arXiv:2304.09821v1 [cs.CY] 17 Apr 2023

# 利用深度强化学习在智能辅导系统中实施元认知干预

MarkAbdelshiheed、John Wesley Hostetter、
Tiffany Barnes 和 MinChi {mnabdels, jwhostet,
tmbarnes, mchi}@ncsu.edu

北卡罗来纳州立大学，美国北卡罗来纳州罗利市 27695

**摘要。** 本研究比较了两种在智能辅导系统（ITSs）中提供元认知干预的方法，以及它们对帮助学生为未来学习做好准备所产生的影响。在连续两个学期中，我们开展了两项课堂实验：实验1采用经典人工智能方法，将学生分类 为不同的元认知群体，并根据其所属群体提供静态 干预。在实验2中，我们利用深度强化学习（*DRL*）提供自适应 干预，以考虑学生元认知水平的动态变化。在两项实验中，学生都接受了这些干预，学习如何 以及何时在支持默认前向链式推理策略的逻辑辅导系统中使用反向链式推理（BC）策略。六周后，我们在没有干预的情况下，使用仅支持BC策略的概率辅导系统对学生进行了训练。我们的结果表明，基于自适应深度强化学习的干预缩小了学生之间元认知技能的差距。相比之下，基于静态分类器的干预仅使一部分事先知道如何 使用BC策略的学生受益。此外，我们的DRL智能体帮助实验组学生为未来学习做好了准备，使他们在两个智能辅导系统中的表现均显著优于对照组同伴。

**关键词：** 强化学习 · 人工智能 · 智能辅导系统 · 元认知干预 · 元认知技能

## 1 引言

学习中一个具有挑战性的理想目标，是持续为未来学习做好准备 [9]。我们不断积累的知识证明了为未来学习做准备这一过程的存在，但这一过程难以预测和衡量 [10]。大量研究发现，帮助学生为未来学习做好准备的一个因素是他们的元认知技能 [7, 26]。我们聚焦于与问题解决策略相关的两类元认知技能：策略意识 [5, 20, 22] 和时间意识 [6, 8], ，它们分别关乎使用每种策略的方式 和时机 。

大量研究表明，针对领域知识或问题解决策略的元认知干预能够加快为未来学习做准备的进程 [7, 19, 22] 并提升策略意识和时间意识 [12, 23, 26]。这类

干预包括提示、反馈、引导式助推、示例习题和直接呈现策略。然而，尽管学生在学习过程中往往会逐渐获得并掌握元认知技能 [17]，这些干预仍然是静态的，或被硬编码 到学习环境中。强化学习（RL） [24] 是智能辅导系统（ITS） [14, 16, 27]中提供自适应支持与脚手架的最有效方法之一。强化学习的深度学习扩展——深度强化学习（DRL）——因其能够更好地支持复杂模型，已被广泛用于智能辅导系统中的教学策略诱导 [15, 21] 。据我们所知，尚无研究利用 DRL 提供自适应 元认知干预，并探究其对智能辅导系统中为未来学习做准备的影响。

在本研究中，我们连续开展了两项实验，以比较提供元认知干预的两种方法及其对帮助学生在智能辅导系统（ITS）中为未来学习做好准备的影响。在实验1中，我们使用随机森林分类器（RFC）将学生划分到不同的元认知群体，并根据分类结果提供静态干预；在实验2中，我们采用基于深度强化学习的方法实施自适应干预，以考虑学生元认知水平的动态变化。此前，我们依据策略意识和时间意识，将学生划分为同时具备策略意识和时间意识的学生（兼具 策略意识和时间意识，$StrTime$）、仅具备策略意识 的学生（$StrOnly$），以及其余遵循默认策略的学生（$Default$） [5, 6]。我们发现，只有$StrTime$学生为未来学习做好了准备，因为他们在不同演绎领域的学习表现显著优于同伴。受此发现启发，我们设计了元认知干预，教学生在支持默认前向链式推理策略的逻辑辅导系统中如何 以及何时 使用反向链式推理（BC）策略。六周后，学生在未接受干预的情况下，转而使用仅支持BC策略的概率辅导系统进行训练。结果表明，$Default$ 和$StrOnly$ 学生从我们的DRL策略中获益相同，并且出乎意料地优于 $StrTime$ 学生。然而，基于RFC的方法仅帮助$StrOnly$ 学生追赶上$StrTime$学生。

## 2 背景与相关工作

### 2.1 策略教学的元认知干预

元认知指的是个体对自身认知的觉察，以及控制和调节认知的能力 [13]。策略意识和时间意识是两项元认知技能，分别涉及应如何 以及何时 使用问题解决策略[6, 8]。许多先前研究强调了策略意识在为未来学习做准备中的作用 [5, 22] ，以及时间意识对学业表现和规划技能的影响 [8, 11]。

大量研究表明，元认知干预能够提升策略意识和时间意识 [12, 23, 26]。我们聚焦于两种元认知干预：直接呈现策略 [12, 23] ，以及提示学生使用该策略 [7, 19, 26]。Spörer 等人 [23] 发现，接受综合阅读策略明确指导的学生表现优于接受以下方式教学的同伴：

教师的文本交互、迁移任务和后续测试。他们还理解了如何、何时以及为何使用每种阅读策略。

Zepeda 等人 [26] 证明，元认知干预会影响学习成果、策略掌握以及为未来学习做准备。接受辅导提示和示例习题的实验组在物理测试中的表现显著优于对照组同伴。他们的元认知判断也更准确，并展现出对如何以及何时运用物理策略的掌握。作为为未来学习做准备的一个例子，实验组学生在一项新颖的自主学习任务（“控制变量”）中的表现优于对照组同伴。

尽管此前已有大量关于元认知干预的研究，但这些干预要么不具备适应性，要么未应用于智能辅导系统（ITS），要么没有评估为未来学习做准备的效果。在本研究中，我们开展了两项实验，以提供元认知干预，并探究其对帮助学生为未来学习做准备的影响，研究范围涵盖多个智能辅导系统。具体而言，我们首先尝试采用基于分类器的方法实施静态元认知干预，随后将其与用于自适应元认知干预的基于深度强化学习的方法进行比较。

### 2.2 智能辅导系统中的强化学习

强化学习（RL）是机器学习的一个热门分支，适用于动作会产生数值奖励、但无需知晓真实值的环境 [24]。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灵活，因此已广泛应用于教育领域 [14, 15, 21]。深度强化学习（DRL）是将强化学习算法与神经网络相结合的领域；例如，深度 Q 学习是 Q 学习在神经网络上的扩展，它使用神经网络来近似 Q 函数 [18]。大量研究已在智能辅导系统中运用强化学习和深度强化学习来生成教学策略 [15, 21, 27]。Zhou 等人 [27] 采用分层强化学习来提升智能辅导系统中的学习增益。他们的研究表明，其策略显著优于专家策略和随机条件。

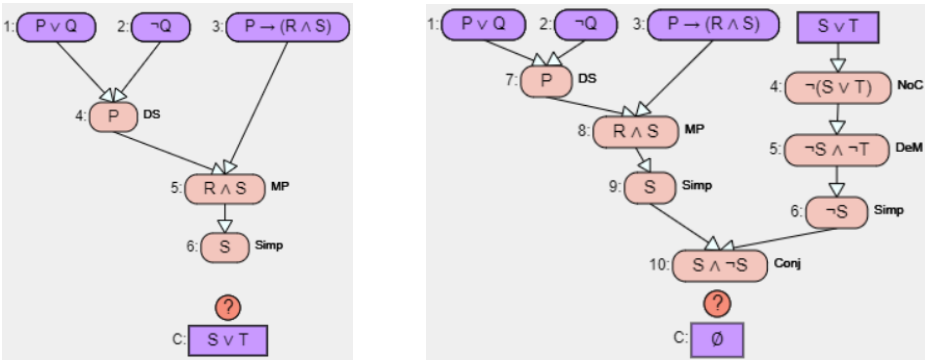
Ju 等人 [15] 提出了一种深度强化学习框架，用于识别关键决策，从而在智能辅导系统中生成关键策略。他们依据两项成功标准评估了这一关键深度强化学习框架：*必要性* 和*充分性*。前者要求在**所有** 关键状态下提供帮助，后者则要求**仅** 在关键状态下提供帮助。结果表明，该框架满足了这两项标准。Sanz-Ausin 等人 [21] 开展了两项连续的课堂研究，将深度强化学习应用于决定由学生还是辅导系统来解答下一道题。他们发现，带有简单解释的深度强化学习策略比专家策略更能显著提升学生的学习表现。

尽管强化学习和 DRL 在 ITS 中得到了广泛应用，但尝试将它们与元认知学习相结合的研究却寥寥无几 [16]。Krueger 等人 [16] 使用强化学习来教授一种元认知技能，即了解应当提前规划多长时间（决定如何决策）。他们的元认知强化学习框架建立在半梯度 SARSA 算法 [24]之上，该算法旨在近似马尔可夫决策过程。他们定义了一个元 Q 函数，称为 $Q_{meta}$，

它以环境的元状态和规划时域动作作为输入。他们在两项规划任务上评估了这一框架；在这些任务中，作者定义了受约束的奖励函数，并且奖励可以提前许多步进行预测，以便制定计划。

在我们的工作中，我们在逻辑辅导系统中诱导并部署了一种元认知干预的DRL策略，并研究了该策略对帮助学生为后续概率辅导系统中的未来学习做好准备所产生的影响。此外，我们没有覆盖任何 DRL 数学定义，例如 Q 函数。相反，我们的 DRL 算法的元认知特性体现在干预的性质上。

## 3 逻辑与概率辅导系统




(a) 前向链接 (b) 逆向链式推理

**图 1.** 逻辑辅导系统问题解决策略

**逻辑辅导系统：** 它通过应用有效的推理规则（如肯定前件）来教授命题逻辑证明，采用标准的前测、训练和后测流程。三个阶段使用相同的界面，但只有训练阶段允许学生寻求并获得帮助。前测有两道题，后测难度更高，共有六道题；前两道题与前测中的题目同构。训练包含五个按顺序排列、难度逐级递增的等级，每个等级包含四道题。每道题都有一个介于 $[0, 100]$ 范围内的分数，该分数根据准确度、用时和解答长度计算。

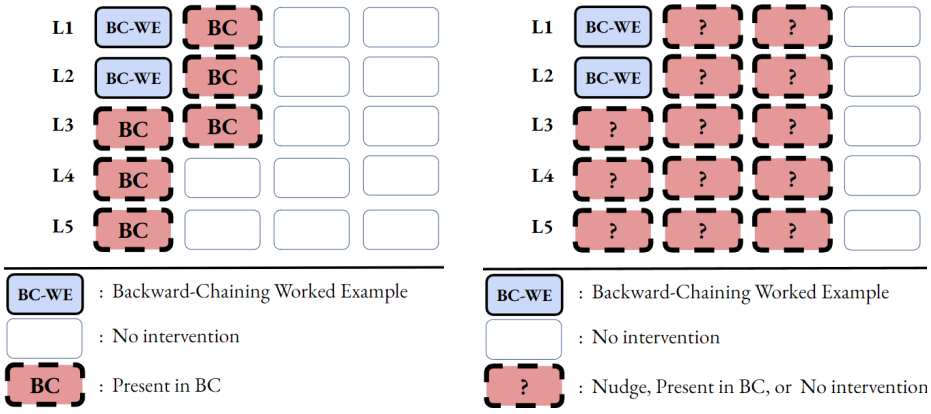
前测和后测分数通过分别对前测和后测的问题得分取平均值计算。学生可以在辅导系统中使用前向链接策略或反向链接（BC）策略解答任何问题。图1a展示了前向链接：必须从顶部给出的条件推导出底部的结论；图1b则展示了BC：学生需要根据给出的条件和结论的否定推导出矛盾。问题默认采用前向链接呈现，但学生可以点击辅导系统界面中的按钮切换到BC。

概率辅导系统：它通过十条原则（例如补集定理）教授如何解答概率问题。该辅导系统由教材、前测、训练和后测组成。与逻辑辅导系统一样，训练是学生获得提示和请求提示的唯一环节，且后测比前测更难。教材介绍该领域的原则；训练包含12道题，每道题都只能通过BC解答，因为学生需要通过列方程并求解，逐步将目标归约为给出的条件。

在前测和后测中，学生分别解答14道和20道开放式题目，其中每道前测题目都有一道与之同构的后测题目。由经验丰富的评分者采用部分得分评分标准进行双盲评分，评分仅 依据准确性，且准确性必须在 $[0, 100]$ 范围内。前测和后测 分数分别为各自部分的平均分。

## 4 方法

由于学生只能在逻辑辅导系统上选择切换问题解决策略，我们的干预在逻辑训练部分提供 [1–3, 5]。研究表明，*StrTime* 学生经常按照预期行为，在逻辑辅导系统上较早切换到*BC*；他们的*StrOnly* 同伴则较晚切换，而*Default*学生不会切换，而是坚持使用默认策略 [4, 6]。此外，我们发现，为*Default* 和*StrOnly* 学生提供建议切换到 BC 的元认知干预（称为助推），或直接以 BC 呈现问题（称为直接呈现），可以帮助他们赶上*StrTime*同伴 [2, 3, 5]。因此，我们开展了两项实验，研究向*Default*和*StrOnly* 学生呈现此类元认知干预的不同方式。




(a) 实验1：RFC–静态　　(b) 实验2：DRL–自适应

**图2.** 在修改后的逻辑辅导系统中进行训练

### 4.1 实验1：RFC—静态

我们使用了一种 RFC，它能够根据逻辑辅导系统中的初始能力，提前预测学生所属的元认知组——*Default*、*StrOnly* 或*StrTime*。此前研究表明，该 RFC 的预测准确率为 96% [5]。

在早期预测之后，我们使用原始逻辑辅导系统对$StrTime$ 学生进行训练，所有题目均默认以$StrTime$ 前向链接方式呈现；而*Default*和*StrOnly* 学生则使用修改后的辅导系统，如图2a所示。具体而言，我们提供了两道关于BC的示例题（WE），辅导系统会逐步展示解题过程；此外，默认以直接呈现的方式提供六道BC题目。我们期望这些示例题和六道题目能教会学生如何以及何时使用BC。请注意，我们根据数据中切换到BC的历史切换行为 [6]，选取了图2a中着色的题目。本实验提供了一种静态元认知干预——直接呈现。由于此前对$Default$ 和$StrOnly$ 学生的实验取得了成功，因此我们选择直接呈现而非助推 [3]。

### 4.2 实验2：DRL—自适应

我们利用DRL提供自适应元认知干预——提示、直接呈现或无干预——而不受RFC对元认知组的预测结果影响。我们使用图2b所示的修改后辅导系统对实验组学生进行训练。图3展示了提示的一个示例：系统会在从先前学生切换行为的概率分布中抽样得到的若干秒后显示提示 [6]。由于我们的干预包含无干预这一选项，因此我们尽可能在更多题目中实施干预。实验1中的示例题予以保留，因为它们对于教会学生如何使用BC至关重要。我们没有在每个级别的最后一道训练题中实施干预，因为该题用于评估学生在该级别上的进步。

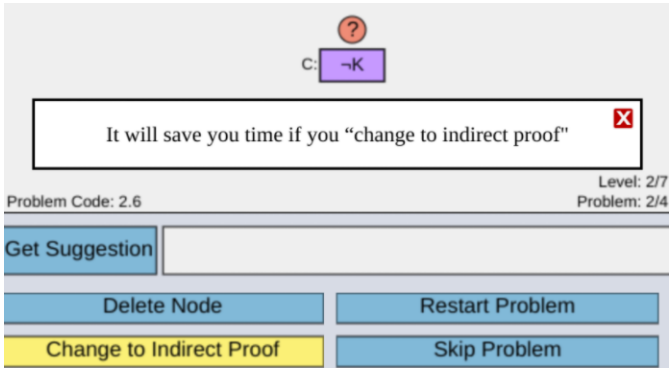



**图3.** 策略切换提示

**训练语料库与策略归纳：**为训练我们的 DRL智能体，我们使用了此前四项研究收集的数据，这些研究共涉及 867 名学生 $[2, 3, 5, 6]$ 并

进行了 $80-20$ 训练测试划分。该数据集包含每位学生针对一道逻辑训练问题的一条记录，以（**状态**、**动作**、**奖励**）表示。状态 是一个由 152 个特征组成的特征向量，用于捕捉时间、准确率和基于提示的行为。动作 可以是提示、直接呈现或无干预。奖励 是学生在逻辑辅导系统中该问题的即时得分，如第 3 节前文所述。

我们的目标是证明 DRL 能与我们的元认知干预配合使用，**而非** 探究 哪种 DRL 算法与我们的干预配合得更好。鉴于强化学习在教育领域已取得广泛成功 [15, 21]，我们选择了 DRL。为了选定 DRL 算法，我们需要避开相对简单的算法，例如深度Q网络（DQN）；该算法会高估动作价值 [18]，并可能导致欠拟合。此外，我们还需要避开复杂的 DRL 算法，例如自编码器和演员-评论家方法，以免 DRL 淡化我们的元认知干预所产生的影响。换言之，若复杂的 DRL 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，而非其提供的元认知干预。因此，我们采用了双重深度Q网络（DDQN）。该算法通过在两个不同的神经网络中**分离**动作选择与动作评估，解决了 DQN 中的高估问题 [25]。由此得到的修正贝尔曼方程如下：

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

其中，$r$ 表示奖励；$\gamma$ 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。具体而言，DDQN 使用**主 ($\boldsymbol{\theta}$)**神经网络为下一状态选择 Q值最高的动作，然后使用**目标($\boldsymbol{\theta}^{-}$)**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，$1e$-3 学习率，$9e$-1 折扣因子，批量大小为 32，主神经网络与目标神经网络之间的同步频率为每 4 步一次，并训练至收敛（$\approx 2000$ 训练轮次）。

## 5 实验设置

这两项实验分别于2022年春季和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开展。参与者将辅导系统作为课程作业，并被告知必须完成才能获得全部学分。在两项实验中，学生均按照标准流程使用逻辑辅导系统：前测、训练（实验1见图2a，实验2见图2b）和后测。六周后，我们对学生进行了概率辅导系统训练，期间没有提供干预。概率训练中的每道题都随机分配给学生，让其独立解答、由辅导系统以示例讲解（WE）的形式呈现，或由双方通过协作式问题解决共同完成。请注意，两种辅导系统中的题目顺序对所有学生都相同。

**实验1（RFC）参与者：**共有121名学生完成了两种辅导系统，并根据RFC被分为47名$Default$、48名$StrOnly$ 和26名$StrTime$。***Default*** 和$StrOnly$ 被随机分配至实验组（***RFC***）和对照组（***Ctrl***）条件。实验组接受了逻辑训练干预（图2a），而对照组和$StrTime$ 未接受任何干预。我们共有24名$Default_{RFC}$、25名$StrOnly_{RFC}$、23名$Default_{Ctrl}$、23名$StrOnly_{Ctrl}$ 和26名$StrTime$ 学生。

**实验2（DRL）参与者：**共有112名学生完成了两种辅导系统的学习，并被随机分配至实验组（***DRL***）和对照组（***Ctrl***）条件。实验组在逻辑训练中接受干预（图2b），而对照组未接受任何干预。为考察DRL是否有助于入学时元认知技能水平不同的学生，我们使用***RFC***确保各条件和元认知群体之间分布均匀，以便进行比较。我们发现，对于$StrTime_{DRL}$ 学生，我们的DRL策略在94%的情况下未提供干预。因此，我们将$StrTime_{DRL}$ 和$StrTime_{Ctrl}$ 合并为$StrTime$。我们有22名$Default_{DRL}$、24名$StrOnly_{DRL}$、22名$Default_{Ctrl}$、22名$StrOnly_{Ctrl}$ 和22名$StrTime$ 学生。

## 6 结果

### 6.1 实验1：RFC–静态

**表1.** 实验1（RFC–Static）结果

| | 实验组 (RFC) | | *Control* | | |
|---|---|---|---|---|---|
| | ***Default*** $_{RFC}$ $(N = 24)$ | ***StrOnly*** $_{RFC}$ $(N = 25)$ | ***Default*** $_{Ctrl}$ $(N = 23)$ | ***StrOnly*** $_{Ctrl}$ $(N = 23)$ | ***StrTime*** $(N = 26)$ |
| 逻辑辅导系统 | | | | | |
| *Pre* | 57。8 (20) | 56.5 (17) | 55.9 (20) | 55.7 (21) | 57.4 (20) |
| *Iso. Post* | 76.5 (14) | **83。9** (9) * | 72.3 (16) | 73 (15) | 81.8 (11)* |
| *Iso. NLG* | 0.18 (.15) | **0.32** (.**11**)* | 0.14 (.29) | 0.16 (.32) | 0.31 (.17)* |
| *Post* | 73.6 (13) | **80.5** (**9**)* | 69.1 (13) | 71。6 (11) | 79.8 (10)* |
| *NLG* | 0。16 (。12) | **0。31** (。**12**)* | 0.12 (.31) | 0.13 (.3) | 0.28 (.16)* |
| Probability Tutor | | | | | |
| *Pre* | 75.5 (16) | 74。2 (15) | 73.8 (14) | 74.8 (16) | 76.1 (16) |
| *Iso. Post* | 75.2 (17) | **93。6** (5) * | 70。9 (14) | 72.6 (15) | 91.2 (7)* |
| *Iso. NLG* | 0.04 (.34) | **0.35** (.**16**)* | -0.03 (.29) | -0.02 (.31) | 0。28 (。18) * |
| *Post* | 74。6 (19) | **92.7** (**7**)* | 69。5 (16) | 70。9 (17) | 90.3 (8)* |
| *NLG* | 0.02 (.37) | **0.32** (.**18**)* | -0.07 (.32) | -0.04 (.36) | 0。25 (。17) * |

同一行中，粗体表示最高值，星号表示与无星号值相比具有统计显著性。

表1比较了实验1中各组的表现。我们展示了前测和后测成绩、同构题成绩以及以标准化学习增益（自然语言生成） [6, 14] 表示的学习结果

为（$NLG = \frac{Post-Pre}{\sqrt{100-Pre}}$），其中100为测试最高分。我们分别将前测、后测和自然语言生成分数称为前测、后测 和自然语言生成。RFC在对未接受任何干预的学生进行分类时准确率为97%——$Default_{Ctrl}$、$StrOnly_{Ctrl}$ 和 $StrTime$。对于每种辅导系统，以组别为因素进行的单因素方差分析发现，逻辑方面的前测得分差异不显著：$F(4,\ 116) = 0.21,\ p = .93$；概率方面的前测得分差异也不显著：$F(4,\ 116) = 0.38,\ p = .82$。

为衡量学生在同构问题上的进步，我们以 {前测, $Iso.\ Post$} 为因素，进行了重复测量方差分析（每种辅导系统分别针对每个组进行一次）。结果发现，在两种辅导系统中，$StrOnly_{RFC}$ 和 $StrTime$ 均有显著学习进步，$p < .0001$；而 $Default_{RFC}$ 和对照组在任一种辅导系统中，$Iso.\ Post$ 的表现均未显著高于前测。这些发现验证了 RFC 的准确性：尽管两组均未接受任何干预，$StrTime$ 仍在两种辅导系统中显著进步，而对照组则没有。

在两种辅导系统中，以前测为协变量、组别为因素进行的单因素协方差分析发现，组别对后测有显著影响：$F(4,\ 115) = 7.4,\ p < .0001,\ \eta^2 = 0.56$（逻辑）；$F(4,\ 115) = 8.2,\ p < .0001,\ \eta^2 = 0.63$（概率）。采用邦费罗尼校正进行的后续成对比较（$\alpha = .05/10$）表明，在两种辅导系统中，$StrOnly_{RFC}$、$StrTime > Default_{RFC}$、$Default_{Ctrl}$、$StrOnly_{Ctrl}$。使用方差分析考察自然语言生成 时，也发现了类似的模式。总而言之，$StrOnly_{RFC}$ 与$StrTime$之间虽无显著差异，但这两组各自的表现均显著优于其余三组。

## 6.2 实验 2：DRL−自适应

**表2.** 实验2（DRL−自适应）结果

| | 实验组 (DRL) | | *Control* | | |
|---|---|---|---|---|---|
| | $Default_{DRL}$ ($N = 22$) | $StrOnly_{DRL}$ ($N = 24$) | $Default_{Ctrl}$ ($N = 22$) | $StrOnly_{Ctrl}$ ($N = 22$) | $StrTime$ ($N = 22$) |
| Logic Tutor | | | | | |
| $Pre$ | 55.6 (21) | 56.1 (21) | 55。2 (19) | 56。4 (23) | 58.2 (19) |
| $Iso.\ Post$ | 91。9 (5)* | **92.1 (4)*** | 72.7 (18) | 74.1 (17) | 83.4 (12)* |
| *Iso. NLG* | 0.46 (.12)* | **0。48 (。09)*** | 0.18 (.3) | 0.14 (.27) | 0.35 (.11)* |
| $Post$ | **87.7 (5)*** | 87.6 (5)* | 70 (15) | 69。7 (16) | 80.2 (11)* |
| $NLG$ | [illegible] | 0.44 (.14)* | 0。16 (。33) | 0.1 (.31) | 0。31 (。15)* |
| 概率辅导系统 | | | | | |
| $Pre$ | 76.9 (15) | 74.6 (16) | 75.2 (15) | 76.7 (13) | 78.6 (14) |
| *Iso. Post* | 94。5 (5)* | **96.1 (3)*** | 73.9 (10) | 71.4 (14) | 89。1 (7)* |
| *Iso. NLG* | 0。36 (。11)* | **0.43 (.13)*** | -0.02 (.18) | -0.06 (.21) | 0.24 (.15)* |
| $Post$ | 94。1 (6)* | **95.6 (4)*** | 71.8 (11) | 68.6 (17) | 87.7 (8)* |
| $NLG$ | 0.34 (.13)* | **0.39 (.16)*** | -0。07 (。24) | -0。1 (。25) | 0.22 (.19)* |

在同一行中，最高值以粗体标示，星号表示与无星号的数值相比具有统计显著性。

表 2 比较了实验 2 中各组的表现。RFC 在正确识别未接受干预的学生方面的准确率为 98%——$Default_{Ctrl}$,

$StrOnly_{Ctrl}$ 和 $StrTime$。以组别为因素进行单因素方差分析后发现，各组在 前测 上的差异不显著：$F(4,\ 107) = 0.18,\ p = .95$（逻辑），以及 $F(4,\ 107) = 0.45,\ p = .77$（概率）。我们针对每个辅导系统中的每个组，以 {前测，$Iso.\ Post$} 为因素进行了重复测量方差分析。结果发现，$Default_{DRL}$、$StrOnly_{DRL}$ 和 $StrTime$ 在两个辅导系统中的学习效果均显著（$p < .0001$），而 $Default_{Ctrl}$ 和 $StrOnly_{Ctrl}$ 则并非如此。

以前测为协变量、组别为因素进行的单因素协方差分析发现，两种辅导系统的后测均存在显著效应：逻辑方面 $F(4, 106) = 9.3,\ p < .0001,\ \eta^2 = 0.61$；概率方面 $F(4, 106) = 10.6,\ p < .0001,\ \eta^2 = 0.74$。随后进行的经邦费罗尼校正的分析（$\alpha = .05/10$）显示，在两种辅导系统中，$Default_{DRL}$、$StrOnly_{DRL}$ > StrTime > $Default_{Ctrl}$、$StrOnly_{Ctrl}$。例如，在逻辑方面，$Default_{DRL}$ 的后测显著高于 StrTime（$t(42) = 3.9,\ p < .001,\ d = 0.88$）；在概率方面也显著更高（$t(42) = 3.7,\ p < .001,\ d = 0.91$）。此外，$Default_{Ctrl}$（逻辑方面 $t(42) = 6.6,\ p < .0001,\ d = 1.6$；概率方面 $t(42) = 7.1,\ p < .0001,\ d = 2.5$）。使用自然语言生成的方差分析也观察到了类似模式。简言之，两个 DRL 组从我们的策略中获益相同，显著优于其对照组同伴，并且出人意料地超过了 StrTime 学生。

### 6.3 事后分析

为了比较两项实验的结果，我们针对表1和表2中每个组的每项指标进行了 Shapiro-Wilk 正态性检验。结果未显示非正态性的证据（$p > .05$）。因此，我们对表1和表2中对应的每一对组进行了独立样本t检验。具体而言，两张表中的 $StrTime$ 组之间没有显著差异[1]。同样，$Default_{Ctrl}$ 组之间以及其 $StrOnly_{Ctrl}$ 同类组之间也未观察到此类差异。

主要目标是比较我们的基于静态 *RFC* 的干预与基于自适应 *DRL* 的干预。首先，我们比较了每项实验中干预的分布。RFC 实验组学生接受静态干预，因此$Default_{RFC}$ 与$StrOnly_{RFC}$的分布相同。对于 DRL 学生，$Default_{DRL}$ 组接受了94（33%）次提示、65（23%）次直接呈现和127（44%）次无干预；而 $StrOnly_{DRL}$ 组接受了82（26%）次提示、74（24%）次直接呈现和156（50%）次无干预。卡方检验显示，实验组中的 DRL 组之间干预分布无显著差异：$\chi^2(2,\ N = 598) = 3.2,\ p = .2$。

其次，我们比较了两种辅导系统上的学习表现。在逻辑辅导系统中，以前测 为协变量，以条件{$RFC$、$DRL$}和元认知组{$Default$、$StrOnly$}为因素进行双因素协方差分析，结果发现对后测的交互效应显著：$F(1,\ 90) = 37.9,\ p < .0001,\ \eta^2 = 0.78$。条件的主效应也显著：$F(1,\ 90) = 28.4,\ p < .0001,\ \eta^2 = 0.59$，其中$DRL$组的表现显著优于其$RFC$ 同伴。随后进行的经邦费罗尼校正的分析（$\alpha = .05/6$）证实，$Default_{DRL}$，

[1]这适用于所有指标，例如前测、后测和自然语言生成。

$StrOnly_{DRL}$ >$Default_{RFC}$、$StrOnly_{RFC}$。例如，$StrOnly_{DRL}$ 的后测 显著高于$StrOnly_{RFC}$：$t(47) = 3.6,\ p < .001,\ d = 0.98$。对自然语言生成使用方差分析也得到了类似结果。

在概率方面，以前测 为协变量、采用相同两个因素进行的双因素协方差分析发现，后测存在显著的交互效应：$F(1, 90) = 29.1, p <.0001, \eta^2 = 0.46$。随后进行的邦费罗尼校正两两比较（$\alpha = .05/6$）显示，$Default_{DRL}$、$StrOnly_{DRL}$、$StrOnly_{RFC} > Default_{RFC}$。例如，$StrOnly_{RFC}$的后测 显著高于$Default_{RFC}$：$t(47) = 6.4, p < .0001, d = 1.3$。使用方差分析考察自然语言生成 时也观察到类似结果。

## 7 讨论与结论

我们表明，利用 DRL 提供自适应元认知干预，缩小了元认知群体之间的差距，并使其表现超过对照组同伴。令人惊讶的是，我们的 DRL策略使实验组学生的表现显著优于$StrTime$ 学生。另一方面，采用基于 RFC 的方法提供静态干预，仅使部分学生受益——$StrOnly$组学生迎头赶上$StrTime$ 同伴。

此外，显然，DRL 使实验组学生为未来学习做好了准备： [9] 他们在两个辅导系统中的表现都优于对照组同伴。换言之，实验组学生凭借在逻辑方面接受的干预，在概率方面的表现优于同伴。这一发现表明，他们在逻辑方面习得了反向链式推理技能，并将其迁移到了概率方面，而在概率方面他们未接受任何干预。

**局限性与未来工作** 我们的工作至少存在两点局限。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，因此我们的干预能力仅限于逻辑辅导系统。未来的工作包括在同一项研究中比较基于 RFC 的方法与基于深度强化学习的方法。此外，我们计划让概率辅导系统支持正向链推理，就像逻辑辅导系统一样。

**致谢：** 本研究得到了 NSF 资助项目：1651909、1660878、1726550 和 2013502 的支持。

## References

1. Abdelshiheed, M., Hostetter, J.W., Barnes, T., Chi, M.：运用深度强化学习弥合陈述性、程序性和条件性元认知知识之间的差距。载于：认知科学大会（2023） 2. Abdelshiheed, M., Hostetter, J.W., Shabrina, P., Barnes, T., Chi, M.：提示的力量：探索智能辅导系统中元认知技能教学的三种干预方式。载于：认知科学大会。第 541–548 页（2022） 3. Abdelshiheed, M., Hostetter, J.W., Yang, X., Barnes, T., Chi, M.：结合逆向链式推理与正向链式推理，以习得和迁移元认知技能。载于：人工智能教育国际会议。第 546–552 页。施普林格国际出版（2022）

4. Abdelshiheed, M., Maniktala, M., Barnes, T., Chi, M.：运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（2022） 5. Abdelshiheed, M., Maniktala, M., Ju, S., Jain, A., Barnes, T., Chi, M.：帮助准备不足的学生为未来学习做好准备。载于：认知科学大会。第 2547–2553 页（2021） 6. Abdelshiheed, M., Zhou, G., Maniktala, M., Barnes, T., Chi, M.：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：认知科学大会。第 945–951 页（2020） 7. Belenky, D.M., 等：探讨操作教具与元认知对参与度、学习和迁移的作用。《问题解决期刊》**2**(2)，6（2009） 8. de Boer, H., 等：元认知策略教学对学生学业表现的长期影响：一项荟萃分析。《教育心理学评论》**24** （2018） 9. Bransford, J.D., Schwartz, D.L.：重新思考迁移：一个简单的提议及其多重影响。《教育研究评论》**24**(1)，61–100（1999） 10. Detterman, D.K., Sternberg, R.J.：迁移接受检验：智力、认知与教学。Ablex 出版社（1993） 11. Fazio, L.K., 等：分数大小比较中的策略使用与策略选择。《实验心理学期刊：学习、记忆与认知》**42**(1)，1（2016） 12. Fellman, D., 等：策略使用在工作记忆训练结果中的作用。《记忆与语言期刊》**110**，104064（2020） 13. Flavell, J.H.：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（1979） 14. Hostetter, J.W., Abdelshiheed, M., Barnes, T., Chi, M.：一种自组织神经模糊 Q 网络：基于离线混合学习的系统设计。载于：自主智能体与多智能体系统国际会议（2023） 15. Ju, S., Zhou, G., Abdelshiheed, M., Barnes, T., Chi, M.：在实际领域评估关键的强化学习框架。载于：人工智能教育国际会议。第 215–227 页。施普林格（2021） 16. Krueger, P.M., Lieder, F., Griffiths, T.：利用奖励结构和反馈增强元认知强化学习。载于：认知科学大会（2017） 17. Kuhn, D.：元认知发展。《心理科学当前方向》**9**(5)，178–181（2000） 18. Mnih, V., 等：通过深度强化学习实现人类水平的控制。《自然》**518**(7540)，529–533（2015） 19. Richey, J.E., Zepeda, C.D., Nokes-Malach, T.J.：提示式和自我报告式类比比较与自我解释的迁移效应。载于：认知科学大会。第 37 卷（2015） 20. Roberts, M.J., Erdos, G.：策略选择与元认知。《教育心理学》**13**，259–266（1993） 21. Sanz Ausin, M., Maniktala, M., Barnes, T., Chi, M.：探讨简单解释和能动性对批量深度强化学习所诱导教学策略的影响。载于：人工智能教育国际会议。第 472–485 页。施普林格（2020） 22. Shabrina, P., Mostafavi, B., Abdelshiheed, M., Chi, M., Barnes, T.：探究逻辑辅导系统中逆向策略学习的影响：通过子目标学习促进问题解决能力提升。《国际人工智能教育期刊》（2023） 23. Spörer, N., 等：提升学生的阅读理解能力：策略教学和互惠教学的效果。《学习与教学》**19**(3)，272–286（2009） 24. Sutton, R.S., 等：强化学习：导论。麻省理工学院出版社（2018） 25. Van Hasselt, H., Guez, A., Silver, D.：结合双 Q 学习的深度强化学习。载于：人工智能促进协会。第 30 卷（2016） 26. Zepeda, C.D., 等：元认知直接教学有益于青少年科学学习、迁移和动机。《教育心理学期刊》**107**(4)，954（2015） 27. Zhou, G., Azizsoltani, H., Ausin, M.S., Barnes, T., Chi, M.：用于教学策略归纳的分层强化学习。载于：人工智能教育国际会议。第 544–556 页（2019）