记忆监测研究综述*

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]中国科学院心理健康重点实验室,北京,100101)

1 前言

对元记忆的研究开展了近40年。国内陆续有人介绍了元记忆的理论框架,元记忆与客体记忆的关系等。近年来,国际心理学届对程序性元记忆的研究更加关注,特别是对记忆监测的研究,积累了大量的数据,同时也发现了众多需要进一步解决的问题,这无疑对推动元记忆研究的开展具有重要意义。

记忆监测是对客体记忆的三个主要阶段进行主观判断,以了解客体记忆的情况,为记忆控制奠定基础。围绕着记忆监测,前人进行了预见性监测和回溯性监测的探讨。预见性监测发生在提取活动之前,包括任务难度的预见(EOL)、学习程度的判断(JOL);回溯性监测发生在提取活动之后,包括提取自信度的判断(JOC)和知晓感的判断(FOK)。

近年来,国内陆续有人对记忆监测进行实验研究,但记忆监测的总体框架、记忆监测的研究进展、记忆监测与记忆控制等的关系以及记忆监测研究中存在的问题等,在国内还没有系统的介绍。本文就近年来围绕记忆监测的研究所揭示的研究结果进行简单的归纳,并对其中的问题进行初步的分析。

2 记忆监测的研究方法

2.1 不同的记忆监测的研究方法

把被试记忆监测的情况外显出来进行测量,经典的方法是口语记录法、访谈法、反应时法和对偶比较法。记忆监测的研究方法在具体应用时有较大的灵活性。总结近年来对不同的记忆监测的研究,我们发现无论对哪一种具体的监测,都存在着两种研究范式:一是对所学习的每一个项目进行逐项(item by item)评定;二是就所学习的项目的总体(overall item)进行评定。同时,人们发现对总体进行评定,往往更加准确[1]。

针对EOL、JOL、JOC以及FOK,各自都有具体的研究方法。

EOL判断有一个典型的逐项评定的程序:给被试按顺序呈现(假定)20个词对,20个词对随机排列,要求被试将最容易的一对挑出来。之后,在剩下的19个词对中,要被试将最容易的词对挑出来,以此类推。被试将20个词对的难易进行了逐一评定,再按照由难到易的顺序评定一次,以消除评定顺序对材料学习的影响。然后要求被试学习这20个词对,学习相同的时间,考察每个词对学习的成绩,与当初的评定进行对照,就可以知道被试的任务难度预见的准确性[2]。

也有要求被试对所学习的材料进行总体难度评定的,即以评定等级作为预测的指标。或者,干脆要求被试直接报告在所要求的条件下,被试能记住多少项,即以记忆项目的绝对值作为评定的指标。

JOL判断的主要范式是:先学习,然后估计自己的记忆成绩,最后提取,比较估计的记忆成绩与实际的提取成绩的一致性程度或差异大小,以此作为JOL判断的准确性指标。对记忆成绩的估计常常采用0%,20%,40%,60%,80%,100%六级评定法。

JOC判断是在提取之后进行的提取自信度的判断。主要方法是对提取的项目进行等级评定,更多的是采取逐项评定的方式,然后考察等级评定的结果与提取准确性之间的一致性,作为JOC判断的准确性的指标。如果提取方式是线索回忆,还常常同时记录反应时,以此作为判断被试自信度的一种依据。

FOK判断发生在提取失败之后,对学习过但没有提取出来的项目进行等级评定。评定之后进行再认。通常用评定等级与再认成绩的相关(常用G相关),来说明FOK判断的准确性。这类研究已经形成了典型的RJR(回忆-判断-再认)范式。其中的等级评定大多也采用6级评定法,以0%为没有知晓感,100%为肯定能再认,每一级相间20%。

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[3]。它被称之为POK(prediction-of-knowing),其实验程序是:先学习,然后尽快进行知晓感的判断(可同时记录反应时),最后进行回忆。统计其知晓感判断与正确回忆成绩的G相关[4]。但POK判断与典型的FOK判断从本质上可能有所不同,因为典型的FOK判断发生在至少

* 本文是中国科学院重要方向项目(批准号:KSCX2-SW-221)、天津市教育科学"十五"规划课题(PE081)和一般课题(CESO16)的部分工作。

** Email: twh2002cn@yahoo.com.cn

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

一次提取之后，POK判断则是在提取行为发生之前进行的。而提取前后的记忆监测在发展速度、影响因素、活动机制等方面均有不同[1,5,6]。因此，近年来，研究者们放弃了POK这一概念，而将提取之前对学习程度的判断统一到JOL的概念上。

2.2 记忆监测的考察方法和统计指标

在描述记忆监测的水平时，有两种考察方法或角度，一种是用自信度的等级作为记忆监测的指标；另一种是用准确性作为记忆监测的指标。准确性的测量又有两种统计思路：一种是考察监测判断与实际记忆成绩的差异，所用的方法是传统的PA(P: predicted performance 是指预测成绩；A: actual performance是指实际成绩)法，即以预见值与实际记忆成绩的差别作为指标，反映记忆监测的准确性；另一种思路是考察监测判断与实际记忆成绩的一致性，用预见值与实际记忆成绩的相关作为考察记忆监测准确性的指标。近年来的研究大多采用相关值作为监测准确性的指标。这一方面是因为PA的计算方法种类繁多，例如，有PA=P-A，PA=(P-A)÷A，PA=|P-A|，PA=|P-A|÷A，PA=2|P-A|÷(P+A)，PA=lgP/A等，难以确定哪种计算方法更为科学合理，同时也给有关研究的比较带来了不小的麻烦；另一方面，Nelson发现，用G相关来计算记忆监测的准确性似乎是最好的，它可以最充分地运用所测量到的数据[4]。

有研究发现，监测判断的等级更多地与人格特点有关[5]，它反映了被试的自信度；但判断的准确性则更多地有赖于对自己的主观加工过程的准确觉察。因此近年来，研究者们发现将判断等级与判断准确性加以区分是非常重要的。

当然，也有一些研究在探讨监测自信度的同时，对记忆监测的准确性进行了考察[7]。这样的报告可以给日后的元分析留有更多的余地。

3 记忆监测的研究维度

3.1 影响记忆监测的因素

人格是影响记忆监测的因素之一。从有限的研究结果来看，人格如果与记忆监测有关联，也应该表现在记忆监测的自信度上。各种监测的等级评定，常常与被试的个性特点有关。记忆监测判断的水平不可避免地受到人格的行为定向性和人格倾向性的影响。有人研究发现，FOK和JOC的判断水平受内外倾、情绪稳定性的影响：外倾者所做的判断水平较内倾者为高；情绪稳定的被试，两种判断水平的值也更少有波动[8]。至于记忆监测的准确性与人格的相关仍然需要进一步的研究。

智力与记忆监测之间的关系究竟如何？有人提出了相应的假设，可以为我们思考相应的问题提供思路。他们认为，认知监控对超常儿童与正常儿童而言都具有一定的难度，任何年龄的儿童都不能体现出超常优势。但Mills等认为，超常儿童在许多技能领域的发展都表现出早慧，所以对儿童的记忆监测与智力的关系进行深入的探讨是势在必行的。

意识清晰程度对记忆监测有影响。Mazzoni和Nelson利用指导语，对被试加工的有意性进行控制，考察其JOL判断在进行有意编码和无意编码时的准确性，发现JOL判断在有意编码情况下更准确[9]。此研究还同时考察了加工深度对JOL判断准确性的影响。实验者将被试分成两个小组，一组对所呈现的词进行愉快度评定，另一组报告词所包含的音节数。显然前者对学习材料的加工深度要高于后者。然后令其进行JOL判断。结果发现，对学习材料进行愉快度评定一组，JOL判断的准确性明显高于报告音节数的被试。说明，加工深度是影响记忆监测准确性的因素。

Leonesio和Nelson控制被试的学习程度，要求被试进行JOK和FOK判断，发现学习程度对JOK判断等级(评定等级范围为：1-20)有较大影响，在非过度学习的情况下，判断等级为6.0，而在过度学习的情况下，判断等级为13.0；学习程度对FOK判断几乎没有影响，判断等级分别为7.5和8.0。这一方面说明学习程度对记忆监测确有影响，另一方面也说明各种记忆监测的性质有所不同[2,5,10]。

记忆难度影响记忆监测的准确性。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(1995)。发现被试对中等难度的课文有非常准确的监测，而不论课文是记叙文还是说明文。刘希平等的研究也发现，材料难度对学习主体的记忆监测的准确性有影响，在反义词对、动宾词对和人为组合词对三种材料下，儿童提取自信度的准确性逐渐降低，而且，这种作用与年龄之间存在交互作用。

3.2 记忆监测的发展

Cultice等的研究发现，即使是学龄前儿童也能意识到，在你知道一个人名字时，仍有可能现在命名不出来。说明儿童的记忆监测发展之早。

当然，不同的记忆监测有发展上的差异。通常来说，越是远离提取行为的监测，其发展越置后[14]。这可能因为处于不同阶段的记忆监测，影响它的因素不同造成的。例如，做EOL判断，更多地受过去经验的影响；而做FOK的判断则既受过去经验的影响，也受当时记忆过程的影响，还受提取是否顺利的影响。

总体而言，记忆监测在小学阶段是快速发展期[14,15]。

有人对儿童记忆监测中出现的监测不准确进行了分析，提出了记忆监测缺损假说(monitoring deficiency hypothesis)，认为儿童不具备相应的知识

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

和技巧,来对自己的记忆监测的情况进行准确判断。这一理论得到了一些实验证据的支持。但是,最近 Schneider 等人研究发现,虽然学前儿童对自己的记忆成绩往往做过高的估计,但他们对同伴的记忆成绩的预见准确性往往要客观得多。说明儿童记忆监测缺乏准确性,并不是由于记忆监测缺损造成的,而与其它因素有关。在上述研究中,显然与儿童的愿望有关,他们总是希望自己能够记忆得更多,于是在预见自己的记忆成绩时就做了更高的估计。而他们对同伴能够记忆多少,似乎并不过多地关注,于是就有了较为客观的估计。这种假设被称之为如意算盘假说(wishful thinking hypothesis)。目前这方面的研究仍然在继续。

有人发现,儿童记忆监测的发展水平与练习有关,特别是对 3、4 年级以上的学生来说,练习的作用更加明显。但是 Schneider 等人(1990)利用"预测——学习——测验——再预测——再学习——再测验"的范式进行研究,并没有发现明显的预见准确性的提高。

3.3　记忆监测的活动机制

各种记忆监测是仅仅所处阶段不同还是性质不同?这是记忆监测研究的一个重要问题。现有的研究表明,分处于不同阶段的记忆监测在影响因素、发展速度、脑电的活动特点等诸多方面表现出了分离现象,似乎说明各种记忆监测之间存在本质差异。先前的神经心理学的研究表明,Korsakoff's 综合症病人的回溯性监测的准确性与正常人没什么区别,但其预见性监测的准确性却明显低于正常被试[5,11]。这说明预见性监测与回溯性监测的神经生理机制可能有所不同。

前几年,大家围绕 FOK 判断究竟是线索熟悉还是目标提取展开了一系列研究,得到大量研究结果。其中有三种较有代表性的说法:一是 FOK 判断有赖于线索熟悉性程度;二是 FOK 判断有赖于靶信息的强度;三是 FOK 判断与靶信息和相关信息都有关系[12]。三类研究结果分别支持了三种理论:线索熟悉性假说、靶项目可提取假说和可接近性模型。但总体而言,支持可接近性模型的报告越来越多。线索和靶项目在不同条件下对 FOK 判断有不同的影响,即靶项目和与其相关的信息都对 FOK 判断产生作用。

最近,国外心理学家更加关注的是 JOL 判断的研究,因为毕竟,JOL 判断发生在学习过程中,监测的结果将对记忆控制产生直接的作用。对 JOL 的研究为揭示程序性元记忆的机制,提高学习和记忆的效果必将产生更大的作用。国内这方面的研究还有待进一步展开。

3.4　记忆监测与客体记忆的相关

3.4.1　记忆监测与外显记忆的相关

按照元记忆假设:记忆监测准确性越高,记忆成绩也越好。W.Schneider(1985 年),统计了 27 篇文章,47 个相关系数,2237 个被试,得到元记忆与记忆水平之间的相关系数为 0.41;后来他又统计了 60 篇报告,其中有 123 个相关系数,7093 位被试,得到元记忆水平与记忆成绩之间的相关系数也为 0.41(1997)。在一系列尚未发表的研究中*,以 $|P-|\div A$ 做指标,分别对任务难度预见、学习程度的判断和提取自信度的判断准确性与记忆成绩之间的相关进行了研究。发现,越是远离提取行为的监测,其准确性与记忆成绩之间的相关越高。韩凯等(1997)的研究也表明 $|P-A|$ 与记忆成绩可以有较密切的关系。

另有人研究了记忆监测对情节记忆和语义记忆的影响,发现记忆监测对不同记忆系统中存储的项目的监测强度是不一样的,对情节记忆的监测明显好于对语义记忆的监测[13]。

但是目前也有一些研究结果否定元记忆假说。Hager1992 年发现,记忆监测的准确性与记忆成绩之间的关系并不密切[14]。

3.4.2　记忆监测与内隐记忆的相关

James 等研究了语音重复启动效应对外显检索和 TOT(就在嘴边感,是 FOK 的最早形式)判断的影响,发现启动效应降低了 TOT 判断水平,却提高了检索成绩[15]。

刘耀中(2001)考察了 FOK 判断与情节记忆、语义记忆和内隐记忆的关系,发现 FOK 判断的准确性对情节记忆比对语义记忆更准确;同时还发现,FOK 判断对内隐记忆的作用是微弱的。

4　存在的困惑

4.1　材料选取的不适用

在研究程序性元记忆的过程中,大家所用的材料更多的是词对(或类似的单个的项目),很自然是沿用了艾宾浩斯的研究假设:一是研究中所涉及的每一个项目都是相互独立的;二是每一个记忆项目在项目集合中占有同等重要的地位。这样做,为我们的研究带来了很大的方便,因为我们在分析结果时,可以将每一个项目都看成等值的,答对 10 个项目所得到的分数是答对一个项目所得分数的 10 倍。

然而用这样的材料进行记忆监测的研究值得商榷。这是因为,艾宾浩斯当初所使用的材料是无意义音节,为了确保无意义音节的无意义性,他曾专门考察了每一个音节的联想值,尽量选择完全无意义的音节。这在最大程度上保证了项目之间的相互独立和在集合中同等的地位。而我们现今使用的单个项目,

* 刘希平等。记忆监测与记忆成绩相关的实验研究,待发表。

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

其意义性要大得多,这样就使得不同的项目首先是不独立,其次是难等值,最后是对不同的被试含义不同。用这样的材料作研究,分析起来要格外小心。

4.2 研究方法的不统一

如前所述,记忆监测的研究方法是不统一的。首先,逐项评定和总体评定都在用。而比较研究刚好发现,逐项评定和总体评定的准确性存在差异,总体评定似乎更准确。究竟用哪种评定方法更合适?用不同的方法必然给若干研究项目的比较带来困难。可是如果都用总体评定,似乎也不太合适。因为第一,有些评定必须逐项进行;第二,逐项评定也许更接近评定的本质。因为总体评定可能平衡了某些波动;第三,总体评定对一些少许的变化可能缺乏敏感性。

但是,即使对总的研究方法进行了协调,仍然存在着其他的问题:记忆监测是考察准确性还是考察自信度(评定等级)?有些研究者已经意识到这一问题的严重性,对准确性和自信度进行了明确区分。但仍然有大量的研究报告对准确性还是自信度含糊其词。

最后,在评定准确性时是用相关系数还是用两点距?也许这只是一个小问题。但也足可以导致比较研究的混乱。例如,用两点距作为衡量准确性的指标,再求与记忆成绩的相关,用以说明记忆监测与记忆行为之间的关系还说得通,但是存在着局限,因为两点距只反映预见情况与记忆成绩之间的距离,而不是位移,这样就使得所得到的数据失去了方向性。但如果用评定等级与记忆行为的相关作为准确性的指标,那么这个相关值就不能反映记忆监测与记忆成绩的关系,而只是反映了记忆监测的准确性。要考察记忆监测的准确性与记忆行为之间的关系,即要对记忆监测的预见效度进行评估,还要用此相关值与另外的记忆行为再求相关才行。

4.3 建立模型的薄基础

要建立记忆监测的理论模型,还缺少以脑损伤病人和各种类型儿童为被试的记忆监测的研究,因为这样的研究无疑会给我们一些直接的证据。同时,记忆监测与神经生理的研究相结合,必将获得更多的突破。现有的研究发现前额叶与元记忆关系密切,与元记忆的本质和先前对额叶机能的认识相一致。如何利用先进的脑认知成像技术对记忆监测进行神经生理研究,是今后记忆监测研究应当更多关注的内容。

4.4 监测地位的难归属

假定上述所有的困惑都已经解决,那么,还有一个非常关键的问题:记忆监测究竟是状态的还是特质的?这等于在说:记忆监测是一种能力呢,还是一种相对无关紧要的心理品质?或者在说,对记忆监测如此投入地研究是否值得?

在记忆监测研究的初期,人们很自然认为记忆监测是一种能力。但后来有人用严格的实验证实了记忆监测的不稳定性,发现记忆监测随情境和任务的改变而不断变化。这给研究者提出了新的研究课题,记忆监测究竟是随状态而改变的,还是稳定的?在什么情况下才是稳定的?如果记忆监测是不稳定的,它的活动机制是什么?与人的能力之间是什么关系?所有这一切,都将是今后短时期内记忆监测研究的主题。

5 参考文献

1 Thomas A. Schreiber. 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

2 R. Jacobb Leonesio, and Thomas O. Nelson. Do different metamemory judgments tap the same underlying aspects of memory?. Journal of Experimental Psychology: Learning, Memory, & Cognition, 1990, 16(3): 464 - 470

3 Annette Farrant, Jill Boucher, and Mark Blades. Metamemory in children with Autism. Child Development, 1999, 70(1): 107 - 131

4 Nelson, T. O. A comparison of current measures of the acuracy of feeling-of-knowing predictions. Psychological Bulletin, 1984, 95: 109 - 133

5 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

6 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

7 William L. Kelemen, Perter J. Frost, and Charles A. Weaver III. Individual differences in metacognition: Evidence against a general metacognitive ability. Memory & Cognition, 2000, 28(1): 92 - 107

8 刘耀中, 郭力平. 记忆监测与人格特征相互关系的实验研究. 心理科学, 2000, 23(1): 20 - 23

9 Giuliana Mazzoni, and Thomas O. Nelson. 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

10 刘希平, 唐卫海. 回忆准备就绪程度的判断的发展. 心理学报, 2002, 34(1): 56 - 60

11 Bauer, R. H., and Kilbey, M. M. Metamemory of alcoholic Korskoff patients. Society for Neuroscience Abstracts, 1984, 10: 318

12 杨治良, 杜建政. FOK: 是线索熟悉, 还是目标提取? 心理学报, 2000, 32(3): 241 - 246

13 罗劲, 林仲贤. 元记忆对情节记忆和语意记忆的监测(心理科学), 1999, 22(6): 481 - 483

14 Hager W, Hasselhorn M. Memory Monitor and Memory Performance: Linked closely or loosely? Psychology Research, 1992, 54: 110 - 113

15 Lori E. James, and Deborah M. Burke. Phonological priming effects on word retrieval and Tip-of-the-Tongue experiences in young and old adults. Journal of Experimental Psychology: Learning, Memory, & Cognition, 2000, 26(6): 1378 - 1391

© 1994-2013 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net