心 理 学 报 2008,40(9):961 ~ 968
Acta Psychologica Sinica DOI:10.3724/SP.J.1041.2008.00961

理解评估与成绩预测:两种不同的元理解监测形式*

陈启山[1,2] 李 利[3]

([1]华南师范大学心理应用研究中心,广州 510631)

([2]香港中文大学教育心理系,香港) ([3]华南师范大学国际文化学院,广州 510631)

摘 要 探讨理解评估与成绩预测与各种强化元理解监测线索的认知任务的关系。结果发现,理解评估与成绩预测的判断值偏离标准测验成绩的程度受监测线索强化方式的调节;主动强化监测线索比被动强化更能提高理解评估和成绩预测的精确性;精确的理解评估或成绩预测所需的线索不同,利用同一线索评估理解或预测成绩,其精确性也不同。这一结果挑战了元理解监测的一维观,表明理解评估与成绩预测涵盖了元理解监测不同方面的心理特征。

关键词 元理解监测;精确性;理解评估;成绩预测

分类号 B842.5

1 引言

从文字符号获得意义的阅读理解不仅是一个复杂的认知过程,还是一个元认知过程[1]。阅读理解中的元认知活动被称为元理解(metacomprehension),它主要关注读者对自己阅读过程及结果的监测(monitoring)与控制(control)。元理解监测是元理解控制的前提,是目前元理解研究的热点。

元理解监测是指读者的元认知水平根据其从认知水平获得的信息对阅读理解过程及结果的评估判断,研究中通常用被试对所读文章理解程度的评估或对阅读理解测验成绩的预测对其进行测量[2,3]。元理解监测是一种学会感判断,即个体对已学材料的掌握程度或学习程度的评估判断,目前普遍认为它是个体运用内部线索、外部线索及记忆线索等做出的推论[4]。元理解监测的主观判断值与标准测验(criterion test)成绩间的一致程度被称为元理解监测的精确性(accuracy):如果几篇文章被评定的等级顺序与它们在标准测验中得分的高低顺序一致,说明读者能辨别或区分各篇文章之间的差异,即精确性高;反之,则精确性低。自我调控学习理论认为学习是主动的建构过程,有效的学习者或读者具备精确的监测力[5]。然而,20余年的元理解研究却表明:读者的元理解监测虽然高于随机猜测水平达到一定程度的精确,但仍不够十分精确,以Gamma系数来表示,其均值仅为0.25左右(最高为1,最低为-1),且受各种变量的影响[6~8]。

元理解监测的研究源于元记忆的学会感判断研究,后者通常是指学完"线索词-目标词"后,被试在呈现线索词的情况下对能正确回忆目标词的可能性的估计,其指向非常明确。相对而言,元理解监测则要复杂的多,从其概念内涵上看,它既可以是被试对所读文章理解程度的评估[2],也可以是对随后阅读理解测验成绩的预测[3],一般的文献综述也未将二者作严格区分[6~8],本研究将它们分别称为理解评估(rate comprehension)与成绩预测(predict performance)。

梳理以往的元理解研究,我们发现研究者不仅在理论上未对理解评估与成绩预测做严格区分,在实证研究中也往往是择其一作为元理解监测的操作指标,如Anderson、Griffin、Schommer、Thiede和Walczyk等人的研究用理解评估来度量被试的元理解监测[9~14];而Glenberg、Maki、Miesner与Weaver等人则用成绩预测作为元理解监测的操作定义[3,7,15~18]。Dunlosky和Rawson等人则交替或混合使用二者作为元理解监测的操作化指标[19,20]。Moore等人虽然区分了理解评估和成绩预测,却将二者合并成为一个指标做进一步的统计分析以考察

收稿日期:2008-03-03

通讯作者:陈启山,E-mail:kaisanchan@yahoo.com.cn

元理解监测与阅读理解的关系,也就是说他们虽然对二者做了区分,但仍然简单的将之合并,坚持了元理解监测的一维观[21]。

元理解监测的一维观认为理解评估和成绩预测是可以互相替换的,本研究不认同这一观点,而主张理解评估与成绩预测是两种不同的元理解监测形式,它们涵盖并体现了元理解监测不同方面的心理特征,对其应加以区分和比较。之所以将理解评估和成绩预测看作元理解监测的两个维度,而不赞成元理解监测的一维观,是因为二者既有共同点,又有很大的差异。

理解评估和成绩预测均是读者借助不同的线索,如内部线索、外部线索和记忆线索做出推论性判断,且都以阅读理解测验作为评价其精确性的标准。二者的差异主要表现为其内涵与指向的不同以及由此决定的判断时所检索、提取和利用的线索的不同。一方面,不同形式的元理解监测所利用的线索不同,这会影响被试的监测判断及其精确性,所以精确的理解评估或成绩预测需要不同的线索;另一方面,利用同一种线索做不同形式的监测判断,被试的判断及其精确性也可能存在较大差异。而这正是理解评估与成绩预测涵盖元理解监测不同方面的心理特征这一主张的逻辑起点。

根据线索模型,操控读者监测判断时所检索、提取或利用的线索会影响其监测判断的精确性。研究发现,读者在阅读时或阅读后从事强化监测线索的认知任务加工,如写摘要、写关键词、自我解释、画概念图、重读、前测等,可以在一定程度上提高元理解监测的精确性[9,10,13~15,22,23]。

Thiede 等人沿用元记忆中学会感判断延迟效应的研究范式,探讨了元理解监测的延迟效应,发现被试读完文章后延迟一段时间总结文章并据此写摘要或关键词可显著提高元理解监测的精确性,Gamma 系数高达 0.7[2,9,12];有趣的是,他们还发现此时让被试思考文章并不能提高元理解监测的精确性[13]。我们认为,思考文章未能提高元理解监测的精确性,可能是由于它既不能保证被试调动足够的认知资源去思考,也不能保证思考的结果能得到有效表征;被试思考的方式、内容与结果若能得到有效控制,则可能会提高监测的精确性。进一步的,我们假设监测判断前能促进读者对文章信息进行建构和表征的认知任务,无论其形式如何,如前测,均有可能强化元理解监测的线索并提高其精确性;而且这些强化监测线索的认知任务对元理解监测精确性的影响会受元理解监测形式的调节。

然而,前测对元理解监测精确性的促进作用并未有一致结论。如 Maki 等人发现前测练习组与无前测练习组的元理解监测精确性并无显著差异[24];在她的另一个探讨元理解监测延迟效应的研究中,Maki 让被试每读完一篇文章后即时或读完几篇文章后延迟预测测验成绩,标准测验也采用即时和延迟两种方式施测,结果发现却延迟组的元理解监测并不比即时组精确[25]。而 Glenberg 等人却发现在监测判断之前做前测练习可以提高被试元理解监测的精确性,而且前测与标准测验在内容上越接近其效果越明显[15];Thomas 等人也发现在监测判断前做排列句子的前测练习利于被试精确地预测概念题的成绩,而做填补字母的前测练习则利于对细节题成绩预测的精确性[26,27]。可见,前测对元理解监测精确性的影响受其他变量的调节,除了研究者所强调的前测与标准测验的相似性、前测的内部或外部反馈等因素外,我们认为,前测所强化的监测线索与监测形式是否一致也是一个重要因素。

本研究将探讨关键词与前测这两种强化监测线索的认知任务对理解评估与成绩预测精确性的影响。我们认为,监测线索与监测形式的一致性对元理解监测的精确性至关重要,关键词和前测所强化的监测线索是不同的,会对理解评估和成绩预测的精确性产生不同的影响:关键词有助于被试做出精确的理解评估但无助于成绩预测,前测则利于被试做出精确的成绩预测,但未必能做出精确的理解评估。

对学习材料进行主动加工时的学习效果比被动加工的效果要好,而且这一效应具有跨学习任务类型、测验类型、被试年龄等实验条件的稳定性,这被称为生成效应(generation effect)[28~30]。监测线索的不同强化方式对元理解监测精确性的影响是否也存在这一效应呢?本研究将关键词与前测分别设计成主动和被动加工的形式,考察理解评估与成绩预测的精确性是否受关键词与前测加工方式的影响,以回答这一问题。主动加工是指被试自己写关键词和做自测,而被动加工是指读已写好的关键词和读带有正确答案的前测题。

简言之,本研究的主要目的是探讨理解评估与成绩预测是否涵盖并体现了元理解监测不同方面的心理特征,我们通过考察强化监测线索的各种认知活动处理是否会对二者的精确性产生不同影响来达到这一目的。若接受同样的强化监测线索处理,理

解评估与成绩预测的精确性出现了差异,或者精确的理解评估或成绩预测所需要的线索不同,则表明二者出现了分离,是两种不同的元理解监测形式。

2　方法

2.1　被试

165 名华南师范大学的本科生,被随机分配到 8 个实验组。实验需时约 60min,被试完成实验后获得人民币 20 ¥ 为酬劳。

2.2　材料

2.2.1　阅读材料　篇幅介于 987 ~ 1180 个汉字的 6 篇说明文,主题涉及咖啡与茶、超声波清洗、太阳风、绿色手机、沙尘暴、疾病与瘟疫等。预实验中,121 名被试用 7 级量表评估各篇文章的加工难易度,结果显示 6 篇文章难度的均值在 4.61 ~ 5.14 之间,评分者信度为 0.81,说明被试对这些文章的领域熟悉性没有差异。

2.2.2　测验材料　72 个四择一选择题,每篇文章有 12 个题,细节题与推理题各半。细节题是基于文章某句话或某段而设计的,主要用于考察篇章理解的文本表征(textbase representation);推理题则是需要被试统筹全文或某些段落借助背景知识进行推理才能正确作答的题目,主要考察读者的情景模型(situation model)[31]。所有题目被分成对称的两半,分别用于部分被试的前测与所有被试的标准测验。

邀请阅读心理学研究者和语文教师对测验题目进行评定,这些题目被认为能有效测量读者对各篇文章的理解,有较好的效度。在 121 名被试中施测 72 个题目,做基于项目的方差分析*,结果表明各篇文章的测验难度没有差异($Fi(5, 66) = 0.95$, $MSE = 0.05$, $p = 0.45$)。将标准测验与前测题分开做同样的分析,也发现各篇文章标准测验的难度差异($Fi(5, 30) = 0.61$, $MSE = 0.04$, $p = 0.69$)以及前测题的难度差异($Fi(5, 30) = 0.65$, $MSE = 0.03$, $p = 0.66$)均不显著。

2.3　设计与程序

采用 2 × 2 × 2 的组间设计,操控三个自变量:(1)认知任务的类型:关键词与前测;(2)认知任务的方式:主动与被动;(3)元理解监测的形式:理解评估与成绩预测。实验程序用 E - prime 写成,具体流程如下。

阶段 1,指导语。各组被试被告知需跟随电脑程序逐步进行,接受各种实验处理。实验前鼓励被试就实验程序提问,以保证每位被试了解实验要求。

阶段 2,阅读文章。随机呈现 6 篇文章供被试阅读,每篇文章按每 400 字 90 秒的速度逐页呈现,被试在此时限内可按键进入下一页阅读;超过此时限,电脑将自动进入下一页,不论被试是否读完;已呈现过的各页,不能返回重看。

阶段 3,强化监测线索的处理。读完所有文章后,写关键词组依次对照各篇文章的标题思考文章大意并写下 5 个关键词,文章标题的呈现顺序跟被试阅读该文的顺序相同;读关键词组则要求认真阅读主试根据预实验制作的关键词并思考文章大意,其呈现顺序与被试阅读文章的顺序一致。做前测组进行常规测试,共 36 个题;读前测组则学习带有正确答案自测题;各篇文章测验的呈现次序与被试阅读文章的顺序一致。

阶段 4,元理解监测。理解评估组对照文章标题用 7 级量表评定对各篇文章的理解程度:越靠近 0 表示越没有读懂,越靠近 6 表示读的越懂。成绩预测组则对照着各篇文的标题按数字键预测测验成绩:0 表示自己没有把握答对该文 6 个题目中的任何一个题目;3 表示有把握答对 3 个题目;6 表示有把握答对全部 6 个题目;依次类推。各篇文章被评估的顺序与被阅读的顺序相同。

阶段 5,施测标准测验。共有 36 题,每篇文章有 6 个题目,细节题与推理题各半。各篇文章测验题的呈现顺序与被试阅读文章的顺序相同。被试按键做答,每个题目均须作答。测验完成后电脑自动反馈测验总成绩,并宣布实验结束。

3　结果

3.1　元理解监测的判断值与标准测验成绩

元理解监测精确性反映的是监测判断值与标准测验成绩的关系,故先对此二者进行考察,并考察了被试的主观判断偏离标准测验成绩的程度,即偏差(bias)。各变量的描述性统计见表 1。

---

* 统计检验的目的是通过样本数据推论总体,被试的选取与刺激项目的选择是两种不同性质的抽样,由此衍生了两类方差分析:基于被试的方差分析(ANOVA by subjects, Fs)与基于项目的方差分析(ANOVA by items, Fi),它们分别以被试与刺激项目为随机变量,结论可推广到各自的总体,前者的使用比后者要广,除非特别标识,本文报告的 F 值均为前者。

表 1 各研究变量的均值与标准误

	n	监测判断值	测验正确率(%)	偏差	Gamma 系数
理解评估					
写关键词	21	4.17(0.17)	66.93(2.10)	0.15(0.18)	0.55(0.11)
读关键词	19	4.44(0.18)	66.96(2.20)	0.42(0.19)	0.11(0.11)
做前测	19	4.04(0.18)	72.66(2.20)	-0.33(0.19)	0.38(0.11)
读前测	22	4.18(0.17)	64.27(2.05)	0.33(0.18)	0.10(0.10)
成绩预测					
写关键词	19	4.04(0.18)	65.35(2.20)	0.12(0.19)	0.36(0.11)
读关键词	21	4.26(0.17)	69.31(2.10)	-0.18(0.18)	0.13(0.11)
做前测	22	4.30(0.17)	66.67(2.05)	0.05(0.18)	0.72(0.10)
读前测	22	4.03(0.17)	64.39(2.05)	0.17(0.18)	0.38(0.10)

方差分析显示任务类型、任务方式与监测形式对监测判断值的主效应及所有交互作用均不显著，说明无论对理解程度的评估还是对测验成绩的预测，被试所做判断的数值大小不受各实验处理的影响。

任务类型与任务方式对标准测验成绩的交互作用显著，$F(1, 157) = 5.98$，$MSE = 92.18$，$p < 0.05$。简单效应检验显示：任务类型为关键词时，主动加工与被动加工时测验成绩的差异不显著；而任务类型为前测时，主动前测时的成绩（$M = 69.66\%$，$SE = 1.50$）显著高于被动前测（$M = 64.33\%$，$SE = 1.45$），$F(1, 157) = 6.53$，$p < 0.05$。简言之，被试写关键词与读关键词对标准测验成绩影响的差异不大；但相对读有正确答案的前测，做前测能显著提高标准测验的成绩。

偏差的计算公式为 $B = \sum (f_j - d_j)/n$，其中 f 为判断值，d 为成绩，偏差的绝对值表示判断偏离测验成绩程度的大小，偏差的正负表示这一偏离是过高还是过低自信[3]。方差分析显示任务方式与监测形式的交互作用显著，$F(1, 157) = 4.45$，$MSE = 0.71$，$p < 0.05$。简单效应检验发现：评估理解时，主动任务加工的偏差（$M = -0.09$，$SE = 0.13$）与被动加工的偏差（$M = 0.37$，$SE = 0.13$）差异显著，$F(1, 157) = 5.98$，$p < 0.05$；而预测成绩时，主动加工的偏差（$M = 0.09$，$SE = 0.13$）与被动加工（$M = -0.01$，$SE = 0.13$）的差异不显著。换言之，评估理解时，被试接受被动的认知任务比主动任务更容易高估自己的阅读理解水平；但预测成绩时，无论接受主动还是被动的任务加工，其预测偏离测验成绩的程度均不大。

3.2 元理解监测的精确性

用 Gamma 系数测量元理解监测的精确性，它是目前使用最广的监测精确性指标[32]。若两篇文章评定值的大小顺序与其标准测验成绩的顺序相同，则这两篇文章为同序对；反之，则为异序对。同序对与异序对的数目用 N_1 与 N_2 表示，则 $\gamma = (N_1 + N_2)/(N_1 + N_2)$，其取值范围为 $[-1, 1]$，值越大精确性越高。Gamma 系数是一种个体内（intrapersonal）相关系数，测量的是被试区分不同项目间差异的能力，各组被试 Gamma 系数的描述统计见表 1 及图 1。

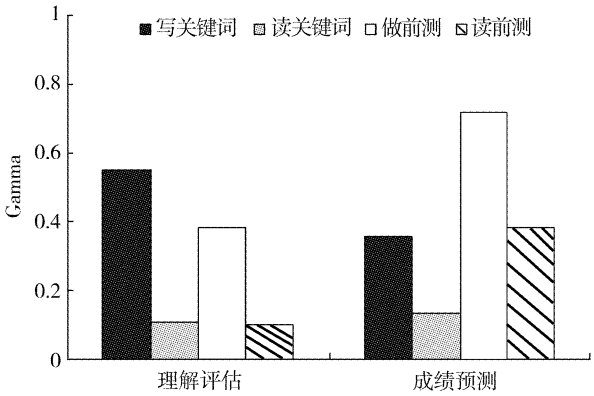

图 1 元理解监测的精确性

方差分析显示任务方式的主效应显著，主动的任务加工（$M = 0.50$，$SE = 0.05$）比被动加工（$M = 0.18$，$SE = 0.05$）更利于元理解监测精确性的提高，$F(1, 157) = 18.10$，$MSE = 0.24$，$p < 0.001$。也即，无论理解评估还是成绩预测，写关键词和做前测时的精确性要高于读关键词和读前测时的精确性，这一结果验证了元理解监测的生成效应假设。

任务类型与监测形式对监测精确性影响的交互作用显著，$F(1, 157) = 6.98$，$MSE = 0.24$，$p < 0.01$，说明理解评估和成绩预测这两种形式的元理

解监测的精确性受任务类型的调节。

对交互作用的进一步简单效应检验显示,当元理解监测形式为成绩预测时,被试接受前测处理时的精确性($M = 0.55, SE = 0.07$)高于关键词处理时的精确性($M = 0.24, SE = 0.08$),$F(1, 157) = 8.39, P < 0.01$。当元理解监测形式为理解评估时,接受关键词处理时的精确性($M = 0.33, SE = 0.08$)略高于前测处理($M = 0.24, SE = 0.08$),但不显著。换一个角度来看,任务类型为前测时,成绩预测的精确性显著高于理解评估,$F(1, 157) = 8.58, p < 0.01$。不但主动前测时成绩预测的精确性高于理解评估的精确性($F = 5.15, p < 0.05$),被动前测时二者的差距也边缘显著($F = 3.49, p = 0.06$)。任务类型为关键词时,理解评估的精确性略高于成绩预测的精确性,但不显著。

综上所述,主动强化监测线索比被动强化更利于元理解监测精确性的提高;理解评估与成绩预测的精确性受强化监测线索的任务类型的调节。

4　讨论

无论理解评估还是成绩预测,各篇文章的判断值没有明显差异,这主要是由于阅读材料间的领域熟悉性或难度对被试而言差异不大造成的。Dunlosky 等人发现成绩预测与材料的加工容易性有关,被试对难度差异较大文章的成绩预测值存在显著差异,但对难度差异较小文章的成绩预测值则差异不大[19,20]。结合本研究的发现,可见不但成绩预测如此,理解评估也是如此。虽然理解评估与成绩预测的判断值无显著差异,但并不能因此而断言二者可互换:一方面,根据差异不显著而得出虚无假设为真的做法并不可靠;另一方面,如果二者的判断值没有差异但其精确性却显著不同,则更能说明它们体现了元理解监测不同方面的特征。

强化监测线索的认知任务对标准测验成绩的影响受其加工方式的调节。一方面,关键词处理不影响标准测验成绩,这与 Thiede 等人的发现类似,即被试延迟写关键词能提高元理解监测的精确性,但不能提高标准测验的成绩[12]。这也正如 van den-Broek 关于篇章理解的风景模型(landscape model)所预测的那样,被试写的关键词是阅读时激活水平最高的及与之有较强关联的概念,是阅读表征中的重要节点,但未必与阅读测验有很强的关联[33,34],因为阅读测验不仅考察深层理解,还涉及具体细节的记忆。另一方面,主动做前测较比动读前测更能提高标准测验的成绩,这一结果可以用测验效应(test effect)来解释,测验不仅可以检验学习效果,更可以促进随后的学习及减少遗忘[35,36]。本研究中被试做前测产生的练习效应与内部反馈对标准测验成绩的提高作用正是这一效应的反映;而读有正确答案的前测题时,被试缺乏主动的加工不能产生有效的反馈,所以在面临新题目时的表现要逊色得多,这一结果对教育教学有启发意义。

理解评估与成绩预测的判断值偏离标准测验成绩的程度受监测线索强化方式调节。在评估理解时,被动的认知任务加工较主动的认知任务加工更容易让被试过高自信,这说明在阅读结束之后让被试从事被动的认知加工任务可能会给被试的理解评估带来一种错觉,将自己没有读懂的文章错认为已经理解,然而,待到正式测验时测验成绩并不理想,由此必然表现出过高自信;而主动的任务加工方式则在一定程度上抵消了这种过高自信。在预测成绩时,无论进行主动的还是被动的认知任务加工,被试的判断偏离测验成绩的程度均不大,这说明被试对阅读理解测验成绩的预测虽然会受任务加工方式的影响出现偏差,但其偏离程度与理解评估并不一样。这一结果初步表明理解评估和成绩预测涵盖了元理解监测的不同方面的心理特征。

监测线索的强化方式对元理解监测精确性的主效应显著,主动写关键词与做前测相对于读关键词与读前测,无论理解评估还是成绩预测,其精确性均有较大提高。这不但说明同一形式的元理解监测接受不同加工方式的强化监测线索的认知任务其精确性会出现不同;还验证了本研究提出的元理解监测的生成效应,即主动的进行强化监测线索的认知活动比被动接受这些强化线索更利于被试对这些线索的有效提取与运用,从而更有效的区分和辨别已读懂的文章和未读懂的文章的差异,提高元理解监测的精确性。可见,生成效应不仅适用于学习、记忆等认知层面[28~30],还适用于元理解等元认知层面。

更为重要的是,强化监测线索的任务类型与监测形式对监测精确性的交互作用显著。从一个角度看,同一形式的监测判断的精确性与两类强化监测线索的认知任务的关系不同:读者做成绩预测前,接受前测处理比接受关键词处理更能提高成绩预测的精确性;理解评估时则有相反的趋势。从另一个角度看,同一类强化监测线索的认知任务对两种形式的监测判断的精确性的影响也不一样:前测处理有助于被试精确的预测其阅读理解测验的成绩,但不

能提高其理解评估的精确性,不仅主动前测如此,被动前测也是这样;关键词处理则有相反的趋势。简言之,精确的成绩预测或理解评估需要借助不同的线索;同一线索对成绩预测和理解评估精确性的影响不同。

无论成绩预测还是理解评估均是被试借助各种线索做出的推论,精确的成绩预测或理解评估需借助不同的线索。成绩预测是前瞻性的(prospective),在不知晓标准测验特征的情况下,读者依据以往的测验经验对标准测验性质、形式及难度等的预期有助于其做出精确的预测;在知晓标准测验特征或类似的前测时,被试对标准测验难度及能否正确作答的估计或前测的内部反馈等均可以帮助读者做出精确的预测。本研究验证了这一点,成绩预测对于前测任务更为敏感,尤其是主动的做前测,这是因为被试可以根据前测以其反馈信息,获取诸如测验难度及能否正确作答类似题目的预期等利于成绩预测的线索。理解评估是回溯性的(retrospective),能提高其精确性的线索与阅读加工及阅读结果的记忆表征密切相关,特别是对情景模型的提取。本研究设计关键词处理来强化理解评估的线索,关键词处理所提取的信息是文章的主要概念读者记忆中的表征,这些线索有助于读者做出精确的理解评估。然而,本研究发现关键词处理时理解评估的精确性高于前测处理时理解评估的精确性但不显著。这一方面可能是因为任务加工方式对冲了任务类型对理解评估精确性的影响,主要表现为做前测一定程度上提高了理解评估的精确性,而读关键词则未能提高理解评估的精确性。另一方面,关键词处理对理解评估精确性的促进作用可能受监测线索与标准测验的相似性这一因素的调节。本研究发现采用与前测类似的标准测验时,前测可以显著提高成绩预测的精确性,那么,如果采用与关键词相似的标准测验(如写摘要或理解问答题),接受关键词处理时的理解评估的精确性可能会高于其他处理,这一假设有待进一步的研究加以检验。

监测线索是读者元理解监测的重要依据,同一类线索对成绩预测和理解评估精确性的影响会不同。根据加工分离法的实验逻辑,如果接受同样的强化监测线索的任务处理时理解评估与成绩预测的精确性出现差异,如该任务影响一种监测形式的精确性但不影响另一监测形式的精确性,或该任务对二者精确性影响的方向不同,如一个为正向而另一个为反向,就表明理解评估与成绩预测出现了分离。本研究中,被试接受前测处理时成绩预测的精确性显著高于理解评估的精确性,即前测处理导致了成绩预测与理解评估的分离,这说明理解评估和成绩预测涵盖了元理解监测不同方面的心理特征,是两种不同的元理解监测形式。监测线索对不同形式的元理解监测的精确性的影响可能受标准测验特征的影响,Maki 等人发现用不同的测验评定同一个元理解监测判断时,它们的精确性不同[16,17],而这也正为我们获得更多的证据来支持理解评估与成绩预测是两种不同形式的元理解监测这一观点指明了研究方向,即不仅要考察理解评估和成绩预测与不同的强化监测线索的认知任务的关系,还应该进一步探讨不同形式的元理解监测与强化线索的任务之间的关系是否受标准测验性质或形式的调节。

综上所述,理解评估与成绩预测的精确性受强化监测线索的任务类型与方式的影响;此外,理解评估与成绩预测的判断值偏离标准测验成绩的程度也受线索强化方式的影响。这一发现挑战了元理解监测的一维观,表明理解评估与成绩预测体现了元理解监测不同方面的心理特征,是两种不同的元理解监测形式,不应简单的合二为一或互换。

5 结论

理解评估时,被动强化监测线索比主动强化更容易导致过高自信;成绩预测时,无论主动还是被动强化监测线索,被试预测的偏差均不明显。

理解评估与成绩预测的精确性受监测线索强化方式影响,主动强化监测线索比被动强化更能提高二者的精确性。

理解评估与成绩预测的精确性受强化监测线索的任务类型的调节:读者预测成绩之前,接受前测处理比关键词处理更能提高预测的精确性;而评估理解之前,接受关键词处理则略高于接受前测处理时的精确性。

理解评估与成绩预测是两种不同的元理解监测形式,它们涵盖了元理解监测不同方面的心理特征。

参 考 文 献

1 Rumelhart D E. Toward an interactive model of reading. In: Ruddell R B, Unrau N J, ed. Theoretical models and processes of reading (5th ed). Newark, DE: International Reading Association. 2004: 1149 ~ 1179

2 Thiede K W, Anderson M C M. Summarizing can improve metacomprehension accuracy. Contemporary Educational Psychology, 2003, 28(2): 129 ~ 160

3 Maki R H, Shields M, Wheeler A E, et al. Individual differences in absolute and relative metacomprehension accuracy. Journal of Educational Psychology, 2005, 97(4): 723 ~ 731

4 Koriat A. Monitoring one's knowledge during study: A cue - utilization approach to judgments of learning. Journal of Experimental Psychology: General, 1997, 126(4): 349 ~ 370

5 Zimmerman B J, Schunk D H. Self - regulated learning and academic achievement: Theoretical perspectives (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, 2001

6 Lin L M, Zabrucky K M. Calibration of comprehension: Research and implications for education and instruction. Contemporary Educational Psychology, 1998, 23(4): 345 ~ 391

7 Maki R H, McGuire M J. Metacognition for text: Findings and implications for education. In: Perfect T J, Schwartz B L, ed. Applied metacognition. Cambridge, UK: Cambridge University Press, 2002. 39 ~ 67

8 Dunlosky J, Lipko A. Metacomprehension: A brief history and how to improve its accuracy. Current Directions in Psychological Science, 2007, 16(4): 228 ~ 232

9 Anderson M C M, Thiede K W. Why do delayed summaries improve metacomprehension accuracy? Acta Psychologica, 2008, 128(1): 110 ~ 118

10 Griffin T D, Wiley J, Thiede K W. Individual differences, rereading, and self - explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. Memory & Cognition, 2008, 36(1): 93 ~ 103

11 Schommer M, Surber J J. Comprehension - monitoring failure in skilled adult readers, Journal of Educational Psychology, 1986, 78 (5): 353 ~ 357

12 Thiede K W, Anderson M C M, Therriault D. Accuracy of metacognitive monitoring affects learning of texts. Journal of Educational Psychology, 2003, 95(1): 66 ~ 73

13 Thiede K W, Dunlosky J, Griffin T D, et al. Understanding the delayed - keyword effect on metacomprehension accuracy. Journal of Experimental Psychology: Learning, Memory, and Cognition, 2005, 31(6): 1267 ~ 1280

14 Walczyk J J, Hall V C. Effects of examples and embedded questions on the accuracy of comprehension self - assessments. Journal of Educational Psychology, 1989, 81(3): 435 ~ 437

15 Glenberg A M, Sanocki T, Epstein W, et al. Enhancing calibration of comprehension. Journal of Experimental Psychology: General, 1987, 116(2): 119 ~ 136

16 Maki R H, Willmon C, Pietan A. Basis of metamemory judgments for text with multiple - choice, essay and recall tests. Applied Cognitive Psychology, 2008, 22 (in press)

17 Miesner M T, Maki R H. The role of test anxiety in absolute and relative metacomprehension accuracy. European Journal of Cognitive Psychology, 2007, 19(4 - 5): 650 ~ 670

18 Weaver C A. Constraining factors in calibration of comprehension. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1990, 16(2): 214 ~ 222

19 Dunlosky J, Baker J M C, Rawson K A, et al. Does aging influence people's metacomprehension? Effects of processing ease on judgments of text learning. Psychology and Aging, 2006, 21(2): 390 ~ 400

20 Rawson K A, Dunlosky J. Are performance predictions for text based on ease of processing? Journal of Experimental Psychology: Learning, Memory, and Cognition, 2002, 28(1): 69 ~ 80

21 Moore D, Lin L A, Zabrucky K M. A Source of metacomprehension inaccuracy. Reading Psychology, 2005, 26(3): 251 ~ 265

22 Wiley J, Thiede K W, Griffin T D. What does it mean to learn from and understand science text? Paper presented at the 2007 Annual Meeting of the American Educational Research Association, Chicago, IL, 2007

23 Dunlosky J, Rawson K A. Why does rereading improve metacomprehension accuracy? Evaluating the levels - of - disruption hypothesis for the rereading effect. Discourse Processes, 2005, 40(1): 37 ~ 55

24 Maki R H, Serra M. Role of practice tests in the accuracy of test predictions on text material. Journal of Educational Psychology, 1992, 84(2): 200 ~ 210

25 Maki R H. Predicting performance on text: Delayed versus immediate predictions and tests. Memory & Cognition, 1998, 26 (5): 959 ~ 964

26 Thomas A K, McDaniel M A. Metacomprehension for educationally relevant materials: Dramatic effects of encoding - retrieval interactions. Psychonomic Bulletin & Review, 2007, 14 (2): 212 ~ 218

27 Thomas A K, McDaniel M A. The negative cascade of incongruent generative study - test processing in memory and metacomprehension. Memory & Cognition, 2007, 35(4): 668 ~ 678

28 Bertsch S, Pesta B J, Wiscott R, et al. The generation effect: A meta - analytic review. Memory & Cognition, 2007, 35(2): 201 ~ 210

29 Bjork E L, DeWinstanley P A, Storm B C. Learning how to learning: Can experiencing the outcome of differential encoding strategies enhance subsequent learning? Psychonomic Bulletin & Review, 2007, 14(2): 207 ~ 211

30 Metcalfe J, Kornell N. Principles of cognitive science in education: The effects of generation, errors and feedback. Psychonomic Bulletin and Review, 2007, 14(2): 225 ~ 229

31 Kintsch W. Comprehension: A paradigm for cognition. New York: Cambridge University Press, 1998

32 Nelson T O, Narens L, Dunlosky J. A revised methodology for research on metamemory: Pre - judgment recall and monitoring (PRAM). Psychological Methods, 2004, 9(1): 53 ~ 69

33 van den Broek P, Rapp D N, Kendeou P. Integrating memory - based and constructionist processes in accounts of reading comprehension. Discourse Processes, 2005, 39(2 - 3): 299 ~ 316

34 Rapp D N, van den Broek P. Dynamic text comprehension. Current Directions in Psychological Science, 2005, 14(5): 276 ~ 279

35 Pashler H, Rohrer D, Cepeda N J. et al. Enhancing learning and

retarding forgetting: Choices and consequences. Psychonomic Bulletin & Review, 2007, 14 (2): 187 ~ 193

36 Roediger H L, Karpicke J D. Test - enhanced learning: taking memory tests improves long - term retention. Psychology Science, 2006, 17(3): 249 ~ 255

Rating Comprehension and Predicting Performance: Clarifying Two Forms of Metacomprehension Monitoring

CHEN Qi-Shan[1,2], LI Li[3]

([1] *Center for Studies of Psychological Application, South China Normal University, Guangzhou* 510631, *China*)

([2] *Department of Educational Psychology, The Chinese University of Hong Kong, Hong Kong, China*)

([3] *College of International Culture, South China Normal University, Guangzhou* 510631, *China*)

Abstract

Metacomprehension monitoring refers to the process of a reader monitoring and evaluating his or her understanding of a text. Participants may make metacomprehension monitoring judgments by either rating how much they understand a text, or predicting how well they can do in a comprehension test. Previous research seems to treat these questions as interchangeable and which one to use as a matter of convenience. Researchers have not investigated whether or not these questions may tap different metacomprehension monitoring processes. In this study, we compared two kinds of monitoring, namely, rating comprehension and predicting performance, by examining the way they were affected by two different orienting tasks that may provide different cues for the monitoring process.

165 college students at South China Normal University participated in this experiment. Three between - subject factors were manipulated. The participants were divided into keyword groups or pretest groups. The keyword group was divided into generating or reading group, the pretest group was divided into taking a pretest or reading the pretest with answers provided after reading all the six expository texts on a computer. Half of each group either rated their comprehension of each text or predicted their performance in a comprehension test on a 0 ~ 6 scale. All participants took a comprehension test on each text. Three questions assessed text - based knowledge (details available within a single paragraph of a text), and three inferential questions assessed knowledge of a situation model.

The results showed that there were no significant effects on magnitudes of JOL. The interaction effect between orienting tasks and activity of the task on criterion test performance was significant. The performance for the taking pretest condition was higher than the reading pretest condition, whereas the generating keyword groups did not differ from the reading keyword groups. As far as monitoring accuracy (Gamma) was concerned, the interaction effect between types of monitoring and orienting tasks was significant. The simple main effect tests revealed that when the participants predicted their test performance, the pretest condition led to more accurate judgments than the keyword condition, whereas when they rated comprehension, the reverse was found.

Metacomprehension monitoring accuracies as found with rating text understanding and predicting test performance were affected by two orienting tasks in different ways. The cues produced by taking a pretest may be useful for predicting the scores of a final test, but may not be useful for estimating how well a text is understood. The cues produced by generating keywords seem useful for readers to judge their understanding of one text relative to another, but may not be useful for them to calibrate their predictions of test scores. The results challenge the view that metacomprehension monitoring is a unitary process. Given our results, it is reasonable to conclude that rating comprehension and predicting performance may tap different aspects of metacomprehension monitoring.

Key words metacomprehension monitoring; accuracy; rate comprehension; predict performance