

改进中国教育评价与教师发展

萨丽·托马斯 (Sally M. Thomas)¹, 彭文蓉¹, 田慧生², 李建忠²

(1.教育研究院, 布里斯托大学, 英国; 2.中央教育科学研究所, 北京 100088)

(论文发表于全国教育管理学科委员会第 11 届学术年会, 2011 年 10 月 南昌)

(论文报告者: 彭文蓉)

摘要: 本文透过英国经济与社会科学研究委员会 (ESRC) 和国际发展部 (UKAID) 联合资助的两项研究课题, 来探讨运用增值评量法评价学校效能对影响和促进相关政策来提高中国教育质量与教师发展上的潜在用途。研究结果指出, 增值评量法普遍被认为可在评价中国教育质量上, 提供一有效且相关的评价方法。研究发现亦指出, 学校对其不同学生群体具有区别性效能。然而, 中国所面临的挑战是如何建立一完整的地区性/全国性的相关信息数据库。另初步研究发现亦指出, 专业学习共同体在教师发展的主要过程中扮演一定的角色。此外, 中国学校的专业学习共同体有其特有的以及不同于英国学校之专业学习共同体的面貌。

关键词: 增值评量法; 教育评价; 多水平统计模型; 教师发展; 学校效能; ESRC

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下面两项相关联课题, 来研究有关中国教育评价议题 [1]。第一项课题: 改进中国教育评价和教育质量 (Improving Educational Evaluation and Quality in China, IEEQC), 此课题最近刚结题, 并延伸成为第二项新课题: 改进中国教师发展和教育质量 (Improving Teacher Development and Educational Quality in China, ITDEQC)。该两项课题由布里斯托大学并与北京中央教育科学研究所合作实施。

「改进中国教育评价和教育质量」课题

「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」概念在学生不同的学习结果上, 来拓展中国现有跨区域范围内对高中学校效能的定义与测量。第二, 探讨政策制定者和教育工作者于考虑当地背景和优先事项的需求之下, 如何采行、适应并发展「增值」评量法在中国的运用。譬如, 以评价学生学习和结果为例, 是评量公民养成或学术成就方面, 评价方法是偏爱定性或定量方式, 以及情境特征对学生和学校表现的影响。广义而言, 这项课题在增进对中国学校效能复杂本质的理解。但重要的是, 研究一个更根本的问题, 也就是, 地方情境因素在界定教育效能和质量方面是扮演什么样的关键作用。再者, 如果能对来自不同社经、政治背景的政策制定者和教育工作者, 譬如来自中国, 是如何理解大部份由英国和欧美国家所发展出来的教育质量概念以及评价方法的方面上, 有一个比较清晰的认识, 则可大大提高学校效能和改进分面的研究 [2] [3]。尤其, 要了解地方和文化方面的议题, 可能对学校和学生的教育价值想法以及其考虑的优先事项所产生的影响, 这类比较研究的证据是必不可缺的。

研究方法

「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, 以及中国可能如何采行、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。一共有超过 90 位包括校长、教

师、学生、地方和国家层面的政策制定者参与个别或团体焦点访谈。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。一共有超过 120 多所来自东西部三个地级县市的高中学校以及超过 90,000 位 2008/2009 学年度的高中学生提供相关信息。

研究结果提要

「改进中国教育评价和教育质量」课题的研究结果指出，虽然受访的利益相关者认可国际上对教育质量的定义，但却非常关注并且强调，在考量中国具体优先事项的情况下，学生的全方位发展（譬如素质教育）和教育公平性是质量的主要目标 [4]。此外，一些受访的利益相关者也认为，目前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。

关于中国学校效能的估测，表一总结了透过前后逐一引入相关变量所建立的各种二阶层多水平模型，来探讨 2008/2009 学年度学校表现的差异性。研究结果显示，跨越三个地级县市，在没有考虑其它因素对学生高考成绩的影响的前提下，2009 年高考总分的总体变异，有 24-27% 归诸于学校间的差异性（见表一，初始模型）。然而，一旦引入各种变量于模型中，学校在原始高考成绩和增值评量的表现，有明显的改变。诸如，当引入中考成绩、学生和学校背景等学校无法控制的变量于模型中（见表一，增值模型三），这些变量的考量解释了 2009 年高考总分之总体变异（total variance）的 43-57% 以及学校间变异（school variance）的 70-91%。其余未被解释的 2009 年高考总分之总体变异中，可归诸于学校间差异性（亦即校内相关系数 intra school correlation）的尚有 5-15%。这些证据展示了，尤其是在中国西部地区，通过学校作用于学生，对其学习成就的影响。一些以往在中国从事的实证研究亦获得类似的结果 [5] [6] [7]。此外，似乎至少有两个中国地区的学校作用，其影响大于英国一些类似研究所得出的学校作用（请见 [8]）。有意思的是，再引入各种学校投入和学校过程变量于模型中，发现只有少许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训练以及校长是否观课。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（见表一，增值模型四）。

研究结果亦显示，参与课题的三个地级县市学校对其不同学生群体，尤其是依中考成绩来区分的学生群体，具有区别性效能（differential effect）。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，特别是在几乎所有的中国高中学校是采混合能力编班教学的情况下，可能会掩盖了学校的区别性效能。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，是否学习能力较差的学生正在苦苦挣扎于跟上对其较为困难的学习内容，又或者，是否学习内容对学习能力强学生并没有足够的挑战性。同样地，提供学校在不同学科的区别性增值评量信息，可以协助学校找出是否其整体策略或招聘有不足或出了问题的地方。

目前，后续项目「改进中国教师发展和教育质量」课题，正在先前参与的三个地级县市采集相对等的 2010-2012 数据和信息，随后之分析将提供有关学校校能增值评量的趋势信息。这类信息可进一步提供另一个关键维度，亦即，随着时间的推移学校表现的变化，这也是用来以鉴定新教育改革的作用为目的的关键所在。

表一：六种用于检测学校表现差异性的多水平模型

解释变量		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%

「改进中国教师发展和教育质量」初步结果

有关教师发展议题，基于第一个课题的研究发现，「改进中国教师发展和教育质量」课题正同样地以包括两个互补性的子研究进行相关质性和量性信息的采集。本课题的目的之一，在了解教师发展的主要过程，特别是探讨「专业学习共同体」(professional learning communities, PLCs)与促进中国教师实践和学生学习结果的相关性。目前已完成大部份对利益相关者的个别或团体焦点访谈以及第一年的量性数据采集。

根据初步分析高中校长问卷调查的结果显示，在 98 份回复的问卷中，有 87%的校长表示同意「专业学习共同体」对中国学校而言是个有意义的概念，有 73%的校长表示同意「专业学习共同体」事实上已存在于中国学校。这结果可能反映了地方文化价值在提倡集体学习和实践上起了关键作用。初步比较中国和英国类似的校长问卷调查结果后，也发现一些有趣的现象。譬如，与英国学校相比，似乎教师团队工作、课堂教学研究和观课等活动较常发生在中国学校。然而，和中国学校相较之下，似乎和父母与社区共享信息、将非教学之辅助人员视为专业学习共同体的一份子，以及为个别学生设定目标等活动较常发生在英国学校。另外，诸如模范课堂的存在和习以为常的教师研究论文发表，可能是中国特有的专业学习共同体的面貌之一。将于 2012 年进行的教师问卷调查，会对这些相关议题做更详尽地探讨，并分析专业学习共同体和教师发展对学生学业成就的影响 [4]。

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，此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，独立的区域评价比全国性评价更能反映当地的具体情况。有关「改进中国教育评价和教育质量」课题的研究发现，在别处另有进一步的报导 [4]。但尽管如此，这里要强调的是，参与本文所提及的两个课题仅仅是三个地级县市，很明显地，如中国如

此幅员广大，需要更多的大规模和具代表性的教育效能研究，来探讨包括各个教育阶段对教育质量和效能的贡献 [9] [10]。

参考文献

- [1] IEEQC (2009) Improving Educational Evaluation and Quality in China [W] . Website accessed 22nd September 2010 <http://ieeqc.bristol.ac.uk>
- [2] Teddlie, C. & Reynolds, D. (2000). International Handbook of School Effectiveness research [M] . Falmer Press.
- [3] Scheerens, J. (2001) Monitoring School Effectiveness in Developing Countries [J] . *School Effectiveness and School Improvement*, 12(4):359-384.
- [4] Thomas, S. & Peng, W. J. (2011). Improving Educational Evaluation and Quality in China [R] . Final report to ESRC.
- [5] Peng, W. -J., Thomas, S. M., Yang, X., & Li, J. (2006). Developing school evaluation methods to improve the quality of schooling in China: a pilot 'value added' study. [J] *Assessment in Education: Principles, Policy & Practice*, 13(2):135-154.
- [6] Ma(马晓强), X., Peng, W. -J., & Thomas, S. (2006). 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] *school effectiveness evaluation with value added method : case study of senior high schools of Baoding Hebei Province. 教育研究 Educational Research*, 10:77-84.
- [7] Ding(丁延庆), Y. Q., & Xue(薛海平), H. P. (2009). 高中教育的一个生产函数研究 [J] A study on the education production function with high school data. *华中师范大学学报(人文社会科学版)Journal of Huazhong Normal University(Humanities and Social Sciences)*, 48(2):122-128.
- [8] Thomas, S. (2001). Dimensions of Secondary School Effectiveness: Comparative Analyses Across Regions [J] . *School Effectiveness and School Improvement*, 12(3):285-322.
- [9] Thomas, S. (2005). 运用「增值」评量指标评估英国学校表现 [J] Using indicators of value added to evaluation school performance in UK. *教育研究月刊 Educational Research Journal*, 26(9):20-27.
- [10] Thomas, S. M., & Peng, W.-J. (2009). Enhancing Quality and Capacity for Educational Research [B] . In D. Stephens (Ed.), *Higher Education and International Capacity Building: Twenty Five Years of Higher Education Links* (pp. 51-77): Symposium Books/Bristol papers in Comparative Education.