

学生提取自信度判断准确性的发展^{*}

唐卫海¹ 刘希平^{1,2} 方 格²

(1. 天津师范大学教育科学学院,天津 300073;
2. 中国科学院心理研究所心理健康重点实验室,北京 100101)

摘 要:采用 3 × 3 × 2(年龄,材料,性别)三因素混合实验设计,按照学习成绩好、中、差分层随机选取小二、初二、大二学生被试各 18 名,其中男女各半,利用不同难度的材料,对被试提取自信度准确性的发展进行了考察。被试的判断等级与被试的记忆成绩之间的两点距随年龄而变化。而这种变化在不同难度间有不同的表现。(1)提取自信度准确性从小二到初二发展迅速,表现出显著差异;初二到大二阶段年龄差异不显著。材料难度对不同年级的被试提取自信度判断准确性的影响不同。材料难度越小,提取自信度准确性发展得越早。(2)提取自信度准确性发展的性别差异不显著。

关键词:记忆监测;提取自信度;准确性

1 问题的提出

记忆监测(Memory Monitoring)是从客体记忆获得信息,对客体记忆过程的每一步骤进行监督,从而为记忆控制奠定基础的过程。它是程序性元记忆(procedural metamemory)研究的重要组成部分。它涉及两项内容:一项是预见性监测(Prospective Monitoring)——是在提取行为完成之前进行的监测活动。包括对任务难度的预见(EOL——easy of learning judgment)和对学习程度的判断(JOL——judgment of learning,也有人称之为回忆准备就绪程度的判断,即 RRA——recall readiness assessment),它为客体记忆中识记和保持过程的决策提供依据,比如使用什么记忆策略,如何分配记忆时间,什么时候终止复习等;另一项是回溯性监测(Retrospective Monitoring)——是在提取过程中进行的、对提取结果准确性的估计,包括提取自信度的判断(JOC——judgment of confidence)和知晓感的判断(FOK——feeling of knowing,但是也有人主张在尝试提取之前对学习项目进行知晓感的等级评定,并把它称之为知晓的预见——POK,即 prediction-of-knowing,详见 Annette Farrant, Jill Boucher, and Mark Blades,

1999。但 POK 从本质上更接近 JOL,所以本文不做特别介绍)。回溯性监测是记忆监测的最终环节,它为提取阶段的决策提供参考,比如选择何种搜索策略,是否改变搜索策略,是否终止搜索等^[1,2]。

前人对预见性监测的发展进行了大量研究:Flavell 对儿童就回忆准备就绪情况的判断水平进行了探讨^[3],发现 5 - 6 岁的儿童常常对自己准备进行的回忆程度做过于乐观的估计,在他们声称已经准备好的时候,他们的回忆成绩往往水平不高,而年龄大些的儿童则有更为客观的估计。另有人就儿童对任务难度的预见的判断进行了探讨,发现即使到了初中二年级,其相应的能力仍然在增长^[4]。有人对回忆准备就绪程度判断的发展也进行了研究,发现与任务难度的预见相比较,回忆准备就绪程度的判断发展的速度稍快些^[5],但从小学二年级到初中二年级之间其发展水平一直有提高。这些研究从预见性监测发展方面对儿童的程序性元记忆进行了考察。

1939 年, D. M. Johnson 首次对被试提取自信度进行了研究,他对被试完成 4 项任务的自信度与所用时间进行了相关研究,证明自信度和人格关系密切^[6]。应该说那是关于自信度研究的开

^{*}基金项目:中国科学院知识创新工程重要方向项目(KSCXZ-SW-221);天津市教育科学“十五”规划课题(PE081)的部分工作。

作者简介:唐卫海(1963 -),男,天津师范大学教育科学学院副教授。

始。之后,人们对自信度的研究涉及了各种认知领域,各种学习方式和各种感觉通道^[7,8,9]。其中大部分研究是为了探讨“记忆错觉”(将另文讨论),而不是针对学生的发展^[8,9]。

尽管直到今天,对JOC判断发展的研究仍然不是很多,特别是在国内,几乎难以查阅到一手的研究报告,但总结国外已有研究的主要范式,大体是,让被试学习一系列材料,然后提取,提取之后对提取出来的项目进行判断,判断自己有多大把握认为提取的项目是正确的,然后将判断的情况与提取项目实际的对错进行比较,从而确定被试的提取自信度判断的准确性。

Shaughnessy对JOC判断的准确性进行了研究,发现被试对多项选择题的提取自信度的判断与测验成绩之间具有很强的正相关^[10]。有人分析,预见性监测和回溯性监测由于处在不同的记忆阶段,所受到的影响是不同的。回溯性监测有更多的参考标准可以利用,所以应该有更准确的判断。

鉴于已有研究较少,对提取自信度研究的一手资料不够丰富,本研究试图就提取自信度判断准确性的发展做基本的探讨,为后续研究积累数据。研究试图探讨如下问题:第一,提取自信度判断准确性在小学、中学和成人身上表现有何不同;第二,材料对提取自信度判断准确性的影响;第三,性别对提取自信度判断准确性的影响。

2 对象与方法

2.1 方案设计

采用3×3×2三因素混合实验设计。A,年龄,被试间设计,有3个水平,即小二,初二和大二;B,材料的记忆难度,被试内设计,也有3个水平,即难度低,难度中等,难度较高;C,性别,被试间设计,有男、女两个水平。

2.2 被试

为了探讨小学生、中学生、成人之间在提取自信度判断准确性上的差别,在小学阶段选择小学二年级学生为研究对象,以较大限度地低限探讨小学儿童的特点;在初中选择初中二年级学生为被试;用大学二年级学生作为成人被试。分别在天津市市区选取小学、初中、大学各一所学校,在小二、初二、大二学生中,按各自所学科目的总成绩排队,把学生分成学习成绩好、中、差三个层次,每个层次各占全班人数的三分之一,在各个层次

上随机选择被试,每个层次上选择男生3人,女生3人,男女各9人,共计54位被试。被试平均年龄分别为8.1岁、14.4岁和20.7岁。

2.3 材料

(1)词表3组共6张:按词对组合方式不同,分成三种材料:第一,反义词对表,如大船-小船、坏人-好人等,属于难度低的材料;第二,近义词对表,如关门-关灯、灭火-灭虫等,属于难度中等的材料;第三,人为组词对表,如月亮-刻字、尺子-抱羊等,属于难度较高的材料。

第1组:在小二被试已经学过的语文课本中选择反义词对20对,从中随机选择5对,构成词表B1',剩余的15对构成词表B1。

第2组:在小二被试已经学过的语文课本中选择近义词对20对,从中随机选择5对构成词表B2',其余的15对构成词表B2。

第3组:在小二被试已经学过的语文课本中选择名词性质的词20个,动宾关系的词对20个,在不构成主谓关系的前提下,组合成20个词对,从中随机选择出5对,构成词表B3',其余的15对构成词表B3。

为了确保材料难度之间的差异(相对于被试而言确实存在),对材料难度进行了检验(每提取一个正确词对记一分)。发现反义词对难度最小,近义词对难度适中,人为组合的词对难度最大;无论是小二还是大二的被试,都表现出了相同的趋势。如表1所示。对表1的结果进行了方差分析,发现,不同年级被试在三种材料上的记忆成绩存在着显著差异,显著性水平为0.001,说明三种材料确实存在记忆难度的差别。

表1 不同年级的被试在3项材料上的记忆成绩得分

年级	材料1	材料2	材料3
大二	8.61	7.11	6.94
初二	6.17	5.78	5.39
小二	4.39	3.50	2.06
总计	6.39	5.46	4.80

(2)评定表格:1式、2式各162张。1式表格用于被试默写所记忆的词对,2式表格用于被试对默写出来的词对进行提取自信度的判断。1、2式表格相背而折。使用时让被试先用1式再用2式。每位被试完成3套材料的学习,所以每位被试需要有3套表格,54位被试共计162套表格(具体表格请见附录)。

(3)记忆仪3台(天津师范大学生产,同型技

术指标校准),笔一支。

2.4 步骤

(1) 主试、被试相向隔桌而坐,被试面前摆放着 3 台记忆仪,分别安排了材料 B1、B2、B3。

(2) 主试宣读指导语 1:现在给你看 15 个这样的词对(用手示意 B1),每个呈现 2 秒钟,间隔 2 秒钟,再看第 2 个词对,以这种速度共看 2 遍,默写时,一个词对中的 2 个词不能颠倒,15 个词对的顺序可以打乱。

主试检查被试理解情况,能理解进行下一项。

(3) 用记忆仪呈现材料 B1(或 B2、B3),每项 2 秒钟,间隔 2 秒钟;呈现 2 遍,被试识记。

(4) 宣读指导语 2。发评定表一式 1 张,要求被试对材料 B1(或 B2、B3)进行即时自由回忆,时间为 5 分钟(正确自由回忆的词对数为实记词对数)。

指导语 2:现在请你把能记住的词对默写下来。

被试默写。

(5) 要求被试将评定表 1 式反折的部分打开,可看到评定表 2 式。

(6) 主试宣读指导语 3,检查被试理解情况。然后被试对已经写出的词对逐项进行评定。

指导语 3:现在请你判断一下,默写的词对有多大把握肯定是对的?用 1~5 个等级来评定,数值越大表明把握越大。

(7) 结果统计:评定为 5 的每一个正确词对记 1 分,评定为 4 的每一个正确词对记 4/5 分,评定为 3 的每一个正确词对记 3/5 分,评定为 2 的每一个正确词对记 2/5 分,评定为 1 的每一个正确词对记 1/5 分。将被试所有的分数相加,得到一个分值,称为检测估计数。

分别统计每个被试的 d 值(两点距),公式为 $d = |\text{实记词对数} - \text{检测估计数}| \div \text{实记词对数}$,d 值越大,表示监测准确性越低。同时统计被试的实记词对数,作为其记忆成绩。

(8) 按上述步骤重复另外两项材料。每项材料间被试休息 1 分钟。

2.5 被试安排

为了控制 3 种材料间由顺序带来的系统误差,采用轮组法。实验均个别进行,时间均为下午。每位被试所用实验时间在 30 分钟左右。

3 结果

3.1 提取自信度准确性的发展

总体而言,儿童提取自信度判断的准确性随年龄增长而提高。到初二阶段达到最高水平。各年级被试的平均 d 值见表 2。

表 2 各年级被试提取自信度判断准确性的平均值

年级	两点距
大二	0.038
初二	0.026
小二	0.166
总计	0.077

对三个年级被试提取自信度判断的准确性进行方差分析,得到年级间差异显著, $F_{(2,408)} = 8.60^{***}$,说明从小学二年级到大学二年级学生之间提取自信度判断准确性存在显著差异。对年级间提取自信度准确性进行了多重比较,得到小二和初二之间差异非常显著, p 值为 0.000;在初二和初二之间差异不显著, p 值为 0.752。说明小学二年级到初中二年级之间,提取自信度判断准确性迅速发展;但初二到大二之间判断水平没有明显变化。

3.2 提取自信度发展的材料差异

对不同年级的被试在三种材料上的提取自信度判断的准确性进行统计,得到表 3。

表 3 不同年级被试提取自信度准确性
在不同材料上的表现

年级	材料 1	材料 2	材料 3	总体
大二	0.03	0.04	0.04	0.04
初二	0.03	0.03	0.02	0.03
小二	0.04	0.14	0.32	0.16
总体	0.03	0.07	0.13	0.08

对上述结果进行方差分析,得到年级的主效应为 $F_{(2,51)} = 8.86^{***}$,材料的主效应为 $F_{(2,102)} = 4.26^*$,年级与材料间的交互作用为 $F_{(4,102)} = 4.29^{**}$ 。

对表 3 中的数据进行简单效应检验,得到材料 1(反义词对)在年级间没有差异, $F_{(2,51)} = 0.06$, p 值为 0.94;材料 2(动宾词对)在年级间差异边缘显著, $F_{(2,51)} = 3.02$, p 值为 0.06;材料 3(人为组合的词对)在年级间差异非常显著, $F_{(2,51)} = 7.34$, p 值为 0.002。说明从材料 1 到材料 3,随着材料难度的加大,提取自信度判断准确性的发展差异表现得越来越明显。反义词对是最容易记忆的,在提取自信度判断的准确性上没有表现出发展的差异,从小二开始,被试的判断准确性已

经达到了相对理想的程度;近义词对较难,提取自信度的发展,虽然也没有显著差异,但显著性水平接近 0.05 (0.057);人为组合的词对是其中最难记忆的,对人为组合词对的提取自信度的判断准确性则表现出了 0.01 水平的显著性。表明,如果材料难度适当,即使是小学二年级的儿童,也可以对自己的提取自信度进行比较准确的监测。同时,这项研究结果也说明,材料难度确实影响儿童记忆监测的发展。

而不同年级的被试在三项材料间提取自信度判断的准确性有不同的差异表现:初中二年级和大学二年级被试在三项材料间差异不显著, F 值分别为 $F_{(2,102)} = 0.02, 0.03$, 对应的 p 值分别为 0.978, 0.972。但小学二年级儿童在三项材料间则表现出提取自信度判断准确性的差异, $F_{(2,102)} = 12.79$, p 值为 0.000。说明对小学二年级儿童来说,他们的提取自信度准确性受材料难度的影响更大。

3.3 提取自信度发展的性别差异

对提取自信度判断的准确性就不同的性别进行统计,得到表 4。

表 4 不同性别的被试提取自信度的准确性

年级	性别	两点距
大二	女	0.03
	男	0.05
初二	女	0.03
	男	0.03
小二	女	0.20
	男	0.13
总计	女	0.08
	男	0.07

对表 4 的结果进行方差分析,发现提取自信度准确性的性别差异不显著, $F_{(1,48)} = 0.26$, p 为 0.610;性别和年级间也不存在交互作用, $F_{(2,48)} = 0.61$, p 为 0.547。

4 讨论

4.1 提取自信度判断准确性发展的年龄差异

前人的研究结果说明,监测准确性应当随年龄增长而有提高,在本实验条件下,提取自信度准确性在小二到初二之间发生了显著的变化,验证了前人的研究结果^[4,11,12]。说明记忆监测的发展几乎遵循着同样的规律,即在小学阶段是儿童元记忆发展的飞跃期。这与很多因素有关:第一,儿童知识范围的扩大。入学后,儿童有时间来集中

接受科学启蒙教育,这为他们很好地客观地认识世界打下了基础,包括对主观世界的认识。第二,随着学习生活时间的延长和学习内容的丰富,儿童的陈述性元记忆知识不断增长^[13,14],他们有更多的机会来对已有的主观经验进行检验,检验的结果为他们提供了充分的反馈信息,这使他们能够随时调整自己的认识,为程序性元记忆的发展提供了必要的条件。第三,提取自信度的判断发生在提取行为之后,因此在判断时,儿童能够对自己的客体记忆过程有更加充分的认识,可以参考更多的信息,例如,材料本身的信息,学习过程的信息,特别是提取过程的信息。这些信息为被试做出准确的判断提供了比较充分的线索^[15,16,17],因此当儿童能够充分地使用这些信息的时候(本实验中的初二儿童始),就能够相对准确地对自己的提取情况做比较客观的估计。

4.2 提取自信度判断准确性发展的材料差异

本研究中发现了提取自信度准确性发展的材料差异,这与预见性监测判断发展的研究结果相吻合,在预见性监测中,材料对相应的判断是有影响的^[4,5]。而另有研究发现,记忆难度影响记忆监测的准确性^[18]。Weaver 等调查了记叙文和说明文的课文难度对记忆监测的影响,发现不论是记叙文还是说明文,被试对中等难度的课文都有非常准确的监测。本实验发现,在所限定的三种材料中,被试在人为组合的词对上记忆监测判断的准确性呈现出显著差异,而在最容易的反义词对上没有表现出显著差异,即小二被试对比较容易的记忆材料,也可以有比较准确的回溯性监测。说明如果材料难度适当,不论儿童是否能够清晰地意识到,即使小学低年级儿童也可以有记忆监测上的良好表现。当然,这种监测上的良好表现是内隐还是外显学习的结果,还有待于进一步的研究探讨。

4.3 在提取自信度判断准确性发展中,性别差异不显著

这与前人有限的研究相一致。也许,在记忆监测中性别差异根本就不存在?但是,这多少使人有些困惑。因为相关的研究发现,记忆监测与个性特征之间是存在相关的^[19],而男女生在个性特点上,显然是有差异的。为什么个性特点与记忆监测存在相关,而性别本身却在记忆监测的发展上没有任何差异呢?是否不同的人格特点会影响到不同的心理特质,而不同的心理特质受人格

特点的影响不同?这还有待进一步的细致研究。

4.4 材料难度的限定

在发展研究中,始终存在一个问题,即材料难度的限定。这里有两层涵义:第一,材料的客观难度用什么做标准?前人在这个问题上有多种选择,有人用词对的语义联系的密切程度^[20],有人用被试对材料的熟悉程度^[21],也有人用被试购买物品的频率^[22]作为衡量学习材料难度的标准。怎样在所选取的被试年龄较小的情况下,有效地区分材料的难度,就成为一个关键而棘手的问题;第二,材料难度相对于不同的主体,可能有不同的涵义。对低龄儿童难度大的材料,对成年人未必难度大;而对成年人难度大的材料,对低龄儿童未必难度大。所以要在涉及到年龄跨度较大的同一个实验中,有效地选择实验材料,就几乎对实验者构成了一种挑战。本实验的材料选择,一方面支持了当初的假设;另一方面为材料难度对儿童提取自信度的判断准确性的影响的分析奠定了基础。尽管这种选择材料的方法,可能还存在某些局限,但在以视觉呈现的方式下、用文字做材料的前提下、在有小学二年级儿童参与的情况下,这种方法还是有效的。当然,如果有机会将国内相关的研究材料标准化,势必为今后的研究奠定很好的基础。

5 小结

(1) 提取自信度准确性从小二到初二发展迅速,表现出显著差异;初二到大二阶段年龄差异不显著。材料难度对不同年级的被试提取自信度判断准确性的影响不同。材料难度越小,提取自信度准确性发展得越早。

(2) 提取自信度准确性发展的性别差异不显著。

参考文献:

- [1] Mazzoni G, Nelson T O. Metacognition and Cognitive Neuropsychology Monitoring and Control Processes. 1998:1 - 22.
- [2] 韩凯. 元记忆研究的理论框架. 心理学动态, 1999, 1: 85 - 87.
- [3] Schneider W. and Pressly M. Memory Development Between 2 and 20. 1989:97 - 112.
- [4] 刘希平. 任务难度预见能力发展的实验研究. 心理发展与教育, 1998, 4: 17 - 21.
- [5] 刘希平, 唐卫海. 回忆准备就绪的判断的发展. 心理学报, 2002, 1: 56 - 60.
- [6] Johnson D M. Confidence and speed in two-category judgment. Archives of Psychology Columbia University, 1939, 241: 52.
- [7] Cabeza R, Kato T. Features Are Also Important: Contributions of Featural and Configural Processing to Face Recognition. Psychological Science, 2000, vol. 11, 5: 429 - 433(5).
- [8] Solso R L, Raynis S A. Transfer of prototypes based on visual, tactual, and kinesthetic exemplars. Am J Psychol. 1982, 95(1): 13 - 29.
- [9] Solso R L, Raynis S A. Prototype formation from imaged, kinesthetically, and visually presented geometric figures. J Exp Psychol Hum Percept Perform. 1979, 5(4): 701 - 12.
- [10] Haughnessy J J. Confidence judgment accuracy as a predictor of test performance. Journal of Research in Personality, 1979, 13(4): 505 - 514.
- [11] Thompson J L W. Funny you should ask, what is the effect of humor on memory and metamemory? Dissertation Abstracts International: Section B: The Sciences and Engineering, 2001 Mar; Vol 61(8 - B): 4442.
- [12] Schneider W, Borkowski J G, Kurtz B E. etc. Metamemory and motivation: A comparison of strategy use and performance in German and American children. Journal of Cross-Cultural Psychology, 1986, 17, 315 - 336.
- [13] Weinert F E, Schneider W. Individual development from 3 to 12: Findings from the Munich longitudinal study. Cambridge: Cambridge Univ. Press, 1999.
- [14] Mazzoni G, Nelson T O. Judgements of learning are affected by the kind of encoding in ways that cannot be attributed to the level of recall. Journal of Experimental Psychology: Learning, Memory, & Cognition, 1995, 21(5): 1263 - 1274.
- [15] Schreiber T A. Effects of target set size on feelings of knowing and cued recall: Implications for the cue effectiveness and partial-retrieval hypotheses. Memory & cognition, 1998, 26(3): 553 - 571.
- [16] Shimamura A P, Squire L R. Long-term memory in amnesia: Cued recall, recognition memory, and confidence ratings. Journal of Experimental Psychology. Learning, Memory, & Cognition, 1988, 14: 763 - 770.
- [17] Weaver III C A, Bryant D S. Monitoring of comprehension: The role of difficulty in metamemory for narrative and expository text. Memory & Cognition, 1995, 23(1): 12 - 22.
- [18] Dufresne A, Kobasigawa A. Developmental difference in children's spontaneous allocation of study time. The Journal of Genetic Psychology, 1988, 149: 87 - 92.
- [19] 刘希平. 回溯性监测判断与预见性监测判断发展的比较研究. 心理学报, 2001, 2: 137 - 141.
- [20] Dufresne A, Kobasigawa A. Children's spontaneous allocation of study time: Differential and sufficient aspects. Journal of Experimental Child Psychology, 1989, 47: 274 - 296.
- [21] Kobasigawa A, Metcalf-Haggert A. Spontaneous allocation of study time by first- and third-grade children in a simple memory task. Journal of Genetic Psychology, 1993 Jun; Vol 154(2): 223 - 235.
- [22] Mazzoni G, Cornoldi C, Tomat L. etc. Remembering the grocery shopping list: A study on metacognitive biases. Applied Cognitive Psychology, 1997, 11: 253 - 267.

