



Thomas, S. M., Peng, W-J., & Li, J. (2015). —. *Educational Research*, 35(7), 64-72.

Publisher's PDF, also known as Version of record

License (if available):
Unspecified

[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/pure/about/ebr-terms>

学校增值表现与教师专业发展关联性探析*

[英]萨丽·托马斯 彭文蓉 李建忠

[摘要] 通过运用不同的多水平模型技术,量化分析中国东西部三个地级市2009—2012年高考考生的高考成绩和中考成绩、学校和教师问卷信息,同时质性分析核心利益相关者访谈质性数据,以探究学校效能的范围、程度、表现趋势以及学校效能和教师专业发展关联性。研究发现,普通高中高考原始成绩和增值表现估测值存在统计上的显著性差异,时间横轴上个别学校的表现缺乏一致性,学校对不同学生群体在高考总分增值表现的影响有差异性,教师专业发展因素和学生学习成就之间具有正向关联性。

[关键词] 增值;专业学习共同体;学校效能;教师专业发展;多水平模型;教育评价;教育质量

[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客座研究员;李建忠,中国教育科学研究院副研究员 (北京 100088)

一、研究背景和方法

(一)研究背景

西方国家广泛开展的学校效能研究证实,学校对学生的学习成绩和进步有着重要的影响,而且学校的影响程度存在着显著的差异。^[1]因此,如何评价学校的实效已是一个重要且不容忽视的议题。中国学生学业的原始成绩和大学升学率常常被看成是衡量学校质量的最重要标准。这可能导致那些学生生源水平较低的学校被不公平评价,而拥有相对高质量生源的学校则处于升学优势地位。在这种情况下,很难鉴别出一所学校实

际的办学水平。事实上,自从40多年前西方开始从事有效学校的实证研究以来,人们不断探索评价学校表现的新方法,这为各国一系列的评价学校政策做出了贡献。

从有效学校实证研究中逐渐衍生出来的增值评价方法为评价学校表现和问责学校提供了一个较为公平的方式,因为这种评价方法可以避免仅仅把关注点放在原始考试成绩上。^[2]更为重要的是,这种方法本质上是在考虑学生之前的学习成绩和其他非学校本身所能控制的相关因素的情况下,用不同学生入学成绩之后的学习成就来估测一所学校的学生与其他学校学生相比的进步情形。因此它所反映的是,与其他同类学校的学生相比,

* 本文系英国国际发展部(DFID)和经济与社会科学研究委员会(ESRC)联合资助的中英合作项目“改进中国教师发展和教育质量”(项目编号:167-25-0428)的研究成果之一。项目英方负责人:萨丽·托马斯;项目中方总负责人:田慧生;项目中方协调人:李建忠。参与本项目的其他成员还有易凌云、冯继。

一所学校促进学生在原有水平上提高的学习水平。因而“增值”的概念既可以是一个学校效能的指标,也可以是一种工具,让校长和员工们用来分析其有效提高学生成绩的程度。

中国还很少有涉及增值方法和学校效能的高质量实证研究报告,这也是为什么在研究学校效能的国际比较研究领域里罕见中国案例的一个重要原因。值得一提的是,在至今少数已发表的有关中国学校效能的研究中,即以学校对学生高考成绩的影响为研究对象的调查里,有一些有趣的发现:调整前和调整后的学校效能在学生高考成绩的总体变异中可能占到50%。^[3]

另外一方面,相关研究证实教师质量和教师专业学习与发展的机会在提高学生学习成绩和进步方面起着非常重要的作用。^[4]同样地,学校改进研究也一再强调课堂教学水平对于提高学生表现的重要性。^[5]由此可见,教师专业发展和质量与学校效能有着密切的关联性。然而,一些研究指出,对于中国的很多教师来讲,专业学习是不发达、不公平和不一致的。^[6]国际上在寻求有效方法来促进教师专业学习与发展方面,美国和英国的相关研究强调专业学习共同体对于提高教师专业实践所具有的价值和作用。^[7]尽管中国一些学者认为,专业学习共同体一般来讲已存在于正规教育体系里,而且大部分以学校教研组的形式出现,然而在文献中还是很少发现有关这类主题的实证研究。^[8]

基于提高教育质量对于减少贫困的重要性^[9],开展教师专业发展及其对学生学习成绩和进步的影响的研究及时且必要。在中国这样一个幅员辽阔、区域发展不平衡的国家,显然需要更多大规模且具代表性的实证研究为学校效能的范围和程度提供强有力的定量研究证据,从而影响相关教育政策的制定,以及充实和丰富有关教育质量的国际比较研究内涵。这一议题对于中国乃至世界其他国家制定教育督导评价政策也显得尤为重要。

有鉴于此,由英国布里斯托大学和中国教育科学研究院共同实施的中英合作项目“改进中国教师发展和教育质量”(简称ITDEQC)旨在提供强有力的新证据,研究分析学生特征、学校情境和教师专业发展等因素对学生在学校的学习成绩的影响,以及这些因素与学校表现的关联性。

(二)数据采集和方法

ITDEQC项目采集了量化和质性两方面的数据。量化数据包括三个地级市2009年到2012年四届学生的高考与中考成绩以及学生和学校的问卷信息,以探讨随着时间推移中国学校效能的本质和影响程度。三个地级市的四届学生的样本量总计有134所学校,303 345名普通高中毕业生。从2009—2012年,每年采集的学生问卷数据包括学生背景及其对学校态度方面的信息。每年的学校问卷都调查了学校情境、过程、文化、教师发展、教与学态度等方面的信息。此外,2012年对项目学校17 000多名教师进行了问卷调查,并将学生问卷、学校问卷与教师问卷调查数据相匹配,探讨学校效能与教师专业发展的关联性。教师问卷还吸收了经济合作与发展组织(OECD)2008年实施的“教与学国际调查”(TALIS)中一些有关教师专业发展的问题,以便提供相对等的的数据。

质性数据的采集则是通过对70多位包括校长、教师、学生、国家和地方教育行政部门官员等在内的核心利益相关者进行了个别采访和焦点小组访谈,并且对采访和访谈文字进行系统记录和分析,以探讨教师学习、评价与专业发展的方法以及专业学习共同体在学校所体现出的关键主题和特征,并辅以教师问卷的量化数据分析。

本文从宏观角度呈现ITDEQC项目部分研究成果,着重运用多水平模型技术进行量化数据分析,来探索随着时间推移学校效能增值的表现,以及增值与教师专业发展的关联性。也就是通过渐进式地考虑由下列不同

类别的变量群组建构的各种不同层次的多水平模型(同时,不分地区和年份,采用相同结构的多水平模型),来估测学校效应的范围和程度:(1)学生中考成绩;(2)学生特征变量如性别、户口所在地、父母教育程度、父母职业、家庭拥有物等;(3)学校情境变量如文理科生比例、母亲为农民工百分比、中考成绩学校均值等;(4)学校投入和过程变量如生师比、相关学校文化以及有关教与学态度因素等;(5)学生努力和态度变量如在家或在学校自学时间、上网时间、有关教与学态度因素等;(6)教师专业发展变量如参加正式培训的时间、各种形式的专业学习等因素。然后,将这些发现与中国高中学校的教师发展程度和专业学习共同体的关键特征进行对比。

二、研究发现和讨论

ITDEQC项目对2009—2012年数据进行分析,结果发现,普通高中学校效能的增值估测值存在统计上的显著性差异,即这些差异不仅因地区而异,也因高考科目而异^[10],而且这些差异会随着时间推移有所变动。研究还显示,学生高考增值表现的差异性不但与学校情境、学校投入与过程、学生努力与态度等变量有所关联,而且也与教师专业发展相关的变量有关。

(一)学校效能存在着显著性差异

就学生的高考原始分数而言,跨届学生(二或四届)、各地区(三个地级市)和各高考科目(如总分、语文、英语和数学),学校之间的差异占到学生分数总体变异的百分比范围为12%~27%(见表1中调整前原始模型,即不考虑任何变量的二阶多水平模型分析所得的校内相关系数)。

与上述的学校在学生高考原始成绩表现的差异性相比,各学校在调整后学生高考情境化增值(adjusted contextualised value added)成绩表现的差异明显缩小,然而仍存在差异性。也就是说,通过建构情境化增值模型^①将学生中考成绩和在学校控制范围以外的学生特征和学校情境因素纳入考虑后,可以解释学生高考成绩18%~77%的总体变异和31%~96%的学校间变异。在尚未被解释的残余总体变异里,仍有2%~22%可归因于学校间的差异(见表1中调整后情境化增值模型分析所得的校内相关系数),从而证实了学校对于学生学习成绩的影响,尤其是在西部,我们可观察到较大的学校效应^[11]。这些发现与以往在中国所做的类似研究的结果相比,有着更好的数据拟合度以及更强有力的估测值。与英国类似研究的结果相比,中国至少有两个地区的学校效应比英国的学校效应略强。^[12]

(二)学校表现的时间趋势存在变动性

我们通过未考虑任何变量因素的二阶多水平原始模型来分析项目学校的每一年高考原始总分表现,结果表明学校表现的时间趋势存在着变异性。图1呈现的是2009—2012年间其中一个项目地区有完整四年数据的32所学校的结果,每一条线代表一所学校,纵坐标是以0为平均分的高考原始分数,横坐标代表从2009年起连续的四届学生群体。由于高考的难易程度随着时间而变化,这种变化也体现在每年的高考录取分数线上,所以我们在建构二阶多水平原始模型时,考虑了跨学校的平均时间趋势。尽管如此,个别学校高考原始总分年与年之间还是出现了明显的变异性,反映在时间横轴上就是高考原始分数的表现缺乏一致性。

^① 基于可以解释更多百分比的总体变异(亦即获得较佳的模型拟合度)的前提下,本文用来分析2009—2012年和2009—2010年数据的情境化增值模型中所控制的学校情境变量与在先前研究里用来分析2009年单年数据的情境化增值模型中所控制的学校情境变量稍微有些许不同。

表1 通过调整前原始模型和调整情境化增值模型分析所得的四种学生高考原始和增值成绩的变异数值比较

| 二阶多水平模型 | 地区一 (2009—2012年) | 地区二 (2009—2010年) | 地区三 (2009—2012年) |
|--|---------------------|---------------------|---------------------|
| 高考总分 | | | |
| 校内相关系数(intra-school correlation)——调整前原始模型 | 22%~26% | 17%~27% | 22%~27% |
| 校内相关系数(intra-school correlation)——调整后情境化增值模型 | 4%~7% | 10%~12% | 4%~14% |
| 调整后情境化增值模型解释了学校间变异的百分比 | 88%~93% | 76%~81% | 71%~89% |
| 调整后情境化增值模型解释了总体变异的百分比 | 58%~68% | 45%~65% | 34%~48% |
| 高考语文分数 | | | |
| 校内相关系数(intra-school correlation)——调整前原始模型 | 12%~19% | 13%~19% | 17%~26% |
| 校内相关系数(intra-school correlation)——调整后情境化增值模型 | 2%~5% | 6%~7% | 8%~22% |
| 调整后情境化增值模型解释了学校间变异的百分比 | 81%~94% | 77%~80% | 31%~66% |
| 调整后情境化增值模型解释了总体变异的百分比 | 39%~50% | 33%~54% | 18%~28% |
| 高考英语分数 | | | |
| 校内相关系数(intra-school correlation)——调整前原始模型 | 20%~24% | 15%~23% | 23%~27% |
| 校内相关系数(intra-school correlation)——调整后情境化增值模型 | 4%~7% | 8%~11% | 3%~12% |
| 调整后情境化增值模型解释了学校间变异的百分比 | 91%~96% | 78%~82% | 72%~94% |
| 调整后情境化增值模型解释了总体变异的百分比 | 66%~77% | 46%~68% | 35%~52% |
| 高考数学分数 | | | |
| 校内相关系数(intra-school correlation)——调整前原始模型 | 19%~23% | 16%~24% | 19%~24% |
| 校内相关系数(intra-school correlation)——调整后情境化增值模型 | 4%~6% | 8%~10% | 2%~10% |
| 调整后情境化增值模型解释了学校间变异的百分比 | 84%~93% | 72%~78% | 71%~93% |
| 调整后情境化增值模型解释了总体变异的百分比 | 45%~61% | 37%~57% | 29%~40% |

注: 显示的数字为跨二或四届学生群体的估测值变化范围。

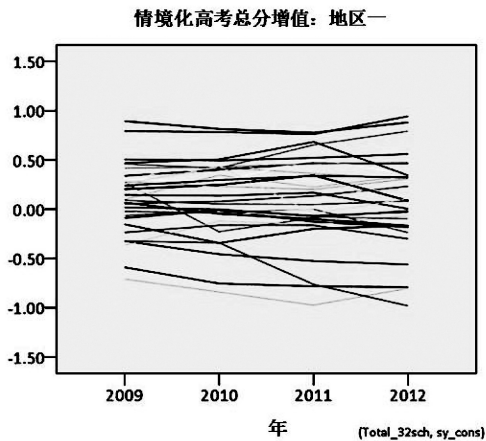


图1 基于调整前原始模型的学校高考原始总分表现之时间趋势图

图2展示的是同一个项目地区的32所学校在2009—2012年的情境化高考总分增值表现的结果,不仅考虑了学生中考成绩、学生特征和学校情境的因素,而且也考虑了跨学校的平均时间趋势。乍看之下,图2与图1相似,但学校时间趋势间的差异明显减少,这是因为学生的中考成绩和其他因素解释了部分

学校间的表现差异。即使如此,学校间的情境化高考总分增值表现时间趋势仍然存在明显差异,如同学校的原始高考总分表现一样,在时间横轴上个别学校的高考总分增值表现也缺乏一致性。相对于这个地区的其他学校而言,某些学校在高考总分增值表现的时间趋势比平均时间趋势更好或者更差。同样,在英国同类研究中也出现了类似结论。

在学校自我评估实践中,学校若能有效运用这类信息则会有助于其反思以往实施了何种程度的政策和实践,对其历届学生的高考原始成绩和增值表现产生了何种程度的重要影响。这些时间趋势的研究结果表明,仅仅依靠单一年份的结果来评价学校的表现可能会有一定的误导性。

我们还通过三阶(学校/年份/学生)多水平模型检验了普通高中在2009—2012年间学校效应的总体估测值(summary estimate of school effect)。表2列出了两个项目地区的

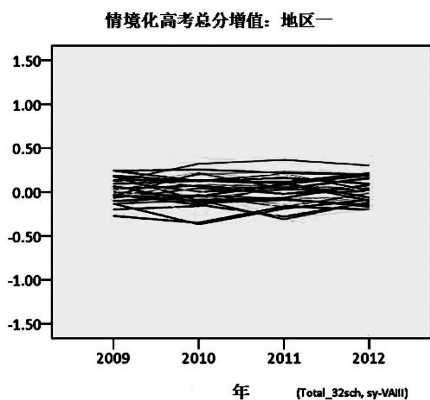


图2 基于调整后情境化增值模型的学校高考总分增值表现之时间趋势图

表2 两个地区自2009—2012年四年间学校高考原始和增值表现的总结估测值

| | 校内相关系数(intra-school correlation)(%) | | | | 年内相关系数(intra-year correlation)(%) | | | |
|-----------------|-------------------------------------|------|------|------|-----------------------------------|------|------|------|
| | 高考总分 | 高考语文 | 高考英语 | 高考数学 | 高考总分 | 高考语文 | 高考英语 | 高考数学 |
| 地区一(2009—2012年) | | | | | | | | |
| 调整前原始模型 | 26.3 | 20.1 | 25.4 | 22.0 | 1.3 | 1.5 | 1.6 | 1.8 |
| 调整后情境化增值模型 | 7.4 | 3.5 | 5.7 | 5.6 | 1.6 | 1.5 | 2.4 | 2.0 |
| 地区三(2009—2012年) | | | | | | | | |
| 调整前原始模型 | 22.7 | 19.7 | 23.4 | 20.3 | 1.6 | 2.0 | 1.7 | 1.4 |
| 调整后情境化增值模型 | 10.4 | 16.4 | 10.4 | 7.0 | 3.5 | 3.4 | 3.0 | 2.4 |

(三)学校内存在着区别性效应

通过个别年份的情境化增值模型的分析结果发现,学校对以中考成绩来区分的不同学生群体在高考总分增值表现的影响有差异性。比如在地区一,在2009—2012年的四年里,中考高分和中考低分成绩的两个学生群体间高考增值表现的相关系数数值范围为0.43~0.69(见表3)。这些研究发现意味着,学校对不同能力学生的学习成绩可能存在着不同程度的影响。如果只用一个总的计量来评价学校增值表现,特别是在中国情境里很多普通高中采用混合能力编班教学的情况下,可能掩盖了学校区别性效能的特性。这类校内区别性效能的证据可以帮助学校和教师判断出是否一些能力相对较差的学生在学习中苦苦挣扎,或者一些有能力的学生没有充分发挥出其学习上的潜能。此外,研究也发现学校对不同的科目如语文、数学和英语的高考增值表现的影响也有较小程度的区别

学校在高考各科目表现的变异来源,包括可以归诸于学校间差异以及可以归诸于年与年之间差异的百分比估测值。其中,从调整后情境化增值模型(该模型亦考虑了随着时间变化而观察到的随机变异,两个地区的估测值约为2%~4%)分析的结果可以看出,项目地区一学生的高考总分增值表现的总体变异可以归于学校间差异的百分比降到7%,项目地区三则降到10%。这两个估测值可以被视为地区一和地区三普通高中在四年里学生高考总分增值表现的总体学校效应。

性效能差异(见表3,任两个学科的相关系数数值范围为0.57~0.89)。

表3 地区一学校区别性效应(differential school effects):2009—2012年不同学生群体间(高考总分)以及不同科目间学校效能的相关系数数值

| 地区一 | 2009年 | 2010年 | 2011年 | 2012年 |
|-----------------|-------|-------|-------|-------|
| 男生与女生 | 0.95 | 0.98 | 0.99 | 0.94 |
| 中考高分成绩组与中考低分成绩组 | 0.43 | 0.60 | 0.69 | 0.65 |
| 理科生与文科生 | 0.82 | 0.81 | 0.80 | 0.79 |
| 城市户口与农村户口 | 0.77 | 0.91 | 0.91 | 0.89 |
| 科目 | | | | |
| 语文与英语 | 0.61 | 0.89 | 0.83 | 0.85 |
| 语文与数学 | 0.57 | 0.80 | 0.78 | 0.81 |
| 英语与数学 | 0.66 | 0.81 | 0.86 | 0.80 |

正如我们所预期的那样,这些学科间增值表现的相关系数数值仍然高于英国类似的研究结果。^[13]我们的研究表明,与英国的学校实践相比较,跨科目的学校政策和教师合作在中国可能是一个明显的优势,尤其是通过中国特有且非常普遍的年级教研组这种组

织形式,教师们经常召开会议,讨论教学策略的改进,从而有助于缩小科目间教学的差距。

(四)学校情境、学校投入与过程、学生努力与态度变量的影响

研究发现,针对不同高考科目的表现,学校情境变量的影响随着时间的推移在不同的项目地区有着很大变化,即在学生中考成绩和学生特征变量的影响之上,再额外解释到的学校间变异的百分比范围为2%~60%。教师专业发展以及学校投入和过程变量也分别有类似的百分比影响范围。这意味着在某些地区,尤其是在不发达的农村地区,这些因素对学校表现产生的影响则更大。有关学生中考成绩和学生特征以及学校情境等个别变量对学生和学校表现的影响已经在之前有过报告^[14]。

以2009年数据为分析对象,在情境化增值模型基础上再将投入和过程变量纳入模型之中,分析结果显示,有少数几个学校投入和过程变量对学校高考总分增值表现呈现统计上显著性影响。随后,通过同样结构的多水平模型去分析随后的三年,即2010年、2011年和2012年数据,结果发现,在三个项目地区以及不同年度间,个别学校投入和过程变量呈现的统计上显著性及其正负向的影响有着轻微变动,也就是说,在某种程度上缺乏一致性。而且,我们的研究还发现,有三个学校投入和过程变量(即学生教师的比例、校长常到课堂听课、教师参与学校各个方面的决策)在所有三个地区都至少有一年或一年以上与高考总分增值表现有着统计上显著性的正向关联。

此外,在上述模型结构的基础之上,再将通过学生问卷调查搜集到的学生努力与态度变量纳入模型中进行分析,研究结果发现,除了一个学生努力变量(每天至少在家或在学校自学两小时)之外,有七个学生态度变量在所有三个地区,都至少有一年或一年以上与高考总分增值表现有着显著的正相关。这

些变量包括:(1)我在班上行为表现良好;(2)在这所学校里我得到公平对待;(3)老师鼓励我课堂提问、思考问题、发言和讨论;(4)老师及时反馈作业情况以提高我的成绩;(5)老师给予我适当的表扬、肯定和帮助;(6)在学校,我感觉很安全;(7)我每天都有足够的食物。我们的研究结果还显示,有一个学生努力变量(学生若每周至少花三小时上网做与学习无关的事)在所有三个地区都至少有一年或一年以上与高考总分增值表现有着显著的负相关。整体而言,这些发现意味着教师的反馈和合作、学生的反馈与支持、学生努力和态度以及资源在提升学生学习成绩中起着关键性的作用。^[15]

(五)教师专业发展对学生增值表现的影响

通过2012年的教师问卷调查对教师专业发展、专业学习共同体和相关问题进行深入剖析,调查涵盖三个地区的一万七千多名教师。教师问卷数据旨在检验专业学习共同体和教师发展因素对学生增值进步的影响。初步调查结果显示,教师专业发展变量确实对学生高考成绩产生一些影响。比如以地区一为例,运用三阶多水平模型(学校/教师/学生)分析学生中考成绩、学生特征、学校情境以及教师专业发展变量,这些教师专业发展变量与2012届学生高考英语、数学和语文科目至少一科或者多科的表现有统计上的显著正向影响,比如参加正式专业发展培训活动的总学时、同事间非正式的改进教学的交流、参加相关课题研究和参加主要教授科目教学策略和方法培训等。研究表明,教师培训的机会、内容与品质以及教师之间的合作和教育资源的提供都与学校效能相关联。

深度分析教师问卷数据,我们发现教师专业发展存在以下关键问题和特点。第一,教师发展的机会存在着地区性差异。在过去的18个月里,三个地区的教师有权享有的正式专业发展培训的时间从34学时到67学时

不等。第二,教师参加专业发展培训活动的类型因地区而异(见表4)。此外,培训活动的类型也因教师的教龄而异,其中青年教师倾向以教学竞赛和观摩课来提升自身专业发展。第三,教师继续专业发展具有实践性、合作性和针对性。实践性是指与教师的实践紧密相关,合作性强调团队合作和相互交流,针对性是指对处于不同专业发展阶段的教师需要量身定制。学科教师最大的培训需求是在所教授科目中对专业知识与理解、教学策略和方法的知识与理解、学生的心理发展等方面。第四,“专业学习共同体”这个概念在中国的认可度较高。84%的教师认同其概念的重要意义及其与国内教育情境的关联性和实

用性。然而,接受访谈的教师认为专业学习共同体的一些特征很具挑战性,如何提升个人反思性专业探究能力,如何提高所有教职员工体现专业学习共同体的包容性,都是需要进一步思考的问题。比较中英学校和教师问卷发现,两国在学校实践上存在一些差异,中国更注重教师团队合作和课堂观察,而英国更注重与家长和社区分享信息,也更欢迎非教学人员融入专业学习共同体,为个别学生设定发展目标。^[16]中国教师专业学习的实践还具有东亚文化特色,普遍实行教研组制度与观摩课制度,强调教师撰写和发表研究论文的重要性,尤其是后者在中国情境下是教师职称晋升的条件之一。^[17]

表4 教师反馈在过去18个月参加过的各种专业发展培训活动的类型的百分比——基于2012年的教师问卷调查

| 过去18个月,教师参加了下列形式的专业发展培训活动的百分比(%) | 地区一 | 地区二 | 地区三 |
|------------------------------------|-------|-------|---------------|
| 进修课程及专题讨论会(例如,着重在所教学科或与教育相关的主题) | 79.50 | 78.12 | 82.88 (74.97) |
| 教育会议和研讨会(有教师或教研员报告其研究成果和讨论相关的教育问题) | 77.54 | 75.82 | 76.50 (67.64) |
| 学历提高课程 | 52.47 | 45.01 | 44.80 (37.98) |
| 教师专业发展专属网络 | 70.97 | 68.37 | 73.08 (62.84) |
| 个人或与他人合作的研究课题 | 65.12 | 60.39 | 67.18 (57.88) |
| 学校安排的同事间的指导、观课或辅导 | 91.33 | 84.81 | 91.56 (80.10) |
| 专业文献(例如,期刊、研究性文章、论文)的研读 | 77.56 | 72.05 | 82.99 (71.83) |
| 同事间非正式的有关如何改进教学的交流 | 91.79 | 83.04 | 91.77 (80.49) |
| 课题研究、论文撰写 | 76.46 | 70.55 | 77.33 (67.70) |
| 教研组、学科组、年级组的活动 | 93.40 | 85.74 | 94.40 (83.68) |

注:(1)样本的大小(教师数/学校数)=地区一(8 093/41),地区二(7 608/45),地区三(1 814/21);(2)就整个样本而言(17 515/107),表中所列每题问卷题目的缺失值百分比低于2%;(3)但是,仅就地区三而言,缺失值的百分比的范围高达9% - 16%,因而也将缺失值视为否定值而计算得出的百分比放在括号中。

以上研究成果,对于如何更好地理解 and 评价教师发展、学校效能和教育质量以及相关实践者如何在中国情境下运用实证研究结论来改善教师实践有很大的启示。它与中国和世界各国教育政策与实践的发展息息相关,尤其是关于情境的特异性、学校改进和问责制度等方面。

三、结语

有鉴于学生享有公平的教育机会和更好的学习结果与提高民族素质和促进经济增长

有明确的关联性,提高教育质量已成为无论是发展中国家还是发达国家的一个重要教育发展目标。^[18]在这种情况下,学校效能研究的结果促使教育政策决策者更加关注这类研究在提升整体教育标准和学生学习成就水平的潜在作用。例如,英国把重点放在鼓励学校和教师运用创新性的评价方法并量化评估学校教育品质,将增值评价结果作为问责制度和督导评价框架的参考。而且,这种方法也被证实与学生学习成就水平的提高有关联性。中国可以进行比较的学校效能实证研究案例还不是很多,也很少有人将学校效能与

教师专业发展结合起来进行研究。中英合作项目的研究结果填补了这方面的空白。特别是在目前正进行教育深化改革的大背景下,本文所呈现的研究证据可对教育督导评估政策的制定提供有益的参考借鉴。此外,课题不仅分析了不同地区的学校效应和教师专业发展在提高学生学习成就方面所发挥的作用,而且也为发展和丰富学校效能研究的国际知识库做出了贡献。

尽管上述研究结果与西方同类研究相比存在差异,但教育效能研究的方法和结论在中国情境下得到了一定程度的验证。从广义上来讲,中国和英国的学校效能有着可比较的、类似的情境化增值结果和时间趋势模式。起源于英美国家的“专业学习共同体”这个概念在中国同样适用,但教师专业发展实践在不同的国家情境下有较大差异。

总体来说,本项目研究为大规模的学校效能研究提供了一个强有力的示例。如果能在中国更大的区域范围内或在国家层面系统地采集相关数据和证据,这种大规模实证研究项目是可行的。本研究的另一个重要意义在于,连续四年的证据显示,情境因素对学校表现的影响明显地有着地区性的差异,这也说明了地方情境因素会在不同程度上以不同的方式展现其影响。至于学校投入和过程因素对学校表现的影响,在某种程度上也有着类似的地区性特性。上述这些研究发现意味着除国家层面的评价体系外,建立单独的、适合地方需要的地方性评价体系可能更有意义而且更重要。

最后需要强调的一点是,本项目研究仅涵盖三个地区,覆盖面相对有限。在中国这样一个幅员辽阔的国家,需要有更多大规模的、具有代表性的教育效能研究,以评测教师专业发展和学校情境对提高教育质量和学校效能所做的贡献。

参考文献:

[1] Charles Teddlie & David Reynolds. *The International Handbook of School Effectiveness Research* [M]. London and New York: Routledge Falmer, 2000. 411.

[2] Sally Thomas, et al. *Modelling Patterns of Improvement over Time: Value Added Trends in English Secondary School Performance across Ten Cohorts* [J]. *Oxford Review of Education*, 2007, (3).

[3] 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); Wen-Jung Peng, et al. *Developing School Evaluation Methods to Improve the Quality of Schooling in China: A Pilot “Value Added” Study* [J]. *Assessment in Education*, 2006, (2); 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); 辛涛,等. 中高考数据链接:对学校进行增值性评价——以某市40所高中2132名学生中高考数据的实证分析为例[J]. *中小学管理*, 2012, (6); 陈育庭. 普通高中教育投入绩效评估实证研究——基于江门市47所高中调查的多层数据[J]. *数学的实践与认识*, 2012, (9).

[4] Ray Bolam, et al. *Creating and Sustaining Effective Professional Learning Communities* [R]. London: DfES and University of Bristol, 2005. 210.

[5] David Hopkins, et al. *Creating the Conditions for Classroom Improvement* [book]. London: David Fulton Publishers, 1997. 128.

[6] 赵明仁,等. 北京市中小学教师参与专业发展活动现状与需求的调查研究[J]. *教师教育研究*, 2009, (1); Bernadette Robinson & Wenwu Yi. *The Role and Status of Non-governmental (‘daike’) Teachers in China’s Rural Education* [J]. *International Journal of Educational Development*, 2008, (1); Wen-Jung Peng, et al. *Emerging perceptions of teacher quality and teacher development in China* [J]. *International Journal of Educational Development*, 2013, (4).

[7] Louis Stoll & Karen Seashore Louis. *Professional Learning Communities: Divergence, Depth and Dilemmas* [M]. Maidenhead: Open University Press/McGraw-Hill. 2007. 232.

[8] Charles Teddlie & Shujie Liu. *Examining Teacher Effectiveness within Differentially Effective Primary Schools in the People’s Republic of China* [J]. *School Effectiveness and School Improvement*, 2008, (4).

[9] Tiedao Zhang & Zhao Minxia. *Universalizing Nine-year Compulsory Education for Poverty Reduction in Rural China* [J]. *International Review of Education*, 2006, (3-4).

[10][14] 萨丽·托马斯,等. 学校效能增值评量研究[J]. *教育研究*, 2012, (7).

[11][17] Sally M. Thomas & Wen-Jung Peng. *Methods to Evaluate Educational Quality and Improvement in China* [A]. Janette Ryan. *Understanding China’s Education Reform: Creating Cross Cultural Knowledge, Pedagogies and Dialogue* [C]. Abingdon: Routledge, 2011. 75—91, 75—91.

[12] Sally Thomas. *Dimensions of Secondary School Effectiveness: Comparative Analyses across Regions* [J]. *School Effectiveness and School Improvement*, 2001, (3); Sally Thomas, et al. *Modelling*

Patterns of Improvement over Time: Value Added Trends in English Secondary School Performance across Ten Cohorts[J]. Oxford Review of Education, 2007, (3).

[13][15] Pam Sammons, et al. Forging Links: Effective Schools and Effective Department [M]. London: SAGE Publications Ltd, 1997. 268, 268.

[16] Sally M. Thomas & Wen-Jung Peng. Evaluating Schools as Effective Professional Learning Communities [Z]. Conference Paper Presented at 2011 AERA sSymposium, New Orleans, 2011.

[18] UNESCO. EFA Global Monitoring Report 2005. Education for All: the Quality Imperative[R]. Paris: UNESCO. 2004. 430.

Study on the Relationship between School Value Added Performance and Teachers' Professional Development

Sally M. Thomas[UK], *Wen-Jung Peng* & *Jianzhong Li*

Abstract: This paper explores time trends in schools value added performance in mainland China and the links between "value added" measures of school effectiveness and teachers professional development. This was done by employing a variety of different multilevel models to analyse data collected from three regions (east/west) and four cohorts as part of a DFID and ESRC funded UK-China collaborative project. The data comprised students' 2009-2012 entrance examination to higher education scores matched to their 2006-2009 prior attainment scores as well as student and school survey information, and 2012 teacher survey information. Alongside this qualitative data-interviews and focus groups with key stakeholders including head teachers, teachers, students, national and local policy makers-was also collected and analysed. The findings reveal significant differences in raw and value added performance in the university entrance examination results between senior high schools in China that vary not only across regions and subject outcomes but also across the time frame of the four student cohorts examined, indicating a lack of consistency in individual school performance. Within schools differential effects were also found for different groups of students. Moreover, in addition to school input and process, and student effort and attitude factors, the paper identifies positive links between teacher professional development factors and students' outcomes in terms of "value added" progress. There is very little comparable empirical research evidence on the range and extent of school effectiveness in China and how this links to teacher development issues. The findings reported in this paper seek to address this crucial gap by contributing new evidence to the international knowledge base of school effectiveness research.

Key words: value added, professional learning community, school effectiveness, teacher professional development, multilevel modeling, education evaluation, educational quality

Authors: Sally M. Thomas, professor of Graduate School of Education, University of Bristol; Wen-Jung Peng, visiting researcher of Graduate School of Education, University of Bristol; Jianzhong Li, associate researcher of National Institute of Education Sciences (Beijing 100088)

[责任编辑:刘 洁]