t(44) = 7.3$, $p < .0001$。在概率方面，我们发现了相同的模式：Decl$_{DRL}$ [0.34(.13)]和Proc$_{DRL}$ [0.39(.16)]分别优于Decl$_{Ctrl}$ [−0.07(.32)]和Proc$_{Ctrl}$ [−0.1(.25)]，：$t(42) = 6.5$, $p < .0001$；$t(44) = 7.9$, $p < .0001$。

策略决策与学生的策略行为

由于学习表现结果显示，在两个辅导系统中，深度强化学习组内均未发现显著差异，我们进一步调查了深度强化学习策略决策的分布；

陈述性$_{DRL}$ 收到94次（33%）提示，65次（23%）以反向链式呈现，127次（44%）为无干预，而程序性$_{DRL}$ 收到82次（26%）提示，74次（24%）以反向链式呈现，156次（50%）为无干预。卡方检验显示，深度强化学习组之间的决策分布没有显著差异：$\chi^2(2, N = 598) = 3.2$, $p = .2$。因此，我们合并了它们的策略决策，并分析了每个逻辑训练水平下的决策分布，如表2所示。

卡方检验显示，策略决策类型与训练水平之间存在显著关联：$\chi^2(8, N = 598) = 81.2$, $p < .0001$。采用邦费罗尼校正的事后两两卡方检验（$\alpha = .05/10$）显示，最后两个训练水平的无干预决策显著更多。例如，第四个训练水平的无干预决策多于第三个训练水平：$\chi^2(2, N = 276) = 33.4$, $p < .0001$。

表2：各训练水平的深度强化学习策略决策分布

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

在一列（训练水平）中，粗体表示最高值。

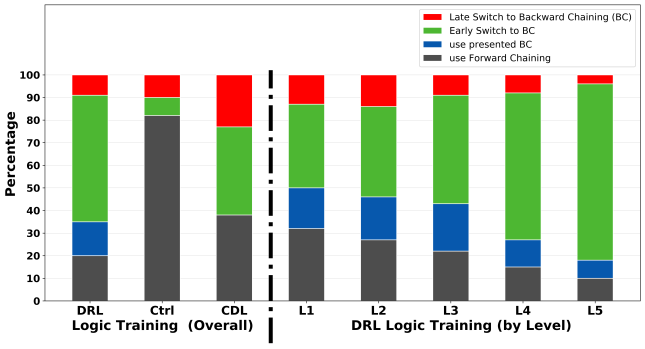


图5：学生的策略决策

学生在逻辑训练中的策略决策如图5所示，用于考察深度强化学习策略对其选择的影响。前三列反映了深度强化学习、对照组和CDL学生在整个训练期间所做的选择，其中*Decl*$_{DRL}$ 和*Proc*$_{DRL}$ 决策的分布相似，因此将二者合并。根据我们先前的研究（Abdelshiheed、Maniktala 等，2022年；Abdelshiheed 等，2020年）的定义，提前切换到反向链式发生在前30个动作内，而延迟切换则发生在此之后。对照组和CDL学生接受的问题中不包含反向链式，因此少了一种选择。Bonferroni校正（$\alpha = .05/3$）卡方检验显示，深度强化学习 >CDL >对照组在提前切换选择方面存在差异[3]。例如，深度强化学习组提前切换的次数显著多于CDL组：$\chi^2(2, N = 794) = 54.9$, $p < .0001$。最后

[3]“使用所呈现的反向链式”未纳入两两比较。

图5中的五列展示了深度强化学习学生在各训练水平上的决策。成对Bonferroni调整（$\alpha = .05/10$）卡方检验显示，深度强化学习组的提前切换选择在最后两个训练水平中显著增加。例如，深度强化学习学生在第四个训练水平中的提前切换显著多于第三个训练水平：$\chi^2(3, N = 276) = 25.1, p < .0001$。

本质上，尽管深度强化学习策略在最后两个训练水平中的干预显著减少（表2），但图5显示，深度强化学习学生在这两个训练水平中更早切换的次数显著增加。这表明，随着训练的推进，深度强化学习学生变得更加**自主**。此外，深度强化学习学生比对照组和*CDL*更频繁地提前切换。

元认知发展的策略适应

我们的目标之一是探究深度强化学习策略是否会随着学生在逻辑训练过程中元认知知识的变化而进行适应。我们利用关联规则挖掘来观察深度强化学习策略产生的频繁模式。具体而言，对于一名学生，每两个连续的策略决策被编码为一个事务，表示为 $\{a_t, c_t, a_{t+1}\}$，其中 $a_t, a_{t+1} \in \{Nud、Prs、No\}$ 分别指当前和下一次策略决策；$c_t \in \{Agree、Disagree\}$ 表示学生是否遵从了当前策略决策（同意/不同意），其中同意和不同意的定义如下：

- 提示（*Nud*）：提前切换到BC表示同意，而延迟切换到BC或使用FC表示不同意。
- 呈现于*BC*（*Prs*）：使用BC和FC分别表示同意和不同意。
- 无干预（*No*）：使用FC表示同意，而使用BC表示不同意。

我们重点研究形如 $\{a_t, c_t\} \Rightarrow a_{t+1}$ 的规则，以根据学生对当前决策的遵从情况，为下一次策略决策提取有意义的关联规则。由于c_t 有两个可能值，而a_t, a_{t+1} 各有三个可能值，因此共有18条可能的规则。每条规则的支持度和置信度传统上计算如下：

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

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

其中总计 $= 598$ [46 *DRL* 学生数 * 13次训练决策]。表3列出了按支持度降序排列的DRL策略的前六条规则。从 $Decl_{DRL}$ 学生和$Proc_{DRL}$ 学生中学习到的规则相似，因此将其合并。其余规则由于支持度和置信度值显著较低而被排除。

解读关联规则 表3揭示了深度强化学习策略的独特视角。从本质上看，前三条规则表明，对于那些事先知道如何以及何时使用反向链式推理的条件性学生，该策略采取了

表3：深度强化学习策略的顶级关联规则

Rank	Rule	支持度（%）	置信度（%）
1	{否，不同意} ⇒ 否	23	76
2	{中立，同意} ⇒ 否	12	61
3	{*Prs*, 同意} ⇒ 否	10	58
4	{不，赞同} ⇒ *Nud*	10	60
5	{*Nud*，不同意} ⇒ *Prs*	7	69
6	{*Prs*，不同意} ⇒ *Nud*	5	66

通过避免在此类情境下进行干预。最后**两**条规则反映出，一旦学生不同意干预措施的效用，策略就会更换干预措施。这一发现表明，学生的元认知知识会在训练过程中发生变化，并证实了我们先前的结果：元认知干预对陈述性学生和程序性学生产生了不同的影响（Abdelshiheed、Hostetter、Yang等，2022年）。**第四**条规则显示，对于此前使用过前向链推理的学生，深度强化学习策略更倾向于建议使用反向链式推理，而不是强制要求使用。

讨论与结论

弥合差距 我们发现，深度强化学习策略使初始元认知知识（陈述性和程序性）较低的学生，表现优于元认知知识最高且未接受任何干预的条件性同伴。换言之，深度强化学习弥合了逻辑辅导系统中的学生（在此系统中提供了干预）与随后概率辅导系统中的学生（在此系统中未提供干预）之间的元认知知识差距。

为未来学习做准备 实验组中陈述性和程序性知识较强的学生在逻辑学习中接受了基于深度强化学习的干预，其表现超过了在逻辑和概率学习中均未接受干预的对照组同伴。这表明，深度强化学习帮助学生为未来学习做准备，因为基于逻辑学习中的干预，他们在概率学习中的表现超过了对照组学生。

自主性与元认知发展 深度强化学习策略适应了学生在陈述性、程序性和条件性元认知发展之间的往复变化。具体而言，关联规则挖掘分析显示，深度强化学习策略会调整干预方式，以适应学生动态变化的元认知知识。因此，即使深度强化学习的干预显著减少，学生也变得更加自主，并能做出有效的策略性决策。

局限性与未来工作 我们的研究至少存在两个需要注意的问题。首先，将学生分为实验组和对照组导致样本量相对较小。其次，概率辅导系统仅支持一种策略，这限制了我们将干预应用于逻辑的能力。未来的工作包括在概率辅导系统中实现前向链推理，比较用于干预的多种深度强化学习算法，并将深度强化学习干预与接受随机干预的更强对照组进行比较。

致谢

本研究得到了美国国家科学基金会以下资助项目的支持：MetaDash：基于实时多通道自我调节学习数据的教师仪表板（1660878）；用于STEM学习环境个性化教学的集成数据驱动技术（1726550）；通过智能辅导系统推广数据驱动技术，以改进个性化STEM教学（2013502）；以及CAREER：改进交互式学习环境中的自适应决策（1651909）。

参考文献

Abdelshiheed, M., Hostetter, J. W., Barnes, T., & Chi, M.（2023年）。利用深度强化学习在智能辅导系统中实施元认知干预。载于《第*24*届国际人工智能教育会议论文集》。Abdelshiheed, M., Hostetter, J. W., Shabrina, P., Barnes, T., & Chi, M.（2022年）。助推的力量：探索在智能辅导系统中开展元认知技能教学的三种干预措施。载于《第*44*届认知科学学会年会论文集》（第541–548页）。Abdelshiheed, M., Hostetter, J. W., Yang, X., Barnes, T., & Chi, M.（2022年）。将反向链式推理与正向链式推理相结合，以促进元认知技能习得与迁移。载于《第*23*届国际人工智能教育会议论文集》（第546–552页）。Abdelshiheed, M., Maniktala, M., Barnes, T., & Chi, M.（2022年）。利用元认知和动机评估能力：时间意识在为未来学习做准备中的作用。载于《智能辅导系统设计建议》（第9卷，第121–131页）。美国陆军作战能力发展司令部士兵中心。Abdelshiheed, M., Maniktala, M., Ju, S., Jain, A., Barnes, T., & Chi, M.（2021年）。为未来学习做好准备的未准备充分的学生。载于《第*43*届认知科学学会年会论文集》（第2547–2553页）。Abdelshiheed, M., Zhou, G., Maniktala, M., Barnes, T., & Chi, M.（2020年）。元认知与动机：时间意识在为未来学习做准备中的作用。载于《第*42*届认知科学学会年会论文集》（第945–951页）。Alam, N., Mostafavi, B., Chi, M., & Barnes, T.（2023年）。探索基于自编码器的特征学习对深度强化学习策略的影响，以确定何时提供主动帮助。载于《第*24*届国际人工智能教育会议论文集》。Anderson, J. R.（1982年）。认知技能的习得。《心理学评论》，*89*（4），369。Anderson, J. R.（2005年）。《认知心理学及其启示》。麦克米伦。Azevedo, R., & Aleven, V.（2013年）。《元认知与学习技术国际手册》（第26卷）。施普林格。Azevedo, R., Cromley, J. G., Winters, F. I., Moos, D. C., & Greene, J. A.（2005年）。自适应人类支架促进青少年借助超媒体进行自我调节学习。*Instructional science*，*33*，381–412。Baker, L.（1994年）。促进元认知发展。《儿童发展与行为进展》，*25*，201–239。Biggs, J., Tang, C., & Kennedy, G.（1999年）。大学中的高质量学习教学。《开放大学，白金汉》。Bloom, B. S.（1956年）。教育目标分类学：教育目标的分类。《手册；认知领域》，*1*。Boden, K., Kuo, E., Nokes-Malach, T., Wallace, T., & Menekse, M.（2018年）。动机在程序性和概念性物理学习中的作用是什么？对自我效能和成就目标的考察。《物理教育研究会议》，60–63。Bransford, J. D., & Schwartz, D. L.（1999年）。重新思考迁移：一个具有多重含义的简单提议。《教育研究评论》，*24*（1），61–100。Brown, A. L.（1987年）。元认知、执行控制、自我调节及其他更加神秘的机制。《元认知、动机与理解》，65–116。Butler, D. L.（1998年）。促进自我调节学习的策略性内容学习方法：三项研究报告。《*Journal of educational psychology*》，*90*（4），682。Case, J., & Gunstone, R.（2002年）。元认知发展作为学习方法的转变：一项深入研究。《高等教育研究》，*27*（4），459–470。Castro-Schez, J. J., Glez-Morcillo, C., Albusac, J., & Vallejo, D.（2021年）。支持主动学习的智能辅导系统：预测分析学习案例研究。《信息科学》，*544*，446–468。Cutrer, W. B., Miller, B., Pusic, M. V., Mejicano, G., Mangrulkar, R. S., Gruppen, L. D., . . . Moore Jr, D. E.（2017年）。促进精通型适应性学习者的发展：指导医学教育中技能习得的概念模型。《学术医学》，*92*（1），70–75。De Backer, L., Van Keer, H., & Valcke, M.（2012年）。探索互惠同伴辅导对高等教育学生元认知知识与调节的潜在影响。《*Instructional science*》，*40*，559–588。Detterman, D. K., & Sternberg, R. J.（1993年）。迁移审判：智力、认知与教学。Ablex Publishing。Dochy, F. J. R. C.（1992年）。先前知识评估作为未来学习的决定因素。开放大学。Flavell, J. H.（1979年）。元认知与认知监控：认知发展研究的新领域。《美国心理学家》，*34*（10），906。Fossati, D.（2009年）。计算机科学辅导系统中程序性知识的自动建模与反馈生成。伊利诺伊大学芝加哥分校。Gao, G., Gao, Q., Yang, X., Pajic, M., & Chi, M.（2022年）。一种受强化学习启发的多变量时间序列分类模式挖掘框架

work for multivariate time series classification. 收录于第*31*届国际人工智能联合会议会刊。Gao, G., Ju, S., Ausin, M. S., & Chi, M.（2023年）。Hope：面向电子学习和医疗保健的以人为中心的离策略评估。收录于第*22*届自主智能体与多智能体系统国际会议会刊。Georgeff, M. P., & Lansky, A. L.（1986年）。程序性知识。《*IEEE*会刊》，*74*（10），1383–1398。Hershkovitz, A., Baker, R., Gowda, S. M., & Corbett, A. T.（2013年）。利用逐时刻学习的定量分析更好地预测未来学习。收录于教育数据挖掘。Hostetter, J. W., Abdelshiheed, M., Barnes, T., & Chi, M.（2023年（a））。利用模糊逻辑使智能辅导系统中由强化学习诱导的教学策略更具可解释性。收录于*2023*年*IEEE*模糊系统国际会议。Hostetter, J. W., Abdelshiheed, M., Barnes, T., & Chi, M.（2023年（b））。自组织神经模糊Q网络：采用离线混合学习的系统化设计。收录于第*22*届自主智能体与多智能体系统国际会议会刊。Jacobs, J. E., & Paris, S. G.（1987年）。儿童对阅读的元认知：定义、测量与教学问题。《*Educationalpsychologist*》，*22*（3-4），255–278。Ju, S., Zhou, G., Abdelshiheed, M., Barnes, T., & Chi, M.（2021年）。在实际领域中评估批判性强化学习框架。收录于人工智能教育国际会议会刊。（第215–227页）。Kiesewetter, J., Ebersbach, R., Tsalas, N., Holzer, M., Schmidmaier, R., & Fischer, M. R.（2016年）。仅有知识不足以解决问题——诊断性知识在临床推理活动中的作用。《*BMC*医学教育》，*16*（1），1–8。Krathwohl, D. R.（2002年）。布卢姆分类法的修订：概述。《*Theoryintopractice*》，*41*（4），212–218。Krueger, P. M., Lieder, F., & Griffiths, T.（2017年）。利用奖励结构和反馈增强元认知强化学习。收录于认知科学。Kua, J., Lim, W.-S., Teo, W., & Edwards, R. A.（2021年）。教育中适应性专长的范围综述。《*MedicalTeacher*》，*43*（3），347–355。Kuhn, D.（2000年）。元认知发展。《心理科学当前方向》，*9*（5），178–181。Larkin, S.（2009年）。幼儿元认知。Routledge。Livingston, J. A.（2003年）。元认知：概述。ERIC。Mbato, C. L.（2019年）。印度尼西亚英语作为外语学习者阅读中的批判性思维：弥合陈述性知识、程序性知识与条件性知识之间的差距。《*Humaniora*》，*31*（1），92。Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., . . . others（2015年）。通过深度强化学习实现人类水平的控制。《自然》，*518*(7540)，529–533。Pintrich, P. R.（2002年）。元认知知识在学习、教学和评估中的作用。*Theoryintopractice*，*41*(4)，219–225。Rittle-Johnson, B., Siegler, R. S., & Alibali, M. W.（2001年）。数学中的概念理解与程序性技能发展：一个迭代过程。*Journalofeduca-tional psychology*，*93*(2)，346。Roberts, M. J., & Erdos, G.（1993年）。策略选择与元认知。*Educationalpsychology*，*13*(3-4)，259–266。Sanz-Ausin, M., Azizsoltani, H., Barnes, T., & Chi, M.（2019年）。利用深度强化学习在智能辅导系统中诱导教学策略。*InternationalEducationalData MiningSociety*。Sanz-Ausin, M., Maniktala, M., Barnes, T., & Chi, M.（2020年）。探究简单解释和能动性对批量深度强化学习诱导的教学策略的影响。载于*AIED*（第472–485页）。Schraw, G.（1998年）。促进一般性的元认知意识。*Instructional science*，*26*(1-2)，113–125。Schraw, G., & Moshman, D.（1995年）。元认知理论。*Educational psychology review*，*7*，351–371。Shabrina, P., Mostafavi, B., Abdelshiheed, M., Chi, M., & Barnes, T.（2023年）。研究逆向策略学习对逻辑辅导系统的影响：通过子目标学习促进问题解决能力的提升。*International Journal of Artificial Intelligence in Education*。Shabrina, P., Mostafavi, B., Chi, M., & Barnes, T.（2023年）。在智能逻辑辅导系统中学习面向子目标的问题解决策略的影响。载于*Proceedings of the 24th international conference on artificial intelligence in educa- tion.*Shabrina, P., Mostafavi, B., Tithi, S. D., Chi, M., & Barnes, T.（2023年）。在智能逻辑辅导系统中，利用数据驱动的块状帕森斯问题学习问题分解与重组。载于*Proceedingsof the16thinternationalconference on educational data mining.*Sutton, R. S., & Barto, A. G.（2018年）。*Reinforcementlearning:An introduction*。MIT Press。Teng, F.（2020年）。元认知知识与调节在调节大学英语学习者写作表现中的作用。*Innovation inLanguage Learning and Teaching*，*14*(5)，436–450。Van Hasselt, H., 等（2016年）。基于双重Q学习的深度强化学习。载于*AAAI*（第30卷）。Yildirim, Z., Ozden, M. Y., & Aksu, M.（2001年）。超媒体学习与传统教学在知识获取和保持方面的比较。*The Journal ofedu-cational research*，*94*(4)，207–214。Zhou, G., Azizsoltani, H., Ausin-Sanz, M., Barnes, T., & Chi, M.（2019年）。用于教学策略诱导的分层强化学习。载于*Proceedings of the 20th inter- national conference on artificial intelligence in education*（第544–556页）。