

《心理学报》审稿意见与作者回应

题目：多篇章学习中提取练习策略对高阶技能的影响

作者：周爱保 杨天成 程晨 马小凤 赵静

第一轮

审稿人 1 意见：研究首次采用多篇文章作为实验材料，进一步深化了提取练习效应的应用范围，并首次验证了提取练习是促进高阶技能连续发展的有效策略，具有一定的创新性。具体建议如下，供作者参考。

意见 1：两个实验中重学组被试都是重新阅读文章，但如何保证被试是认真地重新阅读了文章呢？被试第一遍阅读就已经了解了文章的相关信息，第二遍阅读有可能会敷衍了事，不认真阅读，这样就会造成重学组和提取组在条件上本身的不对等。

回应：感谢审稿专家非常敏锐和周严的问题。“提取练习”和“重复学习”两种学习策略在大量相关研究中被作以比较，究其根本原因是两者“学习”发生阶段不同。人们往往认为学习只发生在编码阶段，即持重复学习方式才是学习的观点。然而，在“Tulving E. The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1967, 6(2):175-184 ”这篇研究中，Tulving 证明学习不只发生在编码阶段，测试也是一种学习。

Carpenter(2009)设计实验，对比了提取组和重学组最终测试成绩。研究对重学组被试的加工深度和专注度的质疑给出回答。其研究中重学组被试在重学之后，还需评定词对材料间相关度。结果表明，被试在重学过程即使进行了积极加工，在最终测试时仍然不能很好的对项目进行回忆，因而重学方式被认为是一种低效的辅助记忆策略。

此外，我们在实验开始前的准备工作也重视了诸多细节。例如，实验前主试向被试仔细讲解大致实验任务，重复学习组在第二次学习之前，屏幕呈现的指导语也明确说明：“请您再一次认真学习，学习之后会安排相应的测试，以便检验您的学习效果。”因此，被试在第二次学习时依然会处于警觉状态。另外，被试在开始实验前也收到了我们准备的一份小礼物，符合实验伦理的要求，实验过程中被试积极性较高，实验的卷入度相对良好。由于行文表达不完善使专家产生了疑问，在此深感抱歉，我们在正文(P4, Line22)用红字做了修改补充。

相关文献：

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.

意见 2: Agarwal(2011)的研究使用了一篇文章作为实验材料，而本文的作者使用了 3 篇或者 5 篇的阅读材料进行研究，怎么能确保 Agarwal(2011)的实验范式适用于本研究的多篇文章阅读？因为多篇阅读中多次提取练习会对被试造成练习效应，因而才导致最后测试时反应时低于重学组，正确率也高于重学组。

回应: 非常感谢审稿专家提出的疑问，下面将对问题依次做出解释，必要处会在文中予以补充修正。

关于范式使用是否恰当问题。涉及测试效应的研究，对刺激的选择十分多样，包括短文(Kang, McDermott, & Roediger III, 2007)，词表(Wheeler, Ewers, & Buonanno, 2003)，词语定义(Cull, 2000)，外语词汇(Kang et al., 2007)，一般性知识(McDaniel & Fisher, 1991)，视频材料(Butler & Roediger III, 2007)，以及词对(Vaughn & Rawson, 2011)。也有研究选取多个学习材料（例如，多个词表）展开讨论，研究证明了提取对后续学习的促进作用（e.g., Szpunar, McDermott, & Roediger, 2008; Tulving & Watkins, 1974）。例如，Szpunar(2008)材料共计 5 个 List。第一组被试在 List1-4 后安排即时测试，第二组被试 List1-4 后安排重复学习，第三组被试在 List1-4 后插入分心任务。最后，三组参加 List5 的学习测试。结果显示，提取练习组在 List5 上回忆出的项目数更多，证明了提取练习策略在记忆上的优势。之后，Pastötter(2011)使用 5 列词表进一步解释提取的优势，具体表现在提高了对于每个 List 的记忆编码。多篇章的学习向学习者提出更高要求，也将提取练习策略的应用范围上升至一个新高度。受 Agarwal(2011)研究启发，本研究立足教学实践，结合 Pastötter(2011)的研究，多个篇章的学习拟合学习者的真实情境。依托于测试效应理论背景及研究框架，通过对范式修改，研究者可探究各种条件下的测试效应，例如呈现刺激（词对、短文等）。另外，为了方便审稿专家理解，可参考“张锦坤，白学军，& 杨丽娟. (2008). 国外关于测试效应的研究概述. *心理科学进展*, 16(4), 661–670.”，文章较为细致梳理了范式使用问题和相关概念定义。综上所述，探讨“多篇章学习中提取练习策略对高阶技能的影响”具备可行性。

关于会出现“练习效应”的问题，很感谢审稿专家的提示。本文在“**2.4 程序(P5, Line1)**”表达存在歧义，已做红字修改。实验一采用三篇文章、实验二增至五篇文章的学习模式，对于提取组和重学组而言，均使用相同材料，相同呈现顺序进行学习，保证了实验条件的匹配；对每组被试而言，所学文章内容间不存在相关，主题不同，但文章结构由正方、反方两种观

点组成，且每个实验所用文章难度相当。可见，在学习阶段，学习材料的主题和内容不相同，提取组中每篇文章后所回答的问题皆不相同。而练习效应定义是：被试由于多次重复同样的程序，实验成绩提高或降低的现象，本研究中提取组需针对不同文章设置的问题给出选择，因此不受经验影响。特别在实验一中，第三篇文章学习后，两组被试进行相同实验处理，即立即作答文章三的选择题，结果表明，提取组成绩显著高于重学组，也直接印证了是实验处理造成的结果差异。

相关文献目录：

- Kang, S. H., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology, 19*(4–5), 528–558.
- Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory, 11*(6), 571–580.
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*(3), 215–235.
- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*(2), 192–201.
- 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.
- 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.
- 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.
- Pastötter, B., Schicker, S., Niedernhuber, J., & Bäuml, K. T. (2011). Retrieval during learning facilitates subsequent memory encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(2), 287–297.

意见 3: 实验一的每篇文章中的提取练习题目与最后的测试的题目是一样的，这必定会导致被试的反应时加快，正确率也提高，虽然作者采用 Duchastel(1981)的研究作解释，但是解释的力度明显不够，其他两种测试形式是什么形式也没有说出。

回应: 感谢审稿专家，这个问题很有价值。不完善之处将在正文使用红字作以修改补充。

按照文章逻辑，引用 Duchastel(1981)研究意在指出实验二目的，即实验一的题目设置为选择题，而实验二问题设置转变为简答题的意义。根据审稿专家的建议，对“**3 实验二(P10, Line7)**”内容做出删除修改，具体见正文。

关于最终测试题目的问题。实验中每篇文章后的题目没有提供正确答案的反馈。举例来说，被试最初回答错误的题目，在最终测试中可能仍将出错，因此不能保证最终测试中提取组的正确率提高和反应时加快，前人也曾使用此程序验证提取练习策略，例如“**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.**”

意见 4: 在多篇章阅读中，提取练习组与重学组条件存在不对等的情况，提取练习在每篇文章后都对被试进行测试，使被试提高警惕，一直很用心的在学习，很可能在学习中就会主动的进行记忆，而重学组的被试只有到最后一篇文章才进行测试，因而警惕性则相对较低，在学习中不会主动的进行记忆。

回应: 感谢审稿专家对文章的细致评审。这个问题可结合我们对第一条建议的回答，相似之处不再赘述。本研究在两组被试开始实验前的总指导语均做以说明：“对所学的所有内容，会安排相应的测试来检验学习效果。”在学习警觉性方面，重学组被试的对于材料的学习专注度良好。

另一方面，实验结果也从侧面证实了提取和重学组保持了“对等”。在实验一中，两组被试在“创造”和“评估”两个维度的差异并不显著。结果证明，连续的篇章学习中，两组被试的差异仅表现在使用学习的策略不同。

意见 5: 如何能证明两个实验中各高阶技能的测试题是测量被试的高阶技能的？作者没有提供任何例子，审稿人难以判断。

回应: 由于文章篇幅字数的限制，在正文我们无法完整展示所有示例，因而造成审稿专家的理解困难，我们深表歉意。为此，我们对每一类高阶技能问题给出具体例子，请审稿专家作为参考。

(1) 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。该分类也可称为迁移。

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？

- 1) 所有人独立自主的社会。
- 2) 没有政府参与的社会。
- 3) 不再需要纳税的社会。
- 4) 所有人都被平等对待的社会。

该题目的正确答案为1)，文中提到正方作者对目前福利型社会的种种观点，题目设置为相反情境，即无福利项目实施的社会。在文章第五段中可以看出，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”，由此推测，将此观点应用到题目设置的新情境中，答案1)为正确选项。

(2) 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。

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”？

- 1) 正方作者。
- 2) 反方作者。
- 3) 都同意。
- 4) 都不同意。

该题目的正确答案为3)，分析类题目要求高于应用类题目，需把握住正、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”，全面掌握后方能从容作答。

(3) 评估性问题：选项中列出四个陈述句（文章中未出现过），被试需要判断哪一项符合作者的观点。

例：哪一项陈述是对正方作者的正确评估和总结？

- 1) 福利项目不会起作用，因为它们都花销太大。
- 2) 福利项目是有害的，因为它们使情况变得更糟。
- 3) 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。
- 4) 福利项目可能有用，但是它们很少满足人们的需要。

正确作答该题目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，很容易误选，没有达到真正的评估文章观点。理解大意“目前大多数的福利项目都不该出台”，也要抓住细节“大多数”，“出发点是好的”，作者的反对，不仅仅是由于花销大、浪费钱的原因，了解其背后用意“渴望个体独立性”，才能正确评估。

(4) 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。

例：请你预测反方作者会支持哪一种税收和支出结构？

- 1) 平等税收；平等支出。
- 2) 富人高税收，穷人低税收；富人多支出，穷人少支出。
- 3) 平等税收；富人少支出，穷人多支出。
- 4) 富人高税收，穷人低税收；富人少支出，穷人多支出。

该题目为高阶技能最高层级问题，需要完整阅读文章中反方作者的观点，找到关键的几个点，“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，只有理解了作者的主旨，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。

意见 6：实验一的被试中男女被试各半，但是每组男女比例为 2:3，那么是否是一组男女比例为 2:3，另外一组男女比例为 3:2？为什么设计为这个比例？为什么不设为 1: 1？

回应：此处确为错误表达，非常感谢审稿专家的指正，我们为文章的表达错误感到抱歉。被试的男女比例问题在实验中至关重要，作为人口学变量通常成为影响结果的因素之一，因而必须受到重视，实验一 2:3 的比例是一个错误表述。本研究实验一结束后，有个别被试并未按要求完成实验（如：自始至终只指定一个选项作答），这些数据我们予以剔除，从而造成男女比例的不匹配（2:3）。但后续我们对实验被试做以补充，以保证样本的完整，实验一最终男女比例为 1:1（男女各 30 名）。“2:3”的写法来源于文章初稿，未能及时更正是我们的错误，在此对审稿专家的严谨作风表示敬佩。正文“**2.1 被试(P3, Line26)**”部分错误表达已更正。

意见 7：实验二中简答题中会根据每道题目预先设定好三个答题点，答题点是怎么设定的？是否具有可信度？

回应：感谢审稿人的建议。对此疑问将以第一篇文章“社会福利弊大于利”为例，分析四道简答题答题点设置。

(1) 应用类简答题：设置一个与文章内容相关的新的问题，要求被试依据文章中对观点作者的解决方式回答题目，有时需要回答正、反两方作者各有什么解决方案，有时仅针对单个一方作者作答。该分类也可称为迁移。

例如：如果没有社会福利项目，请您依据文章分别回答，正、反方作者认为社会将会变成什么样？

该题目要求从两方作者的角度分别探讨。在一个设定的“没有福利项目”的情境下，从文中可以找到，现在美国社会充斥各种福利项目时，正方作者感慨人民被各种缴纳金所累，既未实现福利人民的目的，反而造成人民的依赖，所以被试需回答“正方作者期望社会中人人独立自主”；相对的，反方作者认为福利项目保障了人民的健康，是经济增长和国家安全的基础，所以被试需回答“反方作者认为社会将会一片混乱，失去安定，人民饱受疾苦”。

(2) 分析类简答题：即对两方作者的观点进行分析；题干中陈述一个与文章内容相关的议题，要求被试分析是正、反方作者对这一议题分别会有什么样的态度。

例如：针对“政府对社会发展的作用”这一议题，请您分别用一句话分析正、反方作者的观点。

该题目要求被试把握文章中提到的除了福利项目这一中心主题外，还应注意，正方作者认为虽然政府的出发点都是好的，但议政官员、项目实施人员、从业人员等在福利项目的实施中获得利益，却未对项目的实施负责，所以被试需回答“政府虽值得尊敬但并未让福利项目服务人民”；反方作者认为政府部门的公共作用渗透在我们日常生活的方方面面，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，所以被试需回答“反方作者非常肯定政府的作用，认为没有政府的努力就没有现在安定的社会”。

(3) 评估类简答题：题干中给出一个论述，要求被试评估该论述符合哪一方作者的观点，并简单陈述原因。

例如：请您判断正、反方作者的观点后评估“反对福利项目”的原因。

该题目要求被试首先判断这个题干议题是哪方作者的观点，“正方作者”。然后，根据正方作者的文意梗概：福利项目出台的越多越无法保障人民的基本生活需要、福利项目带来的更多是开支的增长、议政官员们只肯为自己花钱，作者认为福利项目是众矢之的，大多数的福利项目都不该出台。被试需要提炼出文中 1-2 点作者的论点语句，并加以评估“作者正是看到了贫困人民由于对福利项目的依赖，反而变得更加贫困，才反对各项福利项目的出台”。另外，还要求被试能说出正方作者的观点“取消福利项目，或许在短期内，有些人会不得不从事一些低收入的工作。但从长远来看，这才是最人性化的社会形态。”这样才算是对该题目完整的回答。

(4) 创造类简答题：要求被试设计或预测一个新情境可能产生的结果；而被试的回答应以文章中作者陈述的观点为基础。

例如：假设现在要针对“穷人与富人的税收与支出政策”征求正、反方作者的观点，请您预测，两方作者分别会给出什么样的建议。

该题目设定了一个与本文“福利项目”看似不相关的主题——税收与支出，但实际上仍需要被试依据文章的中心观点作答本题。对于正方作者，他认为所有福利项目都没必要出台，人民才能自给自足，不依赖任何人，因此，被试需要回答“正方作者会希望没有税收政策的束缚，每个人都是平等、独立自主的生活”。对于反方作者，他认为政府对人民的帮助是化于无形的，其效用并非流于表面，如果没有政府的治理，社会必将一片混乱，同时，他也主张福利项目对于不同的人在实施过程中是有差异化的。因此，被试需要回答“反方作者可能会建议对于穷人少税收、富人多税收；而穷人多支出、富人少支出的政策”。

本研究答题点设置依据有三，首先是 Andersen 等人(2001)每个高阶技能的分类下均有详细操作指标（如本文图 1 所示），这也是答题点设置最主要的依据；其次是从实验逻辑连贯性考量，参照了实验一中对应选择题选项的设置；最后，与关于篇章学习的研究者 Agarwal 讨论过后，决定了简答题答题点的设置。

意见 8: 讨论部分未能就“两个实验中评估和创造技能均未达到预期水平”的原因进行分析，因为这与以往的研究结果不一致。

回应: 感谢审稿人的宝贵建议。对不完善之处，我们在**讨论 4.3(P15, Line1)**部分用红字做出了补充修改。

意见 9: 实验一的结果分析中出现过“MSE”，而其他地方没有并没有出现，建议作者统一。

回应: 统计值的表述未履行一致性原则是我们的错误，非常感谢审稿专家的这一建议。现将结果分析中的“MSE”予以删除，统一使用效果量 η^2 表述。

审稿人 2 意见: 本研究探讨了提取练习、重复学习对高阶技能的获得的影响，对教育实践具有一定的参考价值。但论文中明显存在一些较为突出的问题：

意见 1: 在导论部分，对布鲁姆教育目标分类学 2001 版的认知过程的理解存在某些偏差。如，在“创造”类别下，第三个亚类是“producing”，其本意是“创造性的产出”，或者直接译为“产出”。本文译为“贯彻”是不妥的。同样地，实验一中“(4) 创造性问题：要求被试设计或预测一个新情境可能产生的结果。被试的选择应该以文章中作者陈述的观点为基

础。”，被试的“设计”大致可以视为“创造”，但“预测”则属于“理解”层级中的“推论”，不属于高阶技能。

回应：感谢审稿专家的意见，您提供了中肯且极具专业性的翻译。对于“producing”的理解，我们参考了理查德等人(2007)的文献，文章阐述了新修订分类学中包含的六个认知过程维度：记忆(remember)；理解(understand)、应用(apply)、分析(analyze)、评价(evaluate)和创造(create)。具体又细分为 19 种认知过程要素，其中创造包括三个认知过程，即生成(generation)、计划(planning)、贯彻(producing)。贯彻(producing)亦可理解为“构建”(constructing)，是指学习者执行计划以解决既定的问题。在贯彻过程中，学习者要根据对目标的功能性的描述并形成一個与其相一致的成果。

经审稿专家点拨，我们查阅了相关文献并请教了英语专业教师，结合 producing 的概念表述，我们认为审稿专家的翻译精准贴合了 producing 的实际含义。据此，将图 1 中“贯彻”更正为“产出”，符合学术表达的规范性和准确性。

对实验一中“(4) 创造性问题”的解释，本文表述欠妥，感谢专家的细致评审。对创造性问题的界定，本研究参照 Anderson 等人(2001)的分类学方法，偏向于创造类目下的“计划(planning)”一项，是指学习者根据特定情境提出解决方法以完成任务。研究中部分题目设置提及“预测”一词，是参照了 Agarwal(2011)对高阶题目设置的表述。由于外语词汇“predict”表达含义较广，单纯解释成预测显得片面，是我们考虑不够周全。但是这样做也是站在被试的角度，便于其理解题目所要求的任务，故在题目中使用了“预测”字眼。思考过后，在正文部分我们将创造性题目的含义作以修改，使其更符合所表达的具体内涵，修改请见“2.2 材料(P4, Line14)”红字部分。

相关文献：

理查德 E, 梅耶, 韩青青, 柯丽丹, & 盛群力. (2007). 面向意义学习的认知过程. *远程教育杂志*, 1(3), 16-20.

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 edition. White Plains, NY: Longman.

Agarwal, P. (2011). Examining the relationship between fact learning and higher order learning via retrieval practice. (Unpublished doctoral dissertation) Washington University.

意见 2：“提取练习”跟“测验效应”密切相关，前者是指学习方式，后者是指学习结果。有时，测验效应干脆被研究者称为“提取练习效应”。在这方面，Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition*, 20, 632–642.这篇经典文献值得参考，作者并没有引用。在全文中也没有出现“测验效应”这一字眼，这是不太妥当的。

回应：关于“测试效应”和“提取练习”间的概念，审稿专家的辨析清楚明确。为了准确阐释此概念，我们梳理了过往研究，并参考了一些权威文献中的表述。通过归纳整理，结果如下：

张锦坤等人(2008)综述了国外测试效应相关研究，测试效应 (testing effect) 指学习某一内容时，进行测试比额外学习能更好地提高后来对它的记忆保持水平，即便在测试无反馈时也是如此。测试的直接作用指测试这一行为本身提高了学习和长时记忆，即测试效应。Carpenter(2009)研究中的表述：测试 (test) 表现出的好于重复学习的优势称之为测试效应 (testing effect)；提取 (retrieval) 是指在记忆测试中回忆出之前所学信息。Pyc 和 Rawson(2011) 认为提取练习 (retrieval practice) 在后续记忆的增强上效果显著。Rawson 和 Dunlosky(2012) 的研究中把提取练习理解为一种有效学习策略。Grimaldi 和 Karpicke(2012)的解释：提取练习是一种认知加工过程。周爱保等人(2013)的研究中的描述：提取练习又称测试效应，是近 10 年来学习与记忆研究的热点问题。提取练习策略相比于传统的学习策略，如重复学习、过度学习，能够产生更长时间的记忆保持。有关测试效应的研究众多，表达相近的不再赘述。

总结了前人研究之后，本文对于“提取练习”和“测试效应”的表述不准确，“提取练习效应”是一个不当表述。审稿专家的提示很关键，我们参考了“Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory and Cognition*, 20, 632–642.”这篇经典文献，已对 **1 引言(P1, Line7)**使用红字进行补充修改，反映结果使用“测试效应”，描述过程使用“提取练习”，**全文统一了表达**，感谢专家的指正。

相关文献：

张锦坤, 白学军, & 杨丽娴. (2008). 国外关于测试效应的研究概述. *心理科学进展*, 16(4), 661–670.

Pyc, M. A., & Rawson, K. A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory?. *Journal of Memory and Language*, 60(4), 437–447.

Rawson, K. A., & Dunlosky, J. (2012). When is practice testing most effective for improving the

durability and efficiency of student learning?. *Educational Psychology Review*, 24(3), 419–435.

Grimaldi, P. J., & Karpicke, J. D. (2012). When and why do retrieval attempts enhance subsequent encoding?. *Memory & Cognition*, 40(4), 505–513.

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.

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.

意见 3: 在两个实验中，作者使用的统计检验方法都是错误的。两个实验中，概念和高阶技能的测验结果都应该视为因变量，作者都作为自变量分析，这是明显的错误。正确的数据分析方法应该是 MANOVA。

回应: 非常感谢审稿专家的批评指正，我们已认识到统计方法上存在的错误。在斟酌了实验设计的表述后，已在正文用红字作以修正。对于本研究应该采用的统计方法，再次感谢专家的提示，已使用 MANOVA 重新进行了数据分析，具体请见 **2.5 结果(P7, Line1)**和 **3.5 结果(P12, Line5)**部分。统计方法上出现失误，我们深感惭愧和抱歉。

意见 4: 本文根据结果还讨论了“提取练习策略对降低记忆负荷”的作用。记忆负荷或认知负荷有其专门的理论和概念界定，本文在引言和研究方案中并没有涉及专门的记忆负荷这一因变量，因此这种讨论显得突兀。

回应: 非常感谢审稿专家的提示。对降低记忆负荷的作用讨论突兀是我们的不足，专家的意见敏锐独到，根据专家的意见在讨论部分删除了这部分内容，对研究涉及的测试效应内部机制以探讨，详细修改请见**讨论 4.2(P14, Line17)**红字部分。

审稿专家的意见使我们在修改过程中受益匪浅，提高了对研究的认识深度。

第二轮

审稿人 1 意见: 论文已经按照审稿意见作了较为细致的修改，基本达到了发表的水平，但仍有一些细小的环节需要修改:

意见 1: 在“引言”部分，作者不是 Andersen 而是 Anderson；教育目标分类学 2001 版是一

个修订团队完成的，不要表述成“Anderson 修订了教育目标分类”，应是“Anderson 等”。

回应：非常感谢审稿专家严谨细致地及时纠正，作者“Anderson”的拼写不当是我们的错误，且仅用 Anderson 一人概括其整个团队的贡献也是欠妥的。现已对全文错误之处作以更正。（绿色字体标注）。

意见 2：在“2.2 材料”部分，对“五类问题”的设置叙述需要再清晰一点，特别是要明确 10 个题目与五类问题的对应关系。

回应：问题设置和对应关系表述不够明确是我们的不足。问题的设置至关重要，是区分概念问题与高阶问题的标尺，感谢审稿专家对行文表述的严格把关。已在“2.2 材料”部分使用绿色字体做出修改。

意见 3：在“总体讨论”部分，最好不要把图 9 放上去，原著没有这样的表述，而且把“记忆”和“理解”放在一个层级也不太合适。

回应：感谢审稿专家的宝贵建议，把“记忆”和“理解”放在一个层级是我们的错误。根据您的建议，文章删除了图 9，并修改了对应的文字表述（详见 4.3 绿色字体部分）。

意见 4：英文摘要的翻译不太专业，建议重新细致准确地翻译一下。

回应：非常感谢审稿专家的提示。仔细阅读《心理学报》英文摘要写作注意事项之后，我们对英文摘要从四个方面（**Introduction, Method, Results, Conclusion**）重新做了修改，并按照要求规范了表达、语法和字数，同时约请母语为英语的专家进行润色修改，修改请见文章最后的英文摘要部分。

审稿人 2 意见：作者针对审稿意见对全文进行了修改，质量明显有提升。但文章中还是一些存在明显的不足：

意见 1：3.5.2 中“进行 4（学习策略） \times 2（高阶维度）的多元方差分析——”这一表述与实验设计中的表述不同。从方差分析结果来看，这一表述明显是错误的。

回应：非常感谢审稿专家的及时指正，帮助我们纠正统计表述的低级错误。按照实验设计的思路，自变量“学习策略”涉及两个水平（提取练习和重复学习）；而自变量“高阶维度”包括四个水平（应用、分析、评估、创造）。正确表述应为“进行 2（学习策略） \times 4（高阶

维度)的多元方差分析...”,为避免类似错误,我们逐条检查了实验设计和结果部分的表述,错误的地方已用绿色字体更正。

意见 2: 图 9 的“教育分类目标分类”表述正确吗?

回应: 这个表述是错误的,非常感谢审稿专家的细心审阅。图 9 为教育目标分类层级图,在参考另一位审稿专家意见后,在正文删除了图 9,并修改了对应的文字表述(详见 4.3 绿色字体部分)。

意见 3: 请作者再解释结论 1 的依据与推理:“提取练习能够有效增强不同内容间的区分度,不仅提高之前所学内容的记忆,更促进了之后即将要学习内容的记忆效果,使得后续学习中的编码如同最开始学习时一样高效。”

回应: 非常感谢审稿专家严谨周密的逻辑推理,结论 1 确实存在过度解释和过度概括化的问题。我们重新对本研究的相关结果进行了梳理和归纳,对结论 1 进行了修改(详见 5 结论绿色字体部分)。

第三轮

审稿人意见: 建议修后采用。

意见 1: 中文摘要中“结果证实了提取练习是高效学习的有效策略,不仅体现在单个内容的记忆性学习,也可扩展到连续学习中高阶技能的发展与促进”中的“不仅体现在单个内容的记忆性学习”表述不准确,什么叫做“记忆性学习”?

回应: 非常感谢审稿专家对语言表述准确性的周严把关。“记忆性学习”的说法不符合学术表达规范,已对表达不恰当之处做了修改,具体修改请见中文摘要紫色字体部分。

意见 2: 引言第二段落的“新修订的分类学包含六个认知过程维度: 记忆(remember)主要涉及“保持”; 理解(understand)、应用(apply)、分析(analyze)、评价(evaluate)和创造(create),这五个维度主要涉及“迁移”,该表述不符合现代汉语的表达习惯。建议修改为“新修订的分类学包含六个认知过程维度: 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; 其余五个维度主要涉及迁移。”

回应：非常感谢审稿专家给出的具体修改示范。根据意见，已修改了对应表述。具体修改请见正文引言紫色字体部分。

意见 3：英文摘要中存在多处语法与表达方面的错误，需要专业人士认真修改。

回应：非常感谢审稿专家的建议。为使英文表述更地道规范，文章的英文摘要委托机构进行了再次修改，如编辑部需要，我们会提供编校证明。修改请见文章最后的英文摘要紫色字体部分。