

# 协商构建省域基础教育质量监测新模式

## ——在监测中应用和发展“第四代评估”理论

何怀金, 龚春燕, 赖寒, 田海林

(重庆市教育评估院, 重庆 400020)

**摘要:** 国家完成了三年为周期的第一轮基础教育质量监测, 在省域层面, 各地也相继对地方基础教育质量监测进行了一轮工作实践。由于实施基础教育质量监测专业性强, 成本高, 因此如何最大限度地提升其服务作用, 充分调动基层县区及学校共同建构、全面参与, 成为进一步实施监测需要探索的重点。文章结合重庆8年来实施基础教育质量监测的经验, 着重分析了在第四代评估理论指导下, 开展基础教育质量监测更好地服务基层的方法、措施, 以期对构建省级基础教育质量监测体系提出建设性意见。

**关键词:** 基础教育质量监测; 第四代评估理论; 省域监测

**中图分类号:** G61, G40-058.1 **文献标识码:** A **文章编号:** 2095-3380(2017)06-0034-04

# A New Pattern of Provincial Elementary Education Quality Assessment Built by Negotiation

## — Application and Development of the Fourth Generation Evaluation Theory

He Huaijin, Gong Chunyan, Lai Han, Tian Hailin

(Chongqing Educational Evaluation Institute, Chongqing 400020)

**Abstract:** The first round of three-year cycle of compulsory education quality assessment has been completed in Mainland of China. At the provincial level, many regions have also carried out a series of practices on the quality assessment. However, since the implementation of compulsory education quality assessment is highly specialized and costly, how to maximize its service function and fully arouse the grassroots regions and schools to jointly construct and participate has become the focus of exploration for further implementation. Based on the experiences of education quality assessment in Chongqing in the past eight years, this paper explores the methods and measures based on the fourth generation evaluation theory to make education quality assessment serve the grassroots better. In addition, the current paper hopes to propose constructive suggestions.

**Keywords:** Compulsory education quality assessment; The fourth generation evaluation theory; Provincial assessment

2007年教育部建立国家基础教育质量监测中心以来,我国基础教育质量监测在政府主导下,成为教育督导机构三大职能之一,被提高到了前所未有的高度,不仅是教育决策的数据支撑,是检验相关政策措施的落实成效和经费效益的有效途径,同时还是促进教育行政部门职能转变、提高管理效能的重要抓手。

## 一、监测与“第四代评估”

“第四代评估”理论提倡协商、共建、共同参与。<sup>[1]</sup>监测是基于数据的量化评估,与第四代评估理论是一致的。监测不同于传统意义上的考试,不具备选拔、淘汰的功能,不对学生个体进行评价,目的是帮助学校或教育行政部门改进教育教学及教育管理。监测是基于数据的评估,强调的是数据准确、真实,方法科学,结论正确,其实质是帮助被评估对象诊断教育教学中存在的问题,<sup>[2]</sup>不是单纯地强调数据的权威性,而是与被评对象的实际相结合,在协商中取得共同的结论,结果由被评估对象根据实际,确定改进方案和措施,主动改进。

### 1. 监测与“第四代评估”两者均强调共同建构

“第四代评估”理论认为评估不是单纯的、中立的、纯客观的过程,认为评估不再是单纯地由评估专家对评估对象进行唯一、客观的测量和判断,而是参与评价的多方人员通过沟通、协商,共同建构统一观点的过程,进而构建共同的评价结果。<sup>[3]</sup>体现在监测上,着重于由各级监测实施单位、各级教育行政部门以及学校各类人员共同构建测量目标、测量工具、测量方式、结果报告等环节(如工具的构建,需经过选择不同的样本,经过多次预试、分析、修订,以保证工具的信效度及科学性),同样重视共同建构。

### 2. 两者均关注全员参与评估

“第四代评估”理论强调在委托人、评估者、评估对象等不同利益相关者之间建立平等关系,主张参与评估的所有人,不管是评估者还是评估对象,都是平等的、合作的伙伴关系,在评估中都有机会表达自己的观点,最大范围地调动全员参与的积极性。事实上,监测旨在尽可能多地采集到数据,参与人员众多,一般包括学生、教师、学校管理干部及家长,方式一般为问卷形式,为保证采集到真实的数据,实施中强调保护参与人员的个人隐私,提供宽松的作答环境,保证能在数据采集中充分表达自己的观点。

### 3. 两者都接受多元价值

“第四代评估”理论对不同评估对象,通过价值协调,缩小分歧,保证评估结果为各方均接受;同样监测允许不同地区、不同人员针对自己的实际,对数据结果以及导致该结果的原因有不同的理解,甚至可以得出完全不同的分析结论。

## 二、省域基础教育质量监测之“困”

基础教育质量监测是基于大规模采集学生发展状况及影响因素数据,进行分析统计而得出的报告。虽然以量化的形式客观报告教育质量状况,有助于教育行政与管理部门决策的科学化、民主化和专业化,然而仍然存在以下几个方面的困难。

### 1. 专业人员配备不足

“基础教育质量监测工作是一个多领域专家共同参与、具有高技术含量的工作,涉及心理与教育测量学、教育评价学、教育统计学、学科教育学、教育心理学、教育技术学、教育经济学以及教育管理与政策分析等多个领域”,目前这些领域的研究人员数量少,研究水平也有待提高;同时因实施投入的财力、物力巨大,专业要求高,也没有像高考中考有几十年的经验。教育部监测中心从2007-2014年,经过研究、试测,依托北师大,建立了一支近200人的高学历、高素质的专业人员队伍,在2015年第一次进行了大规模的正式测试;部分省、直辖市在国家基础教育质量监测中心的帮助下,也相应地建立了省域基础教育质量监测中心,尝试本省域的质量监测,但配备的专业人员不多,很少有超过10人的队伍,在区县层面,更因为专业人员不足,对监测的了解不足,不少地区存在打着监测的旗号,变相进行考试;因此,要形成国家、省、地市、县四级完善的基础教育监控网络,建立起覆盖全国的教育质量监测系统,还任重而道远。

### 2. 实践操作测评实施难以普及

基础教育质量监测不仅测学生的学业发展状况及影响因素,更要测学生的音乐、美育、体质及科学等各方面的动手能力及素养,测评的内容广泛,方法复杂,需不同领域的观测人员对学生的表现进行观察、访谈,与学生实时交流,要求所有的测评人员判断时采用一致的标准,这在大规模实施时,即使使用最精确的抽样方式,因样本学生的数量特别大,需要的专业测评人员数量极大,相应的能力和水平要求高,投入的财力、物力、人力巨大。因而在现阶段,单纯地由省级层面推动对学生的体质、音乐、美术、动手实验等

各方面的能力进行全面测评,困难重重。

### 3. 监测服务功能不足

监测是以诊断教育教学质量发展现状,为教育行政部门适时调整和制定更有针对性的管理措施提供科学依据,最终促进学生发展和教育质量的提升。实施监测,技术门槛高,需要的专业人员数量多,工作量大,采集的数据量大,需经过反复论证,方法专业,每一份报告,每一个数字,每一条结论,都要求准确。在已经实施的省域层面的监测中,一方面,实施的质量监测主要还在国家及省域层面,具有浓厚的行政色彩,基层学校被动参与,结果也主要为国家和省级层面的教育行政决策服务,不能得到区县和学校的自主支持;另一方面,基层学校的教育质量评价理念依然停留在传统甄别、证明、监督考核的层次,限于专业水平和能力不足,仍然以考试为主要评价方式,在指标制定、工具研发中区县和学校只有极少数精英参与,得到的反馈报告相对简单,结果应用重视不足,对区县、学校以及学生、教师等的服务作用非常有限。

### 4. 监测动态性不够,时效性不强

一方面,省域层面的监测需要组织大量的人力物力,操作复杂,一般一年实施一次,每次测试部分学科和指标;国家也是每年测一次,三年一轮回,每年测数学、体质、艺术、阅读、德育、科学等中的两个学科领域。当然,一个地区特定教育问题的存在与解决,可能会经历更长的时间,但是,监测从实施测试到报告发布,快则三个月,慢则一年,各地区、各学校、甚至各班级、各教师、各学生会发生很大的变化。缩短时间跨度,使得测评结果更具有时效性,能更及时地反映教育现状应该是教育质量监测与重点考虑的改进方向;另一方面,监测反映的是群体在一个阶段的发展状况,而对教育教学活动过程的诊断不足,对学生个体、教师个体在教育教学、发展成长中出现的不足不能及时地反馈校正。

## 三、协商共建省(市)、区县、学校协作的监测模式

推进中小学教育质量全面的教育质量监测是一项复杂的系统工程,需要各级部门和学校参与的自觉性和主动性,同时应具备相应的专业力量,建立部门间的协同机制,调动基层参与监测的积极性,改革考试评价,这与第四代评估倡导的“共同建构、全面参与、多元价值”的评价理念相一致。基于此,提出以下

几个方面改进思路。

### 1. 构建共享监测平台,各展所长,实现省域对操作性强的指标进行监测

完整的基础教育质量监测与评价指标体系,除了学业成就之外,还应包括学生综合素质、教师专业化发展水平、学校管理效能水平以及影响学生发展的各种环境因素(社会、学校、家庭、同伴)。而在监测中,受制于客观条件,只能够测量“可施测”和“可大规模施测”的指标,然而学生艺术、体质等方面的能力监测并不是简单的用纸笔就能够测试的。

省级监测机构的优势在于拥有高水平的专业技术人员,监测的专业性、科学性;不足之处在于难以像学校或基层大量教师及教育管理人员一样近距离、长时期地观察学生,对学生进行全方位的评价,对发现的问题和不足及时反馈到教育教学和有关人员;相反基层特别是学校的优势在于监测评价与教育教学全过程紧密结合,及时反馈,不足之处在于技术落后,主观性强,测评碎片化。

建议对于可以大规模测评的指标,由省级层面研发工具,统一组织实施,采集数据,并根据各级部门分层撰写报告,服务于学校和各级教育行政部门;而对于音乐、艺术、体质健康以及科学实验等的测评,基于协商共建的原则,确定测评结果用于改进教育教学和学生学习,不纳入对区县、学校的测评目标。建立对监测数据的真实性、科学性互信的机制,由省级层面组织专家研发统一的指导性的测评指标体系和工具,指导培训区县和学校测评人员,由各区县和学校根据实际具体组织测评,共享测评数据。在省级专业机构的指导下,对测评数据进行分析,根据实际分层撰写监测报告,将结果反馈给学校、学生、教师以及各级教育行政部门,形成大规模统一测评+统一工具及标准下的个性化测评相结合的监测格局。

### 2. 改进各级考试评价,共同建构多方参与的网状评价体系

限于操作复杂性,专业性及测试的目的,国家监测测试对象为四、八年级学生,各省市根据自身实际,监测也主要集中在学段中的一个年级学生,如重庆测试的是九月份的五、九年级,内容主要还是四、八年级完成时的学习内容。在区县层及学校层,一般会在每学期、甚至每半学期测试一次,学校、班级则会每单元甚至每周都有测评反馈,这些测试还是以分数为主的考试。在评价中引入省域专业机构力量,从评价指



标、评价工具改革入手,通过培训各区县测评机构人员,使之能严格按照研发流程和标准,研发测评工具,建立测评工具库,提供给一线学校、老师和测评机构,分为问卷调查、现场观察、个别访谈等多种测评工具。在实施的各级各类测评中,通过共享测评数据,可以达到以下几个目的:

(1)通过等值技术实现不同测评结果的比较;

(2)将专业技术普及到一线师生,改进现有的测评方式方法,提升测评的科学性;

(3)省域可以实时得到监测的最新测评数据,不局限于一年一次测评,同时能对基层测评出现的偏差进行及时调控;

(4)通过共享共建建立一支专业测评队伍。

### 3. 构建基于多元价值标准的监测报告机制

监测报告基于对测评数据及其影响因素数据的统计结果,专业性和技术性都很强。数据本身是中性的,无主观价值判断,但作为行政决策用的报告,需要有一定的价值判断,这要求对当地政策、实际条件、发展态势和教育教学情况有深入的了解。因而在共同认同数据的真实性及测评的科学性的前提下,可以有基于多元价值标准不同解释而撰写的行政决策用报告。以重庆为例,在监测报告撰写中,采用由省(市)监测中心提供全面的数据报告,各区县参照数据报告,结合本区县实际,撰写本区县报告的主要发现、各类专题报告、研究报告,学校在进行培训后,撰写本校的主要发现、专题报告等,所有区县和学校在尊重客观、真实的数据的前提下,可以有不同的思考、判断,如数据报告中存在异议或有进一步分析的需要,市监测中心进一步提供数据分析支持,所有的报告都可以有自身的价值取向,也有不同的侧重点和各自的改进方向、措施,形成一次监测可以有数千份不同报告的监测报告机制。

4. 建立大数据平台,形成基于综合素质评价的即时监测体系

学生的日常行为表现、学校活动、行为习惯、社会实践等综合素质评价记录的表现情况,不仅是学生成长评价的重要方面,同样也是监测需要的非常重要的过程性数据。因此,构建基于大数据的信息化平台,整合监测的主要测评技术,形成基于综合素质评价的即时监测体系,具体包括以下内容。

(1)构建包含学生学业、道德水平、身心健康水

平、学习态度、核心素养与能力、交流与合作及实践创新等指标,以及个体成长背景以及家庭、学校等影响因素指标的监测及综合素质评价指标体系。

(2)借鉴监测的经验与技术,通过论证、编制、预试、修订、专家复审等流程,构建内容全面、参数优良、内容新颖的学生自适应的学业评价工具库,基于心理测量,编制或引入国际通用量表等方式,构建测评学生自尊心、自信心、自我意识、气质、乐观性、羞怯(中学)、自卑等人格、个性特点的测评工具,以及如沟通与合作、抗挫抗压、适应能力、社会交往等社会适应性能力的测评工具体系。

(3)通过对学生成长行为的可记录性进行研究与探索,尤其在不增加教师、学生负担下实现对学生日常行为数据全面采集和进行研究,并就如何挖掘出学生日常行为数据背后隐藏的学生心理、学习习惯等因素进行深入研究,并构建涵盖学生思想品德、学业成长、身心健康、艺术素养、实践创新等方面可操作的、详实的日常行为表现评价工具体系。

针对学生成长海量信息的前期处理和筛选,应用数据挖掘技术构建恰当的模型,挖掘出指导教育教学行为改进的关键性数据以及隐藏的预测性信息,定期形成各级各类监测、评价报告;除及时反馈给学生、教师用于改进教育教学外,也通过挖掘分析形成一系列动态监测的专项报告,弥补现有监测缺乏过程监测、反馈不及时、服务群体较窄等问题。

基础教育质量监测在我国还处于起步阶段,相关的体制、机制建设在各地仍处于探索阶段,还没有形成成熟的测评理念、方式,通过“共同建构、全面参与、多元价值”,对现有的考试、评价进行改革,探索信息化下监测评估的体制、机制建设,构建新型监测体系,是监测未来发展的一个方向。

### 参考文献

- [1] 冯晖. 第四代教育评估理论及其应用效应——以上海高校高峰高原建设方案论证工作为例 [J]. 上海教育评估研究, 2015, 4(5): 29-34.
- [2] 龚春燕. 教育质量监测≠考试 [J]. 人民教育, 2016(2): 39-42.
- [3] 刘青. “第四代评估”从理论到实践的路径选择——中小学校教育评估模式重构 [J]. 教育测量与评价(理论版), 2015(10): 14-17.