

线索强度对测试效应的调节: 时间间隔上的差异*

郑俊 张玮 罗良

(北京师范大学认知神经科学与学习国家重点实验室 北京 100875)

摘要: 本研究采用不同线索强度(强/弱)的中文词对,选取大学生被试 126 人,考察了不同学习方式(重学/测试)在不同时间间隔(5 分钟/2 天)最终自由回忆测试成绩上的差异。结果发现:在强线索强度词对上,学习方式主效应显著,测试条件的回忆成绩在 5 分钟以及 2 天后均高于重学条件;而在弱线索强度词对上,学习方式与间隔时间交互作用显著,5 分钟时重学条件与测试条件的记忆成绩差异不显著,而 2 天后测试条件的记忆成绩却显著高于重学条件。结果表明词对的线索强度能够调节测试对不同时间间隔记忆保持的促进作用,支持了精细提取假设。

关键词: 线索强度; 测试效应; 长时记忆; 精细提取假设; 概念组织

1 问题提出

人们常常认为测试(Testing)仅仅是评估所学知识的一种测量手段,然而越来越多的研究表明,测试不仅是一种评估知识的测量手段,更是能够显著促进知识长时保持的一种学习方式(Roediger & Butler, 2011)。研究者将这种与简单重复学习、过度学习相比,测试更能促进习得知识长时保持的现象称为测试效应(Testing Effect),也叫提取练习效应(Retrieval Practice Effect)(Karpicke & Roediger, 2008; Roediger & Karpicke, 2006)。大量研究证实了测试效应在不同的材料类型,如外语词对(Karpicke & Roediger, 2008; Pyc & Rawson, 2007)、短文(Agarwal, Karpicke, Kang, Roediger, & McDermott, 2008; Roediger & Karpicke, 2006)、单词列表(Carpenter & DeLosh, 2006),以及不同的测试形式,如线索回忆(Pyc & Rawson, 2009)、自由回忆(Butler, Karpicke, & Roediger III, 2007)等中稳定存在。

测试为什么能够促进学习内容的长时保持? 越来越多的研究开始致力于探讨测试效应产生的内在机制(Carpenter, 2009, 2011; Glover, 1989; Pyc & Rawson, 2010)。Bjork (1975) 将测试效应归因于测试可以使潜在的有效提取线索数量增多,他认为测试行为可以创造更多的提取路径,使得测试项目相

对于重学项目更有可能被记住。Glover (1989) 让被试学习文章概念,采用再认、线索回忆和自由回忆三种测试形式对所学内容进行记忆提取练习,结果发现无论最终测试形式如何,学习过程中进行自由回忆测试对记忆保持的促进作用优于线索回忆测试,而线索回忆测试对记忆保持的促进作用又优于再认测试。他认为学习过程中那些包括了更少提取线索的测试(如自由回忆和线索回忆)比包括了更多线索的测试(再认测试)能够唤起更精细的提取操作,从而更有助于记忆保持。Carpenter 和 DeLosh (2006) 认为测试时目标信息不可得,学习者必须搜索记忆以提取出正确答案,在记忆搜索的过程中可能形成了精细化的记忆痕迹,即在线索、目标及其相关信息之间建立了更广泛的联结。因此,在前人研究基础上,Carpenter (2009) 提出了精细提取假设(Elaborative Retrieval Hypothesis),认为从记忆中提取信息导致了记忆痕迹的精细化加工,或者产生了额外的检索通达路径,使得将来成功提取出目标信息的可能性增大。例如,根据线索(如,面包)回忆目标词(如,篮球)将比简单重学线索目标词对(如,面包——篮球)更有可能激活与线索相关的中间信息或概念(如,鸡蛋、圆形),从而建立精细化的语义概念网络(如,面包—鸡蛋—圆形—篮球),这种精细化加工模式有利于对信息的长时保持(刘兆敏,罗良,张玮,2011)。

* 基金项目: 国家自然科学基金资助项目(31100817, 31000506); 教育部人文社会科学研究青年基金项目(10YJCLX032)。

通讯作者: 罗良, E-mail: luoliang@bnu.edu.cn

Carpenter (2009) 的实验中,被试学习不同线索强度的英语词对,并在初始学习后重学其中一半的词对,对另一半词对进行线索回忆测试。5分钟后,最终测试要求被试尽可能多地自由回忆出所学目标词。结果发现被测试项目的最终记忆成绩显著高于重学项目,即出现了测试效应。进一步对测试项目分析发现,初始测试时强线索词对的提取成绩优于弱线索词对;而5分钟后的最终测试上,弱线索词对的记忆成绩优于强线索词对。因此,强线索导致了回忆的初始优势,但从弱线索中回忆的项目实际上随时间保持更好,强线索的优势作用在最终测试时消失或反转。因此,线索强度对测试效应起着调节作用,支持了精细提取假设——学习过程中进行测试时,弱线索强度词对比强线索强度词对激活的精细信息更多,激活的这些精细信息对学习知识的长时保持起到促进作用。

精细提取假设不仅得到了大量行为实验的验证(Kang, 2010; Karpicke & Blunt, 2011),同时也得到了认知神经科学研究的支持(van den Broek, Takashima, Segers, Fernandez, & Verhoeven, 2013; Wing, Marsh, & Cabeza, 2013)。例如,Wing等人(2013)采用fMRI对测试效应的脑活动生理基础进行了考察,结果发现测试效应与前海马区域(Anterior Hippocampus)、外侧颞叶皮层(Lateral Temporal Cortices)、内侧前额叶皮层(Medial Prefrontal Cortices)等相关,并且测试促进了海马与腹外侧前额叶皮层(Ventrolateral Prefrontal Cortices)和中脑区域(Midline Regions)之间的关联。其中,腹外侧前额叶皮层与基于知识表征的提取相关(Devlin, Matthews, & Rushworth, 2003),该部位的损伤导致语义任务水平的下降(Martin & Cheng, 2006);前额叶皮层负责控制语义加工,而海马则将相关信息整合成为更加稳固的记忆表征(Wing et al., 2013)。因此,海马与腹外侧前额叶皮层及中脑区域之间的这种关联反映的是海马调节的精细化效应(Wing et al., 2013)。

然而,精细提取假设试图考察提取中存在的独特的机制或加工过程,解释了为什么测试比重学更有效,但却无法解释测试效应中另一个非常重要的现象——学习条件与间隔时间的交互作用。多数同时考察短时间间隔和长时间间隔的研究都出现了这种交互作用(Congleton & Rajaram, 2011, 2012; Roediger & Karpicke, 2006; Wheeler, Ewers, & Buonanno, 2003),即在短间隔时间时,重学条件的

记忆成绩将优于测试条件,而延长间隔时间后,测试条件的记忆成绩将优于重学条件。例如,Roediger和Karpicke(2006)要求被试分别在5分钟、2天或一周后完成短文的最终记忆测试。结果发现,当最终测试发生在5分钟后,重复学习产生的记忆保持成绩显著高于重复测试;而当延长最终测试的间隔时间时,结果模式恰好相反,重复测试对记忆保持的促进作用将显著优于重复学习。那么,为什么短时间间隔内重学条件表现更好,而提取的优势必须要在间隔一段时间后才能突显出来呢?

Kornell, Bjork和Garcia(2011)提出的二分分布模型(bifurcated model)认为,项目的记忆强度通常是正态分布的,强度越高的项目越容易提取。通过重复学习,学习项目的记忆强度仍是连续的正态分布,所有项目的记忆强度得到了不同程度的增强;而通过提取练习,学习项目的记忆强度不再是一个连续的正态分布:被成功提取的项目记忆强度明显增强,而提取失败的项目记忆强度反而有所减弱。因此,短时间间隔下,重学条件处于阈限以上的项目更多,表现出了重学条件的优势;长时间间隔下,重学条件与测试条件均发生了同等程度的遗忘,由于成功提取项目记忆强度较大,仍处于阈限以上,表现出了测试条件的优势。Congleton和Rajaram(2011, 2012)则进一步提出了概念组织(Conceptual Organization)的概念,认为提取时的稳固回忆伴随着学习项目之间持久而强烈的概念化组织,而重学时概念组织程度却细小而又微弱。他们认为,不仅是项目本身的记忆强度,更是项目之间的关联性加工对学习内容的长时保持至关重要。他们指出学习时存在两种加工方式:一种是特定项目加工(Item-Specific Processing),关注于单个项目不同于其他项目的特征,试图建立更独特、更丰富的记忆痕迹;一种是关联加工(Relational Processing),试图建立项目之间的关联,产生更强的概念化组织。

综上所述,在Carpenter(2009)的实验中,只考察了5分钟间隔下测试对记忆保持的促进作用,对于更长的时间间隔(如2天)并未进行探讨。而人类学习的真正目的在于促进知识更长时间的保持,并且在将来需要时能够成功提取出来。因此,测试的优势效应体现在对所学知识的长时间保持上,而不是更短时间。另外,Carpenter(2009)所采用的词对材料的线索词与目标词之间相关强度都较高,即使弱线索的强度也属于中等强度。那么,如果扩大强、弱线索强度之间的差异,测试是否产生同样的促

进作用也不得而知。另一方面,一些研究(Congleton & Rajaram, 2011, 2012; Roediger & Karpicke, 2006)验证了学习方式与时间间隔的交互作用,即测试条件尤其是对更长时间间隔的记忆保持有优越性,然而却缺乏不同关联强度项目的直接考查。因此,考察线索强度对不同时间间隔(例如5分钟/2天)上的测试效应的调节作用是否存在差异具有重要价值。

为回答上述问题,本研究准备采用高、低相关强度的中文词对作为学习材料,考察不同学习方式(重学/测试)在不同时间间隔(5分钟/2天)最终自由回忆测试成绩上的差异。

2 研究方法

2.1 被试

北京师范大学在校大学生129人,3名被试因没有参加最终测试而被删除,最终有效被试为126人,其中男生43人,女生83人,平均年龄为21.97($SD = 2.74$)岁,母语均为汉语。实验结束后所有被试均可获得一定报酬。

2.2 实验材料

词对材料选自Cai和Brysbaert(2010)的中文词频库。从该词库中,我们选取了200个词语(词频:46.29~0.03/百万)作为目标词,随后选取了与对应目标词具有较高语义关联的200个词语(词频:46.20~0.03/百万)作为强线索词,同时选取了与对应目标词没有太大关联的200个词语(词频:47.16~0.03/百万)作为弱线索词。来自北京数十所高校的不同专业大学生238人参与了词对相关性的评定,他们对线索词与目标词之间的相关程度进行4点评分(1代表没有关系,4代表关系非常密切),并最终建立了由120个线索强度不同的中文词对组成的材料库,所有词语没有重叠。本实验从词库中随机选取了40个词对,其中强、弱线索词对各20个。弱线索词对的平均相关程度为1.24,强线索词对的平均相关程度为3.50。

2.3 实验设计

实验采用2(线索强度:强线索、弱线索)*2(学习方式:重学、测试)*2(最终测试间隔时间:5分钟、2天)的混合设计。其中,线索强度为组内变量,学习方式和间隔时间为组间变量。重学条件的被试

在完成词对的初始学习后,对所有词对再学一遍;而测试条件被试则在完成所有词对的初始学习后进行线索回忆测试。最终测试发生在学习阶段结束后的5分钟或2天。

2.4 实验程序

实验程序采用E-prime 1.0软件编制,在计算机上呈现学习材料,被试个别施测。被试被告知他们将学习40个汉语词对,随后可能再次学习词对,也可能完成线索回忆测试,并将在一段时间后接受最终测试。所有词对呈现在屏幕中间,其中目标词字体加粗并加下划线,要求被试根据线索词来帮助他们记忆目标词。每个词对的呈现时间为10秒,词对的呈现顺序随机,词对之间呈现注视符号“+”,时间为1秒。完成所有词对的首次学习后,所有被试需要完成2分钟的简单计算任务。随后,重学组被试再次学习所有词对,而测试组被试需要根据给出的线索词回忆并口头报告出目标词,录音笔将记录下被试的反应。为了监测被试确实将注意力放在了词对的记忆上,要求被试在学习的同时按键评定词对的相关强度(4点评分,1表示没有关系,4表示关系非常密切)。

最终测试形式为自由回忆测试,要求被试尽可能努力地回忆并键盘输入目标词,每键入一个目标词后按“Enter”键继续。短间隔组被试在5分钟后即刻完成,而长间隔组被试两天后返回实验室完成最终测试。

学习阶段所有词对的线索词与目标词顺序进行了左右平衡,即一半词对的目标词在线索词的左边,一半词对的目标词在线索词的右边。而测试组初始测试时,线索词将呈现在屏幕中间,并在线索词下方呈现“?”指示被试进行线索回忆报告。

3 结果与分析

3.1 相关程度评估

四组被试在初始学习时对词对的相关性程度进行了4点评分,评定结果如表1所示。对相关性程度评分进行2*2(学习方式*线索强度)方差分析发现^①,只有线索强度主效应显著, $F(1,113) = 3623.06$, $p < 0.01$, $\eta^2 = 0.97$,被试对强线索词对的相关性程度评分显著高于弱线索词对。

^① 由于被试在重学条件下对所有词对进行了两次评分,因此分析时取两次评分的平均值。另外,有11名被试因为在学习阶段评分数据缺失较多而被剔除。

表 1 词对平均相关程度评分(4 点量表)

线索强度	测试条件	学习方式	
		重学条件	
		第一次评分	第二次评分
强线索	3.70 (0.31)	3.65 (0.31)	3.71 (0.35)
弱线索	1.16 (0.19)	1.28 (0.39)	1.28 (0.37)

3.2 初始测试

测试条件被试(64 人)在初始测试上的平均提取成绩为 24.19(总记忆项目为 40 个, $SD = 6.30$), 其中强线索词对的提取成绩($M = 16.32$, $SD = 2.59$)显著高于弱线索词对的提取成绩($M = 7.85$, $SD = 4.52$)。配对样本 t 检验结果为 $t(63) = 18.28$, $p < 0.01$, $Cohen's d = 2.30$ 。

3.3 最终测试

各条件下最终测试成绩的平均数和标准差见表 2。对最终自由回忆成绩进行三因素重复测量方差分析(线索强度为组内变量,学习方式和最终测试间隔时间为组间变量)结果发现:学习方式的主效应边缘显著, $F(1, 122) = 3.74$, $p = 0.055$, $\eta^2 = 0.03$ 。测试条件的记忆成绩表现出高于重学条件的趋势。时间间隔主效应显著, $F(1, 122) = 54.13$, $p < 0.01$, $\eta^2 = 0.31$ 。5 分钟后的记忆成绩显著优于 2 天后的记忆成绩。线索强度主效应显著, $F(1, 122) = 5.01$, $p < 0.05$, $\eta^2 = 0.031$ 。强线索词对的记忆成绩显著优于弱线索词对。线索强度和学习条件交互作用显著, $F(1, 122) = 5.83$, $p < 0.05$, $\eta^2 = 0.05$ 。三者交互作用显著, $F(1, 122) = 16.34$, $p < 0.01$, $\eta^2 = 0.12$ 。其他两两交互效应不显著。

由于三者交互作用显著,因此对学习方式和间隔时间在强、弱线索词对上的记忆成绩分别进行方差分析。结果发现:在强线索词对上,学习方式主效应显著, $F(1, 122) = 9.24$, $p < 0.01$, $\eta^2 = 0.07$ 。测试条件记忆成绩显著高于重学条件;而在弱线索词对上,学习方式主效应不显著, $F(1, 122) = 0.11$, $p > 0.05$, $\eta^2 < 0.01$ 。间隔时间在强、弱线索词对的记忆成绩上主效应均显著, $F(1, 122) = 49.66$, $p < 0.01$, $\eta^2 = 0.29$; $F(1, 122) = 29.88$, $p < 0.01$, $\eta^2 = 0.20$ 。5 分钟后的回忆成绩均显著高于 2 天后。学习方式与间隔时间的交互作用在强线索词对上不显著, $F(1, 122) = 2.86$, $p > 0.05$, $\eta^2 = 0.02$;但在弱线索词对上交互作用显著, $F(1, 122) = 6.71$, $p = 0.01$, $\eta^2 = 0.05$ 。简单效应分析发现 5 分钟时重学条件与测试条件记忆成绩差异不显著, $F(1, 122) = 2.47$, $p > 0.05$, $\eta^2 = 0.02$ 。当间隔时间延长到 2 天时,测试条件记忆成绩显著优于重学条件, $F(1, 122) = 4.42$, $p < 0.05$, $\eta^2 = 0.04$ 。

122) = 2.47, $p > 0.05$, $\eta^2 = 0.02$ 。当间隔时间延长到 2 天时,测试条件记忆成绩显著优于重学条件, $F(1, 122) = 4.42$, $p < 0.05$, $\eta^2 = 0.04$ 。

表 2 最终自由回忆测试成绩的平均数和标准差

最终测试	(单位:个)			
	5 分钟		2 天	
	重学	测试	重学	测试
强线索词对	6.10(2.47)	8.43(3.75)	3.45(2.25)	4.12(2.42)
弱线索词对	6.97(2.50)	5.77(3.38)	2.68(2.95)	4.24(3.06)
总成绩	13.07(4.31)	14.20(6.10)	6.13(4.65)	8.36(4.31)

为了考察测试条件下初始提取成功的项目在最终自由回忆时的记忆效果,我们进一步计算了每一种条件下的项目保持率。项目保持率等于初始提取成功且最终自由回忆成功的项目数与初始提取成功项目数的比值。配对样本 t 检验发现,短间隔和长间隔测试时强线索词对的项目保持率($M_{短} = 0.47$, $SD_{短} = 0.21$; $M_{长} = 0.23$, $SD_{长} = 0.02$)分别与弱线索词对的项目保持率($M_{短} = 0.66$, $SD_{短} = 0.04$; $M_{长} = 0.54$, $SD_{长} = 0.25$)差异显著, $t(27)_{短间隔} = -3.50$, $p < 0.01$, $Cohen's d = 1.26$; $t(27)_{长间隔} = -6.65$, $p < 0.01$, $Cohen's d = 1.75$ 。强线索词对的项目保持率显著低于弱线索词对。

4 讨论

本研究采用中文词对材料验证并拓展了 Carpenter (2009) 的实验结果。Carpenter (2009) 分析测试条件的初始测试与 5 分钟后的最终测试成绩时发现,弱线索词在最终测试时保持得更好,而强线索的初始优势作用在最终测试时消失。本实验引入更长的时间间隔(2 天)来进一步比较不同线索强度的词对在初始测试与最终测试时记忆成绩的差异。结果支持了 Carpenter (2009) 的实验结果,发现当最终测试为自由回忆,间隔时间为 5 分钟时线索强度与间隔时间交互作用显著, $F(1, 27) = 52.43$, $p < 0.01$, $\eta^2 = 0.66$ 。间隔时间为 2 天时交互作用同样显著, $F(1, 33) = 191.07$, $p < 0.01$, $\eta^2 = 0.85$ (交互作用图请见图 1)。对于测试项目,强线索导致了初始提取的优势,但从弱线索中回忆出的项目随时间得到了更好的保持,弱线索这种逆转的优势效应随时间延长仍能有效保持。根据精细提取假设(Carpenter, 2009, 2011),弱线索词对的线索词与目标词之间的相关程度低,在初始测试时,被试需要广泛激活各种可能的中间信息以帮助联想出目标词,

促使线索词与目标词之间建立了更丰富的联结通路,进而促进目标信息在最终测试时的成功回忆。

实验发现,测试效应在不同线索强度的词对上表现不同:强线索词对上,测试条件的自由回忆成绩在5分钟以及2天后均显著高于重学条件;而弱线索词对上,学习方式与时间间隔存在显著的交互作用,5分钟时重学条件与测试条件的记忆成绩差异不显著,而2天后测试条件的记忆成绩却显著高于重学条件。出现这个结果,主要因为当学习材料本身联结性较强时(强线索),测试条件使得学习者有机会在材料已有联结基础上进一步进行自我精细加工,确保最终测试间隔为5分钟时,测试条件的记忆成绩优于重学条件,并在更长时间间隔上仍能得到保持。当学习材料本身联结

性较弱时(弱线索),重学相对于测试来说,学习者有机会再次接触学习材料,有利于目标信息的短时记忆,而这种接触材料产生的学习随时间的衰退十分之快,只有那些经过努力提取,成功建立精细联结的项目能够有更持久有效的保持(Carpenter, 2009)。因此,测试效应在强、弱线索词对上的表现出现分离。对测试条件下自由回忆成功率的分析也进一步支持了这一观点:无论在短间隔还是长间隔,强线索词对的项目保持率显著低于弱线索词对的项目保持率。这个结果说明,弱线索强度的项目一旦在先前测试中成功建立起精细联结,那么自由回忆时它的成功回忆可能性更大,其提取效率显著优于强线索词对。

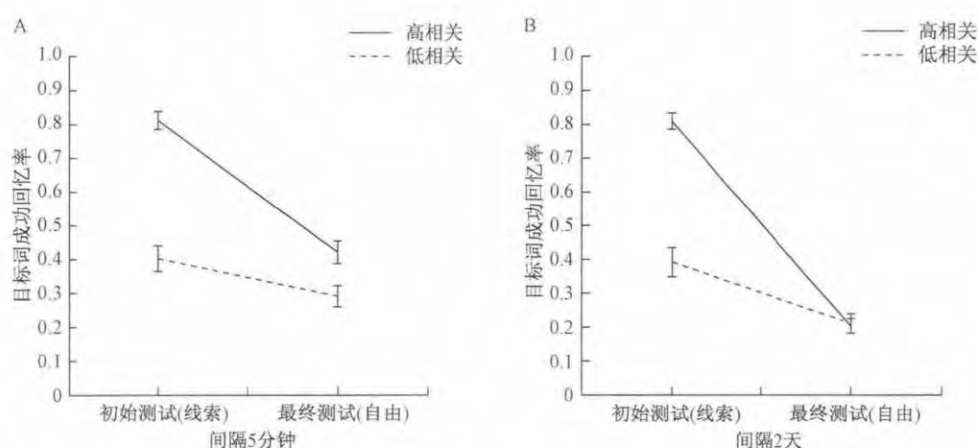


图1 测试条件下初始测试与最终测试交互作用图^②

本实验结果还支持并发展了 Congleton 和 Rajaram (2011, 2012) 的观点。Congleton 和 Rajaram 在 Carpenter 理论观点基础上更进一步,不仅指出了测试促进学习的可能机制,还对两种学习方式下的学习加工方式进行了比较分析。Congleton 和 Rajaram (2012) 认为,概念组织 (Conceptual Organization) 的主动建立对学习内容的长时保持至关重要。他们提出学习时存在两种加工方式,一种是特定项目加工 (Item-specific Processing),一种是关联加工 (Relational Processing)。特定项目加工主要关注于单个项目不同于其他项目的特征,试图建立更独特、更丰富的记忆痕迹,而关联加工则是建立项目之间的关联,产生更强的概念化组织。Congleton 和 Rajaram (2011, 2012) 采用类别词表作为学习材料,考察了学习方式与测试间隔交互作用

的起源。他们认为,即刻回忆时(5分钟),重复学习由于多次接触学习材料,更多地参与特定项目加工,且学习材料自身具有的相关强度诱发了关联加工,因此重学条件的即刻回忆成绩将优于因接触材料次数更少而缺乏必要提取特征的测试条件。而当延迟回忆时(2天),重复测试由于尝试提取而更多地关注学习材料间的关联,产生了更强的概念组织,而恰是这种精细化的概念组织具有长时保持能力,因此测试条件的长时记忆成绩将优于因特定项目加工干扰而缺乏提取组织的重学条件。

本实验扩大了学习材料之间相关程度的差异(线索强度差异显著),发现特定项目加工与关联加工两种编码成分的产生与所占比重由学习方式与学习词对的线索强度共同决定。一方面,在弱线索词对上的自由回忆结果表现出学习方式与时间间隔显

② 注:图中纵坐标是测试条件下被试的初始测试正确回忆率及最终测试正确回忆率,其中初始测试为线索回忆测试,最终测试为自由回忆测试。

著的交互作用,与 Congleton 和 Rajaram (2012) 的观点完全相同:由于词对线索词与目标词之间相关强度低,重学条件重复学习词对时,更多地参与特定项目加工,因此即刻回忆成绩更好;而测试条件具有提取机会,更多地参与关联加工,因此长时保持成绩更好。另一方面,在强线索词对上并没有发现这种交互作用,而是无论间隔时间长短,测试条件自由回忆成绩都优于重学条件。之所以出现这种结果,可能是因为强线索词对线索词与目标词之间的相关强度较高,学习材料本身存在的相关强度诱发了较高的关联加工。当进行首次学习时,被试即获得了词对材料本身所包含的关联强度,随后进行的提取操作更是成功建立了线索与目标之间的额外关联与通达,产生了概念化组织。因此,无论在 5 分钟还是 2 天后的自由回忆测试上,测试条件产生更好的记忆保持,而不产生学习方式与间隔时间的交互作用。

5 结论

本研究以不同线索强度中文词对作为实验材料,考察了重学与测试两种学习方式在不同时间间隔上最终自由回忆成绩上的差异。结果表明:对于强线索词对的自由回忆成绩,无论回忆时间间隔长短,测试条件都高于重学条件;但对于弱线索词对的自由回忆成绩,只在时间间隔为 2 天时,测试条件高于重学条件。这一结果说明线索强度对测试效应有调节作用,且在短时间间隔与长时间间隔上表现不同,验证与发展了精细提取假设。

参考文献:

- Agarwal, P. K., Karpicke, J. D., Kang, S. H., Roediger, H. L., & McDermott, K. B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22 (7), 861-876.
- Bjork, R. A. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena Vol. In R. L. Solso (Ed.). *Information processing and cognition: The Loyola symposium* (pp. 123-144).
- Butler, A. C., Karpicke, J. D., & Roediger III, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, 13 (4), 273-281.
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One*, 5 (6), e10729.
- 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. (2011). Semantic information activated during retrieval contributes to later retention: Support for the mediator effectiveness hypothesis of the testing effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37 (6), 1547-1552.
- 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.
- Congleton, A. R., & Rajaram, S. (2011). The influence of learning methods on collaboration: Prior repeated retrieval enhances retrieval organization, abolishes collaborative inhibition, and promotes post-collaborative memory. *Journal of Experimental Psychology: General*, 140(4), 535-551.
- Congleton, A. R., & Rajaram, S. (2012). The origin of the interaction between learning method and delay in the testing effect: The roles of processing and conceptual retrieval organization. *Memory & Cognition*, 40(4), 528-539.
- Devlin, J. T., Matthews, P. M., & Rushworth, M. F. (2003). Semantic processing in the left inferior prefrontal cortex: a combined functional magnetic resonance imaging and transcranial magnetic stimulation study. *Journal of Cognitive Neuroscience*, 15(1), 71-84.
- Glover, J. A. (1989). The "testing" phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology*, 81(3), 392-399.
- Kang, S. H. (2010). Enhancing visuospatial learning: The benefit of retrieval practice. *Memory & Cognition*, 38(8), 1009-1017.
- 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. (2008). The critical importance of retrieval for learning. *science*, 319(5865), 966-968.
- Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language*, 65(2), 85-97.
- Martin, R. C., & Cheng, Y. (2006). Selection demands versus association strength in the verb generation task. *Psychonomic bulletin & review*, 13(3), 396-401.
- Pyc, M. A., & Rawson, K. A. (2007). Examining the efficiency of schedules of distributed retrieval practice. *Memory & Cognition*, 35 (8), 1917-1927.
- 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.
- Pyc, M. A., & Rawson, K. A. (2010). Why testing improves memory: Mediator effectiveness hypothesis. *science*, 330(6002), 335.
- 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). Test-enhanced learning: taking memory tests improves long-term retention. *Psychological*

- Science*, 17(3), 249 – 255.
- van den Broek, G. S. E., Takashima, A., Segers, E., Fernandez, G., & Verhoeven, L. (2013). Neural correlates of testing effects in vocabulary learning. *NeuroImage*, 78, 94 – 102.
- Wheeler, M., Ewers, M., & Buonanno, J. (2003). Different rates of forgetting following study versus test trials. *Memory*, 11(6), 571 – 580.
- Wing, E. A., Marsh, E. J., & Cabeza, R. (2013). Neural correlates of retrieval-based memory enhancement: an fMRI study of the testing effect. *Neuropsychologia*, 51(12), 2360 – 2370.
- 刘兆敏, 罗良, 张玮. (2011). 记忆提取促进学习: 实验证据与机制解释. *心理发展与教育*, 27(4), 441 – 447.

Cue Strength Regulating the Testing Effect: The Differences between Different Testing Intervals

ZHENG Jun ZHANG Wei LUO Liang

(State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875)

Abstract: The testing effect refers to the fact that in comparison to restudy, involvement in the testing enhances long-term retention. The current study had 126 subjects, and adopted Chinese words pairs with different cue strength (strong/weak) as experimental materials. The experiment was designed to explore how learning methods (restudy/testing) influencing the memory retention on different testing intervals (5 minutes/2 days). The results indicated that: (1) memory of target words with strong cues showed the main effect of learning methods: the target performance in testing group was higher than restudy group. (2) In the situation of weak cues, the main effect of learning methods wasn't significant, but there was an interaction between learning methods and testing intervals: there was no significant difference between the immediate recall of restudy and testing group, however, the performance of testing group was significant higher than restudy group two days later. In conclusion, these results demonstrated that cue strength regulating the testing effect, which supports the elaborative retrieval hypothesis

Key words: cue strength; testing effect; long-term memory; elaborative retrieval hypothesis; conceptual organization