

# 论我国义务教育质量监测系统的完善

商 敏

曲阜师范大学, 山东 曲阜  
Email: shangmin0708@163.com

收稿日期: 2020年12月11日; 录用日期: 2021年1月5日; 发布日期: 2021年1月13日

## 摘 要

相比较美国的NEAP而言, 我国义务教育质量监测起步较晚, 通过对两者间监测对象和范围、工具开发、结果应用等方面进行比较分析, 发现我国义务教育质量监测范围不够全面, 监测工具开发的专业性与技术性有待提高, 对于监测结果的应用亦不够充分。在现有监测基础上增加更多的监测学段, 最重要的是注重提高监测工具开发的专业性和技术性以及加强对监测结果的分析与应用。

## 关键词

学生学业评价, 义务教育质量监测, NAEP

# Discuss the Improvement of the Quality Monitoring System of Compulsory Education in Our Country

Min Shang

Qufu Normal University, Qufu Shandong  
Email: shangmin0708@163.com

Received: Dec. 11<sup>th</sup>, 2020; accepted: Jan. 5<sup>th</sup>, 2021; published: Jan. 13<sup>th</sup>, 2021

## Abstract

Compared with NEAP in the United States, my country's compulsory education quality monitoring started relatively late. Through a comparative analysis of the monitoring basis, monitoring objects, tool development, and application of results between the two, it is found that my country's compulsory education quality monitoring has a positive effect on curriculum standards. The distinction from evaluation standards is not obvious enough, the scope of education monitoring is not

comprehensive enough, the professionalism and technicality of monitoring tool development need to be improved, and the application of monitoring results is not sufficient. While paying attention to the curriculum standards, we have formulated qualified evaluation standards, and added more monitoring disciplines and monitoring sections on the basis of existing monitoring. The most important thing is to focus on improving the professionalism and technicality of monitoring tool development and strengthening the analysis and application of monitoring results.

## Keywords

Student Academic Evaluation, Compulsory Education Quality Monitoring, NAEP

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

学生学业评价是学生评价的重要组成部分。学业质量是教育质量的核心指标，是国际教育界长期关注的焦点[1]，各国教育学界和实务界为此作出了相应的努力。本文以我国的义务教育质量监测和美国的国家教育进展评价(National Assessment of Educational Progress, 下文简称为“NAEP”)为评析对象，利用比较研究法对二者进行深入分析，以期对我国的义务教育质量监测提出切实可行的优化建议。

## 2. 概述我国义务教育质量监测和美国 NEAP

### (一) 我国义务教育质量监测和美国 NEAP

我国义务教育质量监测是为促进我国义务教育质量的提升和素质教育顺利开展实施所进行的监测工作。2007年11月，我国教育部依托北京师范大学成立了基础教育质量监测中心，2007年~2014年，该中心开展了八轮义务教育质量试点监测[2]。2015年根据《国家中长期教育改革和发展规划纲要(2010~2020)》的要求，出台了《国家义务教育质量监测方案》，开始进行义务教育质量监测，2018年，发布了中国首份国家义务教育质量监测报告——《中国义务教育质量监测报告》<sup>1</sup>。

美国 NAEP 是唯一衡量美国学生在全国、州和一些城市地区的不同学科的知识 and 能力的评价，也被称为国家成绩单<sup>2</sup>。自 1969 年以来，NAEP 提供了关于学生学业表现的重要信息。它在美国基础教育领域有着独特的地位，对美国中小学教学改革、课程改革、学生评价改革、考试改革等发挥着不可替代的作用。美国开展全国性质的教育评价比我们国家早几十年，在 20 世纪 60 年代早期美国就对国家评价进行了大量探索，1964 年成立了教育评价进展探索性委员会(ECAPE)，并于 1969 年成功地举行了第一次国家评价。

由此，可以看出二者有一个共同的最终目的，就是提高本国的教育质量，促进学生的发展。另外，NAEP 和我国义务教育质量监测同为地域性的评价，即均为根据本国国情进行的教育评价。一个完善的教育评价项目的诞生，在发展过程中必然会遭到各种困难，既同为地域性的评价，美国 NAEP 比我们国家提前发展这么多年，可以为我国义务教育质量监测的后续发展提供一些启示。

综上所述，将美国 NAEP 和我国义务教育质量监测进行比较分析，对于我国义务教育质量监测的完

<sup>1</sup> 《中国义务教育质量监测报告》[EB/OL] <http://jyt.shaanxi.gov.cn/news/qitawenjian/201807/29/14351.html>, 2020-05-27.

<sup>2</sup> 美国国家教育进展评价所有资料来源于官方网站[EB/OL]. <https://nces.ed.gov/nationsreportcard/>, 2020-05-27.

善和发展是有价值的。通过完善和发展我国的义务教育质量监测,对提高我国教育质量,科学测评学生学业也有重大意义。

## (二) 完善义务教育质量监测对我国学生学业评价的意义

### 1. 促进了我国义务教育评价体系的健康发展

我国关于学生学业评价的方式一直是考试占主导地位,诸如中考、高考等。但考试,仅仅是学生学业评价的其中一种方式,它有着强大的选拔功能,但这种评价方式只关注到了结果,而忽略了评价过程,并不能完全地展现学生经过学习以后取得的成效,自然也就很难对学生的学业情况进行综合判断。我国义务教育质量监测体系能够测查学生掌握知识、技能的程度和分析解决问题的能力等等,是一项比较全面和综合的测评方式。

### 2. 推动了评价方式变革

在以前的评价方式中学生大多是以客体的身份参与评价,当代课程评价试图构建评价者与被评价者之间的关系,不再将评价者和被评价者置于主、客体相互对立的框架中,而是把评价者与被评价者之间看作交互作用的主体[3]。反映在学生学业评价上,我们应该弄清楚我们的评价是“为学生的评价”,在这其中学生不仅是被评价的客体,还可以是评价的主体,我们要通过评价发现每一位学生的独特之处,让学生能够通过有效途径发现并发挥自己的优势,使学生成为学习的主人,从而成为更好的自己[4]。我国义务教育质量监测客观反映了义务教育阶段学生学业质量、身心健康及变化情况,纠正以升学率作为评价学校和学生唯一标准的做法,对于改变传统的评价方式起了重要的推动作用。

### 3. 拓宽了对影响学生学业发展因素的认识

在进行学生学业评价的过程中,学生的学业发展情况不能只靠最终的考试分数来反映,学生的学业质量是受多方面因素影响的。我国义务教育质量监测在监测的过程中,除关注学生学业质量的变化情况外,还会关注影响学生发展的相关因素。比如学习情感态度,从我国监测报告中来看,我国中小学生对学科学习兴趣越高、自身学习自信心越强、学习焦虑程度越低的情况下,学业成绩会越高[5]。义务教育质量监测通过监测学生学业质量以及影响学生发展的相关因素,以此转变教育管理方式以及改变学校教育教学方法,从而引导社会树立正确的教育质量观。

## 3. 比较法视角下的我国义务教育质量监测与美国 NEAP

构建合理的基础教育质量监测评估项目是保障教育质量的重要支撑。因此,诸多知名国际组织已做出榜样垂范,如经合组织(OECD)的 PISA 项目、国际教育成就评价协会(IEA)的 TIMSS 和 PIRLS 项目等[6]。不仅仅是国际组织对基础教育质量监测做出了突出贡献,根据各国的不同国情,以一个国家为单位的教育质量测评也都进行的如火如荼。接下来对我国的义务教育质量监测和美国的 NAEP 进行比较分析,发现了我国义务教育质量监测的待完善之处。

### (一) 监测对象的范围有待扩大

我国监测对象确定为义务教育阶段四、八年级学生,这是根据义务教育各学科课程标准的学段划分情况以及学生认知和学习能力发展的阶段性特征划分的。美国 NAEP 的评价对象是在在所选定的学校 and 要评估的年级中,从选定的年级(4、8 或 12 年级)随机挑选学生参加 NAEP 最多两门科目的测试。每个学生都有被选中的机会——不分种族/族裔、社会经济地位、残疾、英语学习者的地位或任何其他因素。相较于我国,美国的评价范围更广一些。

我国国家义务教育质量监测重点监测学生学业质量,以及课程开发、条件保障、教师配备、学科教学和学校管理等相关因素,这更好的促进了学生的全面发展。美国 NEAP 的评价内容为学生知道什么,怎么做;考察教师如何面对课程评价,如何理解课程评价,如何在课程评价中发挥自己的作用,如何

发展自己的职业能力；被评价的课程需要做出哪些改进；学校为教师的职业发展、学科能力发展等提供了哪些帮助等。其他内容：国家教育统计中心(NCES)包括调查问卷作为NAEP的一部分，调查问卷分为三类：1. 学生填写完整的问卷，提供他们在课堂内外学习的机会、教育经历和其他各种话题的信息，包括社会经济地位和技术使用情况。学生可以跳过任何问题留下空白的回应。2. 负责管理评估的教师完成问卷，收集教师培训和教学实践的信息。3. 学校问卷通常由校长或副校长填写，收集有关学校政策和特点的信息。在监测或评价内容方面，美国与我国相比更加具体的关注学生自身的学业发展情况。

## (二) 监测工具开发的专业性和技术性有待提高

我国监测指标与工具具有良好的信效度，其研发以各学科课程标准为依据，各项量化指标符合测量学要求，监测工具包括学生学业测试卷，学生、教师、校长问卷，体育现场测试仪器设备，音乐演唱测试系统等。在监测工具的设计中，采用矩阵取样设计的方式，这有效防止了因为测试内容太广泛和学生作答时间太有限而作弊的行为；为解决不同年度学生学业表现的可比性问题，采用了测验等值技术，即“量尺分数”<sup>3</sup>。

美国会委托著名的考试评价机构负责评价工具的开发，自1983年以来，美国的评价工具，包括评价项目、文本材料等均由美国教育考试公司(ETS)负责，它是全球著名教育评价机构，先后开发了GRE、SAT等著名考试。考试评价机构在实施评价之前，会组织学科专家、测量学家等专业人士，针对各门学科，开发评价题库，试题经完善以后，还要再选拔大约几千名学生进行小规模的地测试，对其结果进行评分和分析，经过再次完善和修改以后最终确定正式评价中所用的题库，从题库中抽取评价题目来进行评价。同时美国在评价工具开发是非常注重公平性的，会考虑到有多元语言背景以及残疾等考生因素，评价的整个过程也同样注重公平性，评价分数会转化成标准分数进行比较。自2001年以来，NAEP一直在探索新的测试方法和问题类型，以反映技术在教育中的日益广泛的应用。NAEP在评估中引入了各种新的问题和任务类型，以捕捉学生所知道和能够以更真实或直接的方式完成的任务，包括技术和工程知识评估(TEL)中基于情景的任务、交互式计算机任务和混合操作任务。在2016年，NAEP数学和阅读评估在平板电脑上进行试点，平板电脑配有键盘、手写笔和耳塞，并采用了新的测试方法和题型。一些问题包括多媒体，如音频和视频。其他问题允许使用数字工具(比如屏幕上的计算器)来形成一个答案，或者让学生在现实场景中解决问题。美国于2017年过渡到基于数字的评价。NAEP数字评价使用动态和创新技术，为学生提供引人入胜的评价体验，并为教育工作者提供有关学生技能和知识的更有意义的的数据。通过基于数字的评价，学生接收、收集和报告信息。这些新的评价包括通用设计原则，使更多的学生参与。NAEP评价从纸笔到触摸屏平板电脑的转变融合了前沿的学习技术，在每次评价开始时，学生们都会接受一个简短的、互动式的指导，旨在教他们有关评价系统和工具的知识。

在监测或评价工具开发的过程中，我国截至目前，仍然没有形成普遍一致的，专门针对教育质量监测工具的评价指标和标准[7]，在公平性方面提及到了特殊群体，包括西部和少数民族地区，在进行全国范围的教育质量监测时，需要确保所用的监测工具是否考虑到了不同民族的特点。要注意的是，特殊群体还应包括残障人士，他们的测评应该如何进行，也是保证教育质量监测工具的公平性时应该考虑的问题。另外除了体育和音乐的监测需要特定的设备，大多数问卷和测试卷还仅仅是传统的测验工具——纸笔测验，少数通过网络，而美国为了以更有效、更吸引人的方式评价学生，反映他们的课堂经验，越来越多的学校正在使数字技术成为学习环境的一个组成部分，NAEP已经从传统的纸笔形式过渡到数字形式。

<sup>3</sup>量尺分数(scale score): 根据学生作答情况，采用项目反应理论(IRT)模型估计得到的学生能力分数转换而来的测验标准分数即为量尺分数。大型监测一般使用多题本测试，经过等值转换后的量尺分数使得不同试卷或不同时间测试学生的分数具有可比性。目前，教育部基础教育质量监测中心对我国各省市进行基础教育质量监测时即采用量尺分数来评价学生的学习能力。

### (三) 监测结果的分析与应用有待加强

我国义务教育质量监测主要以报告的形式呈现结果,分别为基础数据报告、分省监测报告和国家监测报告<sup>4</sup>,三类报告的目的、内容和阅读对象都是不同。

美国 NAEP 的结果报告全国,在大多数情况下,报告各州以及参加试验性城市地区评价(TUDA)的选定城市地区。成绩以分数和学生达到 NAEP 成绩水平(基础、熟练和高级)的百分比报告。另外,关于学生个人的成绩报告,学生在 NAEP 上的回答是私人的,每个参与的学校和学生的隐私是至关重要的,因此这些结果是不对外公布的。NAEP 监测全国、州、地区以及包括残障学生和英语学习者在内的不同学生群体的整体教育进展。NAEP 不是为学生个人或学校提供结果而设计的。

两国结果均以报告的形式呈现给大众,我国监测报告仅对外公布国家监测报告。我国首份义务教育质量监测报告结果列了十项,涉及学生学业质量、条件保障、教师配备、学校管理等诸多方面,对于结果呈现,报告对监测内容和关键性指标进行了分析诊断,分项给出了监测结论,并相对应的提出了提高义务教育质量的七条建议。美国 NAEP 的报告除国家和各州的结果外,还有关于学生以及每所学校的报告,当然这都是隐私,学生通过一定方式获得一个评价结果,比如电子邮件等,可以从中清楚地知道自己该门课程的学业成就倾向(即学生的学业发展潜力和未来发展趋势),自己在整个学生群体中所处的百分位,以及自己的优势及缺陷。

## 4. 优化我国义务教育质量监测的对策建议

### (一) 构建覆盖基础教育全学段的教育监测体系

我国的监测对象目前仅仅局限于义务教育阶段,对于学前教育以及高中教育并没有涉及。学前教育不仅是现代国民教育体系的不可或缺的组成部分,而且对于人的终身发展和全面发展具有奠基性的作用;而高中教育是基础教育与高等教育的衔接口,在整个教育体系中起着承上启下的作用。经过多年努力,我国义务教育质量监测的发展已积累了丰富的经验,基于学前教育和高中教育的重要性,我国需要进一步加强学前教育质量监测和高中教育质量监测的相关研究,组织开展试点,尽快建立并完善相关工作制度和机制。

### (二) 提高监测工具开发的专业性和技术性

监测工具对于一个监测来说,其重要性不言而喻,监测所涉及的所有环节,都与监测工具有关,监测设计、监测开发、监测评分这几个环节都以监测工具为主要对象[7]。当前我国在监测工具方面做出了很大的努力,也取得了一定的成果,比如在问卷电子填答系统内设立了逻辑判断和合法值范围,减少了不合理数据,大大提升了数据质量。但较美国还是有些落后,我国应着重增强监测工具开发的专业性,可以要求工具研发部门对将要投入使用的工具出具质量报告,确保工具的效度和信度。同时除专家对试题进行多轮次审核外,继续保持进行预测试,确保所有题目具有良好的测量参数。关于监测工具的公平性问题,开发团队中应当包括不同背景的专家,除此之外,还可以增加来自全国各地教育教学一线的优秀教师们;对残障学生的监测需要考虑到他们的特殊情况,需要与专业人士进行沟通,做好背景调查,专门设计试题进行评价,如此可以增加工具开发的公平性和适用性,在测评过程中,必要时可以配备手语老师等等,帮助残障学生更好的理解和使用评价工具,从而得到更确切的评价结果。另外,我国应加快技术发展的步伐,目前利用网络测评的对象仅仅是教师和校长,学生依旧使用传统的纸笔测试。传统测评方式中大量的分拣和数据扫描工作,大大增加了相关工作人员的负担,本着节约纸张、提高工作效率

<sup>4</sup> 我国监测报告分类: 1. 基础数据报告。主要呈现以县为单位的原始数据汇总,供监测评价机构内部分析使用,不对外公开发布。2. 分省监测报告。分省(区、市)呈现学生在学科领域的表现水平,影响该省(区、市)学生学业水平的主要因素,以及相关分析。该报告供各地政府和教育部门参考,不对外公开发布。3. 国家监测报告。主要呈现全国学生学业水平总体状况,影响学生学业水平的主要因素,以及相关分析。该报告向社会公开发布。

和质量的原则,我国要紧跟时代潮流,推进计算机技术的普遍运用。同时还要增加评价的问题和任务类型,比如增加基于情景的任务,通过模拟现实场景观察学生所表现出来的解决问题的能力来评价学生的知识运用能力和动手实践能力等。

### (三) 加强对监测结果的分析与应用

在结果的呈现上,我国仅对于国家以及省(区、市)义务教育的总体水平给出了结果分析和建议,这样一来,监测对于每所学校和学生本身而言影响较小,容易忽视了学校与学生的主体差异性,当然这也与我国的监测目的相关,我国的监测目的中并没有关注到学校与学生本身,监测是为了科学评价我国义务教育的总体质量,但相比较美国 NAEP 中所表述的“评价学生知道什么,怎么去做”来说,给学生个人发布评价卡片,让学生知道自己的学业情况,开设学校和学生个人监测结果的获取渠道,在一定程度上更有利于学校和学生个人进行发展定位,从而促进学校和学生自身的发展,提高一个国家的教育质量。要指导各教育部门和机构正确运用监测结果,改进教育教学,发挥以评促建的作用。比如报告中涉及到的条件保障、教师配备等方面。通过监测,发现教学资源使用方面和教师队伍建设方面的问题,针对问题提出合理的整改建议和措施,合理配置教学资源,提高使用率;加强对教师,特别是中西部教师的培训,提升其职业素养和水平。

## 5. 结语

第一轮义务教育质量监测的完成于我国教育质量监测实践而言具有里程碑的意义。但这只是开始,我们不能停下前进的脚步,在借鉴国际经验的同时,要注意结合中国实际,建立一套具有中国特色的义务教育质量监测体系,特别是在监测工具的开发和监测结果的应用两个方面。监测工具开发要考虑到中国各民族、各地区、各群体的不同特点,监测工具的公平性乃至整个教育的公平性,一直以来是我们所关注和追求的,要通过有效的监测,更好地发现受教育者的不同,从而给受教育者提供有差异和不同特点的教育。监测作用的最终体现要落实到对监测结果的分析与应用上,通过监测,我们能知道教育现状是怎样的,但更重要的是要分析影响我国教育质量现状的关键因素,由此,提出相应的切实可行的完善建议,将结果运用到教育教学改革与实践当中去,从而推进教育改革,提高我国义务教育质量。

## 参考文献

- [1] 张咏梅. 大规模学业成就方法调查的开发: 理论、方法与应用[M]. 北京: 北京师范大学出版社, 2015: 349.
- [2] 赵茜, 辛涛, 刘雨甲. 我国基础教育质量监测与评价的现状与趋势——第二届“中国基础教育质量监测与评价”学术年会综述[J]. 教育研究, 2017(9): 154.
- [3] 罗贵明. 美国中小学课程评价: 理论、实践与借鉴[M]. 武汉: 武汉大学出版社, 2016: 230.
- [4] 辛继湘. 从“对学生的评价”走向“为学生的评价”[J]. 教育测量与评价, 2013(7): 1.
- [5] 2018 年国家义务教育质量监测 数学学习质量监测结果报告[EB/OL]. [http://www.eachina.org.cn/upload\\_dir/editor/20191120171458402.pdf](http://www.eachina.org.cn/upload_dir/editor/20191120171458402.pdf), 2020-05-27.
- [6] 刘建伟. 国际组织开展基础教育质量监测评估项目的经验及其启示[J]. 教学与管理, 2019(33): 114.
- [7] 刘玥, 游森. 教育质量监测工具的公平性研究[J]. 中国教育学刊, 2019(8): 25.