

基于标准的学生学业成就评价*

——以 NAEP 为例

陈 晨

摘要:20 世纪 90 年代以来,美国基础教育评价改革的目标是建立基于标准的评价和教育绩效责任制。本文以 NAEP 为例,介绍了基于标准的评价的背景、外部环境以及实践,在此基础上得出对我国基础教育学业评价的有益的启示。

关键词:基于标准的评价;美国;NAEP

中图分类号:G51 文献标志码:A 文章编号:1673-9094(2010)05-0061-04

“基于标准的评价”即“基于课程标准的评价”。课程标准是国家对学生学习结果的共同的、统一的基本要求,是学生学业成就评价的依据。评价之所以要基于课程标准主要是因为课程标准规范学业成就评价的设计理念、提供学业成就评价的维度框架、限定学业成就评价的内容范围和认知要求^[1]。美国课程标准一般包括内容标准、表现标准和学习机会标准。内容标准划定学习的范围,描述学生应该知道什么以及能做什么;表现标准衡量学生对内容标准的掌握程度;学习机会标准是达到内容标准和表现标准应提供的保障,侧重于教育公平。

从国际范围看,美国可以说是基于标准的评价的先行者。本文通过介绍美国一项权威性的全国教育进展评价(National Assessment of Educational Progress,简称 NAEP)来展示基于标准的评价的具体做法。

一、基于标准的学生学业评价的背景

20 世纪 80 年代,《国家处在危机之中》等一

系列报告从国际比较的角度表明,美国基础教育质量严重下降、不同种族之间教育不公平现象日益突出。教育理论和实践者认为,造成这一现象的原因在于美国基础教育长期实行地方分权管理体制,造成课程上缺乏统一的标准和要求。在此背景下,20 世纪 80 年代末,美国开展了轰轰烈烈的“标准改革运动”。经过十余年的发展,美国各专业协会在教育部的资助下相继制订各学科的标准,于 1996 年完成了所有核心学科的全性课程标准的编制。同时,各州在国家标准的基础上也纷纷制订了州一级的课程标准,截至 1998 年,几乎所有的州都完成了课程标准的制订^[2]。

在制订和实施课程标准的过程中,课程实践者逐渐将改革目光投向学习者的学习质量,以课程标准为基础开发评价工具、编制测验,并在部分年级推行州的学业成就评价,欲建立起与课程标准相匹配的教育评价系统。1990 年,与“标准改革运动”相呼应,美国全国教育进展评价(NAEP)开展了以州为单位的数学评价,并规定以后每两年进行一次州层面的数学和阅读测试^[3];1995

* 本文系全国教育科学规划重点课题“基础教育可靠有效的发展性考试评价制度研究”(DFA030121)的研究成果之一。

收稿日期:2010-02-11

作者简介:陈晨,徐州师范大学教育科学学院(江苏徐州,221116)课程与教学论专业硕士研究生。



年,美国又参加了国际教育成就评价协会(IEA)发起的规模空前的“第三次国际数学与科学研究”(TIMSS)。基于课程标准的学业成就评价日益成为人们关注的焦点。

二、基于标准的学生学业评价的外部环境

2002年,美国颁布了《不让一个孩子掉队》(No Child Left Behind, NCLB)法案,强化了基于标准的学业成就评价,要求学业评价要与州的学科内容标准相一致,将基于标准的学业评价推向一个新的高潮。

20世纪90年代起,作为“标准改革运动”的一部分,各州逐渐建立起教育问责制度,强化学校的绩效责任,建立了和学校相联系的奖惩机制,即对学业评价中表现优秀的学校予以奖励,表现差的学校予以扶助和惩罚。各州的这种政策性导向强化了对基于标准的学业成就评价的深入探讨,确保了教育绩效信息和评价的可信度。

美国的实用主义文化传统决定了美国在基于标准的评价实践中必然选择改进技术这一最具操作性的实践路线。目前,美国已开发出了多种检测评价与课程标准一致性(alignment,又译为“匹配”)分析的制度和工具,为基于标准的评价实践提供了重要的基础。

美国教育部《2002-2007年战略规划》的目标之一就是创造一种“问责文化”或“成就文化”(accountability/achievement culture),从而将联邦教育拨款与学业评价结果的问责联系了起来^[4]。正是美国这种重责任、尚质量、求绩效的文化传统,为基于标准的学业评价提供了良好的环境。

三、基于标准的学生学业评价的实践——以NAEP为例

NAEP是目前美国国内一项权威的中小学生学习成就评价体系。成立40年来,NAEP已周期性地对美国中小学生的阅读、数学、科学、写作、美国历史、公民学、地理等学科进行了全面测试。NAEP收集的数据和报道已被广泛用于美国教育改革的各政策文本中,对美国教育产生了深远的影响。NAEP在基于标准的评价实践中已形成了一套比较完善的评价程序,我们可以通过其评价流程看出NAEP基于课程标准的特点。

1. 评价框架的制定

评价框架是NAEP测试形成的依据,它对各学科所要评价的内容领域、测试的题型、试题难度以及评分规则进行详细说明。1988年,NAEP的政策制定机构——全国评价管理委员会

(NAGB)成立以来,一直负责评价框架的开发。NAEP评价框架设计科学严谨,评价领域的确定、维度的选取以及题目的设计紧紧围绕着课程标准展开。在制定评价框架时,NAGB依据国家课程标准和评估水平,同时考虑州和地方标准,在此基础上确定各年级学生应该达到的知识和能力水平。

NAEP评价框架集中了课程专家、评价领域的专家、州和学校地区的代表、学校管理者、政策制定者、教师、家长以及广大公众的智慧。首先NAGB通过竞标决定框架开发的合同商,如2009年数学评价框架是在CCSSO(Council of Chief State School Officials)与NAGB的契约下,联合CBE(Council of Basic Education)、ASS(the Association of State Supervisors)以及Achieve, Inc.等专业机构合作制定的。接着,合同商与NAGB旗下的各委员会(包括评价框架开发委员会、导航委员会、筹划委员会等)协同工作,经广泛讨论后起草评价框架。草案制定之后,合同商和NAGB成员会举行公共论坛,从商业社团、测试专家、家长、学生等使用者那里获得反馈信息,进行修订。修订好的草案再送到专业机构、各地区学校、商业团体重审。经反复修订和审议,各方意见达成一致后,评价框架才得以成型^[5]。

NAEP评价框架涵盖评价学科的内容标准,各内容领域都附有典型的样例帮助理解。在表现标准的一致性上,NAEP开发了基本(basic)、熟练(proficient)和高级(advanced)三种成就水平,以区分学生对课程标准不同的掌握程度,并对每种成就水平相应的表现进行清晰地界定和说明。在学习机会标准方面,NAEP设定的成就水平也包括学习存在障碍、英语能力有限以及残障的学生。

2. 试题的编制

在评价框架制定完成之后,NAGB成员同考试签约商以及其他专业组织机构中的测验专家在熟悉评价框架的基础上,权衡框架中规定的内容、题型、试题难度、成就水平等编制大量试题。试题内容必须涵盖评价框架中规定的所有内容领域,试题类型要与框架中规定的三种成就水平相匹配,试题难度要与框架中描述的试题复杂度相当。为了保证评价结果客观公正,评价试题必须排除种族、文化、性别以及地域偏见。为了保证试题与框架之间的一致性,NAEP的框架开发人员和试题编制人员至少有20%是相同的^[6]。

在试题编制完成后,有关专业机构对试题进行匹配(alignment of standards and assessment)检验和一系列的试卷分析,然后在全国范围内选择有代表性的学校进行试测,在对试测结果的数



据分析的基础上对试题加以修订。最后,经 NAGB 各委员会的审查和筛选后方可进入 NAEP 的题库 (item pool)。这一系列的程序保证了 NAEP 极高的信度和效度。

3. 评价样本的确定

NAEP 是一种抽测型评价,评价样本包括试题样本和学生样本。

由于 NAEP 试题库很大,不可能让每个参与评价的学生完成所有题目,因而 NAEP 在试题编制完成之后,采用矩阵取样技术抽取部分试题,并将这些试题分成不同的组块 (item blocks), 编订成测验手册。每本手册中的试题内容、题型、难度和题目数量相当,且都满足 NAEP 测试内容和时间上的要求。一般来说,每个年级每门科目至少有 20 套测验手册。

NAEP 的学生样本覆盖各种类型和规模的学校、学区、不同性别和种族的学生,也包括学习存在障碍、英语能力有限以及严重残障的学生,从而保证了样本的科学性、代表性以及评价结果的信度和效度。一般来说,在一个仅有全国评估没有州评估的年份,每个科目通常需要各个年级 6000 到 10000 个学生样本^[7]。

4. 组织实施

为了确保所有学生都能公平、无歧视地参加评价,NAEP 为有学习障碍的学生、英语能力有限者以及严重残障的学生提供各种方便措施,专门为他们制定了“适应性政策”(accommodation),提供不同的试卷版本(不同语言)和考试形式(机试、笔试),并对执行这些特殊评价的人员加以培训^[8],从而使所有学生都有平等的机会表现他们习得的知识和技能,体现了评价的公平性。

5. 评分

NAEP 评分标准界定清晰、可操作性强,并且提供了许多参考答案,有效地控制了评分过程中可能造成的误差。

为了确保评分的可靠性和一致性,NAEP 遵循严格的评分程序。首先,根据教育背景、工作经历和评分经验严格挑选评分者。接着,对评分者进行培训,让评分者明确试题的设计原理以及评分的指南。评分者要在掌握评分原则的基础上给出客观的分数,经检验合格后方能正式参与评分。在正式评分前评分员要再次进行测试,由多位评分人员对同一试卷进行评分,当确认达到一致性要求时,才可以正式评分。在实际评分过程中,每组学生的试卷由两位随机组成的评分员来评,确保评分的准确与公正。另外,还要对每个评分者评分的质量与一致性进行监控与评价,不合

格的要予以解雇^[9]。

在评分技术上,NAEP 基于项目反应理论 (IRT), 并采用学术界一致认可的参数估计软件 (如 BILOG 和 PARSCALE) 进行项目参数估计,然后根据学生的作答情况和各种预测变量 (如性别、年龄、种族等) 估计学生的能力值。

6. 评价结果的报告

NAEP 不仅仅给出一门学科笼统的分数,而且按具体的内容项给分,具体反映学生在规定的学科领域中每一维度应掌握的知识、技能方面的成就,保证分数是基于标准的。如 NAEP2007 年数学报告卡中数学成绩是按照数与运算、测量、几何、数据分析与概率、代数等内容项出现的^[10]。

性别	数与运算	测量	几何	数据分析与概率	代数
男	239	241	238	244	245
女	237	237	239	243	243

注:数学的量尺分数是 0-500 分

除此之外,NAEP 的评价结果是按照民族、种族、社会经济地位、学校地理位置、英语熟练程度和残障状况等分组报告的,严密保护学校和学生的信息,并且通过各州或地区之间的横向对比以及跨年度的纵向比较,检查是否各个群体的学生成绩都得到提高,以及处境不利的学生与其他学生之间的成绩差距正在缩小。这样一方面减轻了学生和学校的压力,另一方面也体现了评价的公正性,有利于更好地向公众报告学生的实际水平。

四、分析与借鉴

基于标准的学生学业成就评价是一种专业性很强的评价行为,美国 NAEP 在基于标准的评价方面为我们提供了很好的操作模式,为我国评价改革提供了借鉴。

1. 进一步完善课程标准是“基于标准的评价”的前提

课程标准是学生学业成就评价的依据,一个好的课程标准可以为评价提供良好的基础。NAEP 在设定每门学科整体内容标准框架的基础上,对每门学科整体内容标准框架的具体内容也做了规定,并详细描述各学科各年级的成就水平,为评价的实施提供了正确的方向。

相比之下,目前我国课程标准还不完善,从评价的角度看,“我国的课程标准看似一个缺少学习成果评价标准的课程内容框架”^[11]。因此,完

善课程标准,使之具体、清晰、可操作,是当前我国评价改革的当务之急。

2.提高评价的专业化程度是“基于标准的评价”的关键

即使课程标准非常完善,要使评价建立在标准之上也需要许多中间环节。美国 NAEP 从评价框架的开发、试题的编制、取样、评分到报告都有着很高的专业要求,需要众多专业机构提供技术上的支持。

反观我国,在学业评价的实践方面还相当薄弱,因而我国应借鉴 NAEP 的经验,一方面学习美国借助专业机构的力量,建立专业化的评价中介机构,另一方面要提高评价参与者对课程标准的解读能力、基于标准的命题能力以及评分能力,提高考试评价的专业化水平。

3.促进学生学习是“基于标准的评价”的最终目的
在新一轮评价改革中,“为学习的评价”逐渐取代“对学习的评价”,成为“基于标准的评价”的指导思想。从 NAEP 的评价流程可以看出其评价的核心理念在于“质量”和“公平”,在评价各环节都注重评价对象的涵盖性(inclusion),考虑到特殊群体学生,关注每位学生的需要和发展。

相比之下,我国大规模考试(中考、高考)过于强调甄别和选拔功能,竞争色彩过于浓重,忽视了评价的发展和监测功能。因此,我国应借鉴 NAEP 的成功经验,明确评价的目的在于改进学习、促进发展,同时要考虑学生的个体差异性,为每个学生提供合适的考试环境,提高教育质量的整体水平。

参考文献:

- [1]崔允漭,夏雪梅.试论基于课程标准的学生学业成就评价[J].课程·教材·教法,2007(1).
- [2]李建华.标准运动:美国 90 年代的基础教育改革述评[J].课程·教材·教法,2000(12).
- [3]Ercikan,K.1997. Linking Statewide Tests to the National Assessment of Educational Progress: Accuracy of Combining Test Results Across States. Applied Measurement in Education 10(2).
- [4]国家教育发展研究中心组译.发达国家教育改革的动向和趋势(第七集)[M].北京:人民教育出版社,2004.22-25.
- [5]Jago,C. A History of NAEP Assessment Frameworks. Paper Commissioned for the 20th Anniversary of the National Assessment Governing Board 1988-2008.5-12.
- [6]NAGB. Item Development and Review. Policy Statement. 2002.
- [7]NCES. An Introduction to NAEP. U. S. Department of Education. 2008.6.
- [8]Pamlmer, E. A., & Barley, Z. A. What States can Learn about State Standards and Assessment Systems from No Child Left Behind Documents and Interviews with Central Region Assessment Directors. Washington, D. C. U. S. Department of Education. Institute of Education Science, National Centre for Education Evaluation and Regional Assistance, Regional Education Laboratory Central. 2008.13.
- [9]Sheida, W., Connie, S., & Alan, V. How Does NAEP Ensure Consistency in Scoring? Focus on NAEP, 2000, 4(2).
- [10]The Nation's Report Card Mathematics 2007-National Assessment of Educational Progress at 4 and 8.[EB/OL]http://nation's report card.
- [11]胡军.学生学习成果标准不能在课程标准中缺失——澳大利亚课程内容与标准给我们的启示[J].课程·教材·教法,2005(9).

Standard-Based Students' Achievement Assessment: Taking NAEP as an Example

CHEN Chen

(School of Education Xuzhou Normal University, Xuzhou 221116, China)

Abstract: Since the 1990s, the core target of reforming American's educational evaluation is students' achievement assessment and accountability mechanisms of education performance based on curriculum standards. This paper selects NAEP as a case to introduce the background of standard-based assessment, its external support system and its practical measures. Thus, several lessons were drawn for the basic education achievement evaluation in China.

Key words: standard-based assessment; American; NAEP