

教育评价是教育质量的守护神吗？ ——一个古今教育评价重心变迁的解析视角

刘 尧

摘 要：教育评价是教育质量守护神吗？回答这个问题需要从教育质量悖论谈起。千百年来“状元难入大家之列，而大家又很难高中状元”的教育之谜背后，隐藏着“高质量教育，难以培养杰出人才”的教育质量悖论。教育评价推崇的高质量教育，为何难以培养杰出人才呢？盖因教育评价历来没有针对教育质量展开。我们梳理古今教育评价重心变迁历史发现，大约有以官员选拔为重心的科举制度、以教育量化为重心的教育测量、以教育目标为重心的泰勒模式、以教育价值为重心的后泰勒模式。一直到了 20 世纪 80 年代之后，人们才对教育评价与教育质量的关系有所反思，开启了以教育质量为重心的增值评价探索。

关键词：教育质量；教育评价；科举制度；教育测量

中图分类号：G40-058.1 文献标识码：A 文章编号：1671-0169(2016)06-0145-07

DOI:10.16493/j.cnki.42-1627/c.2016.06.016

教育评价是教育质量守护神吗？我提出这一问题，一定会被认为是明知故问或是无理取闹。事实上并非如此简单，要回答这个问题，还需要从教育质量悖论回溯整个教育评价重心的变迁历程。2012 年，莫言获得诺贝尔文学奖，填补了中国大陆诺贝尔奖的空白。2015 年，屠呦呦获得诺贝尔生理学或医学奖，实现了中国大陆科学家获得诺贝尔奖的梦想。然而，他们极其平凡的教育经历，再次颠覆了我国社会对教育质量的既定看法。如果说这是极端个例，那么有关科学研究向人们揭示了，自古“状元难入大家之列，而大家又很难高中状元”的教育之谜背后，隐藏着业已存在千余年的“高质量教育，难以培养杰出人才”的教育质量悖论^[1]。教育是培养人才的社会实践活动，能培养出杰出人才的教育才能称为高质量教育。可是，古今教育评价推崇的高质量教育，确实难以培养杰出人才。这一教育质量悖论何以形成？原因虽然有很多，但最重要的是古今教育评价从来不是针对教育质量展开的。笔者对古今教育评价重心进行了梳理，大约有以官员选拔为重心的科举制度、以教育量化为重心的教育测量、以教育目标为重心的泰勒模式、以教育价值为重心的后泰勒模式，一直到了 20 世纪 80 年代之后，才出现了以教育质量为重心的增值评价探索。

一、以官员选拔为重心的科举制度

虽然教育评价产生于近代，但在古代就已经孕育，它就是中国古代的科举制度。科举制度是通过科举考试，即“设科取士、分科举人”的官员选拔制度，始于隋初止于清末，沿用了 1 300 年之久。中国古代秦以前的商周时代，官员选拔主要是贵族世袭制；两汉时期官员选拔主要实行察举

作者简介：刘尧，浙江师范大学教育评论研究所教授、所长（浙江 金华 321004）

制;魏晋南北朝时期,察举制演化成为九品中正制,察举大权完全掌握在豪门世族手中,以致于出现了“上品无寒门,下品无世族”的局面。

隋朝统一中国后,废除了豪门世族把持察举的特权,九品中正制也逐渐废止。隋炀帝大业二年(公元606年)置“进士科”,开启由中央政府设科目、地方举人才、经考试后入仕的科举制度。唐宋元明清选士沿用隋制,科举取士、分科选拔、逐级考试,在考生来源、考试科目与方法、录用程序等方面进一步完善。虽然清末科举制度在方法上已经完备,但在内容等方面却严重阻碍了社会进步,在清朝行将灭亡之际正式退出了历史舞台。

科举制度相对于隋朝之前的世袭制和举荐制,无疑具有极其重大的进步意义。科举制度以自由报考的方式,按科举考试的成绩高低决定取舍为准则,对过去世卿世禄、依门第举荐官员的制度进行了彻底否定。科举制度最根本的进步意义是,通过科举考试选拔官员,最大限度地摒除了权力的恣意干扰,保证了朝廷官员选拔的公平公正,提高了各级官员的知识水平和文化素质,进一步巩固了中央的统治地位。

科举制度不仅对中国社会影响深远,而且对西方社会有较大影响。1896年出版的《中国环行记》一书中,美国传教士丁韪良说,“尽管有其缺陷,科举制度对维护中国的统一和帮助它保持一个令人尊敬的文明水准,起到了比任何其他制度更大的作用”。“在英国、法国和美国正在取得进展的文官考试制度,是从中国经验中借鉴而来的。”^[2]1855年,英国建立了“文职人员委员会”,采取公开考试招募文职人员。1883年,美国成立“文职人员事务委员会”,规定担任公职必须通过考试。

科举制度作为一项官员选拔制度,虽然在清朝末年正式退出了历史舞台,但其精髓依然被西方现代的文官选拔制度、我国现行的公务员选拔制度,以及现代教育考试制度所继承和发展。科举制度作为一种通过考试选拔人才的制度,以其先进的方法和严格的管理,开创了考试选拔人才的新纪元,在人类教育评价史上留下了极为光彩夺目的一页,因此也被西方社会认为是中国除了“四大发明”之外的“第五大发明”。

尽管科举制度不是教育制度,但它与教育关系密切甚至主宰了教育,它将读书与做官紧密结合起来,具有教育考试性质。从隋朝到清朝的教育几乎都是围绕科举制度展开的,这一时期的中国教育可谓是“科举教育”。事实上,中国古代的教育制度总是与当时的科举制度结合在一起的,教育目的主要是为统治阶级培养人才。以致于科举考什么,学校就教什么,官府需要什么人才,学校就培养什么人才。可以说,科举制度主宰着教育,教育从属并服务于科举制度。

当时的学校是科举的附庸,办学方向就是准备参加科举考试。科举注重文辞,学校也注重文辞而少实学;科举注重贴经、墨义,学校也采用这些考试方法,结果学生只重视记诵,不求义理……千余年的科举文化传承,导致古今中国学校的应试教育经久不衰,最大弊病是学子自动地挤上一条狭窄的成才之路。科举考试所限定的科目和内容,不管你喜欢或不喜欢,只要参加考试就得接受特定内容的考核,压抑了个性也桎梏了学子的聪明才智。

二、以教育量化为重心的教育测量

从广义上讲,测量是根据某种法则给事物分派数字。教育测量是依据一定的法则制作量表,对教育现象进行量化描述,即对教育的特征、属性及其运动、发展规律的定量描述。过去的教育测量主要用于对学生的精神特性进行量化测定,今天则是对学生的学习能力、学业成绩、兴趣爱好、思想品德、身体素质以及教育措施上的许多问题的量化测定。教育测量旨在获得有一定说服力的数量事实,是一种以量化为主要特征的事实判断。

1845年以前是主观经验化考试阶段,在这一阶段所采用的测验方法是传统的考试方法,就是

由教师根据主观经验命题、制订评分标准、进行成绩的评定及对考试分数进行解释和分析。1845年，波士顿学校委员会第一次使用试卷大规模地测验学生学习成绩。1864年，美国的费希尔（Fisher）曾广泛搜集学生的书法、拼字、算术、文法、作文、历史、自然、图画、法文等作业样本编成量表集，作为测量各科成绩的标准。

英国的高尔顿（Galton）通过对个体差异的长期研究，于1869年发表了《遗传的天才》一书，拉开了教育测量的序幕。1879年德国的冯特在莱比锡首创了心理实验室，逐步摸索出一套测量方法，对教育测量的发展产生了积极影响。美国的桑代克（Thorndike）1904年出版了《心理与社会测量导论》一书，系统地介绍了统计方法和编制测验的基本原理。桑代克提出“凡存在的东西都有数量，凡有数量的东西都可以测量”的观点，为教育测量奠定了理论基础。

1905年，法国的比纳（Binet）与西蒙（Simon）在《心理学年报》发表的《诊断异常儿童智力的新方法》中，介绍了首个智力量表——《比纳·西蒙量表》。该量表有30个由易到难排列的项目，可以测出高低不同的智力。1908年，美国教育家斯通（Stone）提出了以小学生为对象的客观化测试算术的方法。同时，美国有30到40个大学对各种教育活动进行综合测量，测量采用了客观测验方法，测量内容包括算术、拼字、书法和英语作文等。

20世纪20年代后期到30年代，更多的大学介入到教育测量领域的研究，并指导当地的教育测量运动。教育测量发展成3种不同性质的测验：学历测验、智力测验和人格测验。这一时期，教育评价在定量化、客观化与标准化方面取得了重要进展，其主要工作是编制各种测验量表以测量学生的心理机能与特征。最流行的工具为“智力测验”，它是当时衡量学生的唯一标准。评价等同于“测量”，并以测量结果对学生进行分类。

从19世纪末到20世纪30年代，教育评价本质上是以教育测量的方式，测定学生对知识的记忆状况或某种特质。这一时期，学校被看作“工厂”，学生被看作“产品”，课程、教材、教法都是用来进行“产品”加工的工具，而教师则是操作工，整个教育就是“产品”的生产过程，学校就按照社会需要把学生加工成一件一件标准化的“产品”。教育评价就像对待工厂产品规格一样，按照精确的标准，严格测量学生的知识和心理素质。

虽然，教育测量使考试和测验较以前更客观化和标准化，并把人的智力和人格等变成测量的对象，力求从量上把握这些精神现象，但是，在20世纪20年代后期，还是遭到了美国社会的强烈批评。比如，无论是知识测量还是人格测量，都只能做片断测定，不能完全了解人格与知识的发展过程；测量只注重客观的信度，不足以说明效度；教师采用的学业测量，根本是教科书中心主义；测量容易培养学生的个人主义与被动式学习态度；等等。

三、以教育目标为重心的泰勒模式

教育评价何以从教育测量发展而来？1929年，美国经历了一场严重的资本主义经济危机，经济萧条使许多青少年无法就业而只能进学校，但当时美国的中学在大学入学考试指挥棒制约下，所设置的课程仅为大学入学服务，不能适应青少年的就业需要。因此，美国社会对教育测量提出强烈批评。在此背景下，美国进步主义教育协会从1932年到1940年进行了8年实验研究，旨在编制一套统一连贯的课程，满足学生的不同需要，尤其是满足学生适应社会生活的需要^[3]。

为了对这项实验研究进行评价，成立了以泰勒（Tyler）为首的教育评价委员会。泰勒在改进学校测量方法的实践中认识到，测量要以一定的教育目标为指导，而不能以教科书为中心。于是，他将评价概念应用到教育上，以示与测量的区别。20世纪30到50年代，泰勒提出了以教育目标为核心的评价理论并明确给出教育评价概念，即教育评价是衡量实际活动达到教育目标的程度，评

价的本质是描述教育结果与教育目标的一致性程度。

泰勒以教育目标为核心的评价理论,让人们意识到:通过简单的事实测量不能引发高级心理过程;事实性知识的教学并不能促进还可能干扰教育目标。这一时期,虽然测量还是评价的一个重要工具,但已不再是唯一的评价手段;评价者也不再仅仅是“测量员”,更是一个“描述者”;评价目标已不再是学生本身,而是什么样的学习目标模式对学生最有效;评价范围也涵盖了教学活动的整个过程,涉及事前、事中、事后各个环节。

1949年,泰勒出版了《课程与教学的基本原理》著作,叙述了他的课程设计和评价理论(即泰勒模式)。泰勒模式又称为目标模式或目标达到模式,它以教育目标为导向,把教育目标转化为可测量的行为目标,并根据行为目标编制课程教材或教学方案、开展教学活动,然后依据行为目标对教学活动的效果进行评价,即判断实际教学活动效果达到预期目标的程度。今日广泛应用的教育评价理论,就是在泰勒模式基础上不断完善发展起来的。

泰勒认为,教育目标是有意识地想要达到的目的,也就是学校教师期望实现的结果。教育目标是选择材料、勾画内容、编制教学程序以及制定测验和考试的准则。在泰勒模式中,确定教育目标是非常关键的。首先,要考虑学生的需要、当时的社会生活以及学科专家的建议等多方面的信息,对教育目标做出明智的选择;其次,用教育哲学和学习理论对已选择出来的教育目标进行筛选;再次,陈述教育目标,每一个教育目标要包括行为和内容两个方面。

教育是使人的行为方式发生变化和改进的过程,这些行为方式的变化就是教育效果。人的行为有可量化,也有难量化的,除测量外还需要用其他评价手段来检查教育效果。运用泰勒模式进行评价,就是确定教育效果到达教育目标程度的过程。具体讲就是:确定教育目标;依据预定的教育目标中期望学生变化的那种行为来设计评价情境;选择和编制能够引发教育目标所期望的那种行为的评价工具;再以教育目标为依据判断教育效果。

泰勒把评价引入教育,创建了教育评价理论,他强调用行为动词来为教育目标定义。尽管行为目标清晰、直接、可观察、可测量,对教育目标的描述和外显化有积极作用,但行为目标缺乏完备性,难于覆盖教育产生的全部效果;把教育目标作为评价的出发点和归宿,又缺乏对其合理性的考量;只观察预期教育目标达到度,没有考量非预期教育目标;教育目标是管理者确定的,很少考虑学生需要;注重评价的鉴别、确证和检查功能,忽视形成性功能和决策性功能;等等。

四、以教育价值为重心的后泰勒模式

1957年,苏联人造卫星上天,激起了美国社会对高中教育的批评。1963年,美国政府提出对教育的效能进行评价并拨款支持。此时,在泰勒模式遭遇众多批评的同时,对其进行改进的流派不断涌现。比如:被称为决策类型模式的CIPP模式;反映活动参与者意图的目标游离模式;采取准法律过程评委会审议形式的手模式;以所有与评价利害关系人所关心的问题为中心的应答模式;采用系统分析方法的医疗模式;等等。这些改进后的新模式,在此统称为后泰勒模式。

后泰勒模式认为,评价在本质上是“价值判断”。评价者不仅要运用测量手段去收集各种数据,还要制定一定的判断标准与目标。“第四代教育评价”还提出了“共同构建”、“全面参与”、“多元价值”的思想。这一时期教育评价的特点是:明确提出评价为决策服务;开始探讨教育目标本身的科学性和合理性,即教育目标本身成为评价的对象,同时预期教育目标之外的教育效果也受到关注;正式提出形成性评价的思想,“评价的最主要目的不是为了证明,而是为了改进”^[4]。

克龙巴赫(Cronbach)认为,评价的重点应该放在教育过程之中,对教育决策给予必要的改进,而不是只关心教育目标到达度。1963年,他提出了“评价是为进行决策提供信息的过程”的

思想。1966年，斯塔弗尔比姆（Stufflebeam）在此基础上提出 CIPP 模式，即背景评价（Context Evaluation）、输入评价（Input Evaluation）、过程评价（Process Evaluation）、结果评价（Product Evaluation）。CIPP 模式把背景、输入、过程和结果综合起来评判，为决策过程提供全面的信息，突出了评价的决策功能。

斯克里芬（Scriven）认为，教育评价除了在活动结束后对其最终结果进行的总结性评价外，还要重视在活动过程中进行的形成性评价。1967年，他将评价分成形成性（Formative）和总结性（Summarive）两类。斯克里芬还认为，实际进行的教育活动除了收到预期效果外，还会产生泰勒目标模式中得不到反映的各种重要的“非预期效果”。1972年，他提出了不受预定目标影响的目标游离模式，让评价者注重广泛的可能效果。

1975年，比贝（Beeby）认为，评价在于收集系统而非零散的资料，并将这些资料加以精心地整理和解释，引入评判性的思考。比贝对评价的定义是，“系统地收集信息和解释证据的过程，在此基础上的价值判断，目的在于行动”。他强调评价的本质是价值判断，评价要对教育活动包括教育目标本身作出判断，使评价活动有助于决策的科学化，对实际工作具有指导意义。比贝深化了教育评价的内涵，使评价的主要特征“价值判断”受到人们的关注。

1984年，古巴（Guba）和林肯（Lincoln）出版的《第四代教育评价》一书认为，评价是对被评事物赋予价值，本质上是通过“协商”形成“心理建构”。他们提出了“共同构建”、“全面参与”、“多元价值”思想，把评价看作是一个由评价者不断协调各种价值标准间的分歧、缩短不同意见间的距离、最后形成一致看法的过程。“第四代教育评价”还把评价过程的控制特点与评价对象的伦理要求协调起来，既注重评价事实的客观性，也注重评价方法的道德性^[5]。

针对泰勒模式存在的“浓厚的管理主义倾向”、“忽视价值多元性”、“过分强调科学实证主义方法”等问题，后泰勒模式反对管理主义倾向，坚持价值多元性，尊重被评价者的人格、尊严和隐私……评价被看作是评价者与被评价者，通过事实认识、体验和价值判断活动，进行相互交流、相互理解、相互启发、相互学习、共同创造的学习、交往和双向建构过程。评价的最终目的是提高被评价者的学习与工作效率。

五、关注教育质量的评价重心转移

20世纪80年代以来，西方学者对以往教育评价和教育质量之间的关系进行了深入研究后认为，以往教育评价看重的教育资源与质量之间的相关性比较弱，对于教育资源的要求不足以保证教育质量。以往教育评价虽然覆盖了教育投入、过程和产出等方面，但评价重心在教育资源。这种以教育资源为重心的评价，对改善教育条件有积极作用，而对教育质量提高作用很有限。因此，教育评价重心需要转移，教育评价应该更加重视与教育质量直接相关的学生学习效果。

比如，泰勒模式重视目标到达度的终结性评价，过于强调结果而缺乏对投入与产出的效益比较，只考虑教育产出差距，而不考虑投入差距（生源质量、教育经费、师资水平等方面的差距）。事实上，由于学校的师资、经费、地理位置、学生的家庭背景等许多因素，学校自身无法控制。学校之间出现教育质量的差距，并非完全由学校工作不力造成。如果教育评价将社会、历史等原因造成的教育质量差距归咎于学校，是不利于调动学校教育教学积极性的。

以往这种引起教育资源竞争的评价，在我国还导致学校走上以片面追求升学率为目标、以抢夺优质生源为重点、以换取资源倾斜为动力的畸形发展道路。多少年来，我国形成的以资源优势换取所谓高质量教育的“好学校”，极大地挫伤了其他学校的积极性，其后果是所谓的“差学校”教育质量的下降与整体教育水平的降低。实际上，从教育资源与教育质量之间弱相关的研究成果来看，

所谓的“差学校”在教育上取得的效能可能会更大。

随着对教育产出认识的深化,发达国家开始将评价重心由资源占有向资源有效利用转移,衡量质量的最终落脚点和检验标准更加注重学生的学习效果。英国学者麦尔肯·弗雷泽(Fraze)认为,“教育质量首先是指学生发展质量,即学生在整个学习过程中所‘学’的东西,包括所知、所能做的及其态度”^{[6](P59)}。以学生学习效果为重心的评价,主张从每个学生发展的内在需要和实际状况出发,评价学生个体的发展进步,并通过评价促进学生向更高、更美、更远的方向发展。

《欧洲地区高等教育质量保障准则与标准》第3项指出,“要突出对学生学习产出的测量”。在评价方式上,可以是“诊断性的、形成性的或总结性的”。对学生进行评价的实施者,应当由“充分理解评价在学生进步过程中所起的作用”的人士来担任。他们认为,学生是教育质量的最终体现者,学生评价是教育评价的中心环节。美国的六大区域认证机构的院校认证标准,不仅重视院校使命和培养目标的确立以及办学资源的投入,更加重视教育的产出,即学生的学习效果。

美国以法律条款的形式,对学生的学习效果加以约定。有关法律指出,学院和大学而不是政府,应该对学生的学习效果负主要责任。学校要确立对学生学习效果的期望,评价者要从学校所确立的对学习效果期望的水准与该期望水准已经实现的证据两方面把握学校的效能责任^[7]。对学生学习效果的评价主要依据直接与间接证据。直接证据包括“核心课程”的成绩、专业和实习成绩等材料,间接证据包括学生的学习档案袋、作业等材料。

英国高等教育质量保障局的院校审核流程要求学校提交的书面材料,包括学生是否获得预期的学业效果,在专业学习的过程中面临的问题,学生成绩评定方式,学生作为学习者的体验,学生参与学校教育标准制定和质量管理的情况等。2001年,英国高等教育质量保障局公布了适用于全国的学位标准,在阐述各级学位的质量和水平时,用了一半的篇幅列举各级学位持有人应有的学习产出。他们认为,教育质量是为帮助学生获得某一学位或学历提供给学生的学习条件所达到的效果。

六、以教育质量为重心的增值评价探索

教育评价如何处理客观存在的教育资源差距等社会因素,进而提高教育质量呢?许多国家进行了有益的探索,增值评价就是在这一背景下产生并发展起来的。增值评价思想源于1966年的科尔曼报告(Coleman Report)。1964年美国的科尔曼(Coleman)带领一个研究小组收集了美国各地4000所学校60万学生的数据,在对这些调研材料进行分析的基础上,撰写了《关于教育机会平等》报告。1966年,科尔曼向国会递交了该报告(即科尔曼报告)。

科尔曼报告把学生学业成就引入到教育机会平等的研究领域,通过对学校的调查综合分析教育投入与产出的关系。该报告认为,评价学校的成功与否,只能检查它是否能减少学生对社会出身的机会依赖,教育机会均等的学校不仅意味着资源均等,还包括教育效能均等。学校不仅是被动提供平等的教育资源,而且要提供一个解放孩子潜能、使其免于因为出身和社会环境而带来不平等的教育环境。科尔曼报告的研究结论引发了世界范围对学校作用的争论,并催生了增值评价^[8]。

20世纪70年代以来,以科尔曼报告为起点,增值评价研究在全世界范围内逐渐发展起来。增值评价是运用统计技术的评价,评价水平的提高依赖于统计技术的发展。由于受到统计技术发展水平的限制,20世纪80年代中期以前,增值评价没有得到很好的应用。到80年代末,多水平分析模型的完善,为增值评价提供了精确的统计分析方法。由于英美两国统计技术的发展以及强大的现实需要,多水平分析模型和增值评价方法率先在这两个国家得到广泛的应用。

1983年,《国家处于危机之中:教育改革势在必行》报告出台,美国对教育状况的关注水平空前提高,各地纷纷以“卓越、效率、平等”为目标进行教育改革。20世纪90年代,美国在部分州

推广学校增值评价。1992年，田纳西州最高法院要求学校的资助系统更加平等，州政府在增加教育经费的同时，要求加强教育效能考核。他们最终采用增值评价系统作为田纳西州教育改进法案中的一部分。随后，宾夕法尼亚等21个州的数百个学区也将增值评价用于学校评价。

2001年，美国政府实施的《不让一个孩子掉队法》强调，“政府在教育中的作用不是为体系服务，而是为孩子们服务”。“转变联邦政府在教育中的作用，不让一个孩子掉队。”“联邦政府能够、而且必须帮助缩小处境不利学生与同龄人之间的成绩差距。”^[9]明确规定在考评地区或学校的效能中使用年度增值（Yearly Value-added），要求各州每年对学生学业成就进行追踪监测，并将学生学业成就与政府拨款相挂钩。自此，增值评价逐渐成为美国教育评价的重要方式。

20世纪80年代末，英国政府每年都要公布所有学校以原始分数表示的学生成绩排名表。由于原始分数没有考虑影响学生成绩而学校自身又难以控制的社会因素，因而不能客观公正地反映学校的效能。对此，英国学者也提出了“增值”概念。英国的增值评价同美国类似，也是先从地区水平上发展起来的。英国政府于20世纪90年代接受了增值评价，2002年在全英格兰和威尔士推行学校效能的增值评价，2006年全面开展学校效能的增值评价^[10]。

增值评价通过对学校领导、教师、学生行为的分析，弄清学校工作的优点和缺点，明确离高增值学校的差距所在。这些信息可以引导学校采取有效的对策，使所拟定的改革方案更能符合本校的需求，有助于学校效能增值。实施增值评价后，学校间的竞争就不限于资源获得和占有量的竞争，更重要的是资源利用效益的竞争。增值成为学校共同面临的现实问题与必然的价值诉求，高增值的学校既要努力达成预定教育目标，还要以最有效率的方式达成。

参考文献

- [1] 刘尧. 中国教育质量文化之误区反省[J]. 上海教育评估研究, 2016, (4).
- [2] 刘海峰. 科举存废与文明冲突[J]. 厦门大学学报(哲学社会科学版), 2006, (4).
- [3] 刘尧. 现代教育评价的发展历史与观念嬗变[J]. 高校教育管理, 2005, (1).
- [4] 余文森. 论美国教育评价的历史发展[J]. 福建师范大学学报(哲学社会科学版), 1995, (4).
- [5] 张民选. 回应、协商与共同建构：“第四代评价理论”评述[J]. 外国教育资料, 1995, (3).
- [6] 陈玉琨, 杨晓江, 等. 高等教育质量保障体系概论[M]. 北京: 北京师范大学出版社, 2004.
- [7] 魏红. 重视学生学习效果改善教育评估效能[J]. 中国高教研究, 2009, (10).
- [8] 刘尧. 教育增值评价的教育公平追求[J]. 江苏教育, 2012, (12).
- [9] 雷振海. 不让一个孩子掉队[N]. 中国教师报, 2014-02-12(01).
- [10] 辛涛, 等. 教育增值评价的回顾与前瞻[J]. 中国教育学刊, 2009, (4).

(责任编辑 孙洁)