



大夏书系·教师专业发展

促进教学的测验与评价

赵德成 著

USING
ASSESS-
MENT
TO IMPROVE
TEACHING AND LEARNING



上海市
著名商标

华东师范大学出版社

全国百佳图书出版单位



大夏书系·教师专业发展

促进教学的测验与评价

赵德成 著



华东师范大学出版社
全国百佳图书出版单位

图书在版编目 (CIP) 数据

促进教学的测验与评价 / 赵德成著. —上海: 华东师范大学出版社, 2016. 4
ISBN 978-7-5675-5026-1

I. ①促 ... II. ①赵 ... III. ①教育测验 ②教学评议
IV. ① G424.74 ② G420

中国版本图书馆 CIP 数据核字 (2016) 第 069856 号

大夏书系·教师专业发展

促进教学的测验与评价

著 者 赵德成
责任编辑 任红瑚
封面设计 淡晓库

出版发行 华东师范大学出版社
社 址 上海市中山北路 3663 号 邮编 200062
网 址 www.ecnupress.com.cn
电 话 021-60821666 行政传真 021-62572105
客服电话 021-62865537
邮购电话 021-62869887 地址 上海市中山北路 3663 号华东师范大学校内先锋路口
网 店 <http://hdsdcbs.tmall.com>

印 刷 者 北京密兴印刷有限公司
开 本 700×1000 16 开
插 页 1
印 张 15
字 数 210 千字
版 次 2016 年 7 月第一版
印 次 2016 年 7 月第一次
印 数 6 100
书 号 ISBN 978-7-5675-5026-1/G·9331
定 价 36.00 元

出 版 人 王 焰

(如发现本版图书有印订质量问题, 请寄回本社市场部调换或电话 021-62865537 联系)

推荐序

李希贵

(北京十一学校校长, 兼任国家督学, 中国教育学会副会长)

认识赵德成老师, 是缘于本世纪初的第八次课程改革。那时我正在潍坊担任教育局长, 听过赵老师一个有关评价改革的报告, 如坐春风, 耳目一新。自然, 我也施加自己的影响力, 把赵老师的研究成果运用到全市的教育教学实践之中。

2006年, 我到教育部基础教育质量监测中心主持工作。为了确保质量监测工作的科学性, 中心高度重视理论研究和技术改进, 我们邀请国内外知名专家开展培训, 将项目反应理论、概率比例抽样、增值性评价、多水平分析等最新相关技术引入到监测实践中, 提高了监测工作的技术水平。但必须承认的是, 监测工作仍然面临挑战, 最主要的一个挑战在于监测试题编制。新课程强调学生的多元发展, 监测内容不能局限于知识与技能, 要结合真实生产生活情境设计表现性任务, 以有效考查学生的批判性思考能力、问题解决能力与创新精神等核心素养。可是, 由于我国教研员对新兴评价方法接触不多, 命题经验缺乏, 我们面向全国范围征集来的多数题目不能真正联系生产生活情境, 评价任务实际上仍然是去情境化的, 所考查的仍然是孤立的知识与技能。加强命题研究, 切实发挥评价对基础教育课程改革的导向、诊断和发展性功能, 是学科教育质量监测工作的突破口。

在学校层面, 评价研究的迫切性更是不言而喻。我到十一学校工作以后, 为增加课程的多样性和选择性, 促进学生的自主学习与个性化发展, 在学校全面推进育人模式构建, 让每位学生都有一张自己选择的课表。随着改革的

持续深入，在我校，师生关系、教师教学方式、学生学习体验都有了很大的变化，学生们学得自主，学得高效。但在选课教育背景下，如何让学生了解自己的学习进展，从而把握自己的学习节奏和调整选课安排？如何贯彻中高考改革新理念，在评价实践中聚焦学生的核心素养？如何评价教师的教学绩效，为教师专业发展注入新的动力？评价问题仍然困扰着我们。

教育评价是深入推动课程改革的瓶颈。只有教育者把握教育评价的本质，掌握新兴评价方法和相关分析技术，能在实践中准确、有效地评价学生的多元发展，才能强有力地引领教师的教与学生的学，切实促进教师教学方式变革。教师需要学习，需要专业支持。

我们通过多种渠道搜集了一些可供参考的书籍，研读后发现国内学者编撰的有些著作理论性很强，但联系实践不够，而国外学者编撰的著作能联系实践，但联系的是国外实践，还需要本土化。正当我们如饥似渴的时候，赵老师出版了他的新作《促进教学的测验与评价》。

赵老师长期在高校从事教育评价领域的教学与研究，著述颇丰，是教育评价领域的权威专家。读他的新作，发现这本书具有三个特点：（1）科学性。赵老师在书中不仅阐释了教育评价的基本理论、常用模式和新兴方式，而且强调评价也需要被评价，尤其要高度重视评价的效度。无论是大规模高利害测验，还是教师自编测验，只有当测验能准确评价要评价的学习目标的达成程度，能基于评价数据做出准确有效的推论时，评价才能切实发挥其发展性功能，为改进教师的教与学生的学提供有价值的信息。（2）实践性。赵老师学问做得很精深，但他没有局限于象牙塔内，而是经常深入中小学，躬身实践，与一线教师研究和解决评价实践中面临的问题，具有丰富的实践指导经验。他的新作理论联系实践，提供了大量鲜活的案例，极富启发性，对中小学教师、教研员和命题者等各种关心教育评价的人指导性很强。（3）可读性。赵老师的课观点明确，逻辑清晰，且风趣幽默。赵老师写书也是如此，他心中始终装着读者，以第一人称与读者“对话”，语言朴实、简练和流畅，有吸引力。

这本书的出版恰逢其时。希望更多的教育者、研究者，乃至师范生能发现这本好书，能反思与改进自身评价实践，并进而深化课程与教学改革，推动教育创新。

自序

在我国，教育评价思想源远流长。早在西周奴隶制时期（公元前 1046—前 771 年），为教育贵族子弟而设立的太学就已建立了系统的评价制度。《礼记·学记》有载：“比年入学，中年考校，一年视离经辨志，三年视敬业乐群，五年视博习亲师，七年视论学取友，谓之小成。九年知类通达，强立而不反，谓之大成。”这一制度，不仅规定了大学实施评价的时间，隔年进行一次，而且规定了评价的内容与标准。及至隋唐，我国开始科举取士，在考生来源、考试科目、考试办法、录用程序等诸方面已形成一整套较完备的制度。

然而，中国古代深厚的教育评价思想，却因种种历史原因，未能在 20 世纪伴随西方教育测验与评价运动的兴起而率先发展起来。世界上第一个现代教育评价概念诞生于美国，由当代著名教育家泰勒（R. W. Tyler）在八年研究期间（1933—1941 年）提出。泰勒认为，评价是课程管理中的一个重要环节，它在本质上是确定课程和教学大纲在实际上实现教育目标程度的过程。无论是一门课的教学，还是一个单元乃至一节课的教学，教师都要对照预期目标考查学生达成目标的程度。这一概念以及由此衍生的目标本位评价模式（Goals-Oriented Evaluation Model），奠定了教育评价学科的基础。此后，随着教育思潮及哲学思想的演变，CIPP 评价模式、目标游离评价模式等众多新模式相继涌现，但其中多数模式都根植于目标本位模式，是对目标本位模式的

修订与发展。

相对而言,现代意义上的教育评价研究在我国起步要晚得多。20世纪80年代,一些学者陆续从国外及我国港台地区引介有关文章及专著,其中比较有代表性的书籍有邱渊等人翻译、布卢姆等人编写的《教育评价》,以及陈玉琨、赵永年选编的教育学文集第16卷《教育评价》,这些文献资料全面介绍了世界教育评价改革的动态与趋势,推动了我国教育评价实践活动和研究工作有序展开。1984年,我国正式加入“国际教育成就评价协会”(International Association for the Evaluation of Educational Achievement,简称IEA);1985年,中共中央发布《关于教育体制改革的决定》,指出“衡量任何学校工作的根本标准不是经济收益的多少,而是培养人才的数量和质量”,教育评价作为教育改革中的重大问题开始受到广泛关注,有关制度建设工作的渐次启动。自此,我国教育评价事业才开始进入快速发展时期。

评价在整个教育质量保障体系中扮演着相当重要的角色,具有导向、督促、诊断、激励和发展等多重功能,是教育改革推进中的关键问题。进入21世纪以来,第八轮基础教育课程改革不仅在课程功能、结构、内容及实施方式上提出了新的理念,在教育评价领域也进行了整体设计。《基础教育课程改革纲要(试行)》明确指出,要“改变课程评价过分强调甄别与选拔的功能,发挥评价促进学生发展、教师提高和改进教学实践的功能”,建立适应课程改革需要、符合素质教育要求的发展性教育评价体系。

新课程实施以来,广大教育工作者和研究者在理论、政策和实践层面都对评价进行了大量深入的探索,取得了积极的进展。特别是教育评价实践层面出现了许多可喜的变化,主要有:评价功能由甄别选拔走向激励发展,无论是对学生的评价,还是对教师和学校的评价,都力图体现育人为本,都将学生发展放在第一位,更加强调诊断、激励与改进;评价内容走向多元化,不仅重视学生的知识与技能,而且关注学生的复杂能力和情感发展;评价主体由自上而下的单一主体评价向多主体协商对话转变,评价的民主性与公平性得到加强;评价方法改革持续跟进,成长记录袋、表现性评价等一些新兴评价方式逐渐走进教师视野,成为评价改革的亮点。

最近几年,随着课程改革的深入推进,我国中高考改革的力度也逐步加

大。除了考试科目和形式上的改革，命题改革开始提上日程。许多省市在命题中呼应课改方向，越来越强调考查学生运用所学知识独立思考与分析问题、解决问题的能力，加强表现性评价的应用。

以北京市为例，该市于2013年底公布的高考改革方案强调各学科要突出基础性、应用性与工具性，使评价与学生生活及现代社会联系起来，考查学生在真实情境中分析和解决问题的能力。比如，语文学科命题要“充分体现语文的基础性和作为母语学科的重要地位，注重考查内容与社会生活实践的联系”；又如，理科综合命题要“从学生已有的经验和将要经历的社会生活实际出发，通过生产、生活中的一些真实情境和实验观察、自然体验，考查学生联系实际深化、应用知识的策略和学科素养”。这种改革代表了当前评价与考试改革的新方向。

中高考是“指挥棒”，中高考考什么，中小学就教什么；中高考怎么考，中小学就怎么教。可以说，命题改革直击课程改革的核心，对于深化课程改革和推进素质教育意义重大。当前，中高考命题改革已经取得了突破性进展，中小学教育实践也崭露出很多积极的变化。

教育评价改革在逐步推进。但必须承认的是，实践中仍然存在很多亟待关注和改进的问题。比如，教育者在测验编制过程中关注了内容效度、难度和区分度的分析，但对结构效度的论证、审查重视不够。一份测验或一道测验题目是否测到实际想要测量的特质，基于学生在测验中的表现能否做出有效的推论，不少教育者不知道如何分析和改进。又如，中高考命题者有意识地创设一个真实的生产生活情境，试图考查学生发现、分析和解决实际问题的能力，但由于情境与问题缺乏实质性联系，相互脱离，最终考查的还是孤立的知识与技能。

一名普通的中小学教师如何在教学实践中评价学生学习进展，诊断自己教学的成效与得失，使评价有效引领、支持和辅助自己的教学？一位教研员或命题者如何编制出一份符合测量学要求、经得起元评价考验的试题，以确保关键考试的命题质量？一所学校如何构建体现育人为本理念、能有效促进学生发展的教育评价制度？一个地区如何做好区域教育质量监测，及时发现当地教育教学实践中的问题，并予以有针对性的补救，建立健全教育质量

保障体系?……一系列问题有待理论工作者与一线教育工作者联手进行深入研究。

2000年,第八轮课程改革启动之初,教育部成立了国家基础教育课程改革“促进教师发展与学生成长的评价研究”项目组,由时任北京师范大学副校长的知名心理学家董奇教授担纲主持,召集全国各地有关专家就教育评价领域的一些重要问题开展专题研究。项目组编译出版“新课程与教育评价改革译丛”,陆续译介《有效的学生评价》、《测验的反思:对高利害测验的建议》等十几本国外权威著作或教材,系统介绍了国外教育评价领域的新成果、新方法和新实践。同时,项目组还及时总结本土研究成果,编辑“新课程与教育评价改革丛书”,先后出版了《成长记录袋的基本原理与应用》、《小学日常评价与学业考试改革的探索》等多部著作,为新课程背景下的教育评价改革提供了有力的专业支持。

作为“促进教师发展与学生成长的评价研究”项目组核心成员,我积极参与各种项目工作,并经常深入中小学和教研机构,与一线教育工作者一起开展行动研究,解决实践中面临的突出问题,积累了大量鲜活的实践案例。我还及时总结研究成果,与人合作出版了《成长记录袋的基本原理与应用》(陕西师范大学出版社,2002年)、《新课程与学生评价改革》(高等教育出版社,2004年)等多部著作,独立发表了《减负背景下的评价与考试改革》、《成长记录袋应用的回顾与反思》、《表现性评价应用中的问题——基于2015年多省市中高考命题的分析》等二十余篇学术文章,成果不可谓不多。但纵观我以及其他研究者的诸多成果,我发现适合一线教育工作者阅读的专著非常少。于是,我萌生一个心愿,就是基于我在北京师范大学多年来从事“教育测量与评价”课程教学以及在全国范围内面向教育工作者开展教育评价专题培训的经验,整合有关研究成果,撰写一本专著,希望能深入浅出,理论联系实际地阐述评价原理,帮助教育者掌握评价基本技能,让实践中的评价更加有效,积极促进教师的教以及学生的学。

经过接近一年时间的写作及反复修改,现在这本专著已呈现在大家面前,取名为《促进教学的测验与评价》,很朴实,正如本书的写作风格,质朴又不失生动,平实又不失深刻,实实在在。全书由七部分构成,不仅细致阐释了

测验与评价的基本原理，而且对本轮新课程实施中的热点问题（如表现性评价、成长记录袋、情意领域评价）进行了深入讨论与案例分析。为便于阅读，我在每个部分开头不仅撰写了简短的导读，而且提供了概览，让读者能先对本部分内容有一个整体的了解，再根据个性化需要决定阅读的起点和节奏。如果您是一名研究者，想对某些问题进行细致推敲，每部分末尾我都列举了参考文献，您可以拓展阅读，深入探讨。

测验与评价是教学有机的组成部分，是教学质量保障体系中的重要一环。评价与教学相互联系、相互促进，相辅相成。优秀的教师不能只会教，不会评，“评价”是教师专业能力中的重要成分。优秀教师既要精通教学，又要研究评价；既是一名教学专家，又是一名评价专家。衷心希望本书能够给大家多样化、深层次和成长性的启发，切实引导教育者的观念更新及行为转变，促进教师专业化发展，并有力推动相关制度建设、政策完善和实践创新。

当然，尽管我在撰稿过程中一直本着严谨的态度不断修改和订正，但由于本人才疏学浅，书稿中可能仍有纰漏之处，恳请各位专家和广大读者不吝赐教。如果您对本书有什么意见与建议，或者想就某些问题与我进行深入探讨，欢迎与我联系。我的电子邮箱是：zhaodecheng@bnu.edu.cn。

2016年5月于北京师范大学

CONTENTS 目录

推荐序 (李希贵).....	1
自序.....	3
第1章 走近测验与评价.....	1

教学离不开测验与评价。教师只有在教学中不断通过测验与评价分析学情,诊断学生达成预期目标的程度,才能使教学更有针对性和吸引力,更能有效地促进学生发展。

概览.....	2
初步认识教育评价.....	3
几种经典的教育评价模式.....	9
教学中的测验与评价.....	16
为什么重视测验与评价.....	25
参考文献.....	30

第2章 如何设计测验与评价33

测验与评价需要精心设计。教师只有以严谨的态度设计或选编测验,确保测验的准确性和有效性,测验才能为教学提供有价值的信息。

概 览	34
明确测验与评价的目的	35
编制测验细目表	38
选择评价任务类型	48
设计评价任务	56
汇编测验	63
参考文献	66

第3章 什么样的评价是好的评价69

好的测验与评价首先要确保其效度,要能测到所欲测的目标学习成果,能依据所收集的信息做出准确而有效的推论。此外,好的评价还需在信度、难度、区分度、公平等方面符合测量学要求。

概 览	70
效 度	71
信 度	81
难度和区分度	90
公 平	95
参考文献	98

第4章 表现性评价99

表现性评价克服了传统测验仅能测试低水平知识和孤立技能的弊端,能测量出学生在真实世界中的复杂成就和情意表现。加强表现性评价的应用,是当前我国中高考命题改革的基本方向,也是世界范围内教育评价学科发展的共同趋势。

概 览	100
什么是表现性评价	101
表现性评价的历史性分析	104
好的表现性评价案例	107
表现性评价的优势与不足	112
我国表现性评价实践中的常见问题分析	114
未来进一步推进表现性评价的建议	125
参考文献	129

第 5 章 成长记录袋 131

成长记录袋是一种新兴的质性评价方式。这种评价方式将评价与教学有机整合起来,能有效考查学生在真实、复杂情境中自主思考、积极建构、分析问题和解决问题的能力。

概 览	132
成长记录袋的定义及特点	133
成长记录袋的优势与不足	135
成长记录袋的创建与使用	140
案例:阅读成长记录袋的创建与使用	145
实践中应用成长记录袋的主要问题	153
成长记录袋的未来发展方向	156
参考文献	160

第 6 章 情意领域的评价 163

新课程实施以来,情感态度与价值观的发展成为课程目标的重要组成部分,它涉及学习情感、对学科的态度、对自我的态度,以及价值观等关键变量。教师要掌握观察、访谈、里克特量表等多种常用评价方法。

概 览	164
情意领域的目标	165
情意评价的重要意义	172
常用的情意评价方法	175
学习情感评价：一个实例分析	183
情意评价的改进建议	188
参考文献	190

第7章 评价结果报告191

在发展性评价理念指导下，评价结果报告方式也要推陈出新，与时俱进。好的结果报告要采用定量评价与定性分析相结合的方式，向阅读者描绘学生学习与发展的主要轮廓与关键细节。

概 览	192
学业成绩评定	193
评语写作	200
评价报告单	212
家长会	218
参考文献	224



走近测验与评价

第1章

长期以来，教学相长的道理已经被教师们广泛理解和接受，但评价与教学相互依赖、相互促进的关系却较少被人注意到。实际上，评价是教学的有机组成部分。教师只有在教学中不断通过测验与评价分析学情，诊断学生达成预期目标的程度，才能使教学更有针对性和吸引力，更能有效地促进学生发展。

从这个意义上来说，教师仅钻研教学，即便已经很精通，但如果不懂测验与评价，也很难成长为一名专家型教师。教师既要做教学专家，又要做评价专家。

概 览

1. 评价是基于标准对学生发展变化所进行的综合评判，它对学生发展水平的描述既有定量的，又有定性的；它不仅有描述，还基于定量和定性数据进行价值评判。相对于测验而言，评价是一个更加全面、涵盖面更广的术语。

2. 教育评价模式将复杂的教育评价工作简单化、程序化和操作化。由泰勒提出的目标本位评价模式是当前世界范围内影响最大、应用最广泛的一种模式。教师要熟悉这种模式，能参照这种模式设计与实施评价。

3. 在教学实践中，教师经常对学生的评价情况进行评价。根据测评应用的时间点和所发挥的功能，可以将教学中的评价区分为诊断性评价、形成性评价和终结性评价。

4. 诊断性评价是教学工作一个不可或缺的重要环节。教师在计划学期教学、单元教学，乃至一节课教学的时候，都要重视诊断性评价，全面、客观地开展学情分析，以提高教学的针对性和吸引力。

5. 形成性评价的本质特点是评价所收集的信息主要用于改进。形成性评价不能简单地等同于平时分数记录簿。教师只有对平时作业和测验进行深入分析，并在此基础上改进教与学，形成性评价才能发挥其应有的作用。

6. 终结性评价指在某项计划或方案结束后对其最终结果进行的评价。从性质上来看，它主要属于标准参照测验，要严格依照课程标准、教学大纲，或某个客观标准编制测验任务，以考查学生达成标准的程度。

7. 测验与评价是教学的有机组成部分。优秀的教师不能只会教，不会评，“评价”是教师专业能力中的重要成分。

初步认识教育评价

什么是教育评价？教育评价的本质是什么？如何理解它与测量、测验的关系？这些是教师首先需要弄清楚的问题。在实践中，之所以有些教师在评价实践中出现这样或那样的问题，往往与他们对基本概念理解的偏差有关。

评价不能简单等同于表扬

新课程实施以来，很多教师都开始意识到评价的重要性，特别是在教学中要加强形成性评价，通过评价对学生的学习予以诊断、反馈和激励。但是不少教师将评价简单地等同于表扬，以为只要在课上表扬了学生，就算加强了评价。有一位小学语文教师的公开课十分典型。她给1年级学生讲了《司马光》一课。以下是她刚开始上课时的导入片段。

“同学们，你们好！古时候，有一个大文学家，叫司马光。他小时候很爱学习，也很聪明。有一天，他跟几个小朋友在花园里玩。花园里有个假山，假山下面有一口大水缸，缸里装满了水。有个小朋友很淘气，爬到假山上去玩，一不小心，掉进了水缸。水缸里水很深，要是再不把他救出来，他的生命就有危险了。”老师声情并茂地介绍了故事的背景，然后问道，“如果你是司马光，你会怎么办？”

一个小学生举手说：“我拿一个吸管一样的长棍让那个小朋友先呼吸着，再想其他办法去救他。比如说，找大人。”

“你是个善良又充满智慧的孩子，可是要找一个你说的木棍可能来不及了。”老师对学生的回答做了及时的回应。

“我找一块大木板，让那个小朋友趴上去浮起来。”第二个学生举手回答道。