

试题难度的事前认知任务分析

邵志芳* 余 岚

(华东师范大学心理学系, 上海, 200062)

摘 要 试题的难度常常用试题的通过率来表示, 这种事后检验的方法虽然简单易行, 其数量指标严格确定, 但是对考生水平相当敏感, 也不能在试题设计中起到主导作用。本研究根据认知任务分析的基本原则, 设计了对试题难度进行事前评定的框架和严格的实施程序, 并以高等数学试题为分析对象进行了检验。根据数学试题的特点, 设定了要素数量、要素辨识难度、原理数量、原理辨识难度和认知负荷等指标。结果显示, 各试题事前难度值与其通过率之间存在极其显著的关联。评定指标之间、评定者之间也存在极其显著的一致性, 说明本研究的事前评定方法具有充分的效度和信度。

关键词: 试题难度 认知任务分析 事前难度

1 引言

关于试题难度的研究有非常重要的理论意义和实践意义, 国内外许多心理学家从事这方面的研究, 提出了不少看法和理论。关于试题难度的标定就有多种方法。这些方法大致可以分为事后标定和事前标定两大类。

目前得到广泛运用的是事后标定法中的通过率法。这是测量学家惯用的方法, 以试题的通过率来表示该试题的难度。事后检验法简单易行, 数量指标严格确定, 但是对考生水平相当敏感, 因为同一试题在不同水平的考生中得到的通过率并不相同。这样带来的弊端是显而易见的: 每次考试都不可避免地有一些题目缺乏效度和信度。另外, 事后标定不能在试题设计中起到主导作用。因此, 很多心理学家试图设计试题难度的事前标定方法。

Newell 和 Simon(1972) 提出了问题搜索空间描述法^[1], 其中, 问题的搜索空间是影响问题难度的最重要的因素。但是, 他们没有提出搜索空间大小的准确计算方法以及与问题难度的数量关系。Funke(1991) 提出用以下 6 个特征来表达问题的复杂性或难度: 不透明性、多目标性、情境的复杂性、变量间的联系、动态性和延迟性^[2]。信息加工心理学家提出的负荷估计法常常根据解题过程中的认知负荷或记忆负荷来说明问题难度。例如, Kotovsky, Hayes 和 Simon(1985) 提出问题难度与记忆负荷有关。他们还认为, 影响问题难度的因素还包括规则学习的难度、规则运用的难度、现实世界知识(实际生活经验)与问题规则的吻合程度等^[3]。另外, 还有一些心理学家(尤其是数学心理学家)则用题目的结构特点、语言陈述方式、图式的可得性等解释问题难度。例如, De Corte 和 Verschaffel(1981) 从关系复杂性的角度分析问题难度^[4]。而问题的关系复杂性必然影响主体对问题的表征和解题过程, 尤以对表征的影响最大。Mayer(1981) 认为, 任何应用题都可以描述为由一系列独特命题组成的模板^[5]。这些命题包括赋值命题、关系命题和问句命题, 其回忆难度都是不同的。后来, Low 和 Over(1992) 进一步探讨了这种模板的等级顺序问题^[6]。

国内也有一部分研究者关注问题难度的标定。邵志芳等(1992, 1996) 陆续发表了关于概念难度和问题难度的研究报告^[7, 8]。辛自强(2003) 在综合前人研究的基础上提出了关系-表征复杂性模型。该模型区分了问题难度和问题解决难度两个概念, 认为二者的本质分别是关系复杂性和表征复杂性^[9]。

以上列出的研究都考察了影响问题难度的各种因素, 并提出了这些因素的变化特点, 以期从某个侧面来标定问题的难度。但是这些研究重在考察概念或问题的某些维度与其难度之间的关系, 并没有设计出标定问题难度的数量化指标。

解答试题是一个复杂的认知加工过程。如果能够从设计试题开始就对每道试题所需的认知加工有一个细致的分析, 将大大提高对于试题难度的事前分析的准确性, 同时也可以为教师从学生认知心理出发设计教法提供重要的指引。而能够帮助进行上述分析的, 就是近些年来兴起的认知任务分析(cognitive task analysis, 以下简称 CTA)。CTA 指的是一系列的认知心理学技术, 用于分析可见的任务行为, 获取作为该行为的基础的有关知识、思维过程和目标结构, 并帮助设计更有效地处理复杂情境的认知技能。DuBois 和 Shalin(2000) 在区分了不同种类知识的基础上提出了一个分析框架^[10]。这个框架要求区分陈述性知识、程序性知识、生成性知识、团队/自我知识(关于个人能力、策略或偏好等的知识)等。张璐璐(2005) 进一步探讨了如何用 CTA 建立认知过程的模型^[11]。

CTA 在对试题难度的评定和学生认知负荷的测定中正起到越来越大的作用。本研究试图建立对于试题难度进行事前认知任务分析的系统方法。

2 方法

2.1 问题选择

本研究从结构化程度较高的高等数学试题入手, 初步建立试题难度的事前分析框架。抽取华东师范大学 2004 级和 2005 级的高等数学(B) 期末试卷为试题来源。为了保证题目得分的准确性, 尽可能排除答题正确率中的猜测成分, 分别

* 通讯作者: 邵志芳, 男。E-mail: zfsiao@psy.ecnu.edu.cn

从两张考卷中剔除填空题和选择题的部分,最终选择 2005 级试卷(以下称“05 卷”)的 16 道试题(最高通过率为 94.29%,最低通过率为 12.38%)和 2004 级试卷(以下称“04 卷”)的 10 道试题(最高通过率为 79.03%,最低通过率为 14.52%)进行逐题分析。

2.2 指标体系

为了对试题难度进行事前认知任务分析,需设定一系列数量指标。根据数学试题的特点,设定如下指标:(1)要素数量——题目中给出的所有能使用的要素个数;(2)要素辨识难度——题中各要素能被解题者从题目中辨别的难度(1-5 分,分别表示“非常容易”到“非常难”);(3)原理数量——解题中需要的原理个数;(4)原理辨识难度——解题所需定理、公式、试题原型能被个体提取的难度(1-5 分,分别表示非常容易到非常难);(5)认知负荷——解题过程中认知能力各个成分(包括注意、记忆、表象、概括、推理、发散思维和决策能力等)的需要程度(1-5 分,分别表示“几乎不需要”到“非常需要”)。其中(2)(4)(5)三项为主观评定。

以上指标的总和就成为各试题的事前评定难度指标。

2.3 被试和评定者

华东师范大学心理学系 2005 级共 105 人完成 2005 学年第一学期高等数学期末试卷;2004 级共 62 人完成 2004 学年第二学期高等数学期末试卷。

2 名华东师范大学心理学系人类智能心理学方向研究生经培训后对上述试题难度进行评定。

2.4 评定步骤

对于每一道试题,2 名评定者均分别独立完成以下步骤:

(1)列出其标准解题过程;(2)分析其中涉及的已知条件、原理、公式等;(3)评价从试题表面信息中辨识出原理及其组成要素的难度,评价解题过程中对于各种认知能力的需求程度。

3 结果

2 名评定者对各题都给出事前评定难度,学生考试实际通过率则作为各题事后评定难度,这样,事前与事后难度的相关系数可以看作本研究评定方法的效度系数。结果见表 1。从中可见事前评定的试题难度与通过率呈极其显著的正相关($p < 0.01$)。

表 1 两名评定者对于 05 卷和 04 卷试题事前难度评定的效度系数

评定者	05 卷		04 卷	
	积差相关	等级相关	积差相关	等级相关
评定者 A	-0.775	-0.782	-0.683	-0.794
评定者 B	-0.816	-0.821	-0.833	-0.862
合并计算	-0.866	-0.890	-0.792	-0.832

为了检验各个评定指标间的一致性程度,分别计算两位评定者的各个指标间的肯德尔和谐系数,结果是:评定者 A 为 0.804(05 卷)和 0.791(04 卷);评定者 B 为 0.832(05 卷)和 0.907(04 卷),均达到极其显著相关($p < 0.01$)。

评定者间一致性系数(等级相关)为 0.701(05 卷)和 0.786(04 卷),均达到极其显著相关($p < 0.01$)。

4 讨论

表 1 中的相关系数体现了两位评定者对于 05 卷和 04 卷

试题事前难度评定的效度。结果是,试题的事前评定难度值越高,其通过率越低。这说明此种评定方法具有较高的预测效度。

解题者必须识别出试题所用的原型以及所用原理的过程。这一过程受到多种因素的影响,包括问题本身有哪些要素?其表征与教材的差别(识别难度)?解题过程涉及哪些概念、原理和技巧?其识别难度又如何?另外,解题过程中需要一些重要的认知能力相互配合,例如注意——细节洞察能力;记忆——短时记忆、长时记忆;思维——概括能力(转译)、逻辑推理能力、发散思维(尝试错误)与评判与决策能力等。这些方面都纳入本研究认知任务分析的指标体系中,从而保证了评定的效度。

从各指标间一致性和评定者间的一致性可以看出本研究中评定方法的可信度。这两个方面的一致性程度均达到极其显著的水平。

对于试题难度的认知任务分析是一个相当专业化的工作。评定者必须兼具认知心理学和试题所属学科两方面的素质,或者是来源于上述两个领域人员。本研究的两位评定者中,一位是心理学系本科毕业生,另一位是数学系本科毕业生,两人现都在攻读心理学硕士,主修认知心理学。同时,认知任务分析需严格按照程序进行。本研究获得的较高的效度和信度很大程度上源于上述两个原因。

今后的考试和教学实践中,对每一道试题,可以针对以上各方面指标,由多名心理学家和学科命题专家做出评价,确定该试题(进而整份试卷)的覆盖面、能力要求、认知负荷和难度等质量指标。这样在试卷编制阶段就能较好地把握试卷质量,甚至对教学工作产生一定的指导作用。

5 结论

本研究根据认知任务分析的基本原则,设计了对高等数学试题难度进行事前评定的框架和严格的实施程序。根据数学试题的特点,设定了要素数量、要素辨识难度、原理数量、原理辨识难度和认知负荷等指标。结果显示,各试题事前评定的难度值与其通过率之间存在极其显著的相关。评定指标之间、评定者之间也存在极其显著的一致性,说明本研究的事前评定方法具有充分的效度和信度。

6 参考文献

- Newell, A., & Simon, H. A. Human problem - solving. Englewood Cliffs, NJ: Prentice Hall, 1972.
- Funke, J. Solving complex problems: exploration and control of complex systems. In Sternberg, R. J. & Frensch, P. A. (Ed.), Complex Problem Solving: Principles and Mechanisms. Lawrence Erlbaum Associates, Inc. 1991, 185 - 222
- Kotovsky, K., Hayes, J. R., & Simon, H. A. Why are some problems hard? Evidence from Tower of Hanoi. Cognitive Psychology, 1985, 17: 248 - 294
- De Corte, E., & Verschaffel, L. Children's solution processes in elementary arithmetic problems: Analysis and improvement. Journal of Educational Psychology, 1981, 6: 765 - 779
- Mayer, R. E. Frequency norms and structural analysis of algebra story problems into families, categories and templates. Instructional Science, 1981, 10: 135 - 175

- 6 Low, R., & Over, R. Hierarchical ordering of schematic knowledge relating to area - of - rectangle problems. *Journal of Educational Psychology*, 1992, 1: 62 - 69
- 7 邵志芳, 杨治良. 影响概念难度的两个基本因素. *心理科学*, 1992, 6: 8 - 12
- 8 邵志芳等. 关于问题难度的实验研究. *心理科学*. 1996, 5: 278 - 281
- 9 辛自强. 关系 - 表征复杂性模型的检验. *心理学报*, 2000, 35 (4): 504 - 513
- 10 DuBois D., & Shalin, V. L. Describing Job Expertise Using Cognitive Oriented Task Analyses (COTA). In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.) *Cognitive Task Analysis*, 2000, 41 - 55. Mahwah, NJ: Lawrence Erlbaum Associates
- 11 张璐璐. 简历筛选认知过程的模型探索. 硕士论文(上海:华东师范大学), 2005

The Prior Cognitive Task Analysis of Test Item Difficulty

Shao Zhifang, Yu Lan

(Department of Psychology, East China Normal University, Shanghai, 200062)

Abstract The passing rate of a test item often serves as an index of its difficulty. This posterior index, though easy to obtain and definite, is susceptible to students' learning level. It cannot play a leading role in designing test items. The present research tried to design a prior-to-test index of test item difficulty. According to the principles of cognitive task analysis, a frame and a procedure to be strictly executed were mapped out for the prior assessment of higher mathematics test items. Taking into account the characteristics of math problems, we designed such indexes as number of elements in a problem, element identification difficulty, number of principles used in answering, principle identification difficulty and cognitive load. The results show that prior difficulty is most significantly correlated with the passing rates of math test items. High correlation also exists among sub-indexes in assessment and among evaluators, indicating sufficient validity and reliability of the prior assessment method developed by the present research.

Key words: test item difficulty, cognitive task analysis, prior difficulty

(上接第 688 页)

Envy from Social Comparison : Based on Different Measures

Yang Lixian¹, Zhang Jinkun^{2,3}

(¹ Department of Psychology, East China Normal University, Shanghai 200062)

(² Academy of Psychology and Behavior, Tianjin Normal University, Tianjin 300074)

(³ Department of Psychology, Fujian Normal University, Fuzhou 350007)

Abstract The authors used self-reported measures and Implicit Associated Test (IAT) to measure envy in a scenario in which the participants competed with two groups (high-achievers and average-achievers). The results show that there is no significant difference in explicit envy. The effect of IAT is significant, which means that the subjects have a higher level of envy of the average achievers than of the high-achievers. Compared with the high-achievers, the females have a higher level of implicit envy of the average achievers than the males, while there is no significant difference in explicit envy between them. No significant correlation has been found between implicit envy and explicit envy.

Key Words: Social comparison, Envy, Self-reported measure, IAT