

Impact case study (REF3b)

Institution: University of Bristol
Unit of Assessment: 25 – Education
Title of case study: Bristol research leads to better ways of evaluating schools and promoting learning, achievement and improvement in the UK and internationally
<p>1. Summary of the impact (indicative maximum 100 words)</p> <p>Since 2008, UK and overseas policies, practices and tools aimed at evaluating and promoting quality in schools and supporting student learning, attainment and progress have been profoundly influenced by research conducted at the University of Bristol. The work began in 2001 in the Graduate School of Education; from 2005, the School's efforts were complemented by those of the Centre for Multilevel Modelling. The research has generated original knowledge about school performance measures and school, teacher and context factors which promote student learning. This knowledge has transformed government and institutional policies and practices. New improved methods of evaluating schools and interventions in education (and other sectors) have been demonstrated and widely disseminated, thereby enhancing public understanding of institutional league tables and facilitating the scaling-up of new approaches nationally. The development of statistical methodology and MLwiN software and training has enabled more rigorous and sensitive quantitative analysis of educational datasets around the world, as well as wider take-up of this methodology by non academics.</p>
<p>2. Underpinning research (indicative maximum 500 words)</p> <p>The research comprises studies of educational quality, effectiveness and improvement pioneering innovative “value added” measures of school performance to report original knowledge on the nature and extent of school effectiveness in a range of contexts. Sophisticated (multilevel) methodology and software for statistical modelling (MLwiN), have also been developed, extended and utilised to provide new evaluation tools and substantive findings. The research has involved the creation of new and detailed longitudinal datasets in the UK and overseas (including measures of student academic and attitude outcomes) to analyse, measure and evaluate school performance as well as the influence of other levels within education systems (eg regions, within school departments). Thomas and Peng (University of Bristol staff since 2001), Goldstein and Steele (Bristol staff since 2005), Rasbash (Bristol staff 2005-10) and Leckie (Bristol staff since 2009) have built upon previous studies by extending earlier datasets, analyses and research. For example, the China 2009-2012 and Lancashire 1992-2006 large-scale datasets were unique in collecting new regional student attainment and related information over 4-14 consecutive cohorts for the first time. The latter enabled time trends in value-added school performance to be examined over a longer period (14 years) than any other studies worldwide. The UK Department for Children, Schools and Families (DCSF) national pupil database (NPD) introduced in 2002 has also been extensively employed, as well as other national surveys.</p> <p>Educational effectiveness and improvement underpinning research since 2001 includes:</p> <ul style="list-style-type: none"> • UK and/or overseas evidence of internal variations in school effectiveness (eg subject/departmental effectiveness; differential effectiveness for different pupil groups or curriculum stages), the impact of pupils moving schools and time trends of school effects, as well as evidence that national and regional differences exist in terms of school effectiveness. This demonstrates that effectiveness is best seen as a feature that is outcome-, context- and time-specific and indicates that school league tables have little to offer as guides to school choice. (Thomas, Goldstein, Leckie, Peng) [3][4][5][6]. • Evidence of the extent to which school input, process and context factors link to school effectiveness in China (Thomas, Peng) [6] and the impact of school resources and parental divorce on pupil attainment (Steele), thereby highlighting relevant factors that need to be considered when evaluating schools. <p>Methodological and software development underpinning research since 2005 includes:</p> <ul style="list-style-type: none"> • New equations have been created for predicting the group effects in repeated cross-section multilevel models in order to predict accurately how schools are likely to perform in future. In addition, new (simulation-based) graphical approaches have been developed for communicating the statistical uncertainty of predicted group effects in multilevel models. These have helped in conveying to parents whether the academic performances of several local schools can be statistically separated from one another (Leckie, Goldstein) [3]. • Multilevel models for complex, non-hierarchical data structures have been developed to

Impact case study (REF3b)

model correctly the effects of schools and neighbourhoods on pupils' academic progress when students are changing schools and moving neighbourhoods during their schooling (Leckie). Multilevel models for segregation and inequality have also been developed, for example to measure the extent to which social or ethnic segregation of students across schools has significantly changed over time (Leckie, Goldstein).

- Modelling of multivariate data with different response types at several levels and procedures for handling correlated measurement and misclassification errors. (Goldstein) [1][2].

3. References to the research (indicative maximum of six references)

- [1] Goldstein, H. (2010) *Multilevel Statistical models*. 4th Edition. Wiley. [Citations 2008-13: 2,480] Listed in REF2
- [2] Rasbash, J., Steele, F., Browne, W.J. and Goldstein, H. (2012) *A User's Guide to MLwiN*, v2.26. Centre for Multilevel Modelling, University of Bristol. [Citations 2008-13: 1,070] <http://www.bristol.ac.uk/cmm/software/mlwin/download/2-26/manual-web.pdf>
- [3] Leckie, G. and Goldstein, H. (2009) The limitations of using school league tables to inform school choice, *Journal of the Royal Statistical Society: Series A*, 172, 835-851. Listed in REF2
- [4] Thomas, S.M., Peng, W.-J., Gray, J. (2007) Value added trends in English secondary school performance over ten years. *Oxford Review of Education*, 33 (3: 261 - 295). <http://dx.doi.org/10.1080/03054980701366116>
- [5] Thomas, S. (2001) Dimensions of Secondary School Effectiveness: Comparative Analyses Across Regions, *School Effectiveness & School Improvement Journal*. Vol 12(3): 285-322. <http://dx.doi.org/10.1076/sesi.12.3.285.3448>
- [6] Thomas, Sally et al (2012) 学校效能增值评量研究. [Research on Value Added Evaluation of School Effectiveness]. Jiaoyu yanjiu. [Educational research]. 33 (7), pp. 29-35 Beijing: Zhongyang Jiaoyu Kexue Yanjiusuo. Listed in REF2

Related research grants supporting and evidencing quality of publications

These were awarded, for example, by the ESRC and the Department for International Development following a rigorous process of review by the respective agencies.

- Steele, F., Goldstein, H. and Leckie, G. (2011-2013) *Longitudinal Effects, Multilevel Modelling and Applications (LEMMA III)*, ESRC: 750K
- Steele, F., Goldstein, H. and Leckie, G. (2008-2011) *ESRC: STRUCTURES for Building, Learning, Applying and Computing Statistical Models (LEMMA II)*, ESRC: 700K
- Rasbash, J., Steele, F. and Thomas, S. (2005-2008) *Learning Environment for Multilevel Modelling Applications (LEMMA)*, ESRC: 650K
- Goldstein, H. (2003-2005) *Developing Multilevel Models for Realistically Complex Social Science Data*, ESRC: 300K
- Thomas, S. and Peng, W. -J. (2010-2013) *Improving Teacher Development and Educational Quality in China [ITDEQC]*, ESRC/DfID: 500K
- Thomas, S. and Peng, W.-J. (2008-2011) *Improving Educational Evaluation and Quality in China [IEEQC]*, ESRC/DfID: 250K
- Thomas, S. and Peng, W. -J. (2001-2007) *Lancashire LEA: Value Added Project* (this was a continuation of a project that began in 1992 and moved to the University of Bristol): 200K
- Thomas, S. et al (2002-2004) *Effective professional learning communities*, DfES: 600K

4. Details of the impact (indicative maximum 750 words)

The impact of this research on a wide range of beneficiaries (policy, practitioner, NGO, public) worldwide has reach and is significant in two ways since 2008. First, it informed and underpinned new policy and, practice. Second, it has generated methodological developments in key areas. Each of these is outlined below.

Impact on UK and international policy and government thinking relating to measuring educational effectiveness and school performance

In the UK, Goldstein and Thomas' research [1][3][4][5] has contributed evidence to inform and influence key national policies such as the utility of school self-evaluation, national pupil databases (eg the Pupil Level Annual Schools Census (PLASC)), contextualised value-added measures of school performance (introduced by the DCSF in 2006 and almost identical to measures used in Lancashire LEA, 1993-2006) and separate value-added measures for different student groups (introduced by the DfE in 2011). The research has also promoted the use of a

Impact case study (REF3b)

wider range of outcomes and measures by the DfE/DCSF/OFSTED/LSC [a][b]. Goldstein's research was referenced as underpinning evidence in a 2012 Northern Ireland Assembly Research and Information Service Research Paper, "Providing information on pupil and school performance". He was also a member of the UK government select committee invited seminar to advise on Accountability and League Tables (2013). In addition, Goldstein co-directs the PLASC Users Group, set up in 2006 with support from the DfE. Since then regular meetings have been held with 40-plus participants, involving researchers who have used, or are interested in using the PLASC/NPD datasets. Civil servants from the DfE also attended and frequently returned to report on current developments and participate in discussions [c].

Advice on evaluating educational quality has also been frequently sought by policymakers internationally, which demonstrates the reach and significance of this work. This has resulted in citations in OECD publications that provide guidance to member states [a], as well as invitations to speak in many international contexts, often introducing new ideas on school evaluation to non-academics for the first time and elucidating the input, process and context factors associated with school effectiveness (eg Goldstein (2011) Queensland University of Technology [attended by Australian Government officials]; Thomas (2010) Chilean Ministry of Education; Thomas (2008) EU education conference for the French Presidency).

Impact on UK and international educational and school practices and public understanding relating to evaluating educational quality, improving school effectiveness and best practice in school self-evaluation and use of data

Leckie and Goldstein's research [3] demonstrates the limitations of using the government's school league tables to inform school choice. Since 2008 this has promoted stakeholders' and the public's understanding of the problems with league tables through widespread national and international communication to non-academics via popular articles and other media, including interviews for the BBC Radio 4 programmes "Analysis" and "The Learning Curve", and articles in the Financial Times, the Daily Telegraph and the Times Education Supplement. Goldstein, Leckie and Thomas' work on critiquing school performance measures demonstrates impact in terms of both reach and significance. It has been cited by numerous UK NGOs (eg NUT, RSA, RSS, the Institute for Government) and overseas NGOs and governments seeking to evidence the complexity, dangers and limitations of school performance measures; thereby influencing public thinking and new policy development on educational accountability and improvement initiatives [e][f][g].

Pilot school evaluation studies using value-added techniques have been conducted in UK and several countries worldwide (eg China, Africa) [3][4][5][6] and this has raised the awareness of policymakers and teachers and resulted in new evaluation practices by schools [d]. Professor Xiaoman Zhu, President of the National Institute of Education Sciences (NIES) until 2010, Ministry of Education, Beijing, has emphasised the contribution of the IEEQC research [6] and collaboration between the University of Bristol and NIES to better understanding the concept of educational quality and evaluation methods in the Chinese context, as well as to capacity-building for NIES researchers [h]. Mr Xiaoqiang Ma, NIES and IEEQC project researcher has gone on to publish a 2012 book, "Value added evaluation: a new perspective on school evaluation". The head teacher of a Chinese senior secondary school participating in the IEEQC project stated that "This approach [value added method] is particularly good. From next year [2009], we will use this approach to evaluate our key schools. That is, taking account of the intake of senior high school year 1 when comparing schools with college entrance examination results. We [as a school] particularly welcome the method" [i]. Thomas has also applied methods to evaluate schools' performance alongside other key elements in creating and sustaining English schools as professional learning communities (PLCs), resulting in new tools used by school leaders to develop their schools as PLCs.

Expansion of use of quantitative methods in educational research and social sciences more broadly, which in turn shapes research that influences policy and practice

The impact of new statistical methodology [1][2] has been achieved through further development of the user-friendly MLwiN software, the REALCOM-Impute software for multiple imputation and through dissemination and training events. Since 2008 the MLwiN software (available free to UK academics), together with extensive user guides, has been downloaded by 3,846 new users and it has been purchased by 5,518 overseas academics and 613 non-academic users. Moreover, 67 organisations have purchased MLwiN site licences (50 users) since 2008; of these 8

Impact case study (REF3b)

organisations hold extension licences (250 users).

The CMM website is widely acknowledged as the premier resource for research and training in multilevel modelling. There are around 1,100 page-loads and 360 unique visitors per day (65% from outside the UK). The LEMMA Virtual Learning Environment (launched in April 2008) has around 10,000 registered users, of whom 70% are international and 14% are non-academic, thereby demonstrating the reach and significance of the research impact. Some training events are targeted at non-academics, eg a session on multilevel modelling given at Ofsted in 2008 (Steele). UK non-academic beneficiaries and users include the Departments of Education, Health, and Work and Pensions, the Scottish Executive and the Office for National Statistics [b]. For example, Trevor Knight (consultant statistician to DfE) reported in July 2010 that MLwiN was used by DfE statisticians to calculate Contextual Value Added (CVA) and other value-added school performance measures, employed as an integral part of the OFSTED school inspection process and used to construct the Learning Achievement Tracker – a tool for schools and FE colleges to appreciate progress made by students since the end of compulsory schooling. MLwiN was also used in the National Evaluation of the Sure Start Local Programmes for the DfE [j], the NatCen (2009) report for the Department for Environment, Food and Rural Affairs on educational attainment in rural areas and by Higher Education Funding Council for England and others in conducting new analyses to support higher education institutions in developing “contextualised” admissions policies and equality and diversity policy for REF2014 submissions. Overseas non-academic MLwiN users include Statistics Canada, Statistics Norway, the Netherlands Bureau of Statistics, UNESCO and the World Health Organisation.

5. Sources to corroborate the impact (indicative maximum of 10 references)

- [a] Evans, H. (2008) [Value-Added in English Schools](#). A DfE paper updated from the OECD Project on the Development of Value-Added Models in Education Systems and OECD (2008) [Measuring Improvements in Learning Outcomes: Best Practices to Assess the Value-added of Schools](#). Organisation for Economic Co-operation and Development and [Spanish translation](#) cites work by Goldstein and Thomas regarding use/methodology of VA measures.
- [b] Director General, Monitoring and Assessment, UK Statistics Authority provided information (June 2013) about influence of University of Bristol research on government and public understanding and UK policy on school evaluation.
- [c] [PLUG website](#) lists 9 seminars presented by non-academic/government researchers hosted by PLUG since 2008.
- [d] Executive Headteacher, Bradley Stoke Community School and Abbeywood Community School provided information (September 2013) about influence of University of Bristol research on teachers understanding and good practice in student and school evaluation and links to improved student outcomes.
- [e] Wildman, R (2011) [Beware of the Misleading Means and Measures](#). Chapter 3 in Transformation Audit (The Inclusive Economies Project). Policy and Analysis Unit of the Institute for Justice and Reconciliation (IJR), South Africa cites work by Leckie & Goldstein on dangers of school league tables. IJC is dedicated to researching and influencing policy debates around the issue of socio-economic justice in South African and elsewhere on the continent.
- [f] Mulgan, R. (2012) [Transparency and Public Sector Performance](#). Report prepared for the Australia and New Zealand School of Government cites work by Leckie & Goldstein on dangers of school league tables.
- [g] Cipollone, P. et al. (2010) [Value-Added Measures in Italian High Schools: Problems and Findings](#). Bank of Italy Temi di Discussione (Working Paper) No. 754. cites work of Thomas regarding use and methodology of value added measures.
- [h] 2011 Celebration Book on 70th anniversary of National Institute of Educational Sciences, Ministry of Education, Beijing 70周年所庆纪念文集. Text emphasises the influence of University of Bristol research on understanding school evaluation in Chinese context (pg 42).
- [i] ESRC Research Impact Evaluation Report of Improving Educational Evaluation and Quality in China (IEEQC) project: overall rating “Outstanding”, January 2013.
- [j] NESS Team (2010). [The impact of Sure Start Local Programmes on five years olds and their families](#). DfE Research Report RR067. states use of multilevel modeling analysis in conducting the evaluation (page 23)