Institution: University of Bristol

Unit of Assessment: 23) Education

Title of case study: Education policy, practice and school accountability are improved by multi-level models of school performance in UK and internationally

Period when the underpinning research was undertaken: 2009 – 2020

Details of staff conducting the underpinning research from the submitting unit:

Name(s):	**Role(s) (e.g. job title):**	**Period(s) employed by submitting HEI:**
George Leckie	Professor of Social Statistics	10/2009 to present
Sally Thomas	Professor of Education	01/2001 to present
Harvey Goldstein	Professor of Social Statistics	10/2005 to 2020

Period when the claimed impact occurred: 1st August 2013 - 2020

Is this case study continued from a case study submitted in 2014? No

1. Summary of the impact

University of Bristol (UoB) research using multilevel modelling methodology to analyse school value-added performance measures and characteristics of student progress has improved accuracy and objectivity of educational accountability systems as well as strategies for school improvement. The research has informed the UK Department for Education (DfE) 'Progress 8' evaluation methodology and shaped the way 'Progress 8' and 'looked after' children's attainment information is communicated and reported in England, as well as informing the planning of international education evaluation systems. New improved methods of evaluating and communicating student and school performance have been developed, demonstrated and widely disseminated, thereby enhancing practitioner, media and public understanding of the benefits and limitations of school league tables; the links between key explanatory factors such as a student's looked after status and their outcomes; and the international take-up for best practice in measuring school performance.

2. Underpinning research

Value-added measures of school performance attempt to compare pupil *progress* over a given period of schooling and are widely considered fairer and more meaningful than pupil *attainment*, when assessing school effectiveness and accountability. However, the statistical methodology underpinning *progress* measures, such as DfE's 'Progress 8', can dramatically influence outcomes and result in misleading indications of school performance. Since 2001, state-of-the-art 'value-added' models and school performance measures have been pioneered by UoB researchers, using the DfE National Pupil Database and other national and international administrative and survey datasets [6].

In 2011, Leckie and Goldstein introduced new multilevel modelling extensions to school value-added models to identify the extent to which current measures can be used to predict the future performance of schools. The results suggested that published school league tables do not contain enough information to predict the future performance of schools with any degree of precision and questioned their value for school choice purposes [3]. The study also identified the need to make school performance tables clearer and developed new simulation-based methods for both calculating and graphically communicating statistical uncertainty surrounding school value-added scores [3]. Funding from the ESRC [i] also supported a critical examination of the evolution of school performance measures and league tables in England and their high-stakes role in holding schools to account and informing parental school choice [2]. This revealed the statistical strengths

and weaknesses of each measure and highlighted the potentially negative consequences and gaming behaviour they generate, as well as emphasising the role of Government ideology on changes to the official league tables over time [2].

The introduction of the current 'Progress 8' measure in 2016, which ignores school differences in socio-demographics, necessitated an urgent evaluation of the strengths and weaknesses of the statistical methodology. In 2019, Leckie and Goldstein, with further funding from ESRC [ii], showed that the system unfairly penalises schools with educationally disadvantaged intakes and illustrated this by creating 'Adjusted Progress 8' scores [1]. The research argues that the Government should adjust 'Progress 8' for pupil background and that doing so would improve school accountability and choice in England [1]. The fact that over a fifth of schools would move over 500 places in the national league tables and 40% of schools would shift out of the Government's "underperforming" category, emphasises the significance of this finding for current education policy.

Multilevel modelling methodology also facilitated the exploration of factors influencing the educational progress of children who are, or have been, in care; one of the lowest performing groups in terms of educational outcomes. Funded by the Nuffield Foundation [iii] and carried out in collaboration with the University of Oxford, the research linked for the first time, pupil test score data in the National Pupil database, to background information contained in the DfE Looked After Children Database. The analysis revealed, less progress, as well as a greater variation in school value added performance for children in need and looked after, compared to other children [5].

Methodologies and knowledge developed for UK school performance measures, have also been applied by Thomas (ESRC/DFID funded [iv]), to investigate school effectiveness in China. Analyses of Chinese senior secondary schools' value-added performance over four consecutive cohorts reveal significant differences in schools' raw and value-added performance in the university entrance examination results ("Gaokao"), and that these measures differ across regions, cohorts and subject outcomes [4]. Differences were also found in terms of school effects in comparison to teacher effects, as well as differential within-school effects for different student groups. The research demonstrated the need for a variety of separate value-added measures to reflect the full complexity of school performance and to better to inform both teachers' reflective practice and school improvement initiatives. Furthermore, positive links were identified between teacher professional development factors and students' progress [4].

3. References to the research

[1] **Leckie G.** & **Goldstein H.** (2019). The importance of adjusting for pupil background in school value-added models: A study of Progress 8 and school accountability in England. *British Educational Research Journal,* 45, 518-537. DOI:10.1002/berj.3511 [5 citations]

[2] **Leckie G.** & **Goldstein H.** (2017). The evolution of school league tables in England 1992-2016: 'contextual value-added', 'expected progress' and 'progress 8'. *British Educational Research Journal*, 43, 193-212. DOI:10.1002/berj.3264. [53 citations; Winner of 2018 BERJ Editors' Choice Award'; Top 20 most read paper in 2017-2018]

[3] **Leckie G.** & **Goldstein H**. (2011). Understanding uncertainty in school league tables. *Fiscal Studies*, 32, 207-224. DOI:10.1111/j.1475-5890.2011.00133.x [58 citations]

[4] **Thomas S.** (2020). Chapter 9: Educational Effectiveness Research Example from Asia: School and Teacher Value Added Performance and the Relationship with Teacher Professional Development in Mainland China. In J. Hall, P. Sammons, & A. Lindorff (Eds.), International Perspectives in Educational Effectiveness Research Springer.

[5] Sebba J, Berridge D, Luke N, Fletcher J, Bell K, Strand S, **Thomas S**, Sinclair I, & O'Higgins A. (2015). The Educational Progress of Looked After Children in England: Linking Care and Educational Data. Nuffield Foundation [135 citations]

[6] **Leckie G**. & **Goldstein H.** (2009). The limitations of using school league tables to inform school choice. *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 172(4), 835-851. https://www.jstor.org/stable/20622553 [212 citations]

Research grants

[i] **Leckie G.** Multilevel modelling of the Government's new school performance measures, 'Floor Standards' Target and 'Narrowing the Gap' Priority, ESRC, 2013 – 2015, GBP150,601

[ii] **Leckie G.** & **Goldstein H**. How should we measure school performance and hold schools accountable? ESRC, 2018 – 2021, GBP379,433

[iii] Sebba J, Berridge D, **Thomas S**, & Strand S. The Educational Progress of Looked After Children in England: Nuffield Foundation, 2014 – 2015, GBP150,000

[iv] **Thomas S.** & Peng WJ. Improving Educational Evaluation and Quality in China, ESRC/DfID, 2008 – 2011, GBP217,346

[v] **Thomas S** & Peng WJ. Improving Teacher Development and Educational Quality in China (ITDEQC), ESRC/DfID, 2010 – 2014, GBP428,397

4. Details of the impact

Governments in the UK and around the world are increasingly seeking to measure school performance for accountability and improvement purposes, as important levers in raising educational standards. Since 2001, UoB research has identified, and drawn attention to, statistical weaknesses in these measures, highlighting best practice and informing public debate. In the current REF period, this research has; shaped the way 'Progress 8' and 'looked after' children's attainment is reported in England, enhanced practitioner, media and public understanding of school league tables and informed the international take-up for best practice in measuring school performance, thereby improving education policy, practice and school accountability in the UK and internationally (earlier impacts were presented in a REF2014 case study).

Informing and improving school evaluation policy in England

UoB research has '*played an important part in informing and guiding the DfE in the changes and improvements they have made to their tables'* [A]. In 2016, the latest 'Progress 8' school performance measure adopted a range of key features advocated by UoB research [1, 2, 3], including making progress measures the primary measure for accountability, adjusting flexibly for prior attainment, reporting differential value-added scores for different pupil groups, enabling bespoke parent-focused school comparisons, adding advice on predicting future school performance, and improved reporting of 95% confidence intervals with colour coding [A].

The roll out of a new assessment system for primary schools in 2016 sparked widespread debate. The close link between assessment and accountability created a high stakes system with potentially negative impacts for teaching and learning. Goldstein presented research [2], in favour of lowering the stakes associated with school value-added results, at the 2017 Education Select Committee meeting on Primary Assessment [B] which was described as 'invaluable and insightful' by the Chair. The Committee went on to recommend that '*The Government should change what is reported in performance tables to help lower the stakes associated with them'* [B].

Leckie and Goldstein also presented their Progress 8 research [1], to the 2019 Education Select Committee [H], and in related meetings with Emma Hardy (MP for Manchester Central), the DfE (Nov 2017, Apr 2019) and Ofsted (Sep 2018, Feb 2019). These meetings drew attention to problems with the way Progress 8 as it stands, reflects differences in pupil intake, skewing perceptions of schools unfairly, and alternatively proposing 'Adjusted Progress 8' scores [1]. Emma Hardy MP noted that she would '*continue to push this report, and use it in discussions with Education Ministers'* and that the research ensured a '*fairer and deeper understanding of what*

makes a good school' [H]. Collaboration with the Northern Powerhouse Partnership (NPP) [H, I] has further drawn attention to these problems, with MP Lucy Powell stating that this *"research partnership will assist Parliamentarians, including those of us on the Education Select Committee, to hold the government to account for their chosen narrow accountability measure, and to campaign for change"* [H]. This research [1] has also helped inform Ofsted's understanding of limitations of school performance measures, and its recent move (2018) to place less emphasis on the official performance measures in their own ratings [E]. In Bristol, the Cathedral Schools Trust Executive Principal noted that stakeholder events led by UoB had resulted in *"teachers paying more attention to value added progress when identifying students for setting classes, more attention being given to evaluating student progress in different ways and over time, with teachers and support staff involving students in monitoring their own progress and targets for improvement and arranging for additional support and intervention when and if needed."* [E].

Poor academic achievement of looked after children has been a longstanding concern. Thomas and colleagues analysis of value-added progress of looked after children [5], resulted in the DfE reporting for the first time an annual comparison of the educational outcomes of children in need (at home) and those looked after by local authorities [Giii, Giv]. The Minister of State for Children and Families stated that *"the report's findings […] [are] significant in changing the nature of the debate"* [Gi] and "*prompted action*" including launching a largescale DfE Review of Children in Need [Gii].

Enhancing practitioner, media and public understanding of the benefits and limitations of school league tables

UoB research has improved practitioner understanding of school accountability at the national level, for example via a report by General Teaching Council for Northern Ireland (GTCNI) [D] which cites on 16 occasions school league table work by Leckie and Goldstein (e.g., [3]) in relation to the recommendation to government to review educational targets and consider alternative approaches to monitor educational performance. The report is endorsed by Northern Ireland teachers' council and universities' council for the education of teachers.

The collaboration with the NPP noted above, led them to publish online 'Adjusted Progress 8' scores in the form of 'Fairer Secondary School League Rankings (2019)' [Iiii] for all schools in England and NPP have continued to lobby for Progress 8 to be adjusted [H, I]. Lucy Powell MP emphasised the debate in a radio interview and stated "*This is a ground-breaking piece of work and I hope it is taken up wholeheartedly by the DfE, Ofsted and all those working on improving schools. I will continue to push this report and use it in my discussion with Education Ministers*" [H]. At the local level, Leckie has presented these issues [1] to head teachers, data managers and others at the Bristol Cathedral Schools Trust, with the Executive Head Teacher stating that *"The Bristol research evidence acted as a springboard for increasing staff awareness and understanding of the need to use accurate assessment data"* which has *"resulted in direct and positive impact on students in my schools and on teachers' practice, specifically in relation to progress of individual students rather than cohort headline outcomes"* [E].

Extensive dissemination by Leckie and Goldstein including interviews; BBC Radio 4 (Nov 2013), Independent (Dec 2013), BBC website (Dec 2013), Heart and Love Sport radio (Jan 2019), newspaper coverage; Guardian, Independent, Times, TES (Jan 2019), and published letters; Guardian (2013-2019), has improved both media and public awareness of the problems and weaknesses of school league tables. For example, advice by Leckie to The Times and Sunday Times (2014), on their plans to publish dynamic interactive versions of Government school league tables, and the limitations of such information, resulted in the newspapers' decision not to publish.

Informing international education policy

Advice based on UoB research has been sought by policymakers and education departments internationally. Leckie has supported the World Bank and the Bulgarian ministry of education [Fi], in understanding best practice in school value-added evaluation via visits to Sofia, presentations to the Bulgarian Minister of Education and departmental officials, and detailed feedback on World Bank reports (2016-17). Future involvement was invited regarding a planned pilot followed by a national rollout [Fi]. Leckie has also advised on the design and analysis of school value-added evaluation in; Colombia (2014), Australia (2014), Israel Academy of Sciences (2019), Slovak Republic (2015) [Fii], Sweden (2018), and Trinidad and Tobago (2018) [Fiii]. Following support to officials in Trinidad and Tobago, the Chief Education Officer reported that "*The Ministry of Education has continued the value-added initiative with the goal of statistically isolating the effects of schools on growth in student achievement*". Trinidad and Tobago has gone on to pilot a new school performance and accountability system to 100 schools nationally [Fiii]. All of this work has improved these audiences' understanding of best practice in statistical methods for measuring school performance using student test score data [Fi-iii], for example, practitioners from the National Institute for Certified Educational Measurements in Slovakia, stated "*Your advice about how we treat unreliable scores and consider multiple imputation was very valuable in guiding our work forward*" [Fii]. Thomas' ESRC/DFID funded research in China [4], has also informed Chinese Ministry of Education thinking on value added measures and how to best evaluate schools and teachers (2015) and has been "*extremely useful to support the professional development of teachers and policymakers in China*" [Ci]. Recent Ministry of Education policy has explicitly proposed a more scientific and comprehensive evaluation of schools focusing on exploring; *"value-added evaluation—assessing how much progress the students have made from enrollment till graduation. Greater progress indicates more effective education."* [Cii].

5. Sources to corroborate the impact

[A] DfE (2020). Chief Analyst & Scientific Adviser (2012-2019) - Factual Statement

[B] House of Commons Education Committee (2017). Report: Primary assessment

[C] i) National Institute of Education Sciences, China (2016). President - Factual Statement and ii) Beijing Review (2020). Academic evaluation should be reformed to produce well-rounded students

[D] GTCNI (Sept 2013). Striking the Right Balance: Towards a Framework of School Accountability for 21st Century Learning

[E] Bristol Cathedral Schools Trust (2018). Executive Headteacher - Factual Statement

[F] i) World Bank Bulgaria Office (2017). Senior Economist/Task Manager - Factual Statement
ii) National Institute for Certified Educational Measurements, Slovak Republic (2016). Director - Factual Statement
iii) Ministry of Education, Trinidad and Tobago (2019). Chief Education Officer – Factual Statement

[G] i) Minister for Children & Families (2015). Speech: Children in care can achieve and succeed, ii) DfE (2019). Review of children in need, iii) DfE (2016). Children in need census matched to the national pupil database *(p.12),* and iv) DfE (2016). Outcomes for children looked after by local authorities in England, 31 March 2015 *(p.4).*

[H] House of Commons (2019). Education Select Committee - Lucy Powell MP - Factual Statement

[I] i) Co-op Academies (2020). CEO/Member of the Education & Skills Committee, NPP - Factual Statement, ii) NPP (2019). Latest research shows league tables 'punish and reward wrong schools' and iii) NPP (2019). Fairer Secondary School League Rankings (England) 2019