

多篇章学习中提取练习策略对高阶技能的影响^{*}

周爱保 杨天成 程 晨 马小凤 赵 静

(西北师范大学心理学院学生学习指导中心, 兰州 730070)

摘 要 采用多篇章学习方式, 结合测试效应研究的经典范式, 设计两个实验探究提取练习策略对高阶技能的影响。结果表明: (1)连续学习 3 篇文章后, 在最终测试阶段学习策略主效应显著, 提取练习组高阶技能的成绩更优: 提取组在每篇文章学习间隔中插入即时测试, 无论是概念型知识还是高阶技能, 最终测试中的成绩都显著高于重学组, 但其反应时却低于重学组。在提取练习组中, 概念型知识与高阶技能成绩均有提升, 但概念型知识比高阶技能提升幅度明显; (2)将篇章数扩展至 5 篇后, 最终测试中, 提取练习组的成绩仍显著高于重学组。对高阶技能 4 个维度作以分析后发现, 两组在应用与分析维度上差异显著, 但在评估与创造维度上, 提取练习组成绩虽有提升但差异不显著。结果证实了提取练习是高效学习的有效策略, 不仅体现在单独内容的记忆保持, 也可扩展到连续学习中高阶技能的发展与促进。研究支持了情景语境解释对提取练习内部机制的阐述, 并进一步验证了前摄抑制干扰理论和上下文内容转换假说的观点。

关键词 提取练习; 多篇章学习; 高阶技能

分类号 B849:G44; B842

1 引言

高效自主的学习模式为现代教育理念提出了新的要求, 如何提高学习效率就成为了教育改革的核心内容。教育教学中多采用记忆测试的方法检验学习者的学习效果, 测试(test)衡量了记忆准确性和速率。与此同时, 测试也可以被当成一种有效手段来提高记忆, 例如在学习过后进行一些问答题测试可以有利于学生记住课本上的内容。大量的研究表明, 相对于简单的重复学习某一材料, 测试更好地促进对信息的记忆保持(McDaniel, Howard, & Einstein, 2009)。也就是说, 在记忆某种材料时, 在相同时间下区别于传统的重复学习模式, 进行一次或多次测试促进了对材料的记忆保持水平。研究者将这种测试体现出的优势称之为“测试效应”(testing effect) (Carrier & Pashler, 1992; Karpicke & Roediger, 2008), 在测试中回忆所学内容的记忆检索过程则定义为“提取练习”(retrieval practice) (Carpenter, 2009)。提取练习作为一种学习策略, 优

于重复学习的证据是因为提取行为其本身有助于强化记忆痕迹(Carrier & Pashler, 1992)。

记忆保持和学习迁移是教学过程的两个重要环节。提取练习对长时记忆保持的促进是面向意义学习有效手段, 但同样可以帮助学习者应用所学知识, 促进学习的迁移(Karpicke & Blunt, 2011)。Butler (2010)证明提取练习在远迁移上有积极的效应。周爱保、马小凤、李晶和崔丹(2013)也在迁移的维度上得出相似结论, 表明提取练习比建构概念图更有利于记忆保持和迁移。Anderson 等人(2001)修订了 Bloom 的教育目标分类(如图 1 所示)。新修订的分类学包含 6 个认知过程维度: 记忆(remember)、理解(understand)、应用(apply)、分析(analyze)、评价(evaluate)和创造(create), 其中, 第一个维度主要涉及保持; 其余 5 个维度主要涉及迁移。在此之后 Anderson 等人(2001)将后 5 个维度统称为高阶技能, 指的是发生在较高认知水平层次上的心智活动或认知能力。

虽然有大量研究表明提取练习是促进事实性

收稿日期: 2014-09-24

^{*} 西北师范大学青年教师科研能力提升一般项目(自然科学类, 项目编号: NWNLUKQN-13-25)资助。

通讯作者: 周爱保, E-mail: zhoulabpapers@163.com

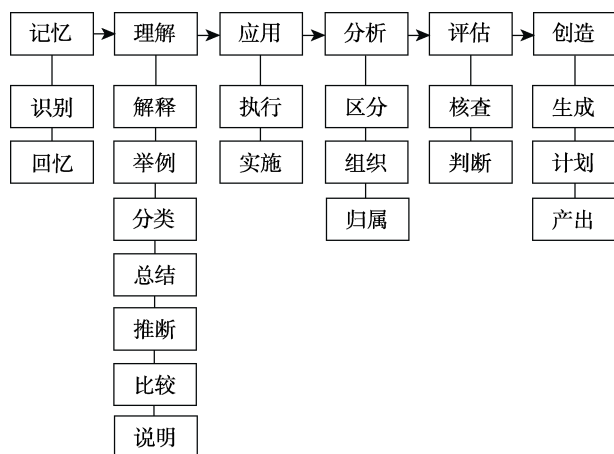


图1 Anderson 等人(2001)对 Bloom 学习分类的修订版

知识学习的有效策略(Roediger, Agarwal, Kang, & Marsh, 2010), 但很少有研究关注提取练习对高阶技能的影响。Jacoby, Wahlheim 和 Coane (2010)发现提取练习对学生分类技能(即学习分类中的理解技能)有积极影响, 结果表明提取练习有助于分类技能的培养。有关高阶技能的其他维度, Agarwal (2011)进行了一系列研究, 并首次探讨了应用、评估、分析和创造这四类主要的高阶技能。Agarwal 认为高阶技能的发展是教育最理想的结果, 证明了提取练习对于高阶技能的发展同样是非常有效的策略。研究表明, 对比重学组和无提取组, 提取组的成绩在事实性知识和高阶技能上均有显著提升。可见, Agarwal 提出了提取练习研究的新思路, 即提取练习对高阶技能各个维度的促进作用, 这对教学实践具有重要的启发意义。但是, 美中不足的是, 与实践教学不同, 在其研究中仅仅使用一篇文章作为实验材料, 而现实情况是学习者通常需要连续学习多篇文章或多列词汇, 从而便于形成完整的知识体系, 也就是多篇章学习(Multiple-list Learning)。再度回顾 Agarwal 的研究, 有违于真实教学情境中的多篇章学习方式, 使其结果的生态效度较低。

那么多篇文章的连续学习中, 提取练习能否促进高阶技能的连续发展就成为一个疑问。近年来, 有研究表明当被试连续学习多列词汇时, 在每列词汇学习之后, 对所学内容的成功提取不仅能促进对该词列的长时记忆, 也能够促进后续词列的记忆(McDaniel, Roediger, & McDermott, 2007)。提取促进后续学习这种更为深入的研究仰仗早期的独立性研究(Leeming, 2002)。Darley 和 Murdock (1971)最先发现并证明, 对新材料的学习很可能会受到之前所学材料是否进行过提取的影响, 其他研究者也得出同样的结论(Sahakyan, Delaney, & Kelley,

2004)。Tulving 和 Watkins (1974)设计两个实验验证前摄抑制效应。结果表明, 若对已学过的 A-B 词对没有进行提取, 就立即呈现 A-C 词对的话, 则学习 A-C 词对会受到严重的前摄抑制干扰。Szpunar, McDermott 和 Roediger (2008)进一步探讨了学习 5 列词表时, 提取对学习的促进作用。结果表明, 提取组记忆成绩表现出明显优势。连续学习语词材料或篇章内容, 是形成前摄干扰的充分条件, 也是前摄抑制的基本实验设计条件(Postman & Keppel, 1977)。而新学习内容成功提取与否和之前所学的数量呈负相关(Underwood, 1957)。除此之外, 根据提取促进上下文内容转换假说(Context-Change Hypothesis)的观点, 提取练习使学习过程中的上下文内容之间顺利转换, 提高了所学内容之间的区分度(Brewer, Marsh, Meeks, Clark-Foos, & Hicks, 2010)。换言之, 在学习材料的过程中, 记忆系统与所学每个项目相关, 构建被试的整体编码系统。提取练习的介入可以促进不同编码流畅转换, 使得每篇文章生成特定的内容线索, 提取一方面增强了篇章间的区分度, 另一方面降低了篇章间的相互干扰(Criss & Shiffrin, 2004)。此外, 最新的脑电研究对此也给予了部分佐证, 提取使得上下文内容顺利转换的同时, 观察到被试 α 脑电波显著降低(Pastötter, Schicker, Niedernhuber, & Bäuml, 2011)。据此推测, 可能是由于编码过程的复位, 认知资源的重置与高效利用, 使后续的学习与之前学习享有同等的编码资源, 有效降低了记忆负荷。

综上所述, 提取练习可以促进多篇章材料的学习。在此基础上, 结合 Agarwal (2011)的研究, 我们假设: 提取练习在多篇文章的连续学习中不仅可以促进概念型知识的学习, 还可以促进高阶技能的连续发展。为此, 研究设计两个实验, 采用多篇章学习形式, 在每篇文章学习间隔中插入与文章主题相关的提取练习, 从而检验提取练习是否可以成为高效学习的有力工具, 能否有效提高学习成绩促进后续文章的学习, 并进一步探讨其应用到高阶技能学习中的有效性。

2 实验 1

2.1 被试

通过校园 BBS 招募西北师范大学本科生 60 名(男女生各 30 名), 分为两组, 每组 30 名被试, 平均年龄 20 岁, 母语均为汉语。被试报告均无神经性疾病史, 视力和矫正视力正常。实验开始前所有被试

签署知情同意书,结束后会有一份小礼物赠予被试。

2.2 材料

使用 3 篇文章,来源于 McGraw-Hill Contemporary Learning Series (McDaniel et al., 2009)。每篇文章约 850 个字。每篇文章都包含同一主题的两相反的观点。

每篇文章之后,依据正、反两方作者的观点设置 10 道选择题,其中 2 道为概念型问题,8 道为高阶技能型问题(应用、分析、评估和创造)。在 10 道题中,5 道题目描述“正方”观点,剩余 5 道题目描述“反方”观点。创立问题的依据及示例如下:

(1) 概念型问题:以记忆和理解为基础,不仅仅将文章出现的信息(如姓名、日期、词语、定义等)直接作为答案,而是针对目标段落的核心内容设置问题。这是为了测试被试对文章基本内容的理解,也属于学习分类中较浅层的类别。实验中每篇文章后的第 1、2 题属于此类问题。以文章一第 1 题为例:

哪一项是正方作者提出的解决办法?

- A 取消医疗补助,但保留所有退伍军人的津贴
- B 保留所有福利项目,但是减少每个项目的开支
- C 取消一些福利项目,同时对保留的一部分福利项目增加投入
- D 取消美国目前所有的福利项目

该题正确答案为选项 D。文中正方作者没有明确指出取消所有福利项目,但多次提到“福利项目不应出台”、“福利项目增长开支”等观点,其意图正是选项 D 的表述。

(2) 应用性问题:设置一个与文章内容相关的新场景或问题,被试依据文章中的解决方式回答。每篇文章后第 3、7 题属于此类问题。以文章一第 3 题为例:

如果没有福利项目,哪种形式的社会是正方作者所期待的?

- A 所有人独立自主的社会
- B 没有政府参与的社会
- C 不再需要纳税的社会
- D 所有人都被平等对待的社会

文中提到正方作者对目前福利型社会的种种观点,题目假设出相反情境,即无福利项目实施的社会。解读文章一第五段后发现,正方作者认为“没有这些福利项目,人们可能更懂得自力更生”,这一观点可应用至题目设置的新情境中,答案 A 为正

确选项。

(3) 分析性问题:对两种观点进行分析。题干中陈述一个与主题相关的观点,要求被试判断是“正方”还是“反方”的观点。每篇文章后第 4、8 题属于此类问题。以文章一第 4 题为例:

哪一方作者会支持该论述“政府对社会的帮助是值得尊敬的”?

- A 正方作者
- B 反方作者
- C 都同意
- D 都不同意

该题目的正确答案为 C,分析类题目要求略高于应用类题目,需同时把握正、反两方作者的观点,之后做出判断。题目设计观点未必仅倾向于某一方作者,如本题答案即为双方作者均支持该论述。从原文中可以看到:正方作者“毋庸置疑,政府的出发点是好的”;反方作者“只有在政府行使行政权和立法权的社会中,才能确保人民安居乐业”,统筹兼顾两方观点是解答的切入点。

(4) 评估性问题:选项中列出 4 个陈述句(文章中未出现过),被试需要判断哪一项符合作者的观点。每篇文章后第 5、9 题属于此类问题。以文章一第 9 题为例:

哪一项陈述是对正方作者的正确评估和总结?

- A 福利项目不会起作用,因为它们都花销太大
- B 福利项目是有害的,因为它们使情况变得更糟
- C 福利项目把纳税人的钱浪费在并非真正需要帮助的人身上
- D 福利项目可能有用,但是它们很少满足人们的需要

解答该题的重点在于剖析正方作者的观点,如果单纯理解为正方作者反对福利项目,则容易误选。正确答案为 D,领会文章核心“目前大多数的福利项目都不该出台”,也不能忽视细节“大多数”,“出发点是好的”,作者的反对不仅仅是由于花销大、浪费钱等原因,了解其背后用意“渴望个体独立性”,才能正确评估。

(5) 创造性问题:要求被试在新情境中设计或计划出可能产生的结果。每篇文章后第 6、10 题属于此类问题。以文章一第 10 题为例:

请你预测反方作者会支持哪一种税收和支出结构?

- A 平等税收;平等支出

B 富人高税收, 穷人低税收; 富人多支出, 穷人少支出

C 平等税收; 富人少支出, 穷人多支出

D 富人高税收, 穷人低税收; 富人少支出, 穷人多支出

此题属于高阶技能最高层级类别, 正确答案为 A。需要完整阅读文章一中反方作者的观点, 找出关键的几个点, “政府出台的计划很大范围保证了人民的健康”、“认同自由贸易的利益”、“不同阶层贡献不同”等等, 所谓“创造”, 不等于凭空捏造, 只有理解了文章主旨, 推测出作者所要表达的深层内涵, 才能在新情境、新问题中选出已学内容涉及到的正确做法, 这也是对一个人能力的最高形式评估。

五类问题的设置参考了 Agarwal (2011) 的实验材料, 经由本校英语系老师翻译修订。

2.3 设计

采用 2×2 混合设计。其中, 学习条件为被试间变量, 包括提取练习条件和重学条件; 知识类型为被试内变量, 包括概念和高阶技能。因变量分别为: 第三篇文章学习后的概念和高阶技能的成绩、最终测试概念和高阶技能的成绩以及被试阅读第三篇文章后的答题反应时。

2.4 程序

遵循测试效应研究范式, 分为三大阶段: 学习阶段、分心任务阶段和最终测试阶段。在学习阶段开始之前, 告知被试有关实验的基本信息及实验过程中的注意事项。实验指导语要求所有被试认真学习每篇文章的内容, 明确指出学习之后会安排相应的测试, 以便检验学习效果。所有程序内容使用心理学编程软件 E-prime 2.0 编制。

学习阶段, 被试学习 3 篇文章, 重学组和提取组在每篇文章学习后安排两种不同的实验处理, 现以学习文章一的过程为例, 如图 2 所示。

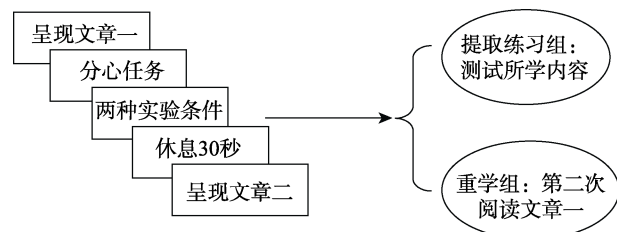


图 2 学习阶段文章 1 的实验流程

两组被试都会学习主题不同, 内容不同, 难度相仿的 3 篇文章, 文章呈现顺序均保持一致。不同之处在于, 重学组对每篇文章进行两次阅读, 时限

分别为 5 min 和 2.5 min。需要注意的是, 重学组第三篇文章的任务与提取组相同, 即阅读文章 5 min 后, 完成针对文章三的 10 道单项选择答题任务。

以此流程两组被试依次学习 3 篇文章, 在学习完第三篇文章后, 完成 5 min 的数学题的分心任务。之后, 开始 30 道单项选择题的最终测试阶段。所有题目的呈现为随机顺序, 每道题目的 4 个选项也是随机排序, 防止被试刻意记忆选项。整个实验大约需要 35 min。

2.5 结果

2.5.1 即时测试实验结果

提取组 3 篇文章的即时测试结果, 描述统计数据矩阵见表 1 所示。

表 1 实验 1 提取组即时测试正确作答的题数

被试	每篇文章作答正确的题目数(道)			
	文章一	文章二	文章三	总体
1	7	5	7	19
2	6	6	6	18
3	5	6	6	17
...	...	...	...	...
30	5	5	6	16
$M(S_d)$	6.75(1.14)	6.36(1.13)	6.96(0.84)	20.07(2.32)

根据描述统计数据, 3 篇文章的正确率分别是 67.50%, 63.57%, 69.64%, 均保持在较高水平(多项选择题基线水平为正确率 50%)。分别比较 3 篇文章, 计算每种题型的正确率, 文章一中概念型题型 85.71%, 高阶技能型题目 62.95%; 文章二中概念型题型 80.36%, 高阶技能型题目 59.36%; 文章三中概念型题型 77.59%, 高阶技能型题目 67.24% (如图 3 所示)。

描述统计结果如表 2 所示。进行 2(学习策略) ×

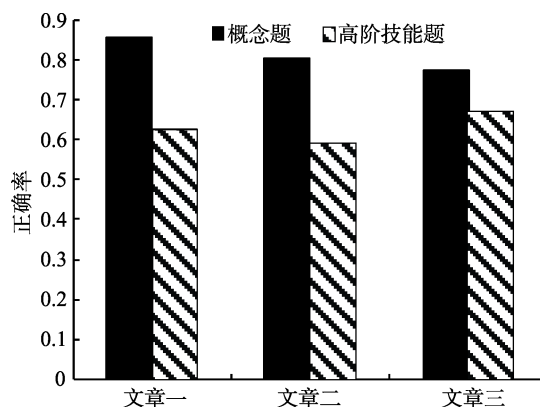


图 3 实验 1 三篇文章中概念题与高阶技能题正确率比较

2(知识类型)的多元方差分析(MANOVA)。结果表明:学习策略的主效应显著, $F(1, 58) = 35.01, p < 0.05, \eta^2 = 0.38$, 知识类型的主效应显著, $F(1, 58) = 17.83, p < 0.05, \eta^2 = 0.24$, 知识类型与学习策略的交互作用不显著, $F(1, 58) = 0.33, p > 0.05, \eta^2 = 0.01$, 如图 4 所示。

表 2 实验 1 提取组与重学组文章三的即时测试数据

被试	文章三即时测试正确题目数(道)	
	提取练习组	重学组
1	7	4
2	6	4
3	6	5
...	...	...
30	6	3
$M(S_d)$	6.96 (0.84)	4.12 (0.97)

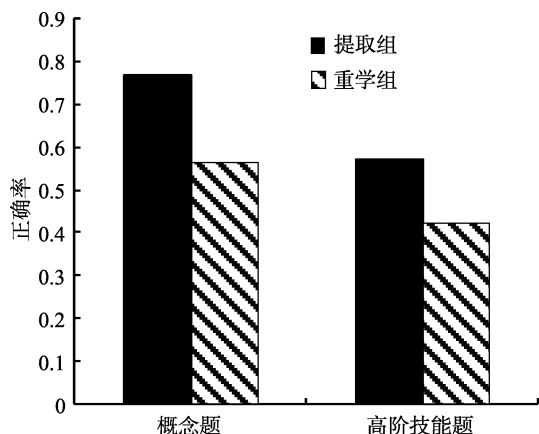


图 4 实验 1 提取组与重学组在知识类型上成绩的比较

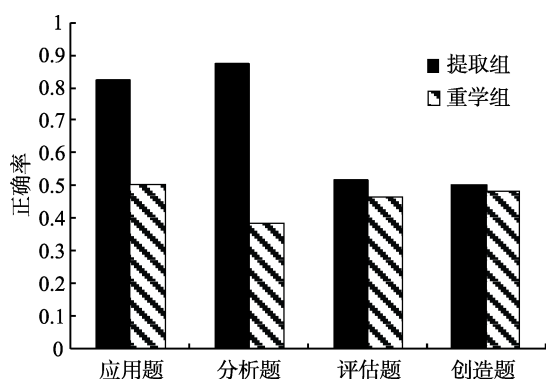


图 5 实验 1 提取组与重学组在高阶技能各维度上成绩的比较

高阶技能的知识类型上, 8 道题目分为 4 类不同维度的高阶技能, 进行 2(学习策略) $\times$ 4(高阶维度)的多元方差分析(MANOVA), 高阶维度与学习策略的交互作用不显著, $F(1, 58) = 2.78, p > 0.05, \eta^2 = 0.05$; 学习策略主效应显著, $F(1, 58) = 146.04, p <$

$0.05, \eta^2 = 0.72$; 高阶维度的主效应显著, $F(3, 174) = 13.25, p < 0.05, \eta^2 = 0.19$, 如图 5 所示。进一步分析提取组和重学组高阶各维度, 应用类问题的成绩(正确率分别为 82.14%, 50.00%)差异显著, $F(1, 58) = 23.32, p < 0.05, \eta^2 = 0.29$; 分析类问题的成绩(正确率分别为 87.50%, 38.00%)差异显著, $F(1, 58) = 44.09, p < 0.05, \eta^2 = 0.43$; 评估类问题的成绩(正确率分别为 51.78%, 46.32%)差异不显著, $F(1, 58) = 1.16, p > 0.05, \eta^2 = 0.02$; 创造类问题的成绩(正确率分别为 50.00%, 48.00%)差异不显著, $F(1, 58) = 2.01, p > 0.05, \eta^2 = 0.03$ 。可见, 不同学习策略仅影响应用和分析问题的成绩, 而评估和创造问题的成绩不因学习策略的改变而有所提高。

2.5.2 最终测试实验结果

提取组和重学组在学习阶段后, 经过 5 min 的分心任务, 都要完成 3 篇文章, 共 30 道选择题的最终测试阶段。描述统计分析如表 3 所示。进行 2(学习策略) $\times$ 2(知识类型)的多元方差分析(MANOVA), 结果表明:学习策略主效应显著, $F(1, 58) = 81.71, p < 0.05, \eta^2 = 0.59$, 知识类型主效应显著, $F(1, 58) = 17.48, p < 0.05, \eta^2 = 0.23$, 知识类型与学习策略交互作用不显著, $F(1, 58) = 1.57, p > 0.05, \eta^2 = 0.03$, 如图 6 所示。进一步分析提取组和重学组高阶各维度, 结果表明, 应用类问题的成绩差异显著, $F(1, 58) = 57.22, p < 0.05, \eta^2 = 0.50$; 分析类问题的成绩差异显著, $F(1, 58) = 52.02, p < 0.05, \eta^2 = 0.47$; 评估类问题的成绩差异不显著, $F(1, 58) = 3.11, p > 0.05, \eta^2 = 0.05$; 创造类问题的成绩差异不显著, $F(1, 58) = 3.74, p > 0.05, \eta^2 = 0.06$, 如图 7 所示。

表 3 实验 1 提取组和重学组最终测试数据

被试	最终测试正确题目数(道)	
	提取练习组	重学组
1	19	13
2	17	14
3	21	16
...	...	...
30	20	9
$M(S_d)$	20.14 (2.05)	11.64 (4.22)

2.5.3 反应时实验结果

在实验中同时记录被试答题的反应时。第三篇文章学习后的即时测试中, 提取组被试的反应时平均数为 14.83 s, 重复学习组被试反应时平均数为 29.07 s, 提取组被试回答问题的速度更快。

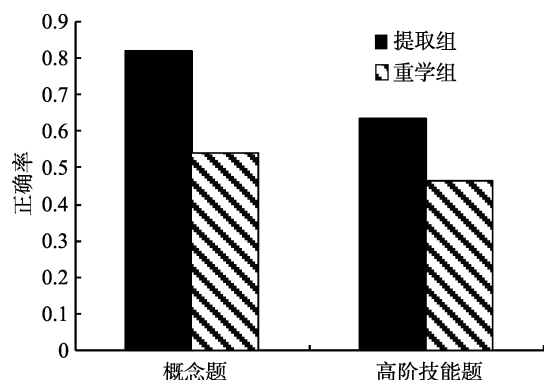


图 6 实验 1 提取组与重学组最终测试中知识类型的成绩比较

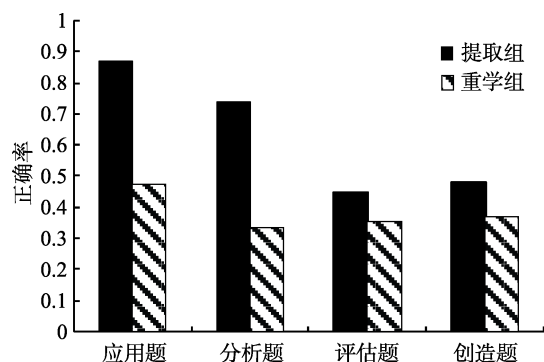


图 7 实验 1 提取组与重学组在高阶技能各维度上成绩的比较

同样,比较各个维度上题目的反应时差异,概念型问题的反应时差异显著, $F(1, 58) = 19.52, p < 0.05, \eta^2 = 0.25$; 高阶技能型问题的反应时差异显著, $F(1, 58) = 49.16, p < 0.05, \eta^2 = 0.46$ 。其中,应用类问题的反应时差异显著, $F(1, 58) = 17.50, p < 0.05, \eta^2 = 0.23$; 分析类问题的反应时差异显著, $F(1, 58) = 42.73, p < 0.05, \eta^2 = 0.42$; 评估类问题的反应时差异显著, $F(1, 58) = 11.09, p < 0.05, \eta^2 = 0.16$; 创造类问题的反应时差异显著, $F(1, 58) = 44.01, p < 0.05, \eta^2 = 0.43$ 。结果表明,提取组在反应时指标上优于重学组。

2.6 讨论

实验 1 中,第三篇文章的即时测试显示出提取组比重学组更高的正确率,再次证明了在多篇章学习中,学习中穿插提取练习能有效促进后续内容的记忆,但在重复学习策略下,没有提高学习成绩和后续内容的记忆(Szpunar et al., 2008)。最终测试中,提取组依旧表现出较高的正确率,即使在一段间隔后,学习者对所学内容仍然能够很好的掌握(Roediger et al., 2010)。

区别于以往研究,将多列词汇变为多篇文章连

续学习,一方面,篇章的学习更符合一般学习情境,更重要的是,篇章的学习对学习者提出更高要求,不再局限于对记忆水平的考察,亦能够验证高阶技能的发展。与 Agarwal (2011)的研究相似,当即时测试的内容(概念或高阶)与最终测试的内容相同时,提取练习对学习促进作用尤为显著。而对于重学组被试,即使给予双倍的学习时间,在正确率、反应时等指标均没有显著提高(Callender & McDaniel, 2009)。

对于作答反应时,研究中仅针对第三篇文章的即时测试记录反应时,因为对于提取组和重学组而言,都是初次接触题目,且两组对等。统计结果证明提取练习组反应速率更快。

3 实验 2

实验 1 已经证明在多篇章学习中,提取练习不仅能提高学习成绩,对高阶技能的提升也十分显著。本研究意图探讨学习连续性的问题,前人关于提取促进后续记忆的研究发现,在第五列词表学习后依旧表现出超过 50%的高回忆率(Pastötter et al., 2011)。那么,当被试学习 5 篇文章时,高阶技能的发展是否依旧持续、稳定。

实验 1 中提取组被试学习阶段每篇文章后所做的 10 道单项选择题,与最终测试的 30 道选择题题目相同,尽管题目的出现顺序是随机的,每道题目的 4 个选项也都随机排列。如前文所述,涉及提取练习相关研究的几十年中,外界存在一种质疑:进行不断的提取是否仅仅强化了学习者记忆选项的能力,并非真正实现对知识的理解运用,这涉及测试效应的机制解释。来自元认知研究(Benjamin & Bird, 2006)的观点认为,学习者在提取练习中会对自身学习行为进行有效监测,判断哪些内容没有很好掌握,进而在之后的学习中有的放矢、提高效率。根据此观点,实验 1 提取组在测试中,会依据上一篇章的问题解答(即使没有正确答案的反馈),潜意识的在元认知层面实时监控自身学习,从而在下一篇章学习时合理分配时间精力,提高回答的准确性。由于随机化的选项设置依然不能排除被试刻意记忆选项的情况,为明确验证提取练习对高阶技能的影响,实验 2 将对测试形式加以改变。测试形式包括再认、线索回忆和自由回忆等,在具体教学测试中又体现为不同的题型,如正误题、选择题、填空题、简答题、论述题等(罗良,张玮,2012)。而不同形式的测试会对记忆产生不同效果,Carpenter 和 DeLosh (2006)让被试记忆词语,最初测试和最终

测试采用 3 种形式：再认、线索回忆、自由回忆。结果表明，无论最终测试采用哪种形式，最初测试为自由回忆时，被试在最终测试的表现最好。Butler 和 Roediger (2007) 在真实学习的情境下研究了这个问题，实验中大学生被试学习一篇文章初始测试中采用多项选择题和简答题，研究发现，简答题条件下的被试在最终测试上表现较好。由于选择题选项中包含正确答案，被试可依赖的线索强，主动性较差。此外，选择题选项中也包含错误信息，这容易误导被试形成错误记忆，在没有反馈的情况下这些错误记忆将持续存在 (Marsh, Roediger, Bjork, & Bjork, 2007)。

因此在实验 2 中安排被试连续学习 5 篇文章，提取组被试在每篇文章学习后完成简答题测试，以验证提取练习影响高阶技能的稳定性程度。

3.1 被试

通过校园 BBS 征募西北师范大学本科生 60 名 (男女生各 30 名)，平均年龄 20 岁，母语均为汉语。被试报告均无神经性疾病史，视力和矫正视力正常。实验开始前所有被试签署知情同意书，结束后会有一份小礼物赠予被试。

3.2 材料

与实验 1 材料基本相同，将 3 篇文章扩展到 5 篇文章，每篇文章约 850 个字。与实验 1 不同的是，每篇文章后，会依据正、反两方作者的观点设置 4 道高阶技能型简答题，在 4 道简答题中，两道题目来自“正方”观点，剩余两道来自“反方”观点。实验 2 中剔除概念型问题，对研究主题“高阶技能、多篇章”进行深入、贴近真实情境的探讨。

3.3 设计

采用 2×4 混合设计。其中，学习条件为被试间变量，包括提取练习条件和重学条件；高阶技能各维度为被试内变量，包括应用、分析、评估和创造。因变量分别为：最终测试高阶技能的成绩和作答反应时。

3.4 程序

与实验 1 基本相同，提取组被试阅读一篇文章的间隔中，需要回答与文章主题相关的 4 道简答题；重学组被试再学习一次文章。学完全部 5 篇文章后，经过一个 5 min 的分心任务，在最终测试阶段进行所学 5 篇文章的单项选择题测试。整个实验大约需要 85 min。

3.5 结果

由于实验 2 的学习阶段提取任务变为简答题，

因而需要辨别与实验 1 的差异，以及高阶技能中 4 个维度的差异。

3.5.1 即时测试实验结果

对于实验 2 的即时测试结果，收集到被试纸笔简答题答案。分析被试的作答情况，根据每道题目预先设定好的 3 个答题点，被试答对两个或两个以上答题点则判定该题目回答正确，描述统计结果如表 4 所示。所有 5 篇阅读后的简答题正确率为 75% (Vaughn & Rawson, 2011) 以上的作为有效被试，可进入下一阶段实验。最后，74% 的被试正确率达到 75.73%，基本都能正确回答出两个答题点。

表 4 实验 2 提取组被试简答题作答情况

被试	每篇文章简答题回答正确的题目数(道)					
	文章一	文章二	文章三	文章四	文章五	总体
1	3	4	3	2	3	15
2	4	2	3	3	4	16
3	3	2	2	3	2	12
...	...	...	...	...	...	...
30	3	4	2	2	2	13
<i>M</i>	2.75	2.92	2.56	2.67	2.75	13.65

3.5.2 最终测试实验结果

被试完成 30 道选择题。描述统计结果如表 5 所示。

表 5 实验 2 两组被试最终测试数据

被试	最终测试正确题目数(道)	
	提取练习组	重学组
1	19	9
2	21	12
3	22	12
...	...	...
30	19	7
<i>M</i> (<i>S_d</i>)	18.38 (1.77)	9.77 (2.06)

进行 2(学习策略)×4(高阶维度)的多元方差分析(MANOVA)。结果表明，学习策略主效应显著， $F(1, 58) = 191.12, p < 0.05, \eta^2 = 0.77$ ，高阶维度主效应显著， $F(3, 174) = 25.34, p < 0.05, \eta^2 = 0.30$ ，高阶维度与学习策略交互作用显著， $F(3, 174) = 32.89, p < 0.05, \eta^2 = 0.36$ 。进一步分析提取组和重学组高阶各维度，应用类问题的成绩差异显著， $F(1, 58) = 127.98, p < 0.05, \eta^2 = 0.69$ ；分析类问题的成绩差异显著， $F(1, 58) = 131.08, p < 0.05, \eta^2 = 0.69$ ；评估类问题的成绩差异不显著， $F(1, 58) = 2.41, p > 0.05, \eta^2 = 0.04$ ；创造类问题的成绩差异不显著， $F(1, 58) = 1.89,$

$p > 0.05$, $\eta^2 = 0.03$ 。如图 8 所示。

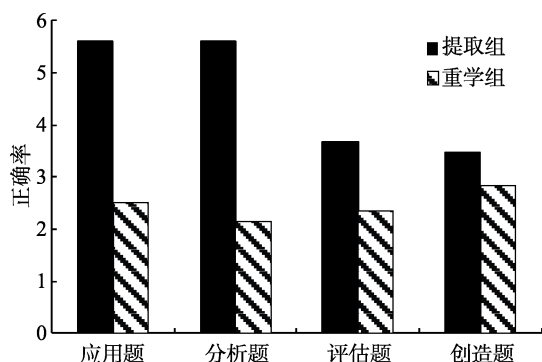


图 8 实验 2 最终测试中提取组与重学组在高阶技能各个维度上成绩的比较

3.5.3 反应时实验结果

提取组与重学组反应时平均数分别为 13.64 s 和 25.74 s。与实验 1 相比, 提取组平均用时缩短, 证明最开始的提取难度越大, 越有利于记忆的保持与迁移(Roediger & Karpicke, 2006)。

3.6 讨论

实验 2 在实验 1 的结果上进一步验证了假设, 当材料增加至 5 篇文章时, 提取练习仍能有效促进高阶技能的发展, 对不同维度的高阶技能产生影响。正如 Bloom 和 Anderson 学习分类的观点, 随着学习深度变化, 所消耗的认知资源越多, 推论至高阶技能上, 就出现不同维度间的差异。相较于重学组, 实验 2 中提取组在应用、分析维度上均有较好发展, 评估和创造维度不受学习策略影响。初步推测, 中国大学生应用、分析能力较为成熟, 能够运用已学概念、理论解决实际问题, 实施相关方案等等。

另一方面, 实验 2 的正确率有所提高。学习内容的增加, 被试的专注程度可能成为实验的干扰因素。实验 2 提取组最终成绩保持了较高水平, 这涉及提取深度的问题。必要难度(desirable difficulty)理论认为, 与记忆保持相联系的提取强度(retrieval strength)与存储强度(storage strength)呈反向相关, 唯有努力的提取才能促进记忆的存储, 使记忆保持更长久, 促进长期学习的效果(Roediger & Karpicke, 2006)。实验 2 选择题变为简答题, 提取强度增加, 消耗更多努力以提高记忆效果。但注重难度的同时还应关注初次提取的正确率, 实验 2 即时测试的结果也证明了这一点。

4 总讨论

研究设计两个实验, 使用多篇文章作为学习材

料, 证实了提取练习策略不仅增强了对于概念型知识的记忆(Wissman, Rawson, & Pyc, 2011), 同样可以促进高阶技能的连续发展。在验证 Agarwal (2011) 研究的基础上, 进一步揭示提取练习促进后续学习的重要意义。研究重点考察了高阶技能和概念知识学习的差异与高阶技能各个维度的差异, 实验结果相互印证, 反映出提取练习策略的高效和稳定性, 证明了该策略在完善学生知识体系和技能培养方面的独到之处。

4.1 对前摄抑制理论与上下文转换假说的验证

前摄抑制理论强调前后所学内容的差异性, 而上下文内容转换假说则强调所学内容的融合性, 究其本质都在探讨所学内容区分度的问题。

前摄干扰指的是记忆当前信息的能力会大大受到先前所学信息的影响(Underwood, 1957)。在过去的几十年里, 该现象一直被广泛研究, 已证明其具有重要意义。真实教育情境中, 学生常常需要记住连续的、成套的学习材料(如, 词表或文章梗概), 从而在学期末的综合性考试中获得好的成绩(如, 学期末论文综述型考试或学期末年级综合考试)。连续的学习内容势必产生前摄抑制干扰, 降低学习效果(Bäuml & Kliegl, 2013)。实验中每学过一篇文章后会安排即时测试, 然而之后几篇文章的即时测试成绩远没有第一篇文章成绩高, 这与前摄抑制理论相契合。因为之前所学内容占用了大量负荷, 以致于后续内容的理解受阻, 这在实验 2 中体现更为明显, 由于简答题作答不仅仅衡量记忆水平, 对被试在理解基础上应用原理解决类似问题的能力要求较高, 若不能很好区分前后内容差异, 往往是导致作答错误的原因。一段延迟后的最终测试中, 将所学内容涉及的所有题目混合在一起, 排除首因和近因效应的影响, 前摄干扰理论的解释就不再适用。根据上下文转换假说, 提取练习策略实际是对所学内容的流畅转换, 为不同内容给予特定线索, 将线索连贯起来, 增强区分度(Brewer et al., 2010; Divis & Benjamin, 2014)。因此, 提取所学内容, 即对所学进行小结, 学习者可以更加透彻掌握内容的实质, 为每篇文章学到的关键点贴上“标签”, 到最终测试时, 每个“标签”会自动区分出其适用的新情境。研究显示被试在最终测试中, 提取组解决问题准确而高效, 该结果与前人结果一致, 支持了上下文转换假说。因此, 本研究不仅是对前人理论的再次验证, 更进一步融合两个理论的观点, 解释了即时测试与延迟测试条件下, 提取练习提高成绩的原因。

4.2 提取练习中情景语境线索的重要性

基于提取的学习策略其内部机制探讨一直以来都是测试效应研究的热点。当前最具代表性的观点之一, Carpenter (2009)的“精细化提取假设”(elaborative-retrieval hypothesis)认为,在提取过程中,由于提取行为本身激活了对于所回忆目标的相关语义信息,再次提取时语义信息可以精细化线索和目标之间的联系,有助于随后提取。而最新的研究却对此观点表示质疑(Lehman, Smith, & Karpicke, 2014),指出“情境语境解释”(episodic context account)更符合测试效应的潜在机制。

语境在提取练习中起关键作用(Sahakyan & Hendricks, 2012),特别是在篇章学习中,材料的复杂性通常很难使学习者生成有效的中介线索,因而精细化提取假设的解释存在局限。测试效应的优势往往体现在某些特定测试类型下,如自由回忆(实验2的简答题测试),很大程度上依赖于当前的语境线索支持(Brewer et al., 2010)。按照情境语境解释,提取重新构建了上下文内容的表征,从而在之后测试中提高了学习者记忆检索的能力。语境提供的线索有限,避免了记忆线索过多对提取造成干扰。本研究中提取组最终测试成绩优于重学组,推测提取组得益于语境线索而非中介线索,使其在限定范围内进行有效记忆检索以完成任务。此外,情景语境解释认为,提取练习缩小了记忆检索范围、加快提取速率,本研究两个实验提取组平均反应时低于重学组,结果验证了情景语境解释对测试效应内部机制的阐述。

4.3 提取练习对高阶技能的差异化影响

提取练习在促进长时记忆保持上效果稳定,对高阶技能各维度也有一定影响。实验中提取组应用与分析两项,其提升幅度之大或超过概念型题目,但两组被试在评估和创造维度上成绩始终差异不大。原因在于高阶技能的分类方式导致每类技能出现层级化差异。高阶技能最初并没有统一的定义,其划分依据来源于早期的教育目标分类(Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956)。Bloom认为每一类别的分界存在进阶性:若想获得更高阶的能力,学习者必须掌握较低层级的技能。换句话说,在掌握理解、应用、分析技能前,学生首要任务是获取知识。拥有巨大的概念事实类知识库有助于掌握高阶技能,即本文研究提取练习促进高阶技能发展的基础。Anderson等人(2001)提出的修订版教育目标分类,将高阶技能的分类细化,其原则仍采用

“进阶性”划分。其中,创造技能位于最高层级,是高阶技能发展的最高要求。实验中,重学组被试高阶各个维度上的正确率均在50%以下,而提取组被试在应用和分析维度正确率皆高于70%,一方面,足以推断实验中的大学生被试的应用与分析能力已发展成熟,能在遇到相似情境时充分调动记忆中已有知识库做出正确判断;另一方面,提取练习策略对激发高阶技能中应用和分析维度的学习也是有效的。对于评估、创造维度,周新虎和刚志荣(2014)发现,目前国内高校大学生普遍创造力不足,且影响其创造力的几个主要因素有学生不敢质疑、明哲保身、缺乏活力等。江学良和杨慧(2010)认为高校学生基本知识理论虽然比较扎实,但观察分析能力差,独立解决实际问题能力差,创新意识薄弱,更加谈不上创造性的分析问题和解决问题。可见,中国大学生自身现状的局限,可能是实验中评估和创造技能均未达预期水平的原因之一,亦不因学习策略的改变而有所提升或抑制。

此外,测试效应诸多研究已经证实,最终测试前的延迟时间影响记忆保持的时长(Roediger & Karpicke, 2006)。短延迟测试增加了提取的正确性,但所提取的信息却难以保存在长时记忆中(Roediger & Butler, 2011)。本研究采用了短延迟的最终测试,可能是造成评估和创造维度正确率偏低的又一原因,未来研究可以考虑使用长延迟的形式。

总之,研究证明了提取练习策略促进了高阶技能的发展,具体表现在高阶技能多个维度的提升显著,这值得教育工作者借鉴,并考虑逐步落实于实际教学中。

5 结论

(1) 在多篇章提取练习中,所生成的不同情境语境线索有效地促进了概念型知识的记忆和高阶技能的发展。

(2) 提取练习促进高阶技能的掌握,且在学习、应用、分析与评估维度上学习效果显著,是培养学生全面发展的良好策略。

参 考 文 献

- Agarwal, P. K. (2011). *Examining the relationship between fact learning and higher order learning via retrieval practice* (Unpublished doctoral dissertation). Washington University.
- Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., ... Wittrock, M. C. (2001). *A taxonomy for learning, teaching,*

- and assessing: A revision of Bloom's taxonomy of educational objectives (abridged ed.). New York, NY: Addison Wesley Longman, Inc.
- Bäuml, K.-H. T., & Kliegl, O. (2013). The critical role of retrieval processes in release from proactive interference. *Journal of Memory and Language*, 68(1), 39–53.
- Benjamin, A. S., & Bird, R. D. (2006). Metacognitive control of the spacing of study repetitions. *Journal of Memory and Language*, 55(1), 126–137.
- Bloom, B. S. (Ed.), Engelhart, M. D., Furst, E. J., Hill, W. H., & Krathwohl, D. R. (1956). *The taxonomy of educational objectives: The classification of educational goals* (Handbook 1: Cognitive domain). New York, NY: David McKay Company, Inc.
- Brewer, G. A., Marsh, R. L., Meeks, J. T., Clark-Foos, A., & Hicks, J. L. (2010). The effects of free recall testing on subsequent source memory. *Memory*, 18(4), 385–393.
- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1118–1113.
- Butler, A. C., & Roediger III, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19(4–5), 514–527.
- Callender, A. A., & McDaniel, M. A. (2009). The limited benefits of rereading educational texts. *Contemporary Educational Psychology*, 34(1), 30–41.
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563–1569.
- Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & Cognition*, 34(2), 268–276.
- Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20(6), 633–642.
- Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis and Humphreys (2001). *Psychological Review*, 111(3), 800–807.
- Darley, C. F., & Murdock, B. B. (1971). Effects of prior free recall testing on final recall and recognition. *Journal of Experimental Psychology*, 91(1), 66–73.
- Divis, K. M., & Benjamin, A. S. (2014). Retrieval speeds context fluctuation: Why semantic generation enhances later learning but hinders prior learning. *Memory & Cognition*, 42(7), 1049–1062.
- Jacoby, L. L., Wahlheim, C. N., & Coane, J. H. (2010). Test-enhanced learning of natural concepts: Effects on recognition memory, classification, and metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(6), 1441–1451.
- Jiang, X. L., & Yang, H. (2010). An factor analysis of hindering college students' innovation ability raise. *Journal of Innovation and Enterprise Education*, 1(2), 31–34.
- [江学良, 杨慧. (2010). 阻碍大学生创新能力提高的影响因素分析. *创新与创业教育*, 1(2), 31–34.]
- Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science*, 331(6018), 772–775.
- Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science*, 319(5865), 966–968.
- Leeming, F. C. (2002). The exam-a-day procedure improves performance in psychology classes. *Teaching of Psychology*, 29(3), 210–212.
- Lehman, M., Smith, M. A., & Karpicke, J. D. (2014). Toward an episodic context account of retrieval-based learning: Dissociating retrieval practice and elaboration. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1787–179.
- Luo, L., & Zhang, W. (2012). Testing for learning enhancement: Recent advances in memory psychology and educational revelation. *Journal of Beijing Normal University (Social Science Edition)*, 1(1), 43–50.
- [罗良, 张玮. (2012). 利用测试促进学习: 记忆心理学的研究进展与教育启示. *北京师范大学学报(社会科学版)*, 1(1), 43–50.]
- Marsh, E. J., Roediger, H. L., Bjork, R. A., & Bjork, E. L. (2007). The memorial consequences of multiple-choice testing. *Psychonomic Bulletin & Review*, 14(2), 194–199.
- McDaniel, M. A., Howard, D. C., & Einstein, G. O. (2009). The read-recite-review study strategy: Effective and portable. *Psychological Science*, 20(4), 516–522.
- McDaniel, M. A., Roediger, H. L., & McDermott, K. B. (2007). Generalizing test-enhanced learning from the laboratory to the classroom. *Psychonomic Bulletin & Review*, 14(2), 200–206.
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K.-H. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(2), 287–297.
- Postman, L., & Keppel, G. (1977). Conditions of cumulative proactive inhibition. *Journal of Experimental Psychology: General*, 106(4), 376–403.
- Roediger, H. L., Agarwal, P. K., Kang, S. H. K., & Marsh, E. J. (2010). Benefits of testing memory: Best practices and boundary conditions. In G. M. Davies & D. B. Wright (Eds.), *New frontiers in applied memory* (pp. 13–49). Brighton, UK: Psychology Press.
- Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15(1), 20–27.
- Roediger, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, 1(3), 181–210.
- Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor of strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review*, 11(1), 131–136.
- Sahakyan, L., & Hendricks, H. E. (2012). Context change and retrieval difficulty in the list-before-last paradigm. *Memory & Cognition*, 40(6), 844–860.
- Szpunar, K. K., McDermott, K. B., & Roediger III, H. L. (2008). Testing during study insulates against the buildup of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(6), 1392–1399.
- Tulving, E., & Watkins, M. J. (1974). On negative transfer: Effects of testing one list on the recall of another. *Journal of Verbal Learning and Verbal Behavior*, 13(2), 181–193.
- Underwood, B. J. (1957). Interference and forgetting. *Psychological Review*, 64(1), 49–60.
- Vaughn, K. E., & Rawson, K. A. (2011). Diagnosing criterion-level effects on memory: What aspects of memory are enhanced by repeated retrieval? *Psychological Science*, 22(9), 1127–1131.
- Wissman, K. T., Rawson, K. A., & Pyc, M. A. (2011). The interim test effect: Testing prior material can facilitate the

- learning of new material. *Psychonomic Bulletin & Review*, 18(6), 1140–1147.
- Zhou, A. B., Ma, X. F., Li, J., & Cui, D. (2013). The advantage effect of retrieval practice on memory retention and transfer: Based on explanation of cognitive load theory. *Acta Psychologica Sinica*, 45(8), 849–859.
- [周爱保, 马小凤, 李晶, 崔丹. (2013). 提取练习在记忆保持与迁移中的优势效应: 基于认知负荷理论的解释. *心理学报*, 45(8), 849–859.]
- Zhou, X. H., & Gang, Z. R. (2014). The influence factors of college students' creativity and the analysis of training methods. *Jiangsu Higher Education*, (5), 110–111.
- [周新虎, 刚志荣. (2014). 大学生创造力的影响因素及培养途径分析. *江苏高教*, (5), 110–111.]

Retrieval Practice Produces More Learning in Multiple-list Tests with Higher-Order Skills

ZHOU Aibao; YANG Tiancheng; CHENG Chen; MA Xiaofeng; ZHAO Jing

(Guidance Center on Student Learning, School of Psychology, Northwest Normal University, Lanzhou 730070, China)

Abstract

Learning is usually thought to occur during episodes of studying, while the retrieval of information on testing simply serves to assess what was learned. Test-enhanced learning, or the use of tests and quizzes to engage retrieval processes, has been widely demonstrated as an effective strategy for facilitating fact learning. Retrieval practice can not only promote single content learning but also be beneficial for multiple-context. However, the impact of testing effects on sequential item learning remains unclear. According to Bloom's taxonomy of knowledge and skills in education, which Anderson revised, training of higher-order skills has been increasingly valued. To this end, current research has established the generality of test-enhanced learning as well as pointed to conditions under higher-order skills are manifested. We examined whether retrieval practice facilitates higher-order learning in multiple-article test format. Across two experiments, participants engaged in either retrieval practice or restudied with basic concept questions and higher-order questions.

In Experiment 1, 60 college students studied three passages in the initial learning conditions and then answered concept and higher-order multiple-choice questions. After five minutes, participants were administered both concept and higher-order tests that were intended to examine the type of skill differences. In Experiment 2, 60 college students studied five passages in the initial learning conditions and then answered higher-order short-answer questions. After five minutes, participants were administered higher-order tests. In both experiments, higher-order skills were considered to comprise the ability to understand, apply, analyze, evaluate, and create categories of a revised Bloom's taxonomy of knowledge and skills in education.

Experiment 1 indicated a significant main effect for learning strategies. Results also revealed that for each taxonomy level, the retrieval practice group's performance and reaction time were better than those of the restudying group. An in-depth analysis showed that retrieval practice greatly contributed to concept question performance. Experiment 2 replicated the results of Experiment 1 with five articles for reading and more appropriate short-answer questions. Thus, Experiment 2 extends the findings of Experiment 1. Results reflected differences in the final test between the groups, especially for the apply and analyze dimensions. Nonetheless, we found no significant difference in the evaluate and create categories. Results suggested that retrieval practice is an effective method for learning, not by merely embodying the memory of single content but also broadening multiple contexts and promoting the development of higher-order skills. To some degree, it effectively reduced the memory load in learning.

The current study explored the relationship between retrieval practice and multiple-list learning. Results of the analyses of internal mechanisms support the episodic context account. The episodic context account explains the finding that retrieval practice is more likely to be beneficial for more difficult retrieval tasks. In addition, combining the context-change hypothesis and proactive interference theory suggests the superiority of testing effects in multiple-list learning. The development of higher-order skills is a desired outcome of education. Practice testing is one of the most well-established strategies for improving student learning.

Key words retrieval practice; multiple-list learning; high-order skills