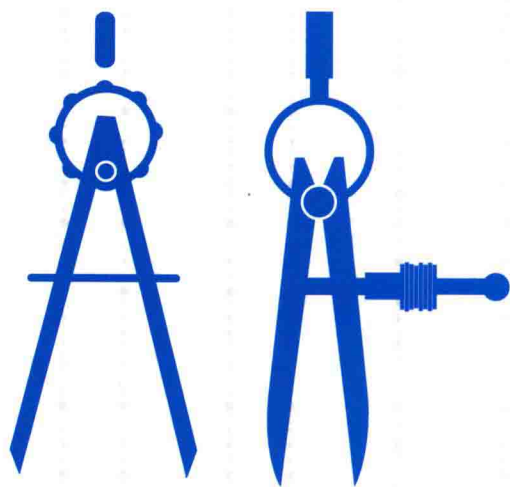


数学学业成就评价 实证分析

沈南山 著

SHUXUE XUEYE CHENGJIU PINGJIA
SHIZHENG FENXI

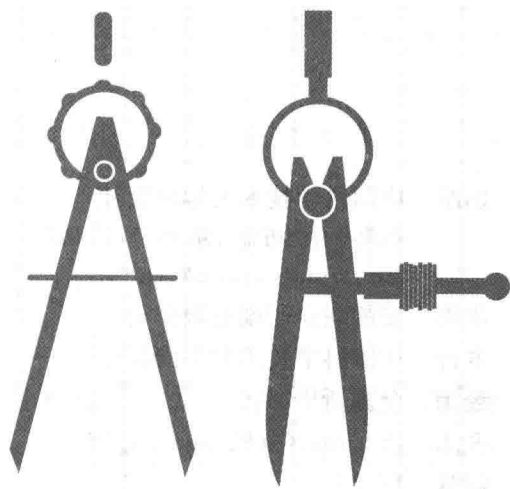


中国科学技术大学出版社

沈南山 著

SHUXUE XUEYE CHENGJIU PINGJIA
SHIZHENG FENXI

数学学业成就评价 实证分析



中国科学技术大学出版社

内 容 简 介

本书是在作者博士论文的基础上改写而成的。本书以国家基础教育质量监测为背景,并基于国家数学课程标准,从理论和实践两个维度,比较系统地探讨了数学学业成就评价调查工具研制的基本问题。本书共分7章,包括问题的提出与研究方法、数学学业成就评价研究综述、研究的理论基础、数学学业成就评价试题编制的基本问题、试题编制检验的多级评分原理与参数分析、试题编制检验的抽样实验与分析、结论与建议等,试图为我国基础教育质量监测数学学业水平测试提供现实的参照坐标。

本书可作为攻读数学课程与教学论专业的研究生、教育硕士的教学用书,也可作为高等师范院校数学与应用数学专业本科生的学习参考用书,还可作为中学数学教师继续教育以及其他各类数学教育工作者的教学参考用书。

图书在版编目(CIP)数据

数学学业成就评价实证分析/沈南山著. —合肥:中国科学技术大学出版社,2017.1
ISBN 978-7-312-04011-5

I. 数… II. 沈… III. 中学数学课—学业评定—研究—高中 IV. G633.602

中国版本图书馆 CIP 数据核字(2016)第 308070 号

出版 中国科学技术大学出版社
安徽省合肥市金寨路 96 号,230026
<http://press.ustc.edu.cn>
印刷 安徽联众印刷有限公司
发行 中国科学技术大学出版社
经销 全国新华书店
开本 787 mm×1092 mm 1/16
印张 13.25
字数 339 千
版次 2017 年 1 月第 1 版
印次 2017 年 1 月第 1 次印刷
定价 36.00 元

序

本书是沈南山在他的博士论文基础上改写而成的,请我作序。审读他的著作,论题是当前我国数学教育研究的热点问题,论域具有学术前沿性和理论深拓性,我乐意为他的著作出版作序。

构建学业成就评价体系是基础教育质量监测体系建设的重要组成部分,是当前我国推进素质教育的一项重要举措。通过基础教育质量监测,可以全面、准确地把握中小學生整体素质的发展状况,促进基础教育改革的不断深入和素质教育的全面实施。在国家基础教育质量监测背景下,研究学生学业成就评价具有现实意义和教育价值。显然,这一选题是针对我国当前基础教育评价领域的薄弱局面展开的。

本书研究的主题是数学学业成就评价的试题编制问题。研究的背景是北京师范大学教育部基础教育监测中心“中国儿童青少年心理发育特征调查项目”的子课题“中小学数学学业成就及其关键环境变量测查工具研制”。

该课题的主要研究问题包括如下方面:

① 数学学业成就评价的范畴、标准以及试题编制的基本原理和相关量纲指标体系。

② 试题编制内容如何体现数学素质教育目标,测试功能如何凸显多种教育功能。

③ 试题编制测查的实验设计。

④ IRT 方法原理在学生学业成就评价领域的应用。

沈南山的博士论文至少有两个方面是值得肯定的:

① 试题编制的指导思想是凸显素质教育功能。在国家层面上进行基础教育数学学业成就评价测查试题编制研究是一种很有教育意义的尝试。试题编制性质是“标准参照测验”的学业水平测试,而不是单一的选拔考试。试题设计内容和形式与传统教育考试有显著的区别,主要功能特征是要实现国家数学课程标准理念和数学素质教育目标,突出数学学业成就评价的多种教育目的,这是基于数学课程标准的“国家课程”测验的基石。

② 研究方法上的创新。项目反应理论在基础教育评价领域的研究现状是国外领先、国内滞后。利用数学方法、数学工具研究基础教育数学学业成就评价问题,应当是方法论意义上的尝试。利用项目反应理论分划试题项目的多级评分、分析试题项目的性状特征、筛选试题等,并在较大样本测试实验的基础上利用 IRT 方法进行数据分析,是项目反应理论在基础教育评价实践中应用的探索,值得研究和推广。

沈南山有较强的数学教育研究学术功底,领悟能力较强,发表了很多高水平的学术研究论文。我祝愿沈南山以这篇论文的出版为新的起点,在数学教育的道路上不断前行、不懈努力,在研究中取得新的成绩。

宋乃庆

2016 年 5 月

于西南大学

目 录

序	(i)
第 1 章 问题的提出与研究方法	(1)
1.1 研究缘起	(1)
1.2 研究思路与方法	(4)
第 2 章 数学学业成就评价研究综述	(6)
2.1 数学学业成就评价	(6)
2.2 发展性学生评价	(8)
2.3 基础教育质量监测	(11)
2.3.1 基础教育质量监测的目标与对象	(12)
2.3.2 基础教育质量监测的学业评价体系	(13)
2.3.3 基础教育质量监测实施制度和政策	(15)
2.4 国内数学学业成就评价研究概述	(17)
2.5 国外数学学业成就评价研究概述	(20)
2.5.1 美国基础教育学业评价体系	(21)
2.5.2 国际学生评价项目	(26)
2.5.3 TIMSS 评价体系	(27)
2.5.4 英国的国家课程测试	(28)
2.5.5 澳大利亚国家评价计划	(30)
第 3 章 研究的理论基础	(32)
3.1 试题编制的教育评价学基础	(32)
3.1.1 现代教育评价理论的发展	(32)
3.1.2 现代教育评价的主要模式	(33)
3.2 试题编制的教育测量学基础	(35)
3.2.1 教育测量的经典测量理论方法	(35)
3.2.2 教育测量的项目反应理论模型	(42)
第 4 章 数学学业成就评价试题编制的基本问题	(47)
4.1 数学学业成就的评价标准	(47)
4.1.1 课程标准与教材分析	(47)

4.1.2	内容标准	(56)
4.1.3	表现标准	(61)
4.1.4	评价标准	(65)
4.2	试题编制的量纲指标体系	(70)
4.2.1	试题编制范畴	(70)
4.2.2	试题编制双向细目表	(71)
4.3	试题编制结构	(73)
4.3.1	试题编制的题型结构	(73)
4.3.2	试题编制的长度结构	(74)
4.3.3	试题编制的组成结构	(75)
4.4	抽样实验	(77)
第5章	试题编制检验的多级评分原理与参数分析	(78)
5.1	测验项目的评分要素	(78)
5.1.1	测验项目的基本概念	(78)
5.1.2	测验项目的样本空间	(79)
5.1.3	二级与多级评分项目	(83)
5.2	测验项目的概率函数	(84)
5.2.1	节点呈包含关系的测验项目的概率函数	(84)
5.2.2	节点呈相互独立关系的测验项目的概率函数	(85)
5.2.3	二参数逻辑斯谛模型下测验项目的概率函数	(86)
5.3	能力参数估计	(87)
5.3.1	能力参数	(87)
5.3.2	能力参数已知情况下的项目参数估计	(88)
5.3.3	项目参数已知情况下的能力参数估计	(89)
5.4	测验项目信息函数与测验信息函数	(90)
5.4.1	测验项目信息函数	(90)
5.4.2	测验信息函数	(91)
5.5	最优评分权	(92)
5.6	测验的有效性检验方法	(94)
第6章	试题编制检验的抽样实验与分析	(96)
6.1	实验设计	(96)
6.1.1	样本取样	(96)
6.1.2	测试工具	(96)
6.1.3	研究假设	(97)
6.2	数据处理与分析	(97)
6.2.1	存储数据文件	(97)
6.2.2	经典测量理论下的项目参数分析	(98)
6.2.3	项目反应理论下的项目参数分析	(101)

6.3 试题编制的素质教育功能分析	(153)
6.3.1 试题编制以数学素质教育为总体目标	(153)
6.3.2 试题编制体现课程标准“三维目标”要求	(156)
6.3.3 试题编制凸显评价的多种教育功能	(157)
第7章 结论与建议	(159)
7.1 研究结论	(159)
7.1.1 试题编制在 CTT 下的各项指标具有良好的可信度	(159)
7.1.2 试题编制在 IRT 下的各项参数反映较大的信息量	(160)
7.1.3 试题编制测验的有效性和显著性检验	(162)
7.2 研究不足	(166)
7.2.1 试题编制理论研究的局限性	(166)
7.2.2 推论范围有限	(166)
7.3 研究建议	(166)
7.3.1 数学素质教育是监测的基本目标	(166)
7.3.2 建立完整的数学学业水平监测系统	(167)
7.3.3 优化数学学业水平评价方案	(167)
7.3.4 推行基于 IRT 的数学学业水平评价实践	(168)
附录1 数学学业成就评价测查试题	(169)
附录2 存储数据文件(dwj. dat)	(174)
附录3 数学学业成就评价关键环境变量调查问卷	(195)
参考文献	(198)
后记	(202)

第1章 问题的提出与研究方法

在数学教育实践中,各级考试名目繁多,所谓的评价也较多,为什么还要进行数学学业成就评价?其意义与价值何在?基于这一追问,我们需要厘清在我国现实的数学教育环境下,学生学业成就评价的意义、范式和实践走向,需要把我国数学教育置于国际数学教育的大背景下进行考察。

1.1 研究缘起

当今教育全球化,教育质量是国际基础教育界共同关注的一个焦点,无论是发达国家还是发展中国家都十分重视基础教育,尤其是数学基础教育。由于数学在基础教育学科中的中心地位和科学发展中所起的重要作用,20世纪60年代以来,以美国为核心的一些国家的教育团体以及全球性的国际经济合作组织(Organization for Economic Cooperation and Development, OECD),出于不同的主体利益诉求和教育目标,开展了一些基础教育学科质量评价项目,如全美教育进展评价(National Assessment of Educational Progress, NAEP),国际学生评价项目(Program for International Student Assessment, PISA),国际数学与科学研究(Trends in International Mathematics and Science Study, TIMSS)等,对学生的学业成就进行全面评价以及国际比较评价研究。聚焦这些项目的评价目标和评价内容,我们不难发现学生学业成就评价的核心价值是改善学生的学习,评价目标与内容凸显两条鲜明的主线:一是重视学生的学业能力,特别是学生的“核心素养”备受关注;二是通过评价促进教学,“不让一个孩子掉队”(No Child Left Behind)的教育问责(Accountability)意识日益彰显。

从国际教育比较的角度来看,鉴于我国的教育国情——基础教育组织结构体系庞大,涉及数以亿计的中小学生、百万计的中小学校,教育发展不平衡,解决基础教育问题历来是一个教育难题。教育评价实践中存在着各种教育利益分配与冲突,基于考试的甄别选拔性评价依然具有十分重要的地位,致使学生学业成就评价的核心价值在一定程度上背离了学生成长与发展的主线,表现在国家通过课程改革推进素质教育的思想并未真正落实,各级学校没有走出应试教育评价的樊篱。教育评价的政策、制度和评价行为坚守学校与个人利益,评价范式推崇工具主义和技术理性。而且,由于长期的“科举式”考试文化思想浸润,教育功利主义深植于评价思维,评价哲学是用考试成绩的符号价值来代替教育价值,因而学生学业成就评价的主流话语中弥漫着狂热的分数崇拜。

从国家利益和教育意志出发,为提高整个民族的教育素质和国际竞争力,大规模的教育改革在强力推进,评价改革也在不断革新。可在现实教育体制和社会、政治、经济、文化因素

的制约下,教育评价方法改革主要是以考试变革为枢纽,虽在素质教育背景下有所改变,如学生学业成就评价倡导“表现性评价”(Performance Assessment)、“真实性评价”(Authentic Assessment)等质性评价形式,但整体上没有离开“高利害性考试”这个大圆圈,中、高考选拔制度仍一直为广大家长所推崇,被视为社会文明的教育基础,也始终是学生学业成就评价的最高追求和永恒主题。教育评价的这种恶性追逐考试竞争,导致素质教育评价取向被考试价值的认同度遮蔽,教育遭遇了社会价值与学生主体发展价值的双重失落,学生学业成就评价处在十字路口。

从基础教育发展的角度来看,基础教育是一个国家政治、经济、文化、科技力量的基石,提高基础教育质量是硬道理。学生学业成就评价路在何方?如何孕育新的教育评价文化?国际教育经验给了我们十分有益的启示:基础教育发展需要有效的教育质量监测的导引,需要国家立法对基础教育质量进行评估和监督,需要建立国家教育质量监测体系来监测各级教育组织舆情,监测各级学校预设的教育标准的达成情况,监测基于国家“课程标准”的落实情况,并进行教育问责等。事实上,实施基础教育质量监测是许多国家的通行做法,体现国家的教育意志和教育责任。

由此,对基础教育形成的一个基本共识是基础教育质量监测有利于“教育问责”评价文化的形成,有利于各级教育行政部门考问教育绩效,即学生学业成就的好坏由谁来负责?怎样负责?我们的基础教育目标是否实现了?教育的未来目标是什么?教育决策是否合理?如何改进?政府在基础教育中应该承担什么样的责任?

不但如此,基础教育质量监测与评估可以帮助学校管理者和教师把握以下问题:我们学校的水平如何?我们学校办学的目标是什么?实现了没有?我们如何达到目标?我们如何判断达到了目标?而且,就学校内部而言,基础教育质量监测与评估可以帮助学校进行自我评估和自我建设。学校自我评估是一个动态的过程,是一个不断反馈的过程,学校管理者和教师可以根据反馈的结果及时有效地调整学校正常教育教学工作的运行,而且基础教育质量监测数据使得不同学校可以互相分享经验,学校联合评估可以打破单个学校的地域限制,不同的学校之间可以共享资源,互相借鉴、互相提高,也利于学校之间团结关系的建立,这不仅能促进学校的发展,也能够让教育决策者们形成对学校未来规划的战略思维。

对于教师来说,教师能从评估中做出教学改进,能够根据评估结果发现问题、解决问题,逐步形成一种自我调节、自我反馈的系统,建立一种崭新的专业发展的文化氛围,大大提高教师的责任心,这样教师、学校、教育主管部门之间的对话更具建构性。

基于此,研究基础教育质量监测与评估,在国家层面上建立学生学业成就评价和监测制度,改革和完善学生学业成就评价体系,对全面推进我国素质教育改革具有十分重要的现实意义。

1. 推进素质教育的需要

学生学业成就评价(以下简称“学业评价”)是素质教育评价改革的核心问题,也是世界各国基础教育改革普遍关注的热点问题。相对于西方学业评价改革而言,我国基础教育学业评价改革起步较晚,成绩也不太理想。自从20世纪80年代我国实施素质教育改革以来,学生学业评价改革以减轻中小学生的学业负担为目的,以培养学生实践能力和提高学生素质为重点,以期逐步达到基础教育评价制度的有效改革。

学业评价直接影响中小学教与学的实践行为,其制度改革是学校最为关注和最具导向作用的改革,基础教育能否真正实施素质教育的关键就在于学业评价制度改革的成功与否。

我国学业评价在监测机制层面上才刚刚起步,大规模教育测评的学业评价还没有广泛开展,也没有引起地方和学校的足够重视,在教育实践中实施的仍是以中、高考为核心的考试,并以考试成绩为标准来衡量学生学业成绩,远远达不到学业评价所要求的科学严谨性^①。有学者指出,学生学业评价无论在理论层面还是在实践层面,都存在不少急需解决的问题,如果不能有效解决,这些问题将阻碍基础教育课程改革和全面实施素质教育的进程。学业评价存在的问题主要包括:学业评价的主要操作方式是考试测评,并以分数来衡量学生学业目标;学业评价追求学校教育的“规模效益”,学业成就目标整齐划一;学业评价重心落在学生的知识技能评价,忽视对学生实践能力和学生素养的评价。所以,学生学业评价改革对素质教育改革具有重要的影响与作用,在国家教育质量监督机制的管理下,推进传统考试制度、考试形式及考试内容的改革,研发中小学生学业评价改革的有效方法与评价模式成为当务之急。

2. 深化课程改革的需要

我国基础教育课程改革正在向深度推进,改革进入了新阶段,对课程标准的理解与执行偏差较大,学术争鸣不断,社会反响强烈。例如,省思课程改革的诸多因素,如课程政策是否正确、课程标准是否完善、课程实施是否有效、课程目标达成度如何等,诸如此类的问题我们只能局限在价值判断层面进行理性思辨和学理论争,缺乏全国性的教育质量测查数据凿以事实诠释,也没有建立国家层面的学生学业评价数据库给予佐证。

事实上,课程改革在课程标准、课程管理、教材选用、教学实施等方面已建立了较为完整的框架,教育和评价理念无疑是正确的,缺乏的是在这种正确的教育和评价理念指导下的对教育教学质量的准确、清晰的描述。从而也就难以保证教育教学的正确方向,无法激励正确的教育教学行为,最终难以保证学生学业全面协调和可持续发展。所以,在基础教育课程改革的深化阶段,迫切需要建立“国家课程”测验检查课程改革的成效,迫切需要建立权威、科学的学生学习质量监测体系为课程改革指引方向。

3. 评价学生发展的需要

学生学业评价是以现代教育评价理论和先进教育理念为指导建立起来的学业评价制度和办法,是新一轮基础教育改革实施素质教育、促进学生发展的重要目标之一。新课程改革关于学生学业评价理念,在《基础教育课程改革纲要(试行)》(以下简称《纲要》)中有明确的表述:“改变课程评价过分强调甄别和选拔的功能,发挥评价促进学生发展、教师提高和改进教学实践的功能……建立促进学生全面发展的评价体系……发挥评价的教育功能,促进学生在原有水平上的发展。”^②

解读《纲要》关于学生学业评价阐释的本质:评价不仅要关注学生的学业成绩,而且要发现和发展学生多方面的潜能,了解学生发展中的需求,帮助学生认识自我,建立自信。主要是评价功能由侧重甄别和选拔转向侧重学生发展;评价目标从过分关注学习结果转向关注学习过程;评价主体强调评价主体多元化,重视自评和互评;评价内容强调评价对象整体素质的综合因素;评价方法强调质性评价和量化评价的结合,以质性评价统整量化评价;评价关系强调民主平等,体现以人为本的主体性评价的价值取向^③。因此,建立以促进学生发展为目标的评价体系,是推动课程改革顺利进行的关键。

① 张布和.我国学业成就评价改革现状及对策[J].中国教育学刊,2009(4):50-53.

② 基础教育课程改革纲要(试行)[EB/OL].(2001-07-27).http://www.edu.cn/20010926/3002911_2.shtml.

③ 新课程改革评价的基本理念是什么?[EB/OL].<http://blog.hnteacher.net/blogs/index.php/2009/08/.../416>.

学生学业评价与学生发展的相关性主要表现在如下方面:

① 学生学业评价为学生的知识学习和技能发展提供了比较明确的目标和方向。学业评价基于国家数学课程标准设定的“三维”教学目标,对学生学业的预期结果做出了明确的描述,有助于学生确立努力方向和调节学习过程,同时也有助于教师调控和改进教学工作,促进学生发展。

② 学生学业评价为学生情感学习和心理发展创造了良好的学习环境和发展空间。学业评价以素质教育为指导思想,力求综合评价学生的全面发展情况,如对情感、态度和价值观学习目标的评价。

③ 学生学业评价为学生的社会学习和个体发展加速了社会化进程。学业评价力求按照适应社会需求确定相应的评价目标和测评要求,全面判断学生的学习结果,为学生将来适应社会生活创造了良好的准备条件。

4. 完善考试制度的需要

我国中小学学业评测体系没有形成统一的参照标准,很不完善。考试性质、种类五花八门,缺乏规范性;而且我国基础教育存在地区差异、文化差异、教师差异、学生差异以及教材差异,需要建立国家(省、市、县(区))常模和标准参照测验,完善中小学教与学评测体系。《中小学评价与考试制度改革的指导意见(讨论稿)》明确指出中小学评价与考试制度改革的原则是:中小学评价与考试制度的改革,应从德、智、体、美多方面全面贯彻党的教育方针,全面评价学生的发展。评价以促进学生的健康成长、教师教学水平的提高以及学校的发展为目的,以提高学生的综合素质为根本目标。应探索多样化的有实效性的评价方法工具,提高教育评价的有效性,要有机地把终结性评价与形成性评价结合起来,使学生成长的过程成为评价的重要组成部分,使管理者、教师、学生、家长共同成为评价主体^①。

综上所述,学生学业评价研究的理论价值在于:学生学习质量的高低直接关系到学生在未来社会中的生存和竞争能力,关系到国民整体素质和综合国力的提高;建立和完善学生学习质量分析与评价系统是提升整体国民素质、推进数学素质教育的战略抓手;根据学习质量分析结果引导中小学树立正确的教育观、质量观和评价观。实践价值在于:明确中小学生学习质量的内涵、评价标准及其影响因素,形成我国中小学生学习质量发展水平的指标与常模;对学生学业成就进行调查与实证分析,形成提高教育教学质量的有效机制,进一步推进课程改革和教育体制改革。

1.2 研究思路与方法

国家基础教育质量监测体系采取何种方式操作值得关注^②。就数学学科而言,目前国家数学学业成就评价测试的内容、框架及功能均处于研究的初始与实验阶段。我们看到,教育部基础教育监测中心等部门对此做了大量的基础性工作:致力于探寻“适合中国文化,符合

^① 中小学评价与考试制度改革的指导意见(讨论稿)[EB/OL]. (2010-02-04). <http://www.tzedu.net.cn/qwexport/78/article.jsp>.

^② 崔允灏,王少非,夏雪梅. 基于标准的学生学业成就评价[M]. 上海:华东师范大学出版社,2008:172.

国家素质要求的课程评价”,发布基础教育监测信息简报,指导基础教育质量监测工作^①。

本研究的思路是“是什么”“为什么”“怎么做”,贯穿内涵界定—缘由分析—试题编制—抽样测查—数据分析的线索,即从数学学业评价试题编制的理论内涵是什么、为什么要这样编制、如何进行试题编制、试题编制的信度和效度如何等层面展开,主要采用文献研究法、问卷调查和结构性访谈法、实验法、教育统计分析法和项目反应理论模型分析法。

1. 文献研究法

文献研究法是指以现存的文献记载资料,如课程标准、教材、论文文献、报告及评估文本等来研究数学学业成就评价的方法。主要分析国内外数学学业评价的现状与发展,揭示数学学业评价试题编制的理论内涵与构成要素指标。其中中文文献主要选取与评价、学业成就以及数学试题编制有关的文献,同时兼顾有关素质教育以及考试评价的文献。英文文献主要选取与 mathematics achievement assessment, curriculum standard, benchmark assessment, item response theory 等词有关的文献。其目的在于通过对国内外文献中有关数学学业评价及其试题编制的内容的梳理,形成对数学学业评价及其试题编制理论内涵的把握;其次是按照本研究对数学学业评价及其试题编制的理解,寻找支撑数学学业评价及其试题编制的理论基础,以期较为合理地形成对数学学业评价试题编制构成要素的量纲指标体系的建构。中文文献源自国内外相关数学课程标准、学业评价、评价文本、评价报告等。英文文献主要源自 Web of Science 数据库、ProQuest 数据库、ProQuest 博硕士论文全文数据库以及网络文献等。

2. 问卷调查和结构性访谈法

问卷调查和结构性访谈法是根据本研究对学业评价的理论思考,形成有效问卷及访谈提纲,对中小学学生、教师、校长和教育行政部门等相关人员进行问卷调查和结构性访谈,以获取有价值的量化数据和资料。

3. 实验法

实验法是研究者从某种假设或者理论出发,为揭示所研究的实验因子(自变量)对某种结果的影响,控制某些实验条件,促使一定的结果现象产生,然后对其结果进行分析,得出相关实验因子对实验结果影响的科学结论。本研究采用调查实验法,主要选择合适的测试区,采集样本,对试题编制评价工具加以检验,验证数学学业评价试题编制的科学性和有效性。

4. 教育统计分析法

教育统计分析法是指为有效地分析试题编制的质量,从教育统计学的角度,观察均值、标准差、信度、效度、分布状态、难度、区分度等各项指标反映考试质量的情况,并提供相关理论依据。

5. 项目反应理论模型分析法

项目反应理论(Item Response Theory, IRT)模型分析法是近二十年发展起来的一种比较先进的心理与教育测试理论,受到国内外许多学者的关注,对于促进我国基础教育学业考试评价方式的改革和发展具有很大的指导意义。本研究主要是对样本的测试数据进行分析,如试题项目的特征曲线、多级评分、信度函数以及卡方拟合性检验等。

本研究采取的技术路线为文献研究→试题编制→抽样测查→数据分析→结论与建议等。研究策略是理论研究与实证分析相结合,以实证分析为主;定量分析与定性分析相结合,以定量分析为主。

① 教育部基础教育质量监测信息简报。

第2章 数学学业成就评价研究综述

2.1 数学学业成就评价

学业成就(Academic Achievement)是一个外来词,一般是指学生在一定时期内学完一门基础或专业课程后所获得的知识技能以及全部被称为“能力”的达成程度^①。因此,简单地说,数学学业成就(Academic Achievement on Mathematics)就是学生在某一阶段学习了某一数学课程之后,所掌握的数学知识、技能以及应用数学知识的能力的总和。

这样的“定义”未免过于笼统,其实,数学学业成就是一个十分复杂的概念,其内涵和外延在不同的评价体系中有不同的含义理解。例如,全美教育进展评价(National Assessment of Educational Progress, NAEP)主要考查学生“国家课程”的达标情况,学业成就偏重于学生基础知识和基本技能的掌握程度,测查数据是评价学生学习和学校教学的依据;国际教育成就评价偏重于学生学科知识、技能的综合掌握,特别是跨学科知识分析问题和解决问题能力的判断,学业成就倾向于学生掌握的知识和技能的应用能力;而国际学生评价项目着眼于15岁学生(相当于我国初中毕业生)在成年后适应成人生活的能力,学业成就倾向于学生生活的能力的测查与判断。

不仅如此,从国际学生评价的发展趋势来看,对学生学业发展水平的认识也在逐步发生变化。一方面,学业评价侧重于学生面对实际问题情景的理解力、判断力以及应用所学知识解决真实问题的能力;另一方面,学生在知识技能学习中的学科情感、态度、价值观等方面的发展水平也逐步受到重视。就是说,学业成就趋向于全面发展的素质。

从我国学生学业评价的现实来看,自实施素质教育以来,伴随着基础教育课程改革的深入和发展,对学生学业成就的认识也在逐步变化。这种变化反映在随着人们教育观念的转变和课程的不断更新,教育质量的标准在发展变化。学生学业成就融入教育教学与课程改革的培养目标,学生的全面发展被关注。例如,针对义务教育阶段学生学业成就,提出素质教育目标要求,注重学生在德、智、体诸方面的全面发展,包括学生思想品德素质、文化科学素质、健康素质和劳动技能素质等方面。而且,在新课程标准下,针对学生学业成就,《义务教育数学课程标准》(以下简称《标准》)明确提出“知识与技能、过程与方法、情感态度价值观”的三维目标要求。所以,从国内外的评价实践来看,要给数学学业成就下一个确定性的定义是不合适的,事实上它是一个概念体系,且具有发展性。因此,关于数学学业成就,我们认为,一个最基本的方面应当是反映学生在一定时期内对国家数学课程标准的掌握程度,应当从数学课程标准的“三维目标”的角度去界定它,这样才能保证界定的科学合理性和依据

^① What is academic achievement? [EB/OL]. (2009-10-26). <http://www.cnm.edu/depts/at/saa/>.

规定性,也才能保证对数学学业成就评价调查结果的有效推论。

概括地说,数学学业成就主要包括数学学习的知识与技能、数学学习的过程与方法以及数学学习的情感态度与价值观三个维度,涵盖学生数学学习质量的主要指标体系,其子维度包括:数学学习知识与技能、数学学习思想方法、数学学习交流与合作、数学学习过程与结果反思、数学学习心理素质、数学创新意识与实践、数学知识应用与问题解决等。据此,我们列出数学学业成就概念体系的基本框架,如表 2.1 所示。

表 2.1 基于课程标准的数学学业成就概念体系的基本框架

数学学业成就	数学学习知识与技能	• 数学基础知识 • 数学基本技能
	数学学习过程与方法	• 数学学习交流与合作 • 数学学习自主探究 • 数学反思性学习 • 数学知识应用(观察、实验、抽象、建模等) • 数学创新意识与实践 • 数学问题解决 • 数学思维方法(推理、数感、符号感、……)
	数学学习情感态度与价值观	• 数学学习兴趣 • 数学学习态度 • 数学学习意志 • 数学学业成就价值观

评价通常有两种译法,一种是评估(Assessment),另一种是评价(Evaluate)。评估主要包括测试,是获得信息、目标或目标进展的方式,以确定一个目标或目标进展是否已经达到,偏向事实判断。评价是旨在提供信息,关于事物情况判断的过程,偏向价值判断^①。

关于评价的研究,布鲁姆(Bloom)是评价研究的奠基人之一。他将人类认知活动过程划分成等级结构模型,评价是该模型的构成要素。在这个模型中,评价是最为复杂的认知活动之一。他认为:“评价就是人们对一定的想法(Ideas)、方法(Methods)和材料(Materials)等做出的价值判断的过程。它是一个运用一定评判标准对事物的各种属性(例如有效性、可靠性、经济性)进行评估的过程。”这就是说,评价就是通过评价者(Evaluators)对评价对象的各个方面进行量化测量和非量化判断的过程,最终得出一个可靠的逻辑结论。

学业评价也是如此,在教育领域,教育评价包括如下两个方面的含义:

- ① 对教育系统的功效和工作状态做出评议和估价的过程。
- ② 根据教育目标对学生行为和个性变化进行的价值判断。^②

显然,学业评价主要蕴涵第二个方面的含义,评价是一个对学生学习价值的判断过程,其处理过程是一个运用综合测试、观察和访谈等方法进行的一个非常复杂的复合分析过程。

一般而言,学业评价(Assessment)有如下三种类型:

- ① 为学习而评价(Assessment for Learning),评价的目的是寻找学习目标与所期望目

① Kizlik B. Measurement assessment and evaluation in education[EB/OL]. (2009-09-01). <http://www.adprima.com/measurement.htm>.

② 卫道治,吕达. 英汉教育大词典[M]. 北京:人民教育出版社,2005:114.

标之间的差距。

② 怎样学习而评价(Assessment as Learning),评价的目的是反思如何学习。

③ 关于学习的评价(Assessment of Learning),评价的目的是使教学人员能够检查了解学生的进步和发展情况。^①

综合分析,学业评价就是根据一定的标准,对学生的学习过程和结果进行价值判断的活动,包括确定学生学业评价的依据、研制学业评价工具、实施评价活动等。

2.2 发展性学生评价

发展性评价(Developmental Assessment,这里指发展性学生评价,下同)源自 20 世纪 90 年代初期,英国开放大学教育学院纳托尔(Latoner)和克利夫特(Clift)等人提出和倡导的开放式教育评价思想,这是一种与惩罚性评价相对立的评价思想,主旨是教育评价要以学生的发展为本,关注学生的全面发展、未来发展和个性发展。“发展性评价”概念提出以后,随着基础教育“素质教育论”的本质诉求,在教育评价领域备受关注,现已发展成为一种以促进学生全面发展为根本目的,面向未来,面向全体,强调合作和自我完善的新型科学评价体系。

发展性评价倡导人性化、人本化的发展性教育(Developmental Education),以发展主义课程观(Developmentalistic View of Curriculum)为基础,强调受教育者总体发展的课程观念。从发展性评价产生的历史时期和社会基础来看,发展性评价可以说是一个“后现代”概念。这种评价理念在某种程度上,是受到 20 世纪后半叶西方社会流行的后现代主义哲学与文化思潮的影响。因为后现代主义在评价方面并未提出实用的评价模式和评价方法,其主要贡献表现在对于现代主义评价观及其评价模式的批判上。因此,对发展性评价思想脉络的解析,重点应放在对现代主义评价思维的解构上。评价范式试图超越理性主义浓厚的以“泰勒原理”为基础的传统评价,走向“人性发展”的评价范式。但到目前为止,它存在着破多于立,理论丰富而实践简单异化的问题。

尽管如此,我们针对现行基础教育学生学业成就的评价现状,以素质教育评价为理论视角,寻觅现代主义评价观的局限性,阐释“发展性评价”概念的现实含义,即发展性评价是依据一定的教育发展目标和发展价值观,由评价双方共同承担目标职责,运用质性评价为主要评价方式,对学生的思想品德、学业成就、身心素质等多方面进行价值判断,使学生在评价活动中,不断认识自我、发展自我、完善自我,逐步实现不同层次的发展性目标的过程。

这是一个朴素的解释,发展性评价思想是一个庞杂的理论体系,蕴涵着深刻的教育意蕴,其价值取向至少包含以下几个方面:

第一,主体取向。发展性评价把教学流程中的全部情况都纳入评价范围,强调对教育现象的整体进行有机的系统评价,在班级、学校乃至社会的“共同体”中,由评价主体和学生共同建构“评价”的意义。这种评价理念体现了建构主义评价思想,关注学生对知识的多元建构。评价主体亦具有多元性,在评价过程中重构师生关系、人际关系、社会关系的价值标准。

^① What is assessment of learning[EB/OL]. (2007-09-26). <http://www.ltscotland.org.uk/assess/for/intro.asp>.

例如,建构“校本评价”,学校是学生发展的主体,把学校看作一个具有蓬勃生命力和发展潜能的生命体。发展性评价倡导有特色、民主、自主的校园文化氛围,强调建设文理相通、中西兼容的校园文化环境,用形式多样、积极健康的校园文化活动对学生的发展产生知识滋养与文化浸润效应。

第二,过程取向。发展性评价强调评价过程本身的价值,重视形成性评价(Formative Assessment)的作用。评价不是为了评比、排名、区分等级,而是为了分析、激励与改进,提高学生的发展水平。换句话说,评价本身就是目的。评价与学生的发展过程相融合,在过程中评价,在评价中发展,体现评价过程的自然性、动态性和生态性。

第三,自由取向。发展性评价把人的自由与解放作为评价的灵魂,以促进学生的自主发展为终极目标。教育思想家和哲学家康德说过,“教育中最重大的问题之一,是人们怎样才能将服从于法则的强制和运用自由的能力结合起来,这是教育发展的灵魂所在”。在我们的教育视野中,“循规蹈矩”“好孩子”仍被尊崇,学生的“天性”被教得丧失,这是教育的不幸。加德纳多元智能理论表明,在教育的背后,每一个学生都有自己的智力特色与独特价值,“存在着关于人类天性的完满性的伟大秘密”“人的天性将通过教育而越来越好地得到发展,而且人们可以使教育具有一种合乎人性的形式”。这描绘的正是发展性教育评价思想的本原图景。

第四,美学取向。发展性评价的真谛在于教育鉴赏和教育批评,是一种导向美学精神的评价。评价双方亦只有真正领悟到评价所蕴涵的美的真谛,才有可能真正地理解评价的内在性本质,获得一种理解的激动和美的享受,评价思想在美学价值观的文化沃土中升华,这是发展性评价的最高境界。

要进行评价,首先要有评价价值尺度,即评价的标准或评价期望目标。评价期望目标是指基于教育目标达成度和教育目标综合度的交互式多目标期望水平,是教育评价的参照体系。有关学校教学效果评价研究表明:教师的期望对学生的学业成就有很大影响。弗鲁姆(Vroom)的“期望理论”(Expectancy Theory)也表明激励与达到目标的概率以及由目标导致期望利益的概率的关系。根据这一评价理论,学生学业成就评价的标准及其相关的期望目标参照体系成为教育评价的重要工具。开发课程评价标准及设立期望目标参照体系是学生学业成就评价的最根本的环节。美国著名的“2061 计划”已经提出了一系列科学评价的标准和样例,这些标准是经过许多科学教师、科学教育研究人员的实际测试得以验证的,是可靠的,对我国教育评价具有参考和借鉴意义。我国基础教育质量监测体系建立的重要任务之一就是开发基础教育质量监测评价标准。

建立评价标准以及评价期望目标参照体系是教育评价的基本出发点。发展性教育评价原理决定了发展性评价标准具有多元开放、过程动态、参与生成等呈现流变或转化的特征,倾向于质性评价方法。目前中小学的发展性评价具有方向性迷茫和许多实践误区,其主要原因之一就是缺乏评价标准和发展性评价期望目标的参照,致使评价无的放矢。另外,对发展性评价的片面理解也是实践偏离理论方向的重要原因所在。所以,需要说明的是,对于发展性评价标准,虽然它不再是固定不变的,随着评价活动的展开而不断变化,但我们不能强调评价标准的动态性和多样性而忽视评价标准的存在,否则就会陷入目标“虚无主义”的危险境地,那样,发展性评价只能是“海市蜃楼”,虚无缥缈,偏离发展性评价的真正本质。我们认为,发展性评价是义务教育课程改革的基调与方向,评价标准是学生参与其中所生成的指引自我行动的目标,它包含的不仅有社会、权威的价值,更有个体的价值。因此,评价目标至