

美国数学评估“测验项目和测验设计”中存在的问题及思考

邢丽丽

【摘要】2008年,美国教育部发布了《国家数学顾问团的报告》。报告中的“教学评估”专题主要基于美国成就测验的广泛应用这一背景,对测验的内容、测验成绩的分类、测验项目和测验设计提出建议。美国数学专家组通过调查研究发现,美国数学教学评估在“测验项目和测验设计”方面存在一些问题,如测验项目“措辞”上的问题、测验程序研发中出现的漏洞、测验项目上的争议等等。笔者对这些问题进行简单介绍,并对其进行思考,进而提出在测验编制和项目设计上的一些建议。

【关键词】美国 教学评估 测验项目 测验设计

【中图分类号】G51/57 【文献标识码】A 【文章编号】1674-1536(2008)12-0053-03

一、报告发布的背景

在20世纪,美国拥有无可匹敌的数学权威,这不仅仅因为美国拥有大量资深的数学专家;而且在当时,团队或个人数学能力的高低是衡量美国工程和科学能否得到长远发展以及财政领导是否具备较强领导能力的标尺之一;美国主流人口接受的数学教育程度也很高。但是许多年来,随着世界经济一体化进程的加快以及各国交流的频繁,美国从海外引入了大量的优秀人才,加上美国在数学领域和工程学领域缺乏本国技术人才,美国越来越依靠海外移民、临时出境后又返回的外侨科学家和工程学家,因此美国在数学、自然科学、工程学方面的独立性和领导地位逐渐被削弱。

与此同时,美国公民所具有的数学素养能力也开始引起人们的关注。根据某调查机构“飞利浦”(2007)的调查,78%的美国成年人不能对怎样计算借款利息作出准确解释,58%的美国人不能正确计算出一顿午餐10%的小费,等等。此外,“从调查研究中我们可以清楚地看出来,很大范围的学生和成年人在‘分数’这一数学技能的理解和应用上也存在困难”(Hecht, Vagi & Torgeson, 2007)。最近的美国国家教育进展评估(NAEP)表明,“27%的八年级学生不能准确地填涂一个矩形的 $\frac{1}{3}$,45%的人不能理解和运用数学字符”,而字符概念是进行分数除法运算所必需掌握的(美国教育部,2004)。更严重的是,

几乎有一半年龄在十七岁左右的孩子的数学能力极其低下,只能当一辈子“蓝领”。

美国政府和教育部意识到,美国在数学领域的领先地位开始动摇。因此,他们组织了一个数学专家组对本国的数学教育进行调查和研究。2008年3月13日,美国数学专家组向美国总统和教育部长提交了一份长达60余页的《国家数学顾问团的报告》,呼吁美国必须采取行动去加强美国公民在数学领域的学习,因为这一领域的成功关系到整个民族的兴衰。它告诉美国学生如何为学好数学作更充分的准备以及怎样才能在中等数学教育中有个好的开始。报告指出,学生为学好代数学而作好充分准备对于每一位学生和家来说都是至关重要的。报告一经发表,立即引起了各界的广泛关注。

美国《国家数学顾问团的报告》主要涉及数学课程、学习过程、教师教育、教学实践、教学资料、教学评估、数学教育研究等方面。报告中的“教学评估”专题主要基于美国成就测验的广泛应用这一背景,对测验的内容、测验成绩的分类、测验项目和测验设计提出建议。数学专家组一再强调:要改善对美国国家教育进展的评估以及对各州的评估,应更加重视对学生学习代数最为重要的相关知识与技能的掌握。专家组还建议:在对八年级及其以下年级学生进行教学质量的监控和评价时,应由教育部批准成立的美国国家教育进展评估机构对各州的考试进行考核,应当力争优化考试命题。通过研究专家组发

邢丽丽/宁波大学教师教育学院硕士生。(宁波 315211)

现,美国“教学评估”专题的“测验项目和测验设计”方面存在一定的问题。笔者对这些问题进行了简单介绍,并提出在测验编制和项目设计上的一些建议。

二、“测验项目和测验设计”中存在的问题

1. 项目“措辞”上的问题

数学专家表示,NAEP和州测验中存在的问题通常和项目“措辞”有关。专家组把测验中出现的问题分为七种主要类型,并提醒测验编制者在编制过程中尽量避免这些问题。也就是说,测验编制者应该特别注意数学测验项目的设计和评价,尤其要弄清楚某些特殊项目所检测的知识点,明确这些项目的范畴。

2. 测验程序研发中出现的漏洞

NAEP和州测验中出现问题的频率表明在测验程序的研发中存在其他可能的漏洞。专家组认为,交流和沟通出现偏差可能是导致失误的其中一个原因。NAEP和州测验的研发者运用了复杂的心理测量学模型和统计方法来进行分析。心理学家为了对被试进行准确地测量,采用了高度复杂的统计模型和数据分析方法,但实际上他们并不十分清楚如何对这些特定结构进行项目设计。相反,专业的项目编写人员和项目评价者拥有较高的学历,但这些人中有一些并没有接受过项目设计方面的专业培训和任务,而且,通常他们从心理测量学家那得到的测量功能反馈很有限。

3. 测验项目上的争议

专家组检查了与多项选择题相比,“构答反应”形式能否测量数学能力的不同方面。很多教育者都相信,“构答反应”项目(例如简答题)在测量数学能力时优于多项选择题,它可以提供一个关于数学技能的更令人信服的测量结果。专家组查阅了研究“构答反应”项目的心理测量学文献,并与多项选择题进行了比较。他们发现,科学文献并不支持与“多项选择题”相比,“构答反应”项目尤其是简答题在测量数学能力的不同方面优势更大的假设。

三、对“测验项目和测验设计”中存在问题的思考

对于美国数学“教学评估”专题中“测验项目和测验的设计”方面出现的以上问题,笔者进行了思

考,并提出了在测验编制和项目设计上的一些建议。

第一,测试应做到公平、可信和有效。只有当学生在一段时间内重复参加某个测验,且分数浮动不是很大时,其测验结果才是可靠的。美国数学专家组提出要特别注意某些特殊项目所检测的知识点,明确这些项目的范畴。因此,能检测到所要测量的特性和内容的测验才是有效的测验;如果所测量的内容是学生没有学习到的内容,或者测验编制者没有提供给学生获得优异成绩的同等机会,那么这样的测验是不公平的。在施测之前,测验编制者应当向学生交代清楚将要测量的内容,并提供均等的机会。还有一点要注意的是,测验的分数必须是直观可靠的,并且能反映出想要测量的能力。美国专家组指出了NAEP和州测验中出现的问题通常和项目措辞有关,因此,我们必须确保测验本身必须是高质量的和有效的,在此基础上,测验才能够公平地进行。

第二,教师应该有效利用测验来提高学生的学习效果。因为测验对学生而言不单单是一项测试,也是学生的一种学习经验,对教师而言,测量出来的结果是对学生学习情况的一种反馈,教师对试卷的分析能够帮助他们了解学生对教学目标的掌握程度,以此来提高教学。教师在分析试卷的过程中,通过讲解、分析和讨论“正确”答案,能够引发学生的认知冲突,帮助学生加深对知识的理解和掌握。优秀的教师会有效地将测试过程整合到整个教学计划中,从而引导、促进学生的学习。

第三,对教师而言,教师通常都会在完成某一单元的教学任务后设计并对学生实施某一测验。那么在测验的编制上,我们应该注意什么呢?美国数学专家组提出要特别注意某些特殊项目所检测的知识点,明确这些项目的范畴。因此,在进行测验编制时,测验项目应该涵盖所有的教学目标和认知领域,测验应该能够清楚地考核教师所教的内容,而且这些内容应该与教师的教学目标一致。每个教学单元包含了不同的学习目标和学习内容,有的是数学代数领域,有的是空间与图形领域,有的是统计与概率领域,有的是综合实践领域,也有的是几个知识领域的交织,对不同的学习目标和内容,我们都应设计相应的测量项目。每个教学目标里都应包含学生必须掌握的不同知识,如记忆性的陈述性知识、理解性的程序性知识。一个性能良好的测验不应该仅仅聚焦于

美国数学评估“测验项目和测验设计”中存在的问题及思考

某一个教学目标,还应该能测量到教师的各种教学目标,应该能测量更为复杂的技能,例如高难度的推理技能。

第四,报告中提出的在测验项目上的争议,让我们认识到,测验编制者应了解相关测验项目的优缺点,这样才能做好测验编制的工作。选择题、填空题和计算题是我国数学测验中常用的测验项目,教师应该了解这些项目的优缺点。

选择题在测量各种类型的知识内容方面有广泛的适用性和灵活性,其评分客观、方便,误差较小,选择题也扩大了试题对知识内容的覆盖面;由于选择题减少了学生的书写过程,使增大题量成为可能,同时,选择题可以采用机器阅卷,减少了阅卷时间,节约了人力。但是,选择题只要求学生从给定选项中挑选出一个或几个正确答案,它不能反映出学生解题的思维过程,因而不大适合测量问题解决能力、组织能力、表达能力、创造能力等高层次能力;另外,选择题的编制费时费力,且存在着凭猜测得分的可能性。

填空题要求学生在不完整陈述或图表中填入适当词语、句子或数据,把该陈述或图表补充完整。填空题的命制灵活,形式多样,知识覆盖面较广,评分也比较客观。但是,填空题往往照搬教材中的句子或段落,过分关注学生对知识细节的记忆,忽视对学生理解能力和思维能力的考查,因而不利于学生系

统、综合地掌握和运用各类知识。

计算题要求学生具有综合运用知识的技能,从定量的角度解决实际问题,它对于考查学生对基础知识的掌握程度、计算能力、逻辑思维能力、空间想象能力以及分析判断能力等都有重要作用。但计算题往往过于强调解题的技能技巧,考查的知识内容比较单一且缺乏开放性和综合性,因而不利于考查学生综合运用知识的技能和分析解决问题的能力,也不利于学生创新精神的培养。

第五,测验研发人员之间应该建立良好的交流与沟通。测验研发人员之间的好沟通是一个测验具有良好效度的重要基础。在现有测验中,除了教师在日常教学中设计的测验,还有一些比较大型的专项测验,这就更要求其具有科学性和有效性。为此,美国数学专家组在报告中一再强调,数学家、课程专家、语言专家、认知心理学家或数学教育者应该参与到测验的研发过程中。

参考文献:

- [1] 美国数学专家组. 国家数学顾问团报告[R]. 2008. 57~60.
- [2] 周勇, 杨明全. 新课程下的评价观与评价策略[M]. 北京: 首都师范大学出版社, 2005. 146~155.
- [3] [美] 理查德·I·阿兰兹著, 丛立新等人译. 学会教学[M]. 上海: 华东师范大学出版社, 2005. 174~204.

责任编辑/王彩霞

(上接第52页)考之所以年年成为社会关注的热点,牵动亿万人心,就是因为机会太少,每年只有一次(曾经一度出现的春季高考因为种种原因并未形成很大的社会影响力)。要有效解决考生机会过少的问题,就必须设法增加考试次数。而新加坡已经形成了大学分数和院系分数的双层评价标准。大学院系通过第二次考试既可以挑选自己满意的考生,又给予在统一考试中发挥不好的学生第二次机会,使严肃公正的考试更加人性化,这在一定程度上能起到遏制高考舞弊的作用。

我国的高考改革已历经数十载,在各个方面都取得了可喜的成就,并最终走向了一条既符合本国国情又与国际考试发展趋向接轨的新型之路。坚持“洋为中用,择善吸收”的改革理念,对缩短我国改革

摸索期、加快完善高考制度的进程有重大意义。

参考文献:

- [1] 杨兴龙. 中俄高考制度之比较[J]. 内蒙古师范大学学报(教育科学版), 2006(6).
- [2] 徐艳平, 张宝昆. 新加坡高等教育入学考试制度及其启示[J]. 教育与考试, 2007(2).
- [3] 冯晓生. 亚洲四小龙课程实践研究[M]. 福州: 福建教育出版社, 1999.
- [4] 康乃美, 蔡炽昌. 中外考试制度比较研究[M]. 武汉: 华中师范大学出版社, 2002.
- [5] 陈国平. 浅谈新加坡大学入学考试与招生制度[J]. 武汉市教育科学研究院学报, 2007(2).

实习编辑/刘建