

DRAFT PAPER - DO NOT QUOTE WITHOUT PERMISSION OF AUTHORS

Using Value Added feedback for Accountability and School Improvement purposes: evidence from China

Sally M. Thomas+, Wen-Jung Peng+, Huisheng Tian# and Jianzhong Li #

+Graduate School of Education, University of Bristol; #China National Institute of Educational Research, Beijing.

Contact: s.thomas@bristol.ac.uk Website: <http://ieeqc.bristol.ac.uk/>

Paper Presented at 11th UKFIET international conference on education and development : Global Challenges for Education: Economics, Environment and Emergency, Theme: Education for good governance and sustainability and Subtheme: Governance, incentives and accountability in education: tackling poor performance and corruption Oxford 13 - 15 September 2011

Abstract

This paper examines the potential use of Value Added measures of school effectiveness and other related data for the purpose of informing and enhancing the impact of strategies to raise educational quality in China. In particular new approaches to accountability and school improvement are explored as a way of tackling poor school and student performance. Evidence is drawn from two recent DFID/ESRC funded projects: “Improving Educational Evaluation and Quality in China” and “Improving Teacher Development and Educational Quality in China”. Qualitative interviews with key stakeholders have been conducted and quantitative evidence of statistically significant differences in senior school effectiveness has been demonstrated in three eastern and western regions. The findings indicate that Value Added measures can provide a valid and relevant measure of educational quality in China, similar to the conclusions drawn from equivalent evidence in other countries such as UK. Moreover, the Value Added concept and measurement approach are seen as a more scientific and welcome addition to current methods of evaluating Chinese schools and teachers, although nevertheless it is essential to take account of local priorities and contexts in China when considering any new evaluation systems. The key question of whether Value Added measures of school effectiveness would be most useful or appropriate within a public accountability framework or as a means to enhance confidential feedback for school self-evaluation and improvement initiatives, or alternatively for both purposes is discussed. The challenges and barriers to introducing new accountability and school improvement measures and reforms in China is also discussed.

Introduction

Many developing countries are worried about the performance of their education systems (Morley & Rassool, 1999) and new improved approaches to school evaluation and pupil assessment are seen key levers to improve educational quality (Gipps, 1998). Largely this is because measures of educational quality and effectiveness are considered to have two critical purposes in raising educational standards. First, to inform accountability mechanisms that hold education systems, schools and teachers accountable for their functioning and performance and also to support democracy in education. Second, to inform improvement

strategies as a mechanism to stimulate improvement in educational outcomes and organisational learning (Scheerens, Glas & Thomas, 2003). The twin strategies of accountability and a continuous focus on improvement, alongside enhanced quality evaluation and student assessment processes, often within a context of decentralisation, are approaches favoured by many western countries aiming to improve educational quality (OECD, 2008). Moreover, valid, reliable and fair student examination and assessment systems play a key role in accreditation - that is to formally regulate desired levels of quality of educational outcomes and provisions - as well as providing critical data to feed into accountability and improvement evaluation systems (Scheerens, Glas & Thomas, 2003).

So irrespective of the particular focus of new educational policy in developing countries in terms of accountability, improvement or accreditation, innovative methods to evaluate quality schooling are needed to provide alternative frameworks for teachers and policy makers to identify best practice in teaching and learning in a variety of contexts. Moreover, similar to the situation in the UK in the early 1990's, alternative approaches are needed to reduce the common practice of evaluating school performance on the basis of raw examination scores (Thomas and Mortimore 1996; Thomas et al, 2007). Using raw performance scores alone, schools with disadvantaged intakes tend to be judged unfairly, while complacency is possible amongst schools with more able pupils, and it is difficult to identify best practice. However, the UK and some other countries worldwide have moved forward in developing value added measures, drawing on School Effectiveness Research (SER) paradigms and methods (Teddlie, & Reynolds, 2000), to provide a fairer and more accurate approach to evaluating school performance than raw examination results. This has involved the establishment of comprehensive and longitudinal datasets and use of sophisticated statistical analysis techniques (multilevel modelling) to create estimates of the relative progress made by pupils in a school, in comparison to pupils in other schools. It is clear that SER has fed directly into identifying new methods to evaluate school performance, and more broadly educational effectiveness at different levels of the system. As a result wide-ranging policy developments in educational evaluation have been introduced in the UK and elsewhere (eg DFES and Ofsted, 2004; Ray, 2006). For example, from 2006 contextualised value added measures have been included in the Department for Education and Skills (DFES) school performance tables for all English schools and school self evaluation is a central element of the national inspection framework. Thus the concept of 'value added' measures of school effectiveness as an indicator of school performance, and related school and educational effectiveness research in the UK, has played a very significant role in focusing the attention of educational policy makers on the potential for raising student achievement. The impact of new approaches to evaluation and accountability has been linked to improved student outcomes (Miliband 2004).

In this paper we examine evidence from China mainland that explores the potential for new educational evaluation systems to enhance educational quality and effectiveness. However, it is important to emphasise that if these methods are to be useful to evaluate quality in different contexts then it is crucial to be aware of both the limitations of the methodology (eg. mainly a quantitative perspective) and the issues of context specificity. In particular, Harber and Davies (1997) have stressed that the contexts for education in developing countries are very different from those of the developed world since they do not have universal enrolment. They contend that, developing countries face little expenditure on education as well as health challenges and poverty which affects capacity to study and learn – especially in rural areas where the majority of people live. It should therefore be acknowledged that there are important priorities in developing countries such as adequate teacher training, school buildings and teaching materials as well as the need to address inequitable distribution of

educational resources and insufficient leadership and expertise to bring about improvements (Stephens, 2007).

Nevertheless, we propose that there is a urgent need for more robust quantitative evidence about the range and extent of school effectiveness in different country contexts, including developing countries. Indeed the application and results of value added methods in delveloping countries may underline important aspects of context specificity such as the underachievement of particular student groups or inequity in the distribution of resources. In other words in spite of key contextual differences, SER methodology remains important as it allows normative within country (or region) comparisons and may help to improve equity in the distribution and provision of quality education through improving evaluation processes, supporting teacher development and self-evaluation and the identification of best practice. Such evidence is vital to inform rapidly changing education reforms and initiatives aiming to raise student outcomes, particularly in rural and socio-economically disadvantaged areas (e.g. DFID Gansu basic education project, Chu and Liu 2005) as well as to promote innovative longitudinal quantitative research methods that will enhance the validity of international comparative research (Goldstein & Thomas, 2008) and increase the SER international knowledge base (Teddlie and Lui 2008).

Policy and Research Context in Mainland China

New educational reforms in China over the last fifteen years have put a strong emphasis on raising educational quality, particularly in rural areas and for girls. As part of these initiatives systematic transformation of curriculum goals, structure and content, teaching and learning approaches, and assessment and administrative structures has occurred and control has increasingly been devolved from the centre to provincial, district and school levels (Chinese Ministry of Education 1996). At the 2009 National People's Congress, the Chinese Premier Wen Jiabao reiterated the need to prioritise educational development and outlined an initial focus on five key areas: (1) promote fairness in education, (2) optimise the education structures to develop vocational education (3) improve the quality of teachers, (4) advance well-rounded education, and (5) implement a program to ensure that all primary and secondary school buildings are safe and promote standardisation in the construction of rural primary and secondary schools (NPC 2009).

The issue of fairness in education is crucial and relates to equal opportunities but also to fair assessment of both students and schools. Therefore one important strand of the new reforms centres on improving pupil assessment as well as the methods used to evaluate quality and improvement in schools. Moreover, this has been an outstanding issue for over fifteen years: in the Ninth Five-Year Plan for China's Educational Development the Chinese government made explicit the need to improve performance measures as well as reduce performance differences between schools (Chinese Ministry of Education 1996). In addition there are particular concerns about the exam-oriented nature of the education system in China, the need for curriculum and examination reform, and the lack of systematic methods evaluate education quality, highlighted by the substantial variations in University admission rates across different areas (Xie 2007; Jiang 2008; Jiang and Ma 2008).

In spite of the Chinese government's desire for new methods to evaluate quality in education, SER is just beginning in China. Raw measures of pupils' academic outcomes and entrance levels to higher education are still viewed as the key indicators of school quality. However, some Chinese researchers have reviewed SER undertaken in the western context (Chen,

2003; Cheng, 1994; Yu, 2005) and others have sought to contextualise the meaning, definition and methodology of school and educational effectiveness in China (Sun & Hung, 1994; Wang & Zheng, 1997). Of the few relevant empirical SER studies that currently exist some studies have examined class and school effects on raw student attainment at one point in time (eg Zhou and Wu, 2008; Xue and Min, 2008; Jiang, Yang and Yao, 2005), but these studies are limited because the data employed are cross-sectional rather than longitudinal (Scheerens, Glas and Thomas 2003). Also due to the small-scale nature of most of these studies and unsatisfactory operational definitions of student ability, Chinese researchers have called for further empirical research, and emphasised the need to introduce prior attainment baseline measures such as Entrance Examination to Senior Secondary School (EESSS) and to systematically establish student databases across cities, provinces and even at national level to enable a fairer evaluation of school effectiveness (Jiang et al. 2005; Tang and Liang, 2005, Sun et al, 2010). This is not the case in Hong Kong and Taiwan where several significant studies of school effectiveness have been conducted (see Cheng, 1999), although given the substantial differences in the education and assessment systems, SER outside mainland China is viewed as a different case.

Mainland China is also largely missing from international comparative studies of school effects (Scheerens 2001), although results for one highly affluent region (Shanghai) have recently been included (OECD, 2011). Moreover, currently there are no contextualised school evaluation criteria or an established indicator system in mainland China, and there is a lack of technical expertise to use and adapt appropriate methodological tools for educational evaluation (Yang & He, 2008). These issues as well as a lack of relevant empirical research on school effectiveness in China to inform policy development (Tang & Liang, 2005; Ding & Xue, 2009) provide the rationale for new research which provides the first rigorous estimates of the range and extent of school effects in China as well as providing a useful model to support the development of new educational evaluation systems.

Improving Educational Evaluation and Quality in China

Two new linked research projects have been funded by the UKAID/ESRC programme to examine the issue of educational evaluation in China (IEEQC 2009), in collaboration with China National Institute for Educational Research, Beijing (CNIER) (see IEEQC, 2009). The first project “Improving Educational Evaluation and Quality in China (IEEQC)” was recently completed and this research has been extended in a new project “Improving Teacher Development and Educational Quality in China (ITDEQC)”. However, this paper focuses on the IEEQC project findings where the aims are twofold. First, to extend current knowledge concerning the definition and measurement of secondary school effectiveness for one student cohort (2009) across a range of regional contexts in China, using a value added approach for different pupil outcomes. Second, to explore how “value added” approaches to evaluating school performance and educational quality may have been adapted and developed by policy makers and practitioners in China to take account of local contexts and priorities. For example, in terms of the type of student learning and outcomes valued such as citizenship versus academic, preference for qualitative versus quantitative approaches, and the kind of contextual features that impact on student and school performance. Broadly the aims of the project are to enhance understanding of school effectiveness in China - but crucially a more fundamental question has been examined - how local context may play a key role in determining definitions of educational effectiveness and quality. Moreover, school effectiveness and improvement research would be very much enhanced by a clearer understanding of why concepts of education quality, and approaches and methods of

educational evaluation largely developed in the UK, Europe and the USA, may be viewed differently by policy makers and practitioners in different social, economic and political cultures such as Mainland China (Teddlie and Reynolds 2000; Scheerens 2001). Comparative evidence of this kind is also essential to understand how contextual and cultural issues may impact on educational values and priorities of both schools and individual students.

Methodology

The IEEQC project involved two complimentary research strands: the first strand collected new qualitative data - interviews and focus groups were conducted with 90+ key stakeholders including headteachers, teachers, students, national and local policy makers - to explore the way educational quality is defined and evaluated in China as well as how international research on school effectiveness, evaluation and self evaluation may have been applied and adapted in the Chinese context. The second strand has investigated school effectiveness in China by using innovative quantitative methodology (multilevel modeling) to analyze examination, prior attainment and other pupil, class and school background data, collected from the 2009 cohort of 90,000+ students in 120+ senior secondary schools in three district education authorities (LEAs) across western and eastern China.

Summary of Findings

The findings from the IEEQC project indicate that although international definitions of educational quality (eg UNESCO, 2004, UNICEF, 2007) are acknowledged in the views of stakeholders, there is very much a concern to take account of and emphasise the specific educational priorities in China where issues of equity and all-round development of students are major quality objectives (Thomas and Peng, 2011). Moreover, some stakeholder's also recognised that current educational evaluation methods in China were not scientific and that improving evaluation processes was a necessary condition for improving policy and practice. With regard to estimates of school effectiveness in China, Table 1 shows a summary of the different models used to explore variations in school performance in 2009, before and after controlling for relevant explanatory factors (an example of explanatory variables employed and detailed results for Value Added Model III are shown in Appendix 1). The findings indicate that in terms of students' raw unadjusted 2009 Total Higher Education Entrance Examination (HEEE) outcome scores, differences between schools account for 24-27 percent of the total variance in student scores, across the 3 LEA investigated (see Table 1 – Raw Model). However, the apparent performance of senior secondary schools changed significantly when comparing raw and value added measures. After controlling for student prior attainment on entry to senior secondary school and other student and school context factors outside the control of the school (Value added Model III), 43 to 57 percent of the total variance and 70 to 91 percent of the school variance in students' 2009 Total HEEE scores was explained. Of the remaining total variance, 5-15 percent was attributable to differences between schools thereby demonstrating a school effect, particularly in western China where the largest school effects were observed. These findings are largely supported by similar comparable results from previous studies in China (Peng et al. 2006, Ma et al. 2006, Ding and Xue, 2009) and in comparison to the UK it seems that at least two Chinese LEAs indicate school effects larger than equivalent results in UK (eg see Thomas, 2001). Interestingly, out of a wide range of school input and process variables additionally tested in the value added models for the three Chinese LEAs, only a few (such pupil-teacher ratio, headteacher training and headteacher observation of class teaching) were found to be statistically significant. These input and process variables were found to explain a further 7-30 percent of the school

variance in students' Total HEEE scores indicating that almost all of the remaining differences between schools were explained (Value Added Model IV). The findings also revealed that in all three LEAs investigated within school differential effects were identified for different groups of students, but especially for groups categorised in terms of prior attainment (see Table 2). These results indicate that the issue of differential school effectiveness may be concealed if only one overall measure is used to evaluate value added performance and is especially pertinent in the Chinese context where almost all senior school students are taught in mixed ability classes. For example, evidence of different value added results within a school – for students of different levels of prior ability - may help schools and teachers identify when less able students are struggling to follow more difficult topics and/or when more able students are not being sufficiently challenged by their academic work. Similarly differential effects by academic subject areas may assist in identifying a lack of whole school policies or particular staffing problems. The follow-up project ITDEQC is currently collecting equivalent examination and other data for 2010-2012 from the same sample LEAs and the subsequent analysis of this data will provide measures of trends over time in schools value added performance. This evidence will provide a further key dimension in terms of examining changes in school performance over time – which is crucial for the purpose of evaluating the impact of new educational reforms.

Table 1: MLM Models used to explore differences in school performance

Explanatory variables		Intra school correlation (Total score)		
		LEA1	LEA2	LEA3
None (Cons)	Raw Model	24.0%	27.2%	25.7%
Prior attainment	Value Added Model I	8.5%	24.3%	19.8%
Student characteristics	Value Added Model II	8.6%	25.1%	24.1%
School context	Value Added Model III	5.1%	13.2%	14.7%
School process School input	Value Added Model IV	1.0%	7.6%	0.0%
Student effort Student attitude	Value Added Model V	1.0%	8.1%	0.0%

Table 2: Differential School Effects across three LEAs: Correlations between school residuals for different student groups (Total HEEE scores) and subject outcomes (2009).

	LEA1	LEA2	LEA3
Gender (girls vs. boys)	0.97	0.96	0.89
Prior attainment:			
middle band vs upper band	0.89	0.83	0.64
middle band vs lower band	0.80	0.72	0.96
upper band vs lower band	0.53	0.47	0.45
Major (sciences vs arts)	0.77	0.76	0.82
Hukou (city vs village)	0.84	0.94	0.90
Academic Subject:			
Chinese vs English	0.72	0.93	0.88
Chinese vs Maths	0.70	0.95	0.81
English vs Maths	0.73	0.93	0.91

Conclusions

In conclusion the IEEQC and ITDEQC projects provide a useful illustration of the kind of large-scale school effectiveness research projects that would be possible if the appropriate datasets were comprehensively available at regional and/or national level in China. One important finding of the research is the apparent differences in the value added model results between regions in China indicating that contextual factors may operate in different ways, particularly for poor rural areas in comparison to affluent urban areas. Consequently, it is possible that separate regional evaluation systems may be more appropriate to reflect local contexts than a national evaluation system. Moreover, the challenges and barriers to introducing new accountability and school improvement measures and reforms in China should not be underestimated given the complex administrative structures in place and the intense focus on raw examination achievement and entrance to higher education. The political will of national and local governments would need to be harnessed in order to reach a consensus and common approach to improving educational evaluation that takes account of local context but also prioritises fairness and equity. The question of whether Value Added measures of school effectiveness if introduced in China would be most useful or appropriate within a public accountability framework or as a means to enhance confidential feedback for school self-evaluation and improvement initiatives, or alternatively for both purposes is also hotly debatable. Useful lessons could be learned from other countries that have introduced one or both of these strategies; for example increasing educational accountability measures has often received the strongest criticism from practitioners and is linked by some research to negative impacts on the quality of student learning (e.g. Gray & Wilcox, 1995; Garipey et al, 2009). However, international contextual and cultural issues also need to be considered so China will need to find the approach most suited to their own unique needs.

Further details of the IEEQC project findings are reported elsewhere (Thomas and Peng, 2011) but nevertheless it is important to emphasise that only three LEAs were examined in this research and it is clear that across a landscape as huge as China, more large-scale and representative educational effectiveness studies are needed, including further examination of the contribution of different levels within the education system to educational quality and effectiveness. (Thomas, 2005; Thomas and Peng 2009).

Note: The research reported in this paper was funded by DFID ESRC research grants RES-167-25-0353 and RES-167-25-0428

References

- Chen(陈丰伟), F.W. (2003). 国外学校效能研究述评 Review on external schools effectiveness. *江西教育科研 Jianxi Educational Research*, 10:33-34&42.
- Cheng, J. K. (1994) Theory and practice of American effective schools, *Global Education*, 3:58–63.
- Cheng, Y. C. (1999) Editorial, Special issue on recent educational developments in South East Asia, *School Effectiveness & School Improvement*, 10(1):3–9.
- Chinese Ministry of Education. (1996). 国家教委关于印发《全国教育事业“九五”计划和2010年发展规划》的通知 Notification of State Council on printing and distributing the Ninth 5-Year Plan for China's educational development and the development outline by 2010. Retrieved 15 July 2009, from http://210.28.182.158/edu/1/law/12/law_12_1082.htm
- Chu, H., & Liu, X. (2005). *Independent Review of the Gansu Basic Education Project (GBEP) on Impacts on Management*: Cambridge Education (CE) and Gansu Provincial Education Department (GPED).
- DFES and OFSTED (2004) *A New Relationship with Schools: Improving Performance Through School Self-evaluation*, Nottingham: the Department for Education and Skills.
- Ding(丁延), Y. Q., & Xue(薛海平), H. P. (2009). 高中教育的一个生函数研究A study on the education production function with high school data. *中范大学学报(人文社会科学版)Journal of Huazhong Normal University(Humanities and Social Sciences)*, 48(2):122-128.
- Garipey, K.D., Spencer, B.L. and Couture, J-C. (Eds.) (2009) *Educational Accountability Professional Voices From the Field*. Rotterdam/Boston/Taipei;Sense Publishers.
- Gipps, C. (1998) Student assessment and learning for a changing society, *Prospects*, XXVIII(1):31–44.
- Goldstein, H., & Thomas, S. M. (2008). Reflections on the international comparative surveys debate. *Assessment in Education: Principles, Policy & Practice*, 15(3):215 - 222.
- Gray, J. & Wilcox, B. (1995) *Good school, bad school: evaluating performance and encouraging improvement*. Buckingham: Open University Press.
- Harber, C. and Davis, L. (1997). *School Management and Effectiveness in Developing Countries: the post-bureaucratic school*, London: Cassell.
- IEEQC (2009) *Improving Educational Evaluation and Quality in China*. Website accessed 20th July 2009 <http://ieeqc.bristol.ac.uk>
- Jiang(姜世健), S. J. (2008). 高考移民现象的成因分析Institutional analysis of the phenomenon of the immigrants for NCEE. *教育展研究Research in Educational Development*, 17:16-20&30.
- Jiang(蒋), W., & Ma(马照云), Z. Y. (2008). 中国高考移民第一案追 Inquiry of the first litigation case of immigration for Entrance Examination for Higher Education. *政府法制Government Legality*, 12:16-17.
- Jiang(蒋莉), L., Yang(杨志明), Z.-M., & Yao(姚), S.-Q. (2005). 学生高考文成影响因素的多性分析Analysis of influencing factors on the Chinese achievement test in the college entrance examination: a hierarchical linear model. *中国临床心理学杂志Chinese Journal of Clinical Psychology*, 13(4):414-416&419.

- 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 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.
- Miliband, D. (2004). Using data to raise achievement. Speech at a Conference of the Education Network, London, 11 Feb 2004. Retrieved 21 May 2007, from http://www.teachers.gov.uk/_doc/6280/Miliband%20speech%20at%20TEN%20Conference%2011Feb04.doc
- Morley, L. and Rassool, N. (1999). *School Effectiveness: Fracturing the Discourse*. Falmer Press.
- NPC (National People's Congress, 全国人民代表大会). (2009). 温家宝作政府工作报告：2009年3月5日在第十一届全国人民代表大会第二次会议上the Report on the Work of the Government delivered by Premier Wen Jiabao at the Second Session of the Eleventh National People's Congress on March 5, 2009 Retrieved 14 June 2010 <http://www.xinhuanet.com/2009lh/090305a/wz.htm>
- OECD (2008). *Measuring Improvements in Learning Outcomes: Best practices to access the value added of schools*. www.oecd.org/publishing/corrigenda.
- OECD (2011). Chapter 4: Shanghai and Hong Kong: two distinct examples of education reform in China. In *Lessons from PISA for the United States: Strong Performers and Successful Reformers in Education* (pp 83-115). <http://www.oecd.org/dataoecd/32/50/46623978.pdf>.
- 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. *Assessment in Education: Principles, Policy & Practice*, 13(2):135-154.
- Ray, A. (2006). School value added measures in England: A paper for the OECD project on the development of value-added models in education systems: Department for Education and Skills.
- Scheerens, J. (2001) Monitoring School Effectiveness in Developing Countries. *School Effectiveness and School Improvement*, 12(4):359-384.
- Scheerens, J., Glas, C. & Thomas, S. (2003) *Educational Evaluation, Assessment and Monitoring. A systemic approach*. Swets & Zeitlinger Publishers.
- Stephens, D. (2007). *Culture in education and development: principles, practice and policy*: Bristol Papers in Education: comparative and international studies. Symposium books.
- Sun(孙涛), M. T., & Hung(洪哲), Z. (1994). 学校效能初探The initiative study of school effectiveness. *教育与经济 Education & Economics*, 3:1-5.
- Sun, H. C. (孙河川), Wang, X. D. (王小栋), Yuan, S. F. (袁书凤), Hao, L. L. (郝玲玲), Liu, W. Z(刘文钊). and Liu, Y. (刘颖) (2010) 中国教育效能研究的近况、不足、发展、展望，发表于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年10月22-23日，沈阳，中国 ‘An Overview of the Status Quo, Weakness and Development of the Educational Effectiveness Research in China’, paper presented at the 2nd International Conference on School Effectiveness and School Improvement and the 3rd Meeting of the Educational Effectiveness Academic Committee in China, 22-23 October 2010, Shenyang, China.
- Tang(林春), L. C., & Liang(梁玲玲), L. L. (2005). 学校效能价值的探索性研究 on school effectiveness using value added method. *上海教育科研 Shanghai Research on Education*, 4:24-26.

- Teddlie, C. & Reynolds, D. (2000). *International Handbook of School Effectiveness research*. Falmer Press.
- Teddlie, C., & Liu, S. (2008). Examining teacher effectiveness within differentially effective primary schools in the People's Republic of China. *School Effectiveness and School Improvement*, 19(4):387-407.
- Thomas, S. (2001). Dimensions of Secondary School Effectiveness: Comparative Analyses Across Regions. *School Effectiveness and School Improvement*, 12(3):285-322.
- Thomas, S. (2005). 运用「增值」评量指标评估英国学校表现 Using indicators of value added to evaluation school performance in UK. *教育研究月刊 Educational Research Journal*, 26(9):20-27.
- Thomas, S. & Peng, W. J. (2011). *Improving Educational Evaluation and Quality in China*. Final report to ESRC.
- Thomas, S., Peng, W. J., & Gray, J. (2007). Modelling patterns of improvement over time: value added trends in English secondary school performance across ten cohorts. *Oxford Review of Education*, 33(3), 261 - 295.
- Thomas, S. M., & Peng, W.-J. (2009). Enhancing Quality and Capacity for Educational Research. 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.
- Thomas, S. and Mortimore, P. (1996). Comparison of Value Added Models for Secondary School Effectiveness. *Research Papers in Education*, 11(1): 5—33.
- UNESCO (2004). *EFA Global Monitoring Report 2005: Education for All. The quality imperative*: Paris.
- UNICEF (2007). *Education Quality Framework*. Accessed on 12th February 2009. http://www.unicef.org/infobycountry/china_statistics.html.
- Xie(秀英), X. Y. (2007). “高考移民”政策限制的合理性分析 On the rationality of policy restriction of "migration for the National College Entrance Examination". *西青年管理干部学院学 Journal of Shaanxi Institute of Junior Managerial Personnel*, 20(1):27-29&35.
- Xue(薛海平), H.-P., & Min(方), W.-F. (2008). 中国西部教育生 函数研究 A study on educational production function in western regions of China. *教育与 Economy*, 2:18-25.
- Yang(会芹), H.-Q., & He(何俊), J.-H. (2008). 多 性模型在心理和教育 向研究中的运用 the application of hierarchical linear model in longitudinal research on the study of education and psychology. *石家庄学院学 Journal of Shijiazhuang University*, 10(3):71-72&102.
- Yu(俞继凤), J.F. (2005). 西方国家学校效能研究的反思及其未来发展 The introspection and future development of school effectiveness research of western countries. *外国教育研究 Studies in Foreign Education*, 6:1-4.
- Wang(王新如), X.R. & Zheng(郑文), W. (1997). 谈学校组织文化与学校效能 Introduction to organisational culture and effectiveness of schools. *教育科学 Education Science*, 3:53-57.
- Zhou(周皓), H., & Wu(巫), X. W. (2008). 流 儿童的教育 效及其影响因素: 多 性模型分析 School performance of migrant children and its determinants: a hierarchical linear model analysis. *人口研究 Population Research*, 32(2):22-32.

Appendix 1: Example of detailed multilevel modeling findings for LEA2 - Total Entrance Examination to Higher Educational (EEHE) Score: Value Added Model III (2009)

	Total EEHE score estimate	Standard Error	
Fixed Part:			
Cons	1.06	0.32	*
Prior attainment:			
Chinese (zscore)	0.04	0.01	*
English (zscore)	0.33	0.01	*
Math (zscore)	0.18	0.01	*
Female (baseline: boy)	0.02	0.01	
Age (zscore in month)	-0.09	0.00	*
Major Arts (baseline: sciences)	-0.09	0.01	*
Hukou registration place (baseline: city):			
township/county	0.10	0.02	*
village	0.13	0.02	*
Tuition fee status (baseline: normal):			
extra full tuition fee	-0.37	0.02	*
full/partial scholarship	0.17	0.04	*
Lived with (baseline: boarder): parents	-0.10	0.02	*
Home school time (baseline: < 15 mins):			
15-30 mins	-0.06	0.02	*
0.5-1 hrs	-0.04	0.02	*
1-2 hrs	-0.03	0.02	
JHS status (baseline: ordinary):			
township/county key school	0.03	0.01	*
city key school	0.01	0.01	
provincial/national model school	0.09	0.02	*
Studied in this school (baseline: SHS Year1):			
since SHS Year2	-0.18	0.05	*
since SHS Year3	-0.10	0.03	*
Father education (baseline: Primary Sch):first degree	0.07	0.03	*
Father occupation (baseline: unit head):			
agriculture and water labour	0.10	0.03	*
production, transport operators	0.09	0.03	*
migrant worker	0.09	0.03	*
teacher	0.11	0.03	*
unemployed	0.10	0.04	*
Mother education (baseline: Primary Sch):			
junior high school	-0.07	0.01	*
SHS, secondary, vocational	-0.07	0.02	*
non-university tertiary	-0.01	0.03	
master and above	0.19	0.07	*
Mother occupation (baseline: unit head):			
agriculture and water labour	0.23	0.04	*
employed household	0.07	0.04	*
Teacher	0.18	0.04	*
unemployed	0.08	0.04	*
Things at home (baseline: no):			
yes - a room of your own	-0.11	0.01	*
yes - computer	0.00	0.01	
yes - mobile phone of your own	-0.30	0.01	*
yes - car (automobile)	-0.07	0.01	*
yes - recorder/CD player/MP3	0.05	0.01	*
yes - colour TV	0.05	0.03	
yes - motorcycle	-0.04	0.01	*
Books at home (baseline: none): >200books	0.08	0.03	*
% student major in art	-0.02	0.00	*
% father education higher than SHS	0.01	0.01	
Zscore of school mean in 3 subjects	-0.04	0.06	
Random Part:			
Level: school	0.09	0.02	*
Level: student	0.56	0.00	*
total variance	0.65		
intra school correlation (x100)	13.17		
% school variance explained	72.20		
% total variance explained	42.60		

Table Notes: School sample size = 54; Student sample size 30,090; Value added Model III included the same fixed set of explanatory variables in the analysis for all 3 LEAs, however only explanatory variables with statistically significant estimates at 0.05 (denoted as *) for at least one LEA are included in this table; All outcome and prior attainment measures have been transformed to normal scores within each LEA cohort (mean=0; standard deviation =1); squared and cubed terms for the prior attainment measures were also included in the model but the results are not presented here.