



Thomas, S. M., Peng, W-J., Tian, H., Li, J., Ren, C., & Ma, X. (2012). 学校效能增值评量研究. *Educational Research (China)*, 33(7), 29-35. <http://english.nies.net.cn/Publications/>

Publisher's PDF, also known as Version of record

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

学校效能增值评量研究*

萨丽·托马斯 彭文蓉 田慧生 李建忠 任春荣 马晓强

[摘要] 改进学校和学生评价的方法是提高教育质量的关键杠杆。研究发现,增值评量普遍被认为可在评价中国教育质量上提供一种有效且相关的度量,这一结论与其他国家,譬如英国的有关研究结果相类似。在中国现有的学校和教师评价体系中,增值评量的概念与其测量的方法被视为是一种较为科学且受欢迎的评价方法。在运用任何新的评价体系时,需要考虑到地方情境以及优先考虑事项。增值评量法是否可在公共问责体制的框架下,为学校效能评价体系提供一种最有用或最适当的方法,向学校提供保密性反馈信息,提高学校自我评价与改进的能力,值得讨论。

[关键词] 增值评量法;教育评价;多水平模型;教育质量;学校效能

[作者简介] 萨丽·托马斯,英国布里斯托大学教育研究院教授;彭文蓉,英国布里斯托大学教育研究院博士;田慧生,教育部基础教育课程教材发展中心主任、研究员、博士生导师(北京 100029);李建忠,中国教育科学研究院副研究员;任春荣,中国教育科学研究院副研究员;马晓强,中国教育科学研究院副研究员(北京 100088)

一、导论

许多发展中国家不但注重发挥其教育体系的功能,而且将改进学校和学生评价视为提高教育质量的关键杠杆。这主要是因为教育质量与效能的测量在提高教育水平上可起到两方面的作用。第一,为质量改进提供决策依据,使教育系统、学校和教师为更好地履行其职责和提升其表现水平而负责,从而促进教育民主。第二,为实施问责制提供决策依据,从而促进改善教育结果和提高学习质量。^[1]通常在分权制情境中,这种问责制和持续关注质量改进的双重策略,以及注重改进质量评价与学生评估的过程,是许多西方国家普遍

采用的做法,以此来达到提高教育质量的目的。有效、可靠和公平的学生考试与评价制度在学校认证中发挥着关键作用,这可有效地管理教育结果与教育供给质量的预期水平,并可为问责制的实施和评价体系的改进提供相关的关键数据。^[2]

发展中国家,不论其有关问责、改进或认证的新教育政策的焦点为何,都需要有评价高质量教育的创新方法,从而为教师和政策制定者提供一种可选择的质量评价框架,以此判断不同情境下教与学的最佳实践。此外,同20世纪90年代的英国国情一样,需要有另一可供选择的方案,来降低以原始考试分数评价学校表现的常见做法^[3]。如果仅仅以原始考试分数作为评价学校的唯一标准,那么可能会使有生源劣势的学校被不公平地

* 本文系英国经济与社会科学研究委员会(ESRC)和国际发展部(UKAID)联合资助的研究课题“改进中国教育评价和教育质量”(课题批准号:167-25-0353)和“改进中国教师发展和教育质量”(课题批准号:167-25-0428)的部分成果。

评价,而生源较好的学校可能会骄傲自满,因而很难判断学校是否具有最佳实践。

英国和其他一些国家已借鉴相关学校效能研究的范例和方法,[4]努力发展增值评量,以便提供一种比原始考试分数更加公平且更准确的评价学校表现的方法。对在提升学生学习成就方面具有潜在作用,以增值评量概念为基础作为学校表现指标的学校效能与教育效能研究受到英国教育决策者的高度关注。事实上,这些评价和问责的新方法已与改进学生学习结果相关联。[5]

在本文中,我们将运用来自中国大陆的证据,探讨新教育评价体系在强化教育质量与效能中的潜在作用。然而,若要使这些方法能够在不同情境下有用地评价质量,关键是了解方法本身的局限性(量化的视角)与情境特异性的问题,尤其是发展中国家,校际和区域间发展差距很大,教育投入不足,其教育情境完全不同于发达国家。发展中国家面临教育经费短缺、卫生条件差以及因贫穷而对学习潜能的影响。应该承认发展中国家有其优先事项的考量,诸如必须提供足够的师资培训、校舍与教材,解决教育资源分配不公,改进领导力,提高校长与教师专业知识水平,等等。[6]

因此,不同国情尤其在发展中国家,在学校效能的范围与程度等信息提供方面,需要更有力的量化证据。显然,增值评量法在发展中国家的运用和结果,有可能会凸显情境特异性的某些重要层面,如特定学生群体的低成就或资源分配不公。尽管存在着关键情境差异,但学校效能研究方法仍然很重要,它允许在国家或区域内进行常态性比较,通过改进评价过程、支持教师发展与自我评价以及总结最佳实践,这些都可帮助促进资源分配的公平性和优质教育的供给。这类证据至关重要,它们不但可以为不断变化着的教育改革以及在提高学生特别是农村和经济落后地区学生学习成就水平所采取的措施方面提供信息,且有助于纵向量化研究方法的创新,强化国际比较的有效性[7],丰富国际学校效能研究的知识库[8]。

二、中国的教育政策和研究情境

近15年来,中国的教育改革把重点放在了教育机会扩大和提高教育质量上,特别是放在了农

村地区和女孩的入学机会和教育质量上。

教育公平不仅与机会平等相关,也与公平地评价学生和与学校有关。因此,改进评价学生以及学校质量与进步的方法,是中国教育改革的核心内容之一。针对这些近15年来一直凸显的问题,政府在《教育规划纲要》中已明确强调需改进评价机制。此外,一些特别受关注的问题还有:以应试为导向的教育体系、课程与考试改革、系统化的教育质量评价,不同地区的高考入学率等。[9]

尽管政府希望有新的方法来评价教育质量,然而,学校效能研究在中国大陆则刚刚起步。学生的考试分数和升学比例仍然被视为学校办学质量的主要指标。但是,已有中国学者对西方学校效能进行了相关文献研究,也有些中国学者在探讨学校与教育效能在中国情境下的内涵、定义以及研究方法论。[10]此外,在目前为数不多的中国相关学校效能实证研究中,也探讨了在一个时间点上班级和学校对学生考试分数的影响。然而,这些研究所分析的是横向而非纵向数据,因此,有其局限性。[11]由于这些研究多数是小规模研究,加上其对学生能力的操作性定义不能令人满意,已有一些中国学者呼吁需要更多的实证研究,并且有必要将学生先前学习能力(prior attainment)基线(譬如,以中考成绩作为基线)评量值引入到分析模型中,而且必须建立各省市甚至是国家层面的学生数据库体系,以此来建立一个更加公平的学校效能评价体系。[12]中国香港和台湾已经有几个大型并且有现实意义的学校效能研究。

最近,上海参加了OECD的PISA项目[13],但整个中国大陆仍未参加各项国际效能学校的比较研究项目,中国大陆目前尚未建立起情境化(contextualized)的学校评价标准或者指标体系,同时,能够熟练地运用适当的教育评价方法论工具的技术专家也很短缺。上述问题加上缺乏学校效能实证研究证据作为中国政府制定相关政策的参考[14],成为本文作者进行有关严谨地评测中国学校效能范围与程度的研究领域,并提供了有用的模式来支持中国大陆发展新的教育评价体系。

三、改进中国教育评价和教育质量

英国经济与社会科学研究委员会(ESRC)和

国际发展部(UKAID)联合资助了布里斯托大学两项相关课题,研究有关中国教育评价的问题^[15]。这两项课题由英国布里斯托大学与中国教育科学研究院合作研究,本文着重呈现第一项课题的研究发现。

“改进中国教育评价和教育质量”这项课题的目的有两方面。第一,运用“增值”概念评价2009届学生的学习结果,对现有跨区域高中学校效能的定义与测量进行拓展。第二,探讨政策制定者和教育工作者在考虑当地背景和优先事项的需求之下,如何接受、适应并推动增值评量法在中国的运用。以评价学生学习 and 结果为例,是侧重评量公民素养养成还是学术成就方面,评价方法是偏重定性的还是定量的,哪种情境特征对学生和学校的表现影响更大,都是需要研究和探讨的。

广义而言,这项课题的目的在于增进对中国学校效能复杂本质的理解,更重要的是研究地方情境因素在界定教育效能和质量方面起着什么样的关键作用。如果来自不同社会经济、政治背景的政策制定者和教育工作者,能对基本上由英国和欧美国家所提出来的教育质量概念和评价方法有一个比较清晰的认识,则可大大提高学校效能和学校改进方面的研究。^[16]特别是对地方情境和文化方面问题的了解,可对学校和学生的教育价值观以及其考虑的优先事项产生重要影响,而这类比较研究方面的证据则是必不可少的。

在研究方法上,“改进中国教育评价和教育质量”课题包括两个互补性的子研究。

第一个子研究是采集新的质性资料,来探讨有关中国如何定义教育质量和其评量方式,以及中国可能如何接受、适应并运用有关学校效能、学校评价和学校自我评价的国际研究成果。本研究中一共有90多位校长、教师、学生、地方和国家层面政策制定者参与了个别或团体的访谈。

第二个子研究是采集量性数据,运用创新性的定量方法(多水平模型)分析学生学业成就和入学时能力、学生与学校背景的关系,来探讨中国的学校效能。

本研究中一共有120多所来自东西部三个地级市的普通高中学校以及9万多名2008/2009学年的高三学生提供相关信息。

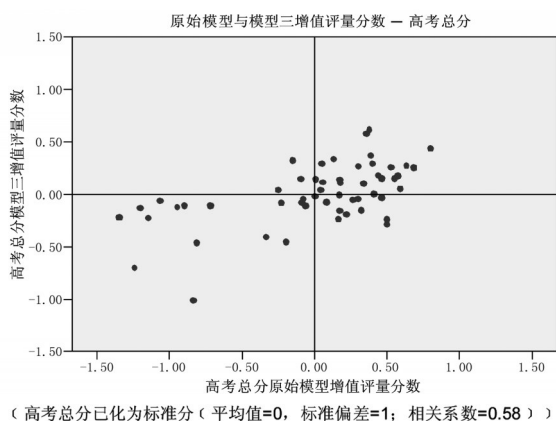
四、主要研究结果

“改进中国教育评价和教育质量”项目的研究结果显示,虽然受访的利益相关者认可国际上对教育质量的定义,^[17]但却非常关注并且强调在中国语境下学生的全面发展(譬如,素质教育)和教育公平成为质量追求的主要目标。^[18]此外,一些受访者也认为,中国现行的教育评价方法并不科学,因此,改进评价的过程是提升相关政策与实践的必要条件。关于中国学校效能估测情况,表1总结了逐一引入相关变量所建立的各种二阶层多水平模型,以探讨2008/2009学年度学校表现的差异性。

表1 6种用于检测学校表现差异性的多层模型

解释变量	模型	校内相关系数 (高考总分)		
		区域1	区域2	区域3
未引入解释变量	初始模型	24.0%	27.2%	25.7%
引入中考成绩变量	增值模型一	8.5%	24.3%	19.8%
再引入学生背景变量	增值模型二	8.6%	25.1%	24.1%
再引入学校背景变量	增值模型三	5.1%	13.2%	14.7%
再引入学校过程和投入变量	增值模型四	1.0%	7.6%	0.0%
再引入学生努力和态度变量	增值模型五	1.0%	8.1%	0.0%

研究结果显示,三个地级市在不考虑其他因素对学生高考成绩影响的前提下,2009年学生高考总分总体变异的24%~27%可归因于学校间的差异性(表1中初始模型)。然而,一旦在模型中引入各种变量,学校在原始高考成绩和增值评量的表现,有明显的改变(见下图)。把中考成绩、学生和学校背景等学校无法控制的变量引入模型中(表1中增值模型三),这些变量解释了2009年学生高考总分总体变异的43%~57%以及学校间变异的70%~91%。其余未被解释的2009年学生高考总分总体变异尚有5%~15%,这可归因于学校间差异性(即校内相关系数)。这些数据说明,尤其是在中国西部地区,通过学校作用于学生,对其学习成就有着影响。这与以往一些在中国从事的实证研究所获得的结果相类似。^[19]然而,似乎至少有两个地级市的学校作用大于英国一些类似



高考总分—原始模型与模型三的增值评量分数比较图

研究所得出的学校作用^[20]。有意思的是,将各种学校投入和学校过程变量再引入到模型中,发现只有少数的变量对学生高考总分具有统计上的显著影响。例如,师生比、校长是否接受过专业培训以及校长是否观课。这些变量的引入又解释了7%~30%的学校间的差异性,也就是说,因学校投入和过程变量的引入,几乎完全解释了在增值模型三中尚未被解释的学校间的差异性(表1中增值模型四)。本文附录列举了纳入增值模型三的所有解释变量以及分析后的详细结果。

研究结果也显示,参与课题的三个地级市学校对其不同学生群体,尤其是以中考成绩来区分的学生群体,具有区别性效能(见表2)。

表2 2009年不同学生群体间(依中考总成绩分类)以及不同学科间学校效能相关系数

	区域1	区域2	区域3
男生与女生	0.97	0.96	0.89
学生组别(依中考总成绩分类)			
中等能力与低能力学生	0.89	0.83	0.64
中等能力与高能力学生	0.80	0.72	0.96
低能力与高能力学生	0.52	0.47	0.45
理科与文科	0.77	0.76	0.82
城市户口与农村户口	0.84	0.94	0.90
学科			
语文与英语	0.72	0.93	0.88
语文与数学	0.70	0.95	0.81
英语与数学	0.73	0.93	0.91

这些研究发现意味着,如果仅以单一的学校整体性评量来评价学校的增值表现,特别是在绝大部分的高中学校采用混合能力编班教学的情况

下,可能会掩盖学校的区别性效能。譬如,提供学校对其不同入学能力学生群体的区别性增值评量信息,可以帮助学校和教师识别出学习能力较差的学生是否能够掌握对其来说较为困难的学习内容。从另一方面说,学习内容对学习能力强较弱的学生来说是否不具足够的挑战性。同样的,提供学校在不同学科的区别性增值评量信息,可以协助学校找出其整体发展战略或在教师聘用方面是否有不足的地方。

目前,“改进中国教师发展和教育质量”后续项目,正在先前参与第一项课题的三个地级市采集相应的2010年至2012年的数据及相关信息,后续分析可提供有关学校效能增值评量的发展趋势。这可考察另一个关键维度,即随着时间的推移学校表现的变化,这也是用来衡量一系列新的教育改革作用的关键所在。

五、结语

总之,只要有相应且全面性的地方或全国性数据库,大规模的学校效能实证研究在中国是可行的,这两项课题为此亦提供了一个有用的例证。另一个重要研究发现是,增值模型的分析结果因中国区域而异,这显示了情境因素可能会以不同的方式运作,特别是穷困与富裕地区的对照。因此,可能的情况是,独立的区域评价体系比单一全国性评价体系更能反映当地的具体情况。此外,鉴于原始考试分数与大学升学率受到社会各界高度关注,在引入新的问责方法和学校效能评量与改革举措时,不可避免地会遇到来自不同利益群体的阻碍。因此,必须考量地方情境和兼顾到公平与公正原则,依靠国家和地方政府的政治意志力,推动各方凝聚共识,达成共享方法推动教育评价改革。

至于增值评量法的引入能否在公共问责体制框架下成为一个最有用或最适当的学校效能评价方法,或者作为保密性反馈信息提供给学校以强化学校自我评价与改进,或者兼具两者功能,都是非常值得讨论的。另外,也可从其他国家实施的策略和有益经验中获得启示。譬如,增加教育问责评量常受到教育工作者的严厉批评,且也有一些研究证据显示,其对学生学习的品质具有负面

作用。^[21]然后,在国际情境和文化议题的考量下,中国也必须找到符合自己独特需求的最佳方案。

有关“改进中国教育评价和教育质量”课题的研究发现,在其他地方有更进一步的报导。但这

里要强调的是,参与这两个课题的仅仅是三个地级市,在幅员广大的中国,需要更大规模和更具代表性的教育效能研究,来探讨包括各个教育学段对教育质量和效能的贡献。^[22]

附录

区域2的增值模型三分析结果:2009年高考总成绩(Z分数)

	估计值(estimate)	标准误(Standard Error)	
固定部分			
截距项	1.06	0.32	*
学生先前学习能力基线(中考成绩)			
语文中考成绩(Z分数)	0.04	0.01	*
英语中考成绩(Z分数)	0.33	0.01	*
数学中考成绩(Z分数)	0.18	0.01	*
女生(参照组:男生)	0.02	0.01	
年龄(Z分数:月)	-0.09	0.00	*
文科(参照组:理科)	-0.09	0.01	*
户口/家庭所在地(参照组:城市)			
县镇	0.10	0.02	*
农村	0.13	0.02	*
缴纳学费(参照组:正常缴费)			
高于正常缴费	-0.37	0.02	*
全额或部分减免	0.17	0.04	*
与父母同住(参照组:在校寄宿)	-0.10	0.02	*
家到学校所需时间(参照组:少于15分钟)			
15分钟至半小时	-0.06	0.02	*
0.5-1小时	-0.04	0.02	*
1-2小时	-0.03	0.02	
初中学校的级别(参照组:普通学校)			
县级示范学校	0.03	0.01	*
市级示范学校	0.01	0.01	
省级或国家级示范学校	0.09	0.02	*
在这所学校的就读经历(参照组:高一至现在)			
高二至现在	-0.18	0.05	*
高三至现在	-0.10	0.03	*
父亲/男性监护人的最高学历(参照组:小学及以下)			
本科	0.07	0.03	*
父亲/男性监护人的职业(参照组:国营、私营单位负责人)			
农林牧渔生产人员	0.10	0.03	*
生产、运输、设备操作人员及有关人员	0.09	0.03	*
农民工	0.09	0.03	*
教师	0.11	0.03	*
无工作	0.10	0.04	*
母亲/女性监护人的最高学历(参照组:小学及以下)			
初中	-0.07	0.01	*
高中、中师、中专、职业高中	-0.07	0.02	*
专科	-0.01	0.03	
硕士及以上	0.19	0.07	*

母亲/女性监护人的职业(参照组:国营、私营单位负责人)			
农林牧渔生产人员	0.23	0.04	*
个体户	0.07	0.04	*
教师	0.18	0.04	*
无工作	0.08	0.04	*
家里有以下事物和条件(参照组:否):			
有自己独立的房间	-0.11	0.01	*
计算机	0.00	0.01	
有你自己的手机	-0.30	0.01	*
家用轿车	-0.07	0.01	*
录音机或CD机或MP3等	0.05	0.01	*
彩色电视	0.05	0.03	
摩托车	-0.04	0.01	*
家里课外读物册数(参照组:没有):多于200册数	0.08	0.03	*
百分比(%)—文科学生	-0.02	0.00	*
百分比(%)—父亲/男性监护人的最高学历高于高中、中师、中专、职业高中	0.01	0.01	
三科中考总成绩学校均值(Z分数)	-0.04	0.06	
随机部分:			
学校层面	0.09	0.02	*
学生层面	0.56	0.00	*
总变异(total variance)	0.65		
校内相关系数(intra school correlation x100)	13.17		
%被解释的学校间变异(school variance explained)	72.20		
%被解释的总体变异(total variance explained)	42.60		

注:学校样本数=54;学生样本数=30 090;增值模型三:三个地级市所纳入考量的解释变量一致,然而,此表仅列出至少在一个地级市的模型中呈现0.05水平的显著变量(以*标示);所有的考试总成绩和各科成绩已转换成Z分数(平均数=0,标准差=1);学生先前学习能力基线的平方和立方项亦纳入模型中,但结果没有列入此表格中。

参考文献:

- [1] [2] [11] Scheerens, J., etc. Educational Evaluation, Assessment and Monitoring. A Systemic Approach [M]. Swets & Zeitlinger Publishers. 2003.
- [3] Thomas, S., etc. Modelling Patterns of Improvement Over Time: Value Added Trends in English Secondary School Performance across Ten Cohorts[J]. Oxford Review of Education, 2007. (3).
- [4] [8] [16] Teddlie, C. & Reynolds, D. International Handbook of School Effectiveness Research[M]. Falmer Press. 2000.
- [5] Miliband, D. Using Data to Raise Achievement. Speech at a Conference of the Education Network [DB/OL]. http://www.teachers.gov.uk/_doc/6280/Miliband%20speech%20at%20TEN%20Conference%2011Feb04.doc
- [6] Stephens, D. Culture in Education and Development: Principles, Practice and Policy: Bristol Papers in Education: Comparative and International Studies[M]. Symposium Books.2007.
- [7] Goldstein, H., & Thomas, S. M. Reflections on the International Comparative Surveys Debate. Assessment in Education: Principles[J], Policy & Practice, 2008.(3).
- [9] 谢秀英.“高考移民”政策限制的合理性分析[J]. 陕西青年管理干部学院学报, 2007, (1); 姜世健. 高考移民现象的成因分析[J]. 教育发展研究, 2008, (17); 蒋伟, 马照云. 中国高考移民诉讼第一案追问[J]. 政府法制, 2008, (12).
- [10] 孙绵涛, 洪哲. 学校效能初探[J]. 教育与经济, 1994, (3); 王新如, 郑文. 谈学校组织文化与学校效能[J]. 教育科学, 1997, (3).
- [12] 蒋莉, 等. 学生高考语文成绩影响因素的多层线性分析[J]. 中国临床心理学杂志, 2005, (4); 汤林春, 梁玲玲. 学校效能评价的尝试[J]. 上海教育科研, 2005, (4); 孙河川, 等. 中国教育效能研究的近况、不足、发展、展望[C]. 沈阳: 中国第二届国际教育效能与学校改进大会暨中国教育效能学术委员会第三届年会, 2010.
- [13] OECD. Shanghai and Hong Kong: two Distinct Examples of Education Reform in China [A]. In Lessons from PISA for the United States: Strong Performers and Successful Reformers in Education[R]. 2012. 83-115.
- [14] 汤林春, 梁玲玲. 学校效能评价的尝试[J]. 上海教育科研, 2005, (4); 丁延庆, 薛海平. 高中教育的一个生产函数研究[J]. 华中师范大学学报(人文社会科学版), 2009, (2).

[15] IEEQC. Improving Educational Evaluation and Quality in China[DB/OL]. <http://ieeqc.bristol.ac.uk>

[17] UNESCO. EFA Global Monitoring Report 2005: Education for All. The Quality Imperative[R]. Paris.2004.

[18] Thomas, S. & Peng, W. J. Improving Educational Evaluation and Quality in China[R]. Final report to ESRC.2011.

[19] 丁延庆, 薛海平. 高中教育的一个生产函数研究[J]. 华中师范大学学报(人文社会科学版), 2009, (2); Jung Peng Wen; Thomas, Sally, etc. Developing school Evaluation Methods to Improve the Quality of Schooling in China: A Pilot “Value Added” Study. [J] Assessment in Education: Principles, Policy & Practice, 2006, (2);

马晓强, 等. 学校效能的增值评价——对河北省保定市普通高中学校的实证研究[J]. 教育研究, 2006, (10).

[20] Thomas, S. Dimensions of Secondary School Effectiveness: Comparative Analyses Across Regions [J]. School Effectiveness and School Improvement, 2001, (3).

[21] Garipey, K.D., Spencer, B.L. and Couture, J-C. (Eds.) Educational Accountability Professional Voices From the Field [M]. Sense Publishers.2009.

[22] 萨丽·托马斯. 运用“增值”评量指标评估学校表现[J]. 教育研究, 2005, (9).

Research on Value Added Evaluation of School Effectiveness

Sally M. Thomas, Peng Wenjung, Tian Huisheng, Li Jianzhong, Ren Chunrong & Ma Xiaoqiang

Abstract: To improve the evaluation methods for the school and student is the key countermeasure to promoting education quality. The research findings indicate that value added evaluation is generally regarded as a kind of effective and related measure supplied to evaluate the quality of Chinese education. This conclusion is similar with the research results of the other countries such as the UK. Secondly, among the existing Chinese school and teacher evaluation system, the concept of value added measures and its measuring methods have been regarded as a kind of comparatively scientific and popular evaluation method. Certainly, when applying any new evaluation system, we need to consider the local situations and the priorities. Whether value added measures can be within a public accountability framework to offer school effectiveness evaluation system a kind of most useful or suitable method, to offer school secretive feedback information, and to improve the capability of school self-evaluation and improvement, which is worthy of discussion.

Key words: value added evaluation approach, educational evaluation, multi-level model, educational quality, school effectiveness

Authors: Sally M. Thomas, professor of Graduate School of Education, University of Bristol; Pen Wenjung, Ph.D. of Graduate School of Education, University of Bristol; Tian Huisheng, Director, senior researcher and doctoral supervisor of National Center for Basic Educational Curriculum and Textbook Development of Ministry of Education (Beijing 100029); Li Jianzhong, associate researcher of National Institute of Educational Sciences; Ren Chunrong, associate researcher of National Institute of Educational Sciences; Ma Xiaoqiang, associate researcher of National Institute of Educational Sciences (Beijing 100088)

[责任编辑:张 平]