

记忆提取练习对学生学习的促进效应

詹 静

福建师范大学心理学院, 福建 福州

收稿日期: 2024年6月12日; 录用日期: 2024年7月31日; 发布日期: 2024年8月9日

摘 要

提取练习是指在学习过程中对新知识进行记忆提取或测试的方法。研究发现, 提取练习可以增强长时记忆和知识迁移。与简单的重复学习相比, 提取练习能够产生更持久的记忆保持和更好的知识迁移效果。如果一个人想要能够在未来对新知识进行提取, 提取练习比重复学习更有效。文章总结了测试效应在各种材料类型、年龄、类型和不同个体的系列实验中的表现并考虑了未来研究的方向。

关键词

测试效应, 提取练习效应, 学生学习

The Promoting Effect of Retrieval Practice Effect on Students' Learning

Jing Zhan

College of Psychology, Fuzhou Normal University, Fuzhou Fujian

Received: Jun. 12th, 2024; accepted: Jul. 31st, 2024; published: Aug. 9th, 2024

Abstract

Extraction practice refers to the method of remembering or testing new knowledge during the learning process. Research has found that extraction practice can enhance long-term memory and knowledge transfer. Compared to simple repetitive learning, extraction practice can produce more persistent memory retention and better knowledge transfer effects. If someone wants to be able to extract new knowledge in the future, extraction practice is more effective than repetitive learning. This article summarizes the performance of the test effect in a series of experiments on various material types, ages, types, and different individuals and considers the direction of future research.

Keywords

Test Effect, Retrieval Practice Effect, Student Learning

Copyright © 2024 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

课本上的新知识巩固应该如何巩固? 大多数人接受和认可的方法是做记号, 反复阅读或者做一些笔记或提纲[1]。但是, 如果最终的目标是能够从记忆中检索那个知识, 对新信息的提取是更好的学习方式[2]-[4]。提取练习是巩固新知识最有效的方法之一, 但是这一事实被大多数学习者和老师低估了[5]。尽管在实验之前, 人们就已经观察到提取练习会对增强长时记忆, 例如, 弗朗西斯·培根爵士提出: “如果你把一篇文章读 20 遍不如读 10 遍的同时不断地背诵, 在提取失败的时查阅文本那样容易地记住它”。然而通常提取练习仅仅被看作是一种对新知识掌握程度的测量手段, 而不是对学习产生影响。在大学生群体中调查时发现大学生缺乏对测试比重复学习更有助于记忆的元认知知识, 大多数学生在学习时反复阅读笔记或课本而不是在学习时进行自我测试或提取练习[6]-[9]。

提取练习效应(retrieval practice effect)是指在一个等值的时间里, 对某个学习材料进行一次或多次的记忆提取(测试)要比重复学习这个材料在将来产生更持久的记忆保持和更优的知识迁移, 也称为测试效应(test effect) [2] [10] [11]。测试效应是一种稳定的现象, 在不同材料, 不同年龄, 不同测试类型中都表现出对长时记忆的促进作用[12] [13]。大多数关于提取学习的研究采用的实验程序可以被描述为有几个离散的阶段。首先, 向参与者提供某种类型的信息, 作为初始学习。在初始学习之后的某个时间, 会出现一个中间阶段, 信息可以重新呈现以进行进一步学习(再学习条件), 或者被试接受记忆测试(测试条件), 提供或不提供测试答案。在一段记忆间隔之后, 对所学的信息进行最终的记忆评估。

事实上, 提取学习可以直接或间接地提高学习和长时记忆成绩[3]。提取学习的直接效应是指提取这一行为本身所产生的影响。提取练习可以将学习者的注意力吸引到他们所不知道的内容上, 增强对知识的元认知理解。提取学习的间接效应是指提取对提取后再编码活动的影响。当学习者未来对知识进行复习时, 先前的测试可以使学习者从重新编码中受益更多, 这种效果被称为测试增强学习[11], 甚至学习更多的新信息, 即测试对新学习的强化[14]-[16]。自从 2006 年 Karpicke 和 Roediger 在 Science 上发表了提取学习比简单的重复学习更加有利于长时记忆保持的报告以后[2], 各类围绕着测试效应的研究硕果累累, 本文将从材料, 年龄, 测试类型, 测试间隔及个体差异这五个方面综述了记忆提取练习对学习的促进效应的相关研究以供相关理论与应用研究者进行参考。

2. 不同实验材料中的测试效应

很多研究工作都关注了测试效应是否在不同的实验材料中稳定存在。Gates (1917)给一组孩子 9 分钟的时间来学习新材料, 无意义的音节或传记。他操纵了花费在背诵上的时间百分比(0%、20%、40%、60%, 80%, 其余时间用于阅读材料)。在记忆无意义音节时, 提取练习的时间的占比越大, 回忆效果越好。

此外, 在文章材料的学习中也发现了测试效应[17]。要求大学生阅读两篇科学短文。学生们学习第一篇文章 7 分钟, 然后再学习第二篇 7 分钟的文章; 对于第二篇文章, 参与者被要求学习 7 分钟, 然后花

7分钟试图记住它(即自由回忆测试)。最后对两篇文章进行自由回忆测试,测试间隔从5分钟到2天或1周不等。当学习5分钟后进行最后的测试时,学习两次比学习一次后再进行第一次测试(没有反馈)更有益。当学习后2天或1周进行期末测试时,出现了相反的模式:单一学习阶段(初始测试条件)与提取练习的结合的学生表现优于两个学习阶段(再学习条件)。实验结果证明了提取练习在促进长期记忆的保持方面显著而有力的作用。

在对非语言材料的学习中同样发现了测试效应的存在[18]-[20]。Carpenter探讨了测试是否可以增强地图中视觉空间信息的学习。50名受试者每人学习两张地图,一份通过传统学习,另一份通过计算机提示测试。在测试中,受试者反复看到同一张地图,但删除了一个特征(例如,一条路或一条河),他们试图秘密地回忆起缺失的特征及其位置。30分钟后的测试成绩表明,与同样时间的常规学习相比,测试学习的被试地图绘制效果明显更好。这些结果表明,测试的效果并不局限于语言反应的记忆类型,而且测试效应会扩展到各种复杂的非语言学习任务中。

同样的在叙事动画中存在测试效应[21]。在Johnson的研究中,282名参与者观看了关于闪电形成的叙述动画,然后再次观看演示(重新学习),进行练习-保留测试(练习-保留),或进行练习-转移测试(练习-转移)。研究者发现多媒体材料中出现测试效应,例如练习记忆组在延迟记忆测试中的表现优于再学习组。在进行练习转移测试时发现了测试效果,例如练习转移组在延迟转移测试中表现优于再学习组。

正如Dunlosky [22]回顾的那样,提取实践的有益效果在各种材料中得到了证明。同时一些研究人员试图解决材料的类型是否影响效益的大小的问题。例如,Gates提出背诵对无意义的音节记忆效果好于传记。因为与不同材料相关的缩放效应以及随之而来的结果测量的差异,这个问题还未有定论。但到目前为止,还没有材料已确定检索练习不能帮助记忆(在没有地板或天花板效应或其他混淆问题)。罗兰[23]的元分析表明,相对于再学习而言,提取练习在散文和词汇配对比单个单词或其他材料的效果更大。

3. 不同年龄被试中的测试效应

与大多数认知心理学领域一样,关于测试效果的文献多依赖于大学生作为实验参与者[13] [24] [25]。Thomas [4]的实验结果表明,当加入了测验(带反馈)时,该材料的考试表现(相对于非测验材料)得到了提高,如果学生被问及事实性问题,他们在应用问题中的表现会有所提高。同样,如果学生被问及应用问题,他们在事实性考试问题上的表现也会有所提高。即使问题类型从测验改为考试,这种好处也会迁移[26]。Lyle [27]在心理学统计学课程的每节课结束时,要求学生们回答一组问题,这些问题要求他们从当天的讲座中提取信息。这种方法叫PUREMEM,是一种增强对重要材料记忆的手段。在使用PUREMEM方法教授的部分课程中,考试成绩明显高于不使用PUREMEM方法教授的学生。

这种效应在小学生[28] [29]中也有体现,检索练习帮助三年级学生学习,有助于四年级学生对单词列表的记忆[29],并提高了四年级和五年级学生对地图位置的记忆[20]。对儿童自传体记忆和目击者记忆的研究中,儿童被要求对父母或调查人员的询问做出回应,以提取记忆。例如,与最初没有接受采访的儿童相比,接受采访的儿童对课堂活动的记忆更好[30]。小学生[12]、学前班学生[13]和婴儿都能从提取练习中受益。

初中生[30]和高中生[31]也很容易发现。Roediger对郊区一所中学的学生做了测试,测验是否能促进社会研究课程的学习和记忆。学生在章节考试和学期考试中的表现在测试后相对于未测试或相对于两次学习的项目都有所提高。该研究结果进一步证明,课堂上的提取练习可以提高学习成绩。McDermott研究了简答题和多项选择题的课堂测试是否能提高高中学生对某个年级课堂考试信息的记忆。对一些信息进行了间歇性的测试(简答或多项选择,都有正确答案的反馈),而其他信息最初没有测试,但在所有其他课堂活动中得到了同等的覆盖。在单元考试和期末考试中,学生们在测试过的信息上比没有测试过的信

息表现得更好。与有反馈的测验相比,重新学习条件对后期测试表现的增强作用更小。经常在课堂上进行反馈测试,可以提高学生的学习能力和记忆力。老年人也可以从检索练习中获益[32]-[34]。这项研究对60名18~25岁的年轻大学生、60名18~25岁的年轻社区成年人和60名55~65岁的中年至老年社区成年人进行了即时和延迟2天的测试。所有群体都从两次测试中受益,这意味着测试可以成为各种学习者的有益终身学习工具。

4. 不同测试形式中的测试效应

测试的形式会影响观察测试效果的大小。自由回忆是一种比识别更有效的提取练习形式。在识别测试中,学习者只需识别答案是否已经被学习过,与识别测试相比,自由回忆时学习者需要承担更多的认知负担[18]。报告证明自由回忆是比识别对记忆保持有更大的促进效应。然而,在没有提供测试反馈的条件下,自由回忆测试的正确率低于识别测试的正确率[29]。如果测试效果由回忆的准确率衡量,识别将是一个更有效的记忆保持方法。一种解释认为越多的检索努力对学习越有帮助,但如果测试太难,导致提取尝试不成功,这样增加的检索努力对长时记忆的保持就没那么有益了(至少在没有正确答案反馈的情况下)。Karpicke认为只有当学习者在提取测试中相当成功时,才能看到提取测试(没有反馈时)的好处[35]。事实上,Rowland的元分析表明,最大的测试效应(相对于重新学习)是在初始测试表现大于75%且没有反馈时看到的;初始测试表现为51%~75%时,为中间效益;当初始测试表现小于50%时,初始测试的效果不比再次学习好。在最初的测试后给予反馈可以促进测试对长时记忆的保持,解决初始测试表现不佳的问题。

最后测试的格式对测试效果产生影响[36]。当初始测试为自由回忆,最终测试也为自由回忆时,测试效果显著;然而,当最后的测试是识别时,有时显现测试效应[1][13][17],但有时没有出现[5][7]。然而,如果将识别测试的组成部分分解为回忆和熟悉度[37],在回忆成分中有测试效果的证据[38]。总的来说,回忆测试(特别是自由回忆测试)显示出更可靠、更强的益处[23]。

在多项选择题、简答题和作文测试中同样出现了初始测试的明显作用,但哪种形式更有效取决于初始测试时是否出现正确答案反馈(有时是在初始测试和最终测试的格式上的匹配)。有研究表明如果在初始测试后提供正确答案,简答题是更有效的测试形式,无论最终测试是采用相同的简答题还是多项选择题[36]。从本质上讲,提供正确答案有助于弥补简答题相对于多项选择题的较差的测试表现。

有研究检验了在实验室中观察到的效果在真实的教室条件下是否成立。当学习者进行多项选择题测试时,错误的答案会导致困惑(特别是在没有立即反馈的情况下)[30]。在完成初始的多项选择题测试后,学生在进行最终的简答题测试时,往往会给出他们之前选择的错误答案[38][39]。在一项研究中,学生回答SAT II考试的问题(生物、化学、历史),总体上看,多项选择题对长时记忆是有益的(即出现了测试效应);但是,先前的多项选择题测试也提高了最后的简答题测试中错误答案出现的概率[38]。简而言之,总的来说,多项选择测试效果是积极的,但当多项选择测试不包括正确答案反馈时,学生可能会选择错误的选项并学习那些错误的答案。

5. 不同测试间隔中的测试效应

研究表明测试间隔是测试效果是否会发生的一个因素。在Roediger和Karpicke 2006年的实验中,当最终标准测试前的保留间隔较短时,重复学习比先前测试更有效[4]。但是也有其他研究表明,即使在很短的间隔时间内,测试也有效果;例如,Runquist发现,之前的测试仅在10分钟后就有了益处,尽管测试比重复学习的效果在1周后要大多得多。测试的好处会随着时间的推移而增加[15],这种模式也可以在元分析中看到[17],这也表明提取练习的效果通常也在较短的时间间隔内存在。

重复学习在早期显示出卓越的效益，但在延迟后被重复测试超越，这是研究中一个引人注目的研究主题。Roediger 和 Karpicke (2006)在两个独立的实验中展示了这种交互作用，并对这种表现模式的显著变化做了很多研究。有研究者认为，尝试确定模式逆转发生时的拐点或延迟可能很有趣。然而，Rowland 和 DeLosh [33]证明，初始测试的成功程度是决定提取练习的优势是否会在短时间内出现的重要因素。当第一次测试的成绩很低时，重新学习的条件下的被试会有不公平的优势，比如重复学习条件下 100% 的项目被重新学习，但在测试条件下只有一小部分项目在第一次测试中被成功回忆起来。因此，信息暴露次数的混淆掩盖了对项目的成功提取带来的益处[12] [13]。基于测试效应通常以较短的间隔出现的发现[23]，这支持了 Karpicke 的观点，即寻找拐点不会提供额外信息。

6. 不同个体间的测试效应

个体差异是否会影响测试效果的大小或存在是最近研究者关注的热点问题[3] [40] [41]。这个问题是在尝试解答检索练习是否对每个人都有益，是否对某些人更有益。显然，这些都是重要的实际问题，而检索练习从整体上看对长时记忆有帮助并不意味着它使每个人受益。有研究表明长期记忆存在个体差异[33]，但这种差异是否延伸到测试效果还有待观察。

Brewer 的研究考察了从长期记忆中检索效果的个体差异(即测试效果)。观察到测试效应的大小有很大的个体可变性。记忆能力差和一般流体智力较低的学生比能力强的学生从记忆测试中受益更多。Minear 的研究进一步证实个体的流体智力对测试效应的调节作用。参与者学习斯瓦希里语 - 英语单词对，重复学习其中一半的单词对，对另一半单词尝试进行回忆，测试时提供反馈。单词对有难有易。总体而言，参与者表现出了测试优于重复学习的优势。然而，结果显示近 1/3 的样本表现出负面的测试效果，学习者从重新学习中受益更多，重复学习的整体表现更好。与从测试中获得更多益处的参与者相比，这些个体自我报告使用较少或较浅的编码策略，但在其他方面没有差异。对于具有积极测试效应的个体，难度对流体智力测量得分高或低的参与者有不同的影响，流体智力高的参与者对困难的测试影响大于简单的项目，而流体智力低的参与者则相反。词汇知识显示了与流体智力相似的模式，词汇量越高，对困难的项目的测试效果越好。这表明提取练习的益处因项目难度和参与者能力的不同而不同。因此，进行提取练习应考虑到学习材料的交互作用和学习者的个体差异。

7. 理论解释

一个被应用于理解提取练习效果的框架是迁移恰当加工理论[18]。基本的主张是，学习活动本身既不是好的，也不是坏的，而是需要在最终的测试环境中被理解。例如，深度或有意义的处理通常是比浅层处理更有效的编码策略(例如，关注编码单词的含义通常比关语音更有效)，但迁移恰当加工理论发现：如果最终的标准测试需要对单词的声音进行记忆，则具有关于语音的编码单词是更好的编码策略。

这个理论最重要的思想是提取练习测试之所以促进记忆成绩是因为在学习或编码阶段的认知加工与提取阶段的认知加工相匹配。更具体地说，提取练习的类型之间的匹配(例如，在简答题或多项选择题测试上回答问题)可能是检索练习效果大小的重要预测因素。然而，如上所述，测试类型中的简单匹配并不总是最理想的情况。相反，有时简答题测试可以更好地帮助学生为以后的多项选择题测试做准备。因此，迁移恰当加工理论在理论上已经取得突破，但尚缺乏足够的实证研究进行证明。

另一个理论是必要难度理论，认为提取努力增强记忆，一个人付出的提取努力越多，为以后的提取做的准备就越好[37]。因此，提取工作被认为是“理想的困难”。例如很多研究比较产生式测试形式(如简答、线索回忆及填空等)与再认测试(如多选)记忆结果来进行验证。研究结论一致认为，产生式测试记忆成绩显著高于再认测试成绩，原因即产生式测试比再认测试需要更多的认知努力或深加工[8]。也许出

于这个原因，它是讨论提取效应的好处的较受欢迎的框架之一。然而，一个主要的挑战是，提取努力的程度不能独立于其对可提取性的影响。同样，事先也不清楚哪些困难是可取的，哪些是不可取的，以及这个问题的答案是否因人而异。

第三种观点是编码可变理论，认为练习提取往往涉及到相关概念的激活。相关概念的这种激活增加了后来召回时可能的检索路线的数量，从而提高了未来召回的可能性[12] [13]。当一个人试图检索(比方说)“胡萝卜”以响应提示“蔬菜”时，在得到正确答案之前，可能会在记忆中搜索可能的答案(“黄瓜”、“生菜”、“羽衣甘蓝”)。然后，在最后的标准测试中，学习者可能会从之前那些无效的检索尝试中受益，这可能有助于得出最终答案。这一逻辑中的一个谜团是，线索过载的原理[34]表明，这些看似合理的答案可能会干扰最终的提取尝试，而不是帮助它，而且还不完全清楚这些可能的答案将如何促进回忆。这一理论似乎也于语言材料最相关，目前还不清楚它能在多大程度上解释非语言提取实践。导致这一假说的效应大小并不是很稳定[8]，直接检验这一假说的实验并没有显示出对该理论的普遍性的支持[12] [13]。

第四个理论情景情境理论借鉴了检索涉及学习情景恢复的想法[12] [13]。这一理论假设(就像许多记忆模型一样)，时间背景是编码到记忆中的相关特征，并且这种背景随着时间的推移而缓慢变化。此外，当试图检索时，人们试图恢复先前的时间上下文以帮助检索搜索过程。因此，在最初的测试中，当人们记住来自较早语境(语境 A)的条目时，该记忆表示将与现有语境(语境 B)组合以形成复合词。最后，当稍后试图再次记忆时，复合上下文表示以某种方式允许对记忆搜索集进行限制，这有利于检索。该理论的局限在于独立定义其核心结构(例如，上下文)是困难或不可能的。此外，与检索单个先前上下文相比，组合上下文将如何简化或限制搜索集并不是完全清楚的。

解释测试效果的最后一个理论是二分模型[33]。该模型表明，人们对记忆材料具有不同记忆强度的原因取决于初始学习(重新学习或测试)期间记忆材料的认知加工深度。提取在初始测试中会有不同的分布，从深度认知处理中获得的信息会被成功提取，记忆强度会更大。相反，浅层认知加工的信息提取失败，记忆强度较小。提取信息的记忆强度远高于重复学习，尽管后者的记忆强度也增加。正是由于不同学习条件下记忆强度增长程度的差异，才使得再学习条件优于即时测试中的测试条件。然而，一段时间后，所有的记忆信息都以相同的频率被遗忘，并且与许多重新学习条件信息的记忆强度低于提取阈值相比，测试条件信息的记忆强度也减弱了，但仍保持在提取阈值之上；时间越长，由于遗忘而导致的分布变化就越大。这个理论也可以解释一些实验的结果，在即时测试中，重复学习条件的回忆分数优于测试条件的回忆分数，延迟测试中测试条件的回忆分数优于重复学习测试的回忆分数[8]。

8. 结论

本研究着重分析了提取练习对学习的促进效应，分别从测试材料，测试年龄，测试类型，测试间隔及个体差异这五个角度探讨了测试效应。综合以往的研究，本文首先总结出在所有的测试材料中都会表现出提取练习效应，讨论了无意义音节、传记、非语言材料、叙事动画等材料中的测试效应，发现提取练习的有益效果在各种材料中得到了证明。其次，归纳了从小学生、初中生、高中生、大学生乃至老年人当中的测试效应，发现所有群体都从提取练习中受益，这意味着测试可以成为各种学习者的终身学习工具。再次，探讨了不同测试类型当中的测试效应，发现回忆测试的总体效果好于识记测试，简答题的效果好于多项选择题。随后，本研究探索了不同测试间隔下的测试效应，当测试间隔足够短，重复学习的效果好于测试，但随着时间的推移，测试的优势逐渐扩大。最后总结了不同个体间测试效应大小的差异，发现测试效应的大小与个体的流体智力与词汇量等因素有关。

在过去的十几年里，关于提取练习如何以及何时促进后期提取的文献已经明显成熟。然而对这一现

象的理论解释缺乏精确度和预测能力。首先, 应关注提取练习是否为所有学习者促进长时记忆的保持, 也许其中一些学习者根本不会从测试中得到任何好处。其次, 应思考如何平衡提取工作与提取成功的问题。对于从提取练习中获益的学习者来说, 其目标是使提取工作变得困难, 但又不至于困难到使提取尝试失败。选定测试项目来达到这个最佳点是一个具有挑战性的问题, 即使这个目标在平均上完成了(对大多数学习者来说), 它也不一定适用于位于分布顶端或底部的学习者, 对于他们来说, 这些项目可能太容易或太难。最后, 一些基本效应背后的机制需要进行进一步研究。例如, 测试可以预防遗忘, 不像其他许多独立因素对遗忘率无影响(如意义) [41], 但为什么出现这种现象还未有合理解释。

参考文献

- [1] Abrams, D., Rutland, A. and Cameron, L. (2003) The Development of Subjective Group Dynamics: Children's Judgments of Normative and Deviant In-group and Out-Group Individuals. *Child Development*, **74**, 1840-1856. <https://doi.org/10.1046/j.1467-8624.2003.00641.x>
- [2] Roediger, H.L. and Karpicke, J.D. (2006) The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science*, **1**, 181-210. <https://doi.org/10.1111/j.1745-6916.2006.00012.x>
- [3] Minear, M., Coane, J.H., Boland, S.C., Cooney, L.H. and Albat, M. (2018) The Benefits of Retrieval Practice Depend on Item Difficulty and Intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **44**, 1474-1486. <https://doi.org/10.1037/xlm0000486>
- [4] Thomas, R.C., Weywadt, C.R., Anderson, J.L., Martinez-Papponi, B. and McDaniel, M.A. (2018) Testing Encourages Transfer between Factual and Application Questions in an Online Learning Environment. *Journal of Applied Research in Memory and Cognition*, **7**, 252-260. <https://doi.org/10.1016/j.jarmac.2018.03.007>
- [5] Karpicke, J.D. and Blunt, J.R. (2011) Retrieval Practice Produces More Learning than Elaborative Studying with Concept Mapping. *Science*, **331**, 772-775. <https://doi.org/10.1126/science.1199327>
- [6] Blasiman, R.N., Dunlosky, J. and Rawson, K.A. (2016) The What, How Much, and When of Study Strategies: Comparing Intended versus Actual Study Behaviour. *Memory*, **25**, 784-792. <https://doi.org/10.1080/09658211.2016.1221974>
- [7] Karpicke, J.D., Butler, A.C. and Roediger III, H.L. (2009) Metacognitive Strategies in Student Learning: Do Students Practise Retrieval When They Study on Their Own? *Memory*, **17**, 471-479. <https://doi.org/10.1080/09658210802647009>
- [8] Kornell, N. and Bjork, R.A. (2007) The Promise and Perils of Self-Regulated Study. *Psychonomic Bulletin & Review*, **14**, 219-224. <https://doi.org/10.3758/bf03194055>
- [9] McCabe, J. (2010) Metacognitive Awareness of Learning Strategies in Undergraduates. *Memory & Cognition*, **39**, 462-476. <https://doi.org/10.3758/s13421-010-0035-2>
- [10] Adesope, O.O., Trevisan, D.A. and Sundararajan, N. (2017) Rethinking the Use of Tests: A Meta-Analysis of Practice Testing. *Review of Educational Research*, **87**, 659-701. <https://doi.org/10.3102/0034654316689306>
- [11] Arnold, K.M. and McDermott, K.B. (2013) Free Recall Enhances Subsequent Learning. *Psychonomic Bulletin & Review*, **20**, 507-513. <https://doi.org/10.3758/s13423-012-0370-3>
- [12] Carpenter, S.K., Pashler, H., Wixted, J.T. and Vul, E. (2008) The Effects of Tests on Learning and Forgetting. *Memory & Cognition*, **36**, 438-448. <https://doi.org/10.3758/mc.36.2.438>
- [13] 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**, 1563-1569. <https://doi.org/10.1037/a0017021>
- [14] Pastötter, B. and Bäuml, K.T. (2014) Retrieval Practice Enhances New Learning: The Forward Effect of Testing. *Frontiers in Psychology*, **5**, Article 286. <https://doi.org/10.3389/fpsyg.2014.00286>
- [15] Szpunar, K.K., McDermott, K.B. and Roediger, H.L. (2008) Testing during Study Insulates against the Buildup of Proactive Interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **34**, 1392-1399. <https://doi.org/10.1037/a0013082>
- [16] Chan, J.C.K., Meissner, C.A. and Davis, S.D. (2018) Retrieval Potentiates New Learning: A Theoretical and Meta-Analytic Review. *Psychological Bulletin*, **144**, 1111-1146. <https://doi.org/10.1037/bul0000166>
- [17] Roediger, H.L. and Karpicke, J.D. (2006) Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, **17**, 249-255. <https://doi.org/10.1111/j.1467-9280.2006.01693.x>
- [18] Carpenter, S.K. and Delosh, E.L. (2006) Impoverished Cue Support Enhances Subsequent Retention: Support for the

- Elaborative Retrieval Explanation of the Testing Effect. *Memory & Cognition*, **34**, 268-276. <https://doi.org/10.3758/bf03193405>
- [19] Kang, S.H.K. (2010) Enhancing Visuospatial Learning: The Benefit of Retrieval Practice. *Memory & Cognition*, **38**, 1009-1017. <https://doi.org/10.3758/mc.38.8.1009>
- [20] Rohrer, D., Taylor, K. and Sholar, B. (2010) Tests Enhance the Transfer of Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **36**, 233-239. <https://doi.org/10.1037/a0017678>
- [21] Johnson, C.I. and Mayer, R.E. (2009) A Testing Effect with Multimedia Learning. *Journal of Educational Psychology*, **101**, 621-629. <https://doi.org/10.1037/a0015183>
- [22] Dunlosky, J., Rawson, K.A., Marsh, E.J., Nathan, M.J. and Willingham, D.T. (2013) Improving Students' Learning with Effective Learning Techniques: Promising Directions from Cognitive and Educational Psychology. *Psychological Science in the Public Interest*, **14**, 4-58. <https://doi.org/10.1177/1529100612453266>
- [23] Rowland, C.A. (2014) The Effect of Testing versus Restudy on Retention: A Meta-Analytic Review of the Testing Effect. *Psychological Bulletin*, **140**, 1432-1463. <https://doi.org/10.1037/a0037559>
- [24] Butler, A.C. (2010) Repeated Testing Produces Superior Transfer of Learning Relative to Repeated Studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **36**, 1118-1133. <https://doi.org/10.1037/a0019902>
- [25] Baayen, R.H., Davidson, D.J. and Bates, D.M. (2008) Mixed-Effects Modeling with Crossed Random Effects for Subjects and Items. *Journal of Memory and Language*, **59**, 390-412. <https://doi.org/10.1016/j.jml.2007.12.005>
- [26] Foss, D.J. and Pirozzolo, J.W. (2017) Four Semesters Investigating Frequency of Testing, the Testing Effect, and Transfer of Training. *Journal of Educational Psychology*, **109**, 1067-1083. <https://doi.org/10.1037/edu0000197>
- [27] Lyle, K.B. and Crawford, N.A. (2011) Retrieving Essential Material at the End of Lectures Improves Performance on Statistics Exams. *Teaching of Psychology*, **38**, 94-97. <https://doi.org/10.1177/0098628311401587>
- [28] Fazio, L.K. and Marsh, E.J. (2019) Retrieval-Based Learning in Children. *Current Directions in Psychological Science*, **28**, 111-116. <https://doi.org/10.1177/0963721418806673>
- [29] Karpicke, J.D., Blunt, J.R. and Smith, M.A. (2016) Retrieval-Based Learning: Positive Effects of Retrieval Practice in Elementary School Children. *Frontiers in Psychology*, **7**, Article 350. <https://doi.org/10.3389/fpsyg.2016.00350>
- [30] Roediger, H.L., Agarwal, P.K., McDaniel, M.A. and McDermott, K.B. (2011) Test-Enhanced Learning in the Classroom: Long-Term Improvements from Quizzing. *Journal of Experimental Psychology: Applied*, **17**, 382-395. <https://doi.org/10.1037/a0026252>
- [31] McDermott, K.B., Agarwal, P.K., D'Antonio, L., Roediger, H.L. and McDaniel, M.A. (2014) Both Multiple-Choice and Short-Answer Quizzes Enhance Later Exam Performance in Middle and High School Classes. *Journal of Experimental Psychology: Applied*, **20**, 3-21. <https://doi.org/10.1037/xap0000004>
- [32] Meyer, A.N.D. and Logan, J.M. (2013) Taking the Testing Effect Beyond the College Freshman: Benefits for Lifelong Learning. *Psychology and Aging*, **28**, 142-147. <https://doi.org/10.1037/a0030890>
- [33] Rogalski, Y., Altmann, L.J.P. and Rosenbek, J.C. (2014) Retrieval Practice and Testing Improve Memory in Older Adults. *Aphasiology*, **28**, 381-400. <https://doi.org/10.1080/02687038.2013.870965>
- [34] Tse, C., Balota, D.A. and Roediger, H.L. (2010) The Benefits and Costs of Repeated Testing on the Learning of Face-name Pairs in Healthy Older Adults. *Psychology and Aging*, **25**, 833-845. <https://doi.org/10.1037/a0019933>
- [35] Karpicke, J.D., Blunt, J.R., Smith, M.A. and Karpicke, S.S. (2014) Retrieval-based Learning: The Need for Guided Retrieval in Elementary School Children. *Journal of Applied Research in Memory and Cognition*, **3**, 198-206. <https://doi.org/10.1037/h0101802>
- [36] Kang, S.H.K., McDaniel, M.A. and Pashler, H. (2011) Effects of Testing on Learning of Functions. *Psychonomic Bulletin & Review*, **18**, 998-1005. <https://doi.org/10.3758/s13423-011-0113-x>
- [37] Yonelinas, A.P. (2002) The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, **46**, 441-517. <https://doi.org/10.1006/jmla.2002.2864>
- [38] Marsh, E.J., Agarwal, P.K. and Roediger, H.L. (2009) Memorial Consequences of Answering SAT II Questions. *Journal of Experimental Psychology: Applied*, **15**, 1-11. <https://doi.org/10.1037/a0014721>
- [39] Roediger, H.L. and Marsh, E.J. (2005) The Positive and Negative Consequences of Multiple-Choice Testing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **31**, 1155-1159. <https://doi.org/10.1037/0278-7393.31.5.1155>
- [40] Brewer, G.A. and Unsworth, N. (2012) Individual Differences in the Effects of Retrieval from Long-Term Memory. *Journal of Memory and Language*, **66**, 407-415. <https://doi.org/10.1016/j.jml.2011.12.009>
- [41] Robey, A. (2019) The Benefits of Testing: Individual Differences Based on Student Factors. *Journal of Memory and Language*, **108**, Article ID: 104029. <https://doi.org/10.1016/j.jml.2019.104029>