King's Research Portal

DOI:
10.1080/02680939.2019.1706103

Document Version
Peer reviewed version

Link to publication record in King's Research Portal

Citation for published version (APA):
Gewirtz, S., Maguire, M., Neumann, E., & Towers, E. (2019). What's wrong with 'deliverology'? Performance measurement, accountability and quality improvement in English secondary education. *JOURNAL OF EDUCATION POLICY*, *0*, 1-26. Advance online publication. https://doi.org/10.1080/02680939.2019.1706103

Citing this paper
Please note that where the full-text provided on King's Research Portal is the Author Accepted Manuscript or Post-Print version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version for pagination, volume/issue, and date of publication details. And where the final published version is provided on the Research Portal, if citing you are again advised to check the publisher's website for any subsequent corrections.

General rights
Copyright and moral rights for the publications made accessible in the Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognize and abide by the legal requirements associated with these rights.

•Users may download and print one copy of any publication from the Research Portal for the purpose of private study or research.
•You may not further distribute the material or use it for any profit-making activity or commercial gain
•You may freely distribute the URL identifying the publication in the Research Portal

Take down policy
If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 03. Aug. 2024

What's wrong with 'deliverology'? Performance measurement, accountability and quality improvement in English secondary education

Authors

Sharon Gewirtz (corresponding author), Centre for Public Policy Research, School of Education, Communication and Society, King's College London (email: Sharon.Gewirtz@kcl.ac.uk; postal address: King's College London, Waterloo Bridge Wing, Waterloo Road, London SE1 9NH, UK)

Meg Maguire, Centre for Public Policy Research, School of Education, Communication and Society, King's College London (email: Meg.Maguire@kcl.ac.uk; postal address: King's College London, Waterloo Bridge Wing, Waterloo Road, London SE1 9NH, UK)

Eszter Neumann, Institute of Advanced Study, Central European University (email: neumann.eszter@gmail.com; postal address: Nádor u. 9., Budapest 1051, Hungary)

Emma Towers, Centre for Public Policy Research, School of Education, Communication and Society, King's College London (email: Emma.Towers@kcl.ac.uk; postal address: King's College London, Waterloo Bridge Wing, Waterloo Road, London SE1 9NH, UK)

This is an original manuscript / preprint of an article published by Taylor & Francis in Journal of Education Policy on 27 December 2019, available online: http://www.tandfonline.com/ [https://doi.org/10.1080/02680939.2019.1706103].

Abstract

Informed by the ideology of 'deliverology', performance measurement has become a core component of how English schools are held accountable for the quality of their provision. A wealth of research conducted in diverse national contexts where this approach has been influential has suggested that the unintended harms it generates - including a widening of inequalities, a test-driven pedagogic culture and a narrowing of the curriculum - may be outweighing its benefits. In an attempt to circumnavigate such perverse effects, the performance measures used in England have become increasingly sophisticated in recent years, giving them a *prima facie* plausibility that has led to them being welcomed by some progressive educators as a means of increasing access to high-status knowledge for disadvantaged students. This paper uses data from a survey of English school teachers to interrogate this plausibility. The analysis suggests that when we drill down into the daily life of schools and its underlying logic it becomes increasingly difficult to be comfortable with the progressive defence of the performance measures currently in use; and that, far from improving educational quality, the measures themselves, and the wider deliverology framework with which they are associated, are in certain fundamental respects incompatible with quality improvement.

Key words

Accountability; deliverology; performance measurement; Progress 8; quality; social justice

Funding details

The research reported in this paper was supported by funding from the National Union of Teachers.

Biographical notes

Sharon Gewirtz is Professor of Education in the School of Education, Communication and Society at King's College London where she also co-directs the Centre for Public Policy Research. Her research is in the sociology of education and education policy with a particular focus on issues of equality and social justice, teachers' work, and the changing culture and values of schooling and higher education in the context of managerial reform.

Meg Maguire is Professor of Sociology of Education in the School of Education, Communication and Society, King's College London. She has a long-standing interest in education policy and practice, social justice, the life and work of school teachers, teacher education, and the challenges of inner-city schooling.

Eszter Neumann obtained her PhD in the sociology of education from King's College London in 2018 and is currently based in the Institute of Advanced Study, Central European University. Her research, which draws on policy sociology, ethnography of education, and action research, concerns education policymaking in England and Hungary, interethnic relations, the production of social inequalities, and normalization processes in schools.

Emma Towers is a Teaching Fellow in Education Policy in the School of Education, Communication and Society at King's College London. Before joining King's, Emma worked for 10 years as a primary school teacher in London schools. Her research interests include teachers' lives and career trajectories, urban education, and teacher identity.

Word count: 12,963

Introduction

This paper uses data collected from a survey of teachers in English secondary schools (Neumann *et al*. 2016) as a springboard to ask some critical questions about performance measurement in education. Reflecting a global movement of educational reform (Sahlberg 2011; Lingard *et al*. 2016) and informed by the ideology of 'deliverology' (Barber, Moffit, and Kihn 2010), in England, as elsewhere, performance measurement has become a fundamental component of the means by which schools and teachers are held accountable to the government and the public for the quality of their provision. Deliverology is a top-down approach to governance that focuses on improving public service delivery processes by identifying goals, and measures to assess progress towards them, that are designed to motivate and inform system improvements with respect to both excellence and equity and ensure that public money is well spent. A wealth of previous research conducted in diverse national contexts where this approach has been influential has suggested that the unintended harms it generates may be outweighing its benefits. In the case of school education, these harms have been found to include a widening of inequalities (Au 2009; Ravitch 2011; Cumming, Wyatt-Smith, and Colbert 2015; Picower and Mayorga 2015; Stuart Wells 2019), forms of 'gaming the system' or outright cheating (Nichols and Berliner 2007; Klenowski and Wyatt-Smith 2012; Lingard and Sellar 2013; Ohemeng and McCall-Thomas 2013; Thompson and Cook 2014; Hofflinger and Hippel 2018; Johnson 2018), a test-driven pedagogic culture and a narrowing of the curriculum (Stobart 2008; Klinger, Maggi, and D'Angiulli 2011; Polesel, Rice, and Dulfer 2014; Hardy 2015; Pinto 2016). In an attempt to circumnavigate such perverse effects, in England the performance measures used have, at least in some respects, become increasingly sophisticated over recent years. This sophistication gives the measures a plausibility which has led to them being welcomed in some quarters by progressive educators (e.g. Smith 2015; Francis, Mills, and Lupton 2017) as

a means of increasing access to a broad curriculum, high status knowledge, and rich pedagogies for disadvantaged students. In what follows, we will use our survey data to interrogate the *prima facie* plausibility of the performance measures currently being applied to English secondary schools, and to raise questions about the wider approach to educational accountability with which they are associated.

We begin by outlining some of the history of how those of us who are caught up in the English education system have got to where we are now, before going on to examine, and question, the underlying rationale for and effects of the performance measures currently in use. We identify four grounds for being sceptical about these measures and, in so doing, draw attention to some fundamental shortcomings of the wider deliverology approach that has informed their development. Specifically, we will argue that, with regard to their stated goal of quality improvement, there are compelling reasons for thinking that both the particular measures in use and the wider approach are, in certain key respects, counterproductive.

Performance measurement in English education: a brief history

History is important in policy analysis, both because it supports an appreciation of processes of policy sedimentation – how contemporary layers of policy build upon previous ones - and because it helps to remind us that sometimes there are sea changes where lines of policy are abandoned or reversed. Both of these things apply in this case. In this section, we sketch out the historical context for the analysis that follows by signalling some key moments in the English history of performance measurement in education. In doing so, a core intention is to chart the increasing sophistication of the performance measures being applied, which, we suggest, plays an important role in legitimising their continued use.

The idea of controlling what goes on in schools so that the state can satisfy itself as to the quality of education is not new. In the 1840s and 1850s in England and Wales, this idea

fuelled a number of schemes involving payment by results, culminating in the 1862 Revised Code. This required every pupil attending schools in receipt of state funds to take an annual exam focused on the '3 Rs' (i.e. the basic skills of reading, writing, and arithmetic), the results of which determined the size of the grant and hence teachers' salaries (Committee of the Privy Council on Education 1862). The key themes characterising the heated debates surrounding the introduction and implementation of the policy at the time – in Parliament and within the inspectorate and teaching profession - will be familiar to a contemporary audience. Whilst supporters, most notably its architect, Robert Lowe, the then Vice-President of the Board of Education, viewed the Revised Code as a positive move on the part of the state to assume responsibility for the quality of education, ensure value for money, and incentivise teachers to raise their game with regard to the teaching of basic literacy skills, critics pointed to a range of deleterious effects. These included a narrowing of the curriculum, a decline in the quality of teaching, a mechanistic approach to both teaching and inspection, and the growth of an instrumental mind-set on the part of teachers, with teaching for understanding replaced by teaching to the basic minimum standard required to pass the test (Arnold 1867; Connell 1950; Rapple 1994). The Code also had knock-on effects on the stability and quality of the teaching workforce. Following its introduction, teachers' salaries declined and teachers began to move from school to school in search of those with better results and hence better remuneration, with many leaving the profession altogether (Rapple 1994).[1]

In 1895 payment by results was abandoned, and in 1926 curriculum prescription effectively came to an end, ushering in what Gerald Grace has described as the 'modern principle of curriculum autonomy' (Grace 1985: 10), or what Denis Lawton (1980: 22) calls 'the golden age of teacher control'. This state of affairs continued relatively undisturbed until the mid-1970s when then Prime Minister James Callaghan's (1976) Ruskin speech called for

[1] The code also had deleterious knock-on effects for the quality of teachers entering the profession and the intellectual breadth and rigour of the teacher training curriculum since drilling students to pass a mechanical test of the '3 Rs' did not require a broad liberal education for teachers (Rapple 1994).

the opening up of the 'secret garden of the curriculum'. However, it wasn't until the 1980s and 1990s that successive Conservative governments reasserted state control over the curriculum. These decades saw the introduction of a highly prescriptive national curriculum with specified attainment levels in each subject, national tests for every state school pupil at four 'key stages' (age 7, 11, 14 and 16) (HMSO 1988), and the publication of the results of these tests (HMSO 1992). At the same time, a market in education was introduced with the idea that the test results would give parents the information they needed to make informed choices (Gewirtz *et al*. 1995). In this way, a neoliberal choice agenda was tied to a neo-conservative curriculum agenda - an approach that was emblematic of the 'free market-strong state' logic of the Thatcherite era (Gamble 1988).

The focus on performance measurement was retained by New Labour administrations from 1997 to 2010, but here the measuring was tied (at least ostensibly) to a social democratic agenda concerned with reversing the inequitable effects of the reforms of previous governments (Gewirtz 2002; West and Pennell 2002; Thrupp and Tomlinson 2005; Sammons 2008). Throughout this period also, an additional ideological strand became gradually more prominent - the *prima facie* non-ideological ideology of deliverology, developed and popularised by Michael Barber, Head of Prime Minister Blair's no. 10 'Delivery Unit'. The deliverology approach was later exported to other countries, including the US, Canada, Chile, Peru, Australia, Indonesia, Pakistan, Sierra Leone and South Africa, by Barber and colleagues via the consultancy organisations McKinsey, Delivery Associates, the US Education Delivery Unit and the Boston Consulting Group's Centre for Public Impact; and publications such as *Deliverology 101 (*Barber*,* Moffit, and Kihn 2010) and the *Art and Science of Delivery* (Dör 2013). Deliverology is a top-down, command and control based technology of governance defined by its proponents as 'a systematic process for driving progress and delivering results in government and the public sector' (*ibid*.: vii). The

approach, which is geared towards 'delivering' both 'excellence and equity' (*ibid*.: vii), sees policies as best enacted through the central setting of aspirations and target metrics, establishing systems for producing the metrics, monitoring progress against them, and ensuring compliance through a system of rewards for institutions - and individuals within them - deemed to be performing well and sanctions for those deemed to be performing below the required threshold.

Under successive New Labour governments, the performance measures (or target metrics) became increasingly sophisticated (Leckie and Goldstein 2017). The original performance tables of test results introduced by the Conservatives in 1993 had used raw performance scores that failed to take into account either 'distance travelled' or the contextual factors that determine how well students perform in tests. To address this, from 2002-5 a system of value added rankings was used and from 2006-10 this was replaced with a yet more sophisticated 'contextual value added' (CVA) measure which controlled for students' prior attainment, gender, special educational needs and care statuses, free school meals eligibility, first language, mobility, age and score on an index that measured the proportion of under 16s living in low-income households in students' immediate localities (as defined by their postcodes)[2]. The CVA measure also incorporated confidence intervals to capture the degree of statistical uncertainty relating to the school-effect (DCSF 2006).

In 2010 the incoming Conservative Coalition dropped the CVA measure, arguing that it was too complicated for people to understand and that, because it took into account factors relating to family circumstances and ethnicity, it 'entrenche[d] low aspirations for children because of their background' (DfE 2010: 68). The Coalition government reverted instead to a value-added measure that only controlled for prior attainment (the 'expected progress' measure) (*ibid*.). It also introduced a new performance measure, the English Baccalaureate

[2]Referred to as the 'IDACI' (income deprivation affecting children index).

(EBacc) (explained below) and a more sophisticated minimum expectation ('floor target') for schools (*ibid.*). Schools below this target of 35% (subsequently raised to 40%) of students passing at least 5 GCSEs[3] at C grade or above (including maths and English) and with a below median proportion of students making the 'expected progress' in English and maths faced the prospect of the removal of their head teacher, closure or forced conversion to academy status (DfE 2010).[4]

Each wave of reform throws up its own imperfections, some of which are a product of schools and teachers learning how to play the system. The key stage 4[5] reforms that were the focus of our study represented the latest government attempt at the time to fine tune the performance measures in order to address perceived shortcomings and 'gaming' associated with the previous measures. These included the tendency for schools to attempt to boost their performance table positions by focusing resources disproportionately on students at the borderline of achieving a C grade at GCSE (because only GCSEs at grades C or above were counted in the performance tables) and by entering students for subjects deemed to be easier to pass (Wilson, Croxson, and Atkinson 2006; Wolf 2011).

The 2016 key stage 4 reforms

In 2016, five new 'headline measures' for secondary schools were introduced:

- The percentage of students attaining a 'good pass' in new, more academically demanding, English and maths GCSEs for which a new numerical grading system replaced the alphabetic one. A good pass was defined as a grade 5

[3] GCSE (General Certificate of Secondary Education) qualifications are taken by students in schools in England, Wales and Northern Ireland usually at the age of 16.
[4] Often referred to as 'forced academisation', this process involves schools being moved outside of local authority control through the issue of an 'academy order' by the Secretary of State for Education. Academies are independent, state-funded schools, which receive their funding directly from central government, and are run by a government-approved sponsor. This can be an existing educational institution, or a private or voluntary sector sponsor.
[5] Key stage 4 is the stage of education in English and Welsh schools when students are aged 14-16. During this stage, most students work towards national qualifications, usually GCSEs.

which was deemed to denote a higher standard than a C in the old grading system.

- The percentage of students attaining the EBacc, which is achieved by students attaining GCSE grades of C (or 5 in the new grading system) or above in subjects deemed to be the 'core' academic qualifications that best equip students for further study and work – English language and literature, maths, the sciences, history or geography, and a foreign language.
- The percentage of students entered for the EBacc.
- Progress across 8 subjects selected from three subject 'baskets' - English and maths; 3 other Ebacc qualifications; 3 remaining Ebacc subjects or other subjects from an approved list. This measure is referred to as 'Progress 8'. Progress 8 is a value added, but not contextual, measure that compares students' progress with others with the same prior attainment. The measure includes 95% 'confidence intervals', which the Department for Education defines 'as a proxy for the range of scores within which each school's underlying performance measure can be said to confidently lie' (DfE 2016b: 12).
- Attainment across these 8 subjects, a measure that is referred to as 'Attainment 8'.

These results were to be published in performance tables. Under the Education and Adoption Act 2016, schools falling below particular thresholds and thereby deemed to be 'causing concern' were made subject to intervention measures, such as additional Ofsted inspections. Forced academisation became a required action in the case of schools categorised as

'inadequate' by Ofsted and a possibility in the case of those identified as 'coasting'[6] or underperforming in some other way, for example because the school was deemed to have fallen below the 'floor standard'[7].

All we have done here is to briefly summarise the range of new measures introduced, but it is important to stress that, as will be seen in subsequent sections, each of these is having a substantial effect on the organisation of, and what is deemed salient, within schools. Indeed, we are now in a position where the technology of deliverology is deeply engrained in school life. Not least because of ongoing processes of refinement, implemented through successive policy adjustments, there is a plausibility to these measures, with even some progressive voices persuaded by aspects of the measures, which look, in some respects, to be more sophisticated than some earlier versions. This plausibility is important because it plays a key role in legitimating the measures and thereby embedding them into taken-for-granted policy common sense. For instance, in an article published in the *Journal of Education Policy*, Becky Francis and colleagues (2017: 426) have written in praise of Progress 8, which they describe as a measure that incentivises and rewards 'a range of just practices':

> For example, it analyses individual pupil progress, rather than raw outcome; recognising uneven starting points and rewarding the contribution a school has made. As outcomes in 8 subject choices can 'count' towards the measure, it incentivises schools to focus on supporting pupils in 8 curriculum areas – encouraging entitlement to a broad curriculum but with 'high status' knowledge at the heart. The stipulation of choices from different

[6] This categorisation applies to schools failing to meet a 'coasting standard' for three consecutive years (DfE 2016a, 2019). The policy was designed to address the assumed 'complacency' of schools, often located in 'leafy suburbs', deemed not to have stretched their pupils sufficiently and likely to have previously 'fallen beneath the radar' because, for example, they had high attaining intakes or had not received a negative Ofsted report (DfE 2015).

[7] A school is deemed to have fallen below the floor standard when its Progress 8 score is below -0.5 and the upper band of the 95% confidence interval is below zero (DfE 2016b, 2019).

'baskets' of subjects ensures that this broad curriculum is also balanced – and can facilitate routes to higher study ... Hence this indicator seeks to support justice in terms of entitlement for all, [and] comprises an impressively 'strong' model in this regard, as it actively incentivises support for lower attainers as well as middle and higher attainers (the former having been disregarded in prior models).

Here Francis *et al*. are drawing attention to two aspects of Progress 8 that can be understood as social justice measures. First, because Progress 8 is a measure of progress rather than raw attainment, to perform well in the measure, schools need to focus on the attainment of all students, not just those whose performance happens to coincide with a particular attainment threshold. And second, by measuring progress across a broad range of subjects that are deemed to be more intellectually demanding,[8] it is less likely that schools will funnel their more socially disadvantaged students into a narrower and less challenging set of qualifications that will effectively disqualify them from accessing higher status post-school study routes, as has so often been the case in the past (see, for example, Anders *et al*. 2017).

A related line of argument is advanced by Laurie Smith (2015) although specifically in relation to the English curriculum. Smith argues that the combination of the introduction of a more demanding curriculum and a measure which incentivises schools to focus on the attainment of all of their students has the potential to facilitate richer pedagogic practices characterised by open-ended exploration and dialogue. This is because, Smith suggests, to perform well in the new qualifications, students are required to show evidence of deep knowledge and higher order cognitive skills, which cannot be taught using spoon feeding or the whole-class, teacher-directed methods that had previously dominated.

[8] Effectively, the subjects that count towards the EBacc are the GCSE equivalent of what the elite group of Russell Group (2017/18) universities used to refer to at A level as 'facilitating subjects', i.e. those that facilitate access to higher study. The list of 'facilitating subjects' was abandoned by the Russell Group in 2019 following criticisms that it was contributing to a devaluing of creative and technical subjects (Sharratt 2019).

If the effects of the new performance measurement regime wholly coincided with the kinds of account provided by Francis *et al*. and Smith - in a nutshell, that Progress 8 will make it more likely that working-class students and other marginalised constituencies will be able to access the 'powerful knowledge' and higher order cognitive skills (Nash 2010; Young and Muller 2013) so often denied to them - it would be difficult to argue against. However, we want to suggest that the problem with this kind of defence is that it operates at a very abstract level and that, when we drill down into the day-to-day life of schools and into the elements of its underlying logic, it becomes increasingly difficult to be comfortable with this generalised defence. In the following sections, we will expand on this argument, drawing on data from our teacher survey.

The study

The study on which this paper draws was commissioned by the National Union of Teachers (NUT)[9] to investigate the early effects of the new GCSEs and the 2016 accountability reforms summarised above. The study consisted of a national survey and case studies of three secondary schools. In this paper, we draw exclusively on the survey data and mainly on teachers' qualitative responses to the open-ended questions to which respondents were invited to provide free-text responses (see Appendix). The survey was distributed to all NUT secondary school members in England (totalling 68,833) in April 2016 via the NUT email database and was completed by 1802 teachers, giving a response rate of 2.6%. Questions focused on GCSE curricular offerings in the respondents' schools; their pedagogic approaches, approaches to data management, and systems of grouping students; the allocation of resources for the teaching of different groups of students; and teachers' perceptions of students' experiences of schooling in the context of the new reforms. The survey was piloted

[9] Since we conducted the research the National Union of Teachers has merged with the Association of Teachers and Lecturers to form the National Education Union. At the time of the research in 2016 (and the merger a year later), the NUT, with approximately 400,000 members, was the largest teachers' union in the UK.

with a range of secondary school teachers in different roles and levels of seniority. Ethics approval was granted by King's College London.

The summary of the sample provided in Table 1, below, suggests that it is broadly representative of the NUT English secondary school membership in terms of gender and school type.

[Table 1 near here]

Although we did not aim for a representative sample of teachers from the English education system as a whole, it is worth noting that, with regard to Ofsted evaluations of secondary schools, the sample is slightly skewed towards schools rated as 'Good' and 'Requires improvement', and that schools rated as 'Outstanding' and 'Inadequate' were slightly under-represented in our sample (see Table 2).

[Table 2 near here]

The majority of the respondents were in middle-managerial roles (43.1%) or classroom teachers (56.2%). Deputy heads and headteachers, who are underrepresented in the NUT membership, represented only 0.7% of the sample. Respondents primarily taught the following subjects as their main subject at key stage 3[10]: science (13%), maths (10%), English language (10%) and modern foreign languages (MFL) (9%). At key stage 4[11], respondents primarily taught English (14%), maths (11%), MFL (9%) and the sciences (altogether 16%). In terms of key stage 4 subject 'baskets', 25% of the respondents primarily taught English or maths, 52% taught 'basket 2' subjects and 21% taught non-EBacc 'basket

[10] Key stage 3 covers the period when students are in the first three years of secondary school (and aged 11-14).
[11] See footnote 5.

3' subjects. With regards to teaching experience, respondents had an average of 13 years of teaching experience, and the mode was 10 years. Overall, the teaching experience of the respondents ranged from 1 to 43 years.

As union members motivated to complete our (10-15 minute-long) survey, it is likely that our respondents will have been more critical of current reforms than the average secondary school teacher. Nevertheless, the consistency of the messages coming through from the data, set alongside some similar findings emerging from other research (see, for example, sources cited in the House of Commons, Committee of Public Accounts [2018] report on teacher retention and development), means that it can be assumed that the views expressed are broadly representative of at least a substantial swathe of secondary school teachers in England.

The Bristol Online Survey service administered the survey and the structured responses were processed and analysed using SPSS. However, as noted above, this paper draws mainly on teachers' qualitative responses to the open-ended questions (see Appendix). These free-text comments (4926 in total across the 11 open-ended questions) were analysed manually using the constant comparative method (Corbin and Strauss 2015), with themes constructed from an inductive reading of the data in dialogue with existing sociological and philosophical analyses of school accountability and the curriculum. Theoretically, our analysis draws inspiration from the work of John Dewey, Paolo Freire, Henry Giroux and others in the critical pedagogy tradition. These writers emphasise the crucial role that education plays in 'the production of agency' (Giroux 2001) and hence the risks associated with authoritarian, technocratic, instrumentalist and/or 'banking' (Freire 2006) models of education that tend to 'kill off the imagination' (Giroux 2014) and suppress students' critical faculties. Work in this tradition emphasises, in particular, the potential for humanities and social science subjects, well-taught, to 'enlarge [students'] perspective[s] on the world [and]

their relationship with themselves and others' (Giroux, cited in França 2019) and help them develop core dispositions, such as compassion and critical awareness, for contributing effectively to society. Such forms of education require substantial autonomy for teachers and students; as autonomy is not only an intrinsic component of human wellbeing, but is also required if pedagogies and curricula are to be sufficiently well-aligned to local social contexts of enactment, for teachers to fulfil their civic function as 'transformative intellectuals'[12] (Giroux 1988) and for students to be able to learn and practice the habits of democracy (Dewey 1916).

One core set of themes that emerged from our analysis – and the one we focus on here - revolved around the deep concerns teachers expressed about the value and constitutive effects of the 'headline' performance measures currently in use. In the sections that follow we elaborate on our respondents' concerns about the value and effects of the new performance measures. Using extracts from their written comments, we will explicate and illuminate four interconnected grounds to be sceptical about the suitability of the measures for driving improvements in the quality of secondary schooling. These grounds for scepticism relate to the validity, reliability, power, and utility of the measures. We devote most space to the last of these, as it is the constitutive effects of the measures that are of greatest concern from an equity perspective.

The validity and reliability of the measures

The suitability of the current headline measures as indicators of educational success is hotly contested. Part of the problem, as Biesta (2009) reminds us, is that not everything that is valuable in education is amenable to quantitative measurement; yet, within the deliverology

[12] Defined by Giroux (1988: 126-8) as 'active, reflective scholars and practitioners [who can] educate students to be active, critical and risk-taking citizens [and] speak out against economic, political, and social injustices both within and outside of schools'.

paradigm, quantitative measures are privileged. But the current measures are also open to the criticisms that they are both measuring the wrong things and that they do not measure the things they are purporting to measure. With regard to the former, survey respondents expressed concerns that the new GCSE qualifications reflect an anachronistic conception of success. This is a conception of success that is aligned with longstanding hierarchies of esteem in English state education, which privilege academic over practical knowledge, the abstract over the applied, 'traditional' subjects over creative ones, and the ability to perform well in examinations over the less tangible yet more durable transformative effects of a good education – what one survey respondent described as 'the shaping of our children into tomorrow's citizens' (MFL teacher and union representative in a local authority school with a 'Good' Ofsted rating). Such concerns were strongly reflected in the survey responses.

For teachers of dance, drama and physical education (PE) there was a particular concern about the increased weight given to written forms of assessment and the reduction in the practical component which is so central to these subjects:

> *For creative subjects, the emphasis on theory is important but the fact that this it now weighed equally to practical ability is not appropriate. Dancers who are amazing DANCERS will not be as successful as average dancers with good subject knowledge. This is not the way that a creative subject should be assessed.* (Dance teacher in a multi-academy trust school with a 'Good' Ofsted rating)

> *The practical aspect, the drama, in my subject is now worth less than written work. How is that a GCSE in Drama?* (Head of Drama in a local authority school with a 'Good' Ofsted rating)

> *In PE they have … changed a practical vibrant subject into a dull classroom based one, with an emphasis on exam skills rather than subject skills.* (Head of MFL and PE teacher in a multi-academy trust school with a 'Good' Ofsted rating)

For teachers of the more traditional academic subjects, their validity concerns[13] related to the ethnocentricity of the new GCSEs and their emphasis on students' ability to memorise facts and formulae over research and evaluation skills:

> *Monocultural GCSE English curriculum for students living in a global and multi-cultural society doesn't make sense.* (English teacher and union representative in a multi-academy trust school with a 'Good' Ofsted rating)

> *The new curriculum is focused almost entirely on knowledge. In the age of Google, I cannot, for the life of me, understand, why we would want to reduce the amount of evaluation that is required.* (Head of Religious Education (RE) in a multi-academy trust school with a 'Good' Ofsted rating)

> *Students should be allowed to take in textbooks etc. to the exam and be examined by using ... the tools they have, just as you would in a job situation. We are still assessing in the dark ages and the way our government leaders were taught. (*Assistant headteacher and Head of Maths in a local authority school with a 'Good' Ofsted rating)

> *Just because a student cannot remember 20 physics equations, it does not mean they are stupid but this is the way students feel!!* (Science teacher in a local authority school with a 'Good' Ofsted rating)

This last quote points not only to a concern that the test questions are measuring the wrong things, but also that they are damaging the confidence of at least some of those young people who struggle to reproduce the kind of rote-learned knowledge that is being called for. This is a point we will return to again below, when we discuss the utility of the measures.

---

[13] In framing anachronism and ethnocentricity as validity concerns, we are not here using validity in a technical sense to mean that there is a lack of alignment between the revised GCSE examinations and the knowledge and skills they are purporting to measure. The suggestion is rather that the GCSE examinations are invalid measures of what our participants considered to be worthwhile knowledge.

However, even if we were to agree that the new GCSEs do invariably measure things that are important, there are good reasons to be sceptical about the validity and reliability of Progress 8 as a measure of student progress. Two key problems with this measure were a matter of particular concern to our survey respondents: first, that it is invalid[14] because it uses key stage 2 test results in maths and English (i.e. those taken at the age of 11 when children are in the last year of primary school) as a baseline measure for progress across all subjects including subjects where there is no obvious relationship between maths or English ability and ability in that subject; and second, that the key stage 2 score data is inaccurate and hence an unreliable predictor of students' attainment at 16, even in the core subjects. Progress 8 is the highest stakes of the new performance measures because it determines the floor standard a school needs to meet if it is to avoid triggering an Ofsted inspection or other serious consequences, which might include the closure of the school or its forced conversion to an academy. As we discuss in more detail below, it is also increasingly being used to inform judgements about the performance of individual teachers; hence the depth of concern expressed by many of our respondents in their free-text responses.

The majority of concerns about the validity of key stage 2 results as a baseline measure came from teachers of MFL and creative subjects. The following typified the kinds of concerns expressed:

> *Targets should be based on individual subjects, not just attainment at KS2 [key stage 2] English and maths. My targets are based on how well pupils did in KS2 English which bears no correlation to their ability in my subject, leaving some pupils with completely unachievable targets.* (Newly qualified MFL teacher in a multi-academy trust school with an 'Outstanding' Ofsted rating)

[14]Here we are using the concept of validity in the conventional technical sense of meaning the extent to which something properly measures what it purports to measure.

The new target grade setting should not be based on comprehension (KS2 reading SATs[15]) when we are now expecting students to produce coherent written work and show application of knowledge in written form. (Head of key stage and MFL teacher in a multi-academy trust school with a 'Good' Ofsted rating)

MFL in particular finds data from KS2 to be irrelevant. In a previous school CEM[16] data proved more reliable an indicator of likely outcome. (MFL teacher in a standalone academy with a 'Good' Ofsted rating)

Pupils are often expected to work with unrealistic targets especially for creative subjects as those targets don't tell me anything about their creative abilities. (Head of Art and Design in a standalone academy school, Ofsted rating not specified)

As a teacher of a creative subject, I fail to see how and why our results should be held accountable against KS2 data when pupils do not have any results in my subject. (Head of Drama in a local authority school with an 'Outstanding' Ofsted rating)

Scepticism about the reliability of the key stage 2 test scores was expressed by teachers in all subjects, but especially maths and science:

KS2 data is pretty unreliable for my core subject of maths. All schools I have taught in have retested kids at start of year 7 anyway. (Maths teacher in a local authority school with a 'Good' Ofsted rating)

The SATs at KS2 are fundamentally flawed, mismanaged and not consistently administered. (Assistant headteacher and maths teacher in a standalone academy school with a 'Good' Ofsted rating)

[15] Standardised Attainment Tests.
[16] CEM (Centre for Evaluation and Monitoring) is a commercial provider of assessment monitoring systems.

When a child has been 'hot housed' to pass a SAT in Maths or English it does not mean that they are consistently attaining that level. (Maths teacher in a multi-academy trust school with an 'Inadequate' Ofsted rating)

Teacher assessed SATs for KS2 are incredibly unreliable. We make all students sit an old Science SAT paper as a baseline test. The average level of incoming students, as judged by KS2 teachers, is 6. The average student achievement in the SATs paper is 4a. (Science teacher and STEM coordinator in a standalone academy school with an 'Outstanding' Ofsted rating)

Basing Progress on KS2 results will be flawed as KS2 results in my subject often seem to be inflated or not directly related. There needs to be some sort of baseline testing in place where students are not primed for the test. (Head of key stage and science teacher in a standalone academy school with a 'Good' Ofsted rating)

KS2 exam results do not reflect what a child is capable of in a subject - they reflect how intensively a child has been schooled in passing those exams. (Newly Qualified science teacher in a local authority school with a 'Good' Ofsted rating)

The KS2 results are inflated by training pupils to pass and teachers being pressured to cheat. (I know this happens as the pressure of teachers is great and I have friends who have done it). (Head of year and design and technology teacher in a local authority school with a 'Good' Ofsted rating)

KS2 falsified/over exaggerated data is the main problem. (Newly Qualified geography teacher in a multi-academy trust school with a 'Good' Ofsted rating)

KS2 teachers cheat during SATs exams to get pupil results, this will increase and create more pressure on KS3/4 than ever before. (English teacher in a standalone academy school with a 'Good' Ofsted rating)

The views expressed here that key stage 2 data is artificially inflated because of widespread practices of coaching and cramming and, in some cases, outright cheating resulting from the high stakes nature of the key stage 2 tests for primary schools resonate with concerns articulated elsewhere (Bew 2011, Allen 2016, Abrams 2017, Teacher Tapp 2018).[17] The practice of secondary schools retesting students on arrival because of scepticism about the reliability of the key stage 2 tests has also been reported by others (Bew 2011, Coldwell and Willis 2017). Overall, 93% or our respondents agreed with the statement that 'key stage 2 SATs do not provide an adequate basis for tracking student progress across the whole range of secondary subjects'.

An additional flaw, not explicitly referred to by our survey respondents but picked up by other commentators and implicit in some of our respondents' responses, is that, while Progress 8 controls for prior attainment, it does not account for the influence of students' socio-economic and demographic characteristics on their progress between the ages of 11 and 16, thereby penalising 'schools serving educationally disadvantaged communities and reward[ing] those serving advantaged ones' (Leckie and Goldstein 2017). Yet, in line with the logic of deliverology, in many schools, responsibility for students' progress is assigned to individual teachers. Not only do such practices fail to recognise the complex ways in which socio-economic and other material factors interact with in-school processes to produce student progress (Maguire et al. 2019), they also fail to take into account the fact that so much of teaching is a collective effort. Pointing to the corrosive effects that the

[17] See also Nichols and Berliner 2007, Klenowski and Wyatt-Smith 2012; Lingard and Sellar 2013, Ohemeng and McCall-Thomas 2013, Thompson and Cook 2014, Hofflinger and Hippel 2018, and Johnson 2018 for similar accounts of such practices operating in other countries where high-stakes testing has been introduced.

individualisation of responsibility for student progress can have on the organisational culture of schools, one teacher commented in response to the survey question, 'Have any other areas of teaching and learning or organisational arrangements been affected by P8?':

> *less collaboration and collective responsibility for students, and strained relationships between colleagues as teachers are seen responsible for all the nuances of achievement.* (Geography teacher in a local authority school with an 'Outstanding' Ofsted rating)

.

The power and utility of the measures

Concerns about the validity and reliability of the Progress 8 measure in a context in which responsibility for students' progress is being routinely assigned to individual teachers were exacerbated for our respondents by worries that the measure would be used, and, in some cases, was already being used, to inform judgements about their capability and decisions about their pay, with 85% of respondents agreeing with the statement that teacher appraisal was becoming increasingly data-focused in response to Progress 8:

> *Appraisal linked to data and results is unfair, as the original basis for P8 is flawed, and progress expected is unrealistic. SATS in year 6 are unreliable to judge progress and all students are different. Teachers' salary linked to such a system is not fair, we can work ourselves ragged and not get anything to show because students didn't progress ... enough.* (Head of key stage and science teacher in a free school with a 'Requires improvement' Ofsted rating)
>
> *Flawed KS2 data does not [allow] for proper predictions but ... staff can be threatened with disciplinary action over Progress 8.* (Head of vocational subject

department in a local-authority school with a 'Good' Ofsted rating)

> *A lot of staff failed performance management last year based on their data targets and it has been made very clear that the same will happen again this academic year even if all other targets are achieved thus blocking pay progression.* (Head of Art and Design in a local authority school with an 'Inadequate' Ofsted rating)

> *It is ridiculous that our pay is determined by KS4 target grades which are created using KS2 data. How can tests in English, maths and science taken at the age of 10 determine the target grade for music or drama etc. at the age of 16? How can students have a target grade that is the same in all subjects? This is a crazy way of deciding teacher pay progression – and particularly unfair when you have a smaller class!* (Head of Music in a multi-academy trust with a 'Good' Ofsted rating)

As these extracts illustrate, there is huge weight attached to Progress 8. Doubts about the validity and reliability of this measure would be less crucial if it was solely being used as an indicator to be considered *alongside* other factors to *inform* decision making in schools. In reality, however, in accordance with the logic of deliverology, this measure is *powerfully decisive* – making or breaking the reputations of schools and teachers, determining whether teachers are rewarded or disciplined, whether a school should be placed in special measures, subject to additional inspections or turned into an academy, and in some cases whether a head teacher should keep their job (Tickle 2017).

It is too soon to conclusively assess whether Progress 8 and associated measures will lead to improvements in educational outcomes, understood in the narrow sense of students being able to perform the tasks required of them in the new GCSE examinations, and whether they will lead to a narrowing of the attainment gap between more advantaged and disadvantaged student groups in relation to these outcomes. However, a significant majority

of the teachers surveyed reported that the more demanding content of the new GCSEs is making the curriculum less accessible to lower attaining students and those from low-income and second or other language-speaking backgrounds. In addition, our survey data suggests that the new performance measures are having deleterious effects on the breadth of the curriculum offer, the quality of pedagogy and relationships in schools and the mental wellbeing of teachers and students alike.

Many of our survey respondents reported that social-class and language related cultural biases in the tests are making success less attainable for low-attaining and low-income students and those who speak English as a second or other language, thereby exacerbating inequalities. 77% of survey respondents strongly agreed that 'the new GCSE curriculum will be less suitable for low attaining students', and phrases such as 'setting students up to fail' were frequently used in this context, with the new English GCSE identified as especially problematic:

> *The SEND[18] students and low attaining students are struggling severely with the new English GCSE. Staff feel as though they are setting them up to fail.* (English teacher in a multi-academy trust school with an 'Outstanding' Ofsted rating)

> *More students are having to study English Literature as well as English Language. This is stressful for students with literacy difficulties/SEN[19] and EAL[20]. We have students in year 10 who are new to English and in tears because they are struggling with Shakespeare and other literature.* (SENDCO[21] in a standalone academy with a 'Good' Ofsted rating)

[18] SEND is an acronym used in English schools for students categorised as having special educational needs and disabilities.
[19] Special Educational Needs.
[20] English as an Additional Language.
[21] Special Educational Needs and Disabilities Coordinator.

The new English language exams are an absolute disaster, especially for lower attaining students. The tasks and texts are inaccessible ... There have been too many factors introduced to increase the difficulty level (set texts, loss of controlled assessment, longer exams, closed book conditions). (English teacher in a standalone academy with a 'Good' Ofsted rating)

The problems described here, like those reported in the previous section regarding the anachronism and ethnocentrism of the revised GCSEs, would no doubt still persist with the removal of Progress 8 and associated threats such as forced academisation. However, the erosion of teachers' professional autonomy associated with the deliverology approach (discussed below) risks exacerbating these problems by restricting the flexibility that would otherwise be available for teachers to deviate from the mandated curriculum, and hence to teach in ways that are relevant, accessible and inclusive for all students regardless of their diverse starting points, capabilities and interests.

Respondents also expressed concerns about the shift away from tiered assessments and a modular coursework-based approach to assessment by terminal examinations because of the detrimental impact that they believed this shift would have, especially on low attaining students. Such concerns were expressed by teachers working in all kinds of schools and, in particular, by teachers of English, science and the humanities:

[The previous] GCSE provided something for everyone ... with its different tiers and well thought out choices of texts. This exam, in its attempt to reach everyone, is failing the low ability. (English teacher in a multi-academy trust school with an 'Outstanding' Ofsted rating)

The single tier aspect means that most will sit an exam that they fail to complete. The message that they have failed will be loud and clear. (Assistant head and English teacher in a multi-academy trust school with a 'Good' Ofsted rating)

The un-tiered system is isolating lower-attaining students. (Newly qualified English teacher in a multi-academy trust school with a 'Requires improvement' Ofsted rating)

Higher academic demand on lower attainers already shows them to be overwhelmed and demotivated. (Deputy head of science and union representative in a local authority school with a 'Good' Ofsted rating)

We have stopped offering science GCSE to [our] lowest attaining students as they would not be able to access [the] exams at all. (Science teacher in a multi-academy trust school with a 'Good' Ofsted rating)

The removal of tiers and the difficult language in the sample papers will disadvantage lower ability learners. (Head of Geography in a local authority school with a 'Good' Ofsted rating)

There is a lot more maths which will make it harder for low attaining students. This will be exacerbated by the change to one paper rather than tiered papers. (Geography teacher in a local authority school with a 'Good' Ofsted rating)

Weaker students will not be entered for exams. (Head of year and Humanities teacher in a multi-academy trust school with an 'Inadequate' Ofsted rating)

A concern expressed by a number of respondents related to the impact of the new GCSEs on students from low-income backgrounds who also attend resource-poor schools, thus resulting in a 'double disadvantage'.[22] For example, one teacher commented:

> *We are losing hundreds of thousands of pounds each year which impacts directly upon the school experience of all pupils but especially disadvantaged and special needs pupils.* (Design and technology teacher in a multi-academy trust school with an 'Outstanding' Ofsted rating)

However, even if we are to be generous and concede that educational outcomes, narrowly conceived, will improve as a result of the new curriculum and its associated performance measures and that attainment gaps will be reduced, there are other significant costs of the tests, which arguably far outweigh any such benefits. Because of their high stakes nature, the knock-on effects of the new measures, beyond simply providing information on what has been learnt, are enormous – these include: a narrowing of the curriculum, as creative and vocational subjects alongside religious education (RE), personal, social and health education (PSHE) and citizenship education become increasingly marginalised; a reduction of opportunities for students to develop higher order cognitive skills, because of the emphasis on teaching to the test (contrary to the arguments of Smith 2015, summarised above); and an increased prevalence of teaching and working practices that dehumanise teaching, threaten the mental health of both students and teachers and reduce opportunities for teachers to model civic and democratic 'habits of mind' (Meier, 2009).[23]

---

[22] This point is supported by our case study data (reported in Maguire *et al*. 2019) which suggests that 'resource poor' schools are more likely to feel under pressure to prioritise the 'core' subjects over 'many of the experiences that do not obviously relate to official indicators of school success but that enrich secondary schooling for young people'.

[23] These findings resonate strongly with those reported in research on other national systems with high-stakes testing regimes. See, for example, Meier and Wood 2004, Lambert and McCarthy 2006, Au 2009, Klinger, Maggi, and D'Angiulli 2011, Ravitch 2011, Polesel, Dulfer, and Turnbull 2012, Polesel, Rice, and Dulfer 2014, Cumming, Wyatt-Smith, and Colbert 2015, Hardy 2015, Picower, and Mayorga 2015, Pinto 2016 and Stuart Wells 2019.

According to our survey respondents, as more time and resources are invested in the ‘core’ subjects (particularly, English, maths and science), creative subjects like drama, music, dance and the performing arts are being removed from the the key stage 4 curriculum altogether and teachers of these subjects are losing their jobs. While these subjects were reported to be the most severely impacted by a withdrawal of resources, respondents also indicated that RE was often removed or relegated to being an option and thus taught in reduced time and that PSHE and citizenship education were negatively impacted as well. Only 4% of the teachers surveyed agreed that the reforms will ensure a broader and more balanced curriculum than before, whilst 75% reported that students had a reduced number of GCSEs to choose from in their schools, with teachers working in schools categorised as ‘Inadequate’ or ‘Requires Improvement’ more likely to report this to be the case. A decrease in examination entry rates in the respondents’ subjects in their schools was reported by 82% of teachers of creative subjects, 84% of vocational subject teachers and 75% of technology teachers. 72% of teachers across all non-EBacc subjects reported that their subject had lost a significant number of students as a result of the key stage 4 reforms.

Comments about the narrowing of the curriculum were often linked to the claim that students are being made to take subjects they are not motivated to study:

> *It's forced students who struggle with academic subjects to pick ... subjects that they don't enjoy. It causes them to lose confidence in their ability which causes their self-esteem to plummet. It's horribly unfair.* (Head of Humanities and history teacher in a local authority school with a ‘Requires improvement’ Ofsted rating)

> *Students are pressured into the EBacc with the result that they are now taking subjects that they 'dislike least'. This has led to demotivated pupils and more behavioural issues for subjects like history and geography.* (Head of department and history teacher in a standalone academy with an ‘Outstanding’ Ofsted rating)

Far more pressure on pupils [who] are pushed into choosing subjects for EBacc but cannot cope academically causing more problems with failure and behaviour/dysfunctionality. (Science teacher and union representative in a standalone academy with a 'Good' Ofsted rating).

Behaviour worsened amongst children who are forced into EBacc. (Head of key stage and history teacher in a multi-academy trust school with an 'Inadequate' Ofsted rating).

92% of maths and English teachers surveyed and 87% of teachers of other subjects agreed with the statement that their classroom practice has become more focused on exam and test preparation as a result of the key stage 4 reforms; and in their free-text responses many expressed concerns about an associated decline in creativity and in teaching quality. Such sentiments closely echo those of Mathew Arnold and fellow critics of payment by results in the 19th century:

I became a teacher to encourage creativity, promote independent learning and develop problem solving skills. To teach students to think for themselves in new and exciting ways NOT JUST EXAM FACTORY FODDER - strictly learning answers does not teach people to think!! (Drama teacher in multi-academy trust with an 'Outstanding' Ofsted rating)

We are merely becoming an exam production unit and we are losing the breadth and depth of knowledge that we ought to be giving students. Also, we seem to be driven by the need to get rid of any in-depth, exciting and innovative teaching and now are solely focused on the PowerPoint driven lesson with reliance on textbook materials which I feel is a real step backwards and is a result of the changes which have taken

place recently. (Geography teacher in a multi-academy trust school with a 'Good' Ofsted rating)

The sheer level of fear about accountability and the difficulty of the new GCSEs and of terminal assessments have meant the ENTIRE curriculum (from Y7) has been reduced to replicated GCSE-style assessment. Kids are doing far more testing. The curriculum has narrowed. Forms of assessment have become narrowed. Teaching is becoming more a form of 'transmission'. It is depressing. Kids feel fatigued and stressed. I feel bored and demotivated. (English teacher in multi-academy trust school with a 'Requires Improvement' Ofsted rating)

Respondents also reported concerns about the dehumanising effects of the current preoccupation with data and quantification that is integral to the deliverology approach:

Students are seen as statistics, not human individuals. I cannot see how such rigorous and constant examination of data can benefit any pupil. (Head of English in multi-academy trust with 'Good' Ofsted rating)

It is less about how well students are doing and more about meeting the school's figures ... Education has lost a lot of the human aspect. (Lead science practitioner in multi-academy trust with 'Outstanding' Ofsted rating)

*It's all robotic now. Data is good but never allows for the human factors. We are forced to apply robotic principles and thinking to human beings. (*Head of History in standalone academy with 'Good' Ofsted rating)

Closely linked themes in the teachers' responses were the increased levels of stress, demotivation and mental ill health respondents observed in their students, with 84% agreeing with the statement that the reforms are strengthening an exam culture which undermines

students' mental health and wellbeing. The following extracts illustrate the kinds of concerns underlying this statistic:

> *Students are dragged in on holidays, weekends... The pressure on them is relentless.* (RE teacher in a standalone academy with an 'Outstanding' Ofsted rating)
>
> *All students have to make at least 4 levels [of] progress. They are getting too high targets. The teacher has to put in extra intervention to achieve this … and pupils are … suffering from stress-related illness because of this.* (Maths teacher in a chain academy with an 'Inadequate' Ofsted rating)
>
> *I am seeing more young people having emotional and mental [health] issues than ever before. How can we say this is progress?* (English teacher in a chain academy school with a 'Good' Ofsted rating)
>
> *We will look back on these days and people will ask how we could do this to our children. It is akin to abuse.* (Music teacher and union representative in a multi-academy trust school with an 'Outstanding' Ofsted rating)
>
> *Mental health issues are rising where children are finding it very hard to cope with the exams and stress put upon them.* (Art teacher in a standalone academy rated 'Good' by Ofsted).
>
> *'[There is an] increase in pressure on students and resulting increase in mental health problems*. (Head of Year and DT teacher in a local authority school with an 'Outstanding' Ofsted rating).

With core subjects also taking precedence over pastoral activities, and teachers now devoting a substantial amount of time to the collection, analysis and tracking of data, teachers have a sense of the time available for attending to the pastoral needs of their students being severely curtailed. As one teacher put it: *'The pastoral, health and wellbeing of students has taken a*

back seat' (Head of year and DT teacher in standalone academy with 'Good' Ofsted rating). Another teacher, in reply to our question asking whether their school was considering any form of workforce restructuring or re-training in response to the EBacc and the GCSE and school accountability reforms, told us that their school's pastoral team was being disbanded altogether (Newly Qualified Teacher of Geography in a local authority school with 'Requires Improvement' Ofsted rating).

Underlying all of these concerns is a deeper anxiety about the erosion of opportunities for teachers to exercise their professional autonomy arising from a combination of workload intensification and a compliance-based performance culture, which was resulting in many teachers considering or actually leaving the profession. 92% of the teachers surveyed reported that their workload had increased as a result of data collection for Progress 8, with a significant majority reporting increased workload related to one-to-one booster classes (78%), data tracking and data collection (91%) and data analysis (90%). 72% agreed that meeting the demands of Progress 8 takes time away from teaching young people:

> *Workload has hugely increased in [the] last five years and … students and teachers are more stressed out than ever before. I know many teachers who are quitting, or who have quit, who were great teachers but constant changes to exam specifications, and a huge decrease in teacher morale due to constant monitoring and accountability measures, which have stifled creativity in the classroom, have led to the very best finding alternative careers, or often quitting with no job to go to, just burnt out and exhausted.* (Head of RE in a local authority school with a 'Good' Ofsted rating).

> *Teaching is no longer about doing the best you can for your pupils; it's about data and numbers. Pupils are not children anymore; they are dots on a graph which must look a certain way. I am leaving teaching because of changes to teaching. I have seen teachers ignore a class because they have to hit data deadlines because that is what is*

> *important these days.* (Head of year and design and technology teacher in a local authority school with a 'Good' Ofsted rating).

One teacher, who reported feeling 'disheartened' about the focus on core EBacc subjects at the expense of a broader curriculum and 'gutted' at her school's decision to stop offering Expressive Arts GCSE, commented very powerfully on her own intention to leave the profession:

> *After 32 years I feel at a loss as to what is expected of me. I am too scared to stay in role as the accountability is too much. The expectations are daunting. We all try our best for all students and the data is just overwhelming, feeling like a stick to beat us with. I am leaving at 54.* (Head of department and drama/expressive arts teacher in a multi-academy trust school - Ofsted rating not specified)

If we take a broad view of the purposes of school education that encompass a civic and critical pedagogic function, then schools ought to be sites where students can learn the habits of democracy and civic agency (Dewey 1916; Freire 2006; Giroux 2001). As Deborah Meier (2009) reminds us, one essential component of this involves students being able to observe their teachers exercising their own professional and civic agency. Whilst no doubt some teachers will strive to continue to exercise their own agency, as far as they can, within the narrow confines of the new order, the increasing likelihood is that students will observe their teachers working as rule-following operatives, with limited scope to use their own judgement let alone question the status quo. The ramifications of this (alongside the withdrawal of resources for citizenship education, mentioned above) for students' ability to learn - at least in school – critical and democratic habits of mind are disturbing.

Conclusion

We began this paper by charting some key moments in the history of the use of performance measures in English state education from the 1862 Revised Code through to the 2016 key stage 4 accountability reforms. Whilst there have been periods of rupture during this period, perhaps most notably in the 1960s and 1970s when teacher autonomy was arguably at its peak, in more recent years, performance measurement has become re-established as a hegemonic project through a cumulative process of policy sedimentation. As each set of measures has thrown up its own imperfections, subsequent iterations have been brought in to remedy them; and since the late 1990s, when the ideology of deliverology first began to appear as a powerful influence on governmental approaches to public service accountability, the performance measures being applied to English secondary schools appear to have become increasingly sophisticated. They have also become increasingly tied to a social justice agenda geared towards equalising access to 'powerful knowledge'. This increased sophistication, coupled with the social justice orientation of much of the rhetoric surrounding the new measures, has given the measures a *prima facie* plausibility that has led to them being positively welcomed by at least some progressive educators and in this way embedded as a new policy 'common sense'. We have argued, drawing on the survey data we have presented, that we need to be sceptical about this common sense.

Whilst the new measures appear in theory to be more aligned with social justice goals than the ones they replaced because they incentivise schools to provide greater support for lower-attaining students and encourage them to study 'a broad curriculum … with "high status" knowledge at [its] heart' (Francis *et al*. 2017), our survey respondents' reports of their experience on the ground present a very different picture. In particular, teachers reported that the anachronism, ethnocentrism and increased difficulty of the new GCSEs, coupled with the removal of tiering and modular assessment, make the qualifications less accessible to lower attaining students, those with special educational needs and those from low-income and EAL

backgrounds, and hence more likely to exacerbate than to narrow the attainment gap at key stage 4. Whilst in a lower stakes accountability environment, there would be greater flexibility for teachers to deviate from the mandated curriculum, and hence teach in ways that are more responsive to students' diverse capabilities and interests, the intense pressure to perform within the current high stakes regime substantially limits teachers' freedom of manoeuvre to mitigate the worst effects of the new measures.

Lying behind all of the concerns expressed by our survey respondents about the validity, reliability, power and utility of the measures is the idea that education is a very complex and contested set of goods. This is evident, for example, in the diverse conceptions of educational success signalled in their responses and in their criticisms of the ways in which the new measures appear to be prioritising traditional academic subjects and capabilities over practical and creative ones, and the rote learning of facts and formulae over the capacity to think creatively and critically and to apply knowledge in real-world situations. It is also evident in respondents' expressions of regret at the reduction in time and resources available for pastoral work, PSHE and citizenship education in their schools and their references to the dehumanising effects of a system of accountability that has generated a pervasive preoccupation with the quantification of student progress and the construction of students as data units.[24] And it is evident in the widely articulated belief that the new measures are seriously damaging the mental health of both students and teachers and will ultimately reproduce existing educational and social inequalities.

We would suggest that a climate in which some of this complexity and contestation is acknowledged, talked about and responded to is much to be preferred over one in which there is just one authoritative measure or set of measures of what counts as success. In the current climate, even where there is still some space to have intelligent conversations about these

[24] A phenomenon referred to elsewhere as the 'datafication' of schooling (Lingard, Sellar, and Savage 2014), about which there is now a burgeoning empirical literature. See, for example, Selwyn, Henderson, and Chao (2015), Finn (2016), Selwyn (2016), Bradbury and Holmes (2017), and Lupton and Williamson (2017).

things, there are high risks and costs attached to them, with any conversation a distraction from the core business of raising attainment against the measures. In a nutshell, the deliverology model constructs what counts in education in an at best contestable way, produces measures which are not anything like as robust as they are made to appear but, despite these things, powerfully defines the day-to-day life of schools and in so doing produce substantial injustices and harms, including inhibiting intelligent discourse about education. The cumulative effect of all of this is to take something that is rational at the level of policy generation and turn it into something that has emotional and social effects on the ground that are, for good reason, experienced by many teachers as monstrous in various ways.

Since we carried out the research in 2016, further adjustments have been made to the performance measures. For example, in 2018 the EBacc attainment measure was replaced by the 'EBacc Average Point Score' (APS), which measures pupils' point scores across the 'five pillars' of the EBacc (English, maths, the sciences, history or geography, and modern languages). This measure was designed with the laudable intention to ensure that 'the attainment of all pupils is recognised, not just those at particular grade boundaries, encouraging schools to enter pupils of all abilities, and support them to achieve their full potential' (DfE 2019: 12). At the same time, the Progress 8 measure was also refined to mitigate 'the disproportionate effect that a small number of extremely negative scores can have on a school's average progress' (*ibid*: 7), and the decision was made that schools falling below the 'floor' and 'coasting' standards could no longer face forced academisation unless also judged 'inadequate' by Ofsted (*ibid*).

Such continual adjustments are enabled by the relative ease with which large volumes of data can now be collected, digitalised, processed and publicly displayed. But what chance is there that these kinds of refinements will ever embody the level of sophistication that is

needed to do justice to the extreme complexity and contestability of educational goods? The long history of performance measurement in education recounted at the start of this paper, coupled with our survey data - and logic - tell us that, however sophisticated the measures, they will never be able to circumvent the inherent problem highlighted here. That is, that working to any externally imposed set of measures with high stakes consequences squeezes out the time, energy and 'head-space' teachers need if they are to be able to meaningfully engage in critical reflection and debate and make independent judgements that are grounded in and properly responsive to their in-depth first-hand knowledge of their students and the contexts in which they are working. As all of these things are integral to the bedrock of good teaching, including protecting the interests of disadvantaged students, it is necessary to conclude that, far from improving educational quality, deliverology is in certain fundamental respects incompatible with quality improvement.

References

Allen, R. 2016. *Submission to the House of Commons Education Select Committee Inquiry on Primary Assessment*. Accessed 14 July 2019. https://ffteducationdatalab.org.uk/wp-content/uploads/2016/12/Primary-assessment-Education-Datalab-submission.pdf.

Anders, J., M. Henderson, V. Moulton, and A. Sullivan. 2017. *Socio-Economic Status and Subject Choice at 14: Do They Interact to Affect University Access*? London: Nuffield Foundation.

Arnold, M. 1862. "The twice-revised code." In *The Complete Prose Works of Matthew Arnold, Vol II*. edited by R. H. Super. Ann Arbor: University Michigan Press.

Au, M. 2009. *Unequal by Design: High-Stakes Testing and the Standardization of Inequality*. London: Routledge.

Barber, M., A. Moffit, and P. Kihn. 2010. *Deliverology 101: A Field Guide for Educational Leaders*. Thousand Oaks, CA: Corwin.

Bew, P. 2011. *Independent review of key stage 2 testing, assessment and accountability*. London: Department for Education.

Biesta, G. 2009. "Good education in an Age of Measurement: On the Need to Reconnect with the Question of Purpose in Education." *Educational Assessment, Evaluation and Accountability* 21 (1): 33-46.

Bradbury, A. and G. Roberts-Holmes. 2017. *The Datafication of Primary and Early Years Education: Playing with Numbers*. London: Routledge.

Coldwell, M. and Willis, B. 2017. "Tests as boundary signifiers: level 6 tests and the primary secondary divide." *The Curriculum Journal*, 28 (4): 578-597.

Committee of the Privy Council on Education. 1862. *Revised Code of Minutes and Regulations of the Committee of the Privy Council on Education*. Accessed 14 September 2018. http://www.educationengland.org.uk/documents/cce/revised-code.html.

Connell, W. F. 1950. *The Educational Thought and Influence of Mathew Arnold*, London: Routledge and Kegan Paul.

Corbin, J. and A. Strauss. 2015. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded* Theory. 4th ed. Thousand Oaks, CA: Sage.

Cumming, J., Wyatt-Smith, C. and Colbert, P. 2015. "Students at Risk and NAPLAN". In *National Testing in Schools: An Australian Assessment*, edited by B. Lingard, G. Thompson and S. Sellar. London: Routledge.

DCSF (Department for Children, Schools and Families). 2006. *Technical Annex*. Accessed 10 February 2019. https://web.archive.org/web/20071110192100/http://www.dcsf.gov.uk/performancetables/pilotks4_05/annex.shtml

Dewey, J. 1916. *Democracy and Education: An Introduction to the Philosophy of Education*. New York: Macmillan.

DfE (Department for Education). 2010. *The Importance of Teaching*. Accessed 9 February 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/175429/CM-7980.pdf.

DfE. 2015. "Hundreds of 'Coasting' Schools to be Transformed". Accessed 9 February 2019. https://www.gov.uk/government/news/hundreds-of-coasting-schools-to-be-transformed.

DfE. 2016a. *Coasting Schools: Provisional Data*, *England*. Accessed 9 February 2019. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/566690/coasting_schools_note.pdf.

DfE. 2016b. *Progress 8 Measure in 2016, 2017 and 2018: Guide for Maintained Secondary Schools, Academies and Free Schools*. Accessed 9 February 2019. https://dera.ioe.ac.uk/27619/1/Progress_8_school_performance_measure_13_Oct_update.pdf.

DfE. 2019. *Secondary Accountability Measures: Guide for Maintained Secondary Schools, Academies and Free Schools*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/772801/Secondary_accountability_measures_guidance.pdf, accessed 9 February 2019.

Dörr, N., ed. 2013. *The Art and Science of Delivery*. New York: McKinsey. Accessed 3 February 2017. http://mckinseyonsociety.com/downloads/reports/Voices/ArtofDelivery-web.pdf.

Finn, M. 2016. "Atmospheres of Progress in a Data-Based School." *Cultural Geographies* 23 (1): 29-4.

Francis, B., M. Mills, and R. Lupton. 2017. "Towards Social Justice in Education: Contradictions and Dilemmas." *Journal of Education Policy* 17 (4): 414-431.

França, J. 2019. An interview with Henry Giroux, 2nd July 2019. Accessed 20 October 2019. http://lab.cccb.org/en/henry-giroux-those-arguing-that-education-should-be-neutral-are-really-arguing-for-a-version-of-education-in-which-nobody-is-accountable/

Freire, P. (2006) [1970]. "The banking model of education." In *Critical Issues in Education: An Anthology of Readings,* edited by E. F. Provenzo. Thousand Oaks, CA: Sage Publications.

Gamble, A. 1988. *The Free Economy and the Strong State: The Politics of Thatcherism*. London: Palgrave MacMillan.

Gewirtz, S. 2002. *The Managerial School: Post-Welfarism and Social Justice in Education*. London: Routledge.

Giroux, H. A. 1988. *Teachers as Intellectuals: towards a Critical Pedagogy of Learning*. Westport, CT: Bergin & Garvey.

Giroux, H. A. 2001. *Theory and Resistance in Education: Towards a Pedagogy for the Opposition*. Rev. and expanded ed., Westport, CT: Bergin & Garvey.

Giroux, H. A. 2014. "When Schools Become Dead Zones of the Imagination: A Critical Pedagogy Manifesto." *Policy Futures in Education* 12(4), 491–499.

Grace, G. 1985. "Judging Teachers: The Social and Political Contexts of Teacher Evaluation." *British Journal of Sociology of Education* 6 (1): 3-16.

Hardy, I. 2015. "A Logic of Enumeration: The Nature and Effects of National Literacy and Numeracy Testing in Australia." *Journal of Education Policy* 30 (3): 335–362.

HMSO (Her Majesty's Stationary Office). 1988. *Education Reform Act*. London: Her Majesty's Stationary Office. Accessed 18 March 2019. http://www.educationengland.org.uk/documents/acts/1988-education-reform-act.pdf

HMSO (Her Majesty's Stationary Office). 1992. *Education (Schools) Act*. London: Her Majesty's Stationary Office. Accessed 18 March 2019. http://www.legislation.gov.uk/ukpga/1992/38/pdfs/ukpga_19920038_en.pdf

Hofflinger, A. and von Hippel, P. 2018. "The Response to High-Stakes Testing in Chile: Did Schools Improve Learning or Merely Inflate Test Scores?" Accessed 20 October 2019. http://dx.doi.org/10.2139/ssrn.2906552

House of Commons, Committee of Public Accounts. 2018. *Retaining and Developing the Teaching Workforce*. Accessed 22 August 2018. https://publications.parliament.uk/pa/cm201719/cmselect/cmpubacc/460/460.pdf.

Hutchings, M. 2015. *Exam Factories: The Impact of Accountability Measures on Children and Young People*. Accessed 9 February 2019. https://www.teachers.org.uk/files/exam-factories.pdf.

Johnson, L. 2018. "'Advocacy Cheating': Teacher Noncompliance in Standardised Literacy Testing." *English Teaching: Practice & Critique*.17 (3): 168-181.

Klenowski, V. and Wyatt-Smith, C. 2012. "The Impact of High Stakes Testing: The Australian Story." *Assessment in Education: Principles, Policy & Practice* 19 (1): 65-79.

Klinger, D.A., S. Maggi, and A. D'Angiulli. 2011. "School Accountability and Assessment: Should we Put the Roof Up First?" *The Educational Forum* 75 (2): 114–128.

Lambert, R, and McCarthy, C. (Eds) 2006. *Understanding Teacher Stress in an Age of Accountability*. Greenwich, CT: IAP.

Lawton, D. 1980. *The Politics of the School Curriculum*. London: Routledge & Kegan Paul

Leckie, G. and H. Goldstein. 2017. "The Evolution of School League Tables in England 1992-2016: 'Contextual value-added', 'expected progress' and 'progress 8'." *British Educational Research Journal* 43 (2): 193-212.

Lingard, B., W. Martino, G. Rezai-Rashti, and S, Sellar. 2016. *Globalizing Educational Accountabilities*. New York: Routledge.

Lingard, B. and S. Sellar. 2013. "'Catalyst data': Perverse Systemic Effects of Audit and Accountability in Australian Schooling." *Journal of Education Policy* 28 (5): 634–656.

Lingard, B., S. Sellar, and G. Savage. 2014. "Re-articulating social justice as equity in schooling policy." *British Journal of Sociology of Education* 35(5): 710–730.

Lupton, D. and Williamson, B. 2017. "The Datafied Child: The Dataveillance of Children and Implications for Their Rights." *New Media and Society* 19 (5): 780-794.

Maguire, M., Gewirtz, S., Towers, E., and Neumann, E. 2019. "Contextualising Policy Work: Policy Enactment and the Specificities of English Secondary Schools." *Research Papers in Education*. DOI: 10.1080/02671522.2019.1601758

Meier, D. and Wood, G. (Eds) *Many Children Left Behind: How the No Child Left Behind Act Is Damaging Our Children and Our Schools*. Boston, MA: Beacon Press.

Meier, D. 2009. "Democracy at Risk." *Educational Leadership* 66 (8): 45-49.

Nash, R. 2010. *Explaining Inequalities in School Achievement: A Realist Analysis*. Edited by H. Lauder. Farnham: Ashgate.

Neumann, E., Towers, E., Gewirtz, S. and Maguire, M. 2016. *A Curriculum for All?* National Union of Teachers. Accessed 18 March 2019. https://www.teachers.org.uk/sites/default/files2014/curriculum-for-all-64pp-10845.pdf

Nichols, S. L. and D. Berliner. 2007. *Collateral damage: How High-Stakes Testing Corrupts America's Schools*. Cambridge, MA: Harvard Education Press.

Ohemeng, F. and E. McCall-Thomas. 2013. "Performance Management and 'Undesirable' Organizational Behaviour: Standardized Testing in Ontario Schools." *Canadian Public Administration* 56 (3): 456–477.

Picower, B. and Mayorga, E. 2015. *What's Race Got To Do With It? How Current School Reform Policy Maintains Racial and Economic Inequality*. New York: Peter Lang.

Pinto, L. 2016. "Tensions and fissures: the politics of standardised testing and accountability in Ontario, 1995–2015." *The Curriculum Journal* 27 (1): 95-112.

Polesel, J., S. Rice and N. Dulfer. 2014. "The Impact of High-Stakes Testing on Curriculum and Pedagogy: A Teacher Perspective from Australia." *Journal of Education Policy* 29 (5): 640-657.

Polesel, J., Dulfer, N. and Turnbull, M. 2012. *The Experience of Education: The impacts of high stakes testing on school students and their families.* Sydney: Whitlam Institute.

Rapple, B. 1994. "Payment by Results: An Example of Assessment in Elementary Education from Nineteenth Century Britain." *Education Policy Analysis Archives* 2 (1). http://epaa.asu.edu/ojs/article/viewFile/664/786

Ravitch, D. 2011. *The Death and Life of the Great American School System: How Testing and Choice are Undermining Education*. New York: Basic Books.

Russell Group. 2017/18. *Informed Choices: A Russell Group Guide to Making Decisions about Post-16 Education*. 6th ed. Accessed 19 March 2019. https://russellgroup.ac.uk/media/5686/informed-choices-2018-1-6th-edition-final.pdf

Sahlberg, P. 2011. *Finnish lessons: What Can the World Learn from Educational Change in Finland?* New York: Teachers College Press.

Sammons, P. 2008. "Zero Tolerance of Failure and New Labour Approaches to School Improvement in England." *Oxford Review of Education* 34 (6): 651-664.

Selwyn, N. 2016. "'There's So Much Data': Exploring the Realities of Data-Based School Governance." *European Educational Research Journal* 15 (1): 54 - 68.

Selwyn, N., M. Henderson, and S.-H. Chao. 2015. "Exploring the Role of Digital Data in Contemporary Schools and Schooling - '200,000 lines in an Excel spreadsheet'." *British Educational Research Journal* 41 (5): 767 - 781.

Sharratt, C. 2019. "Russell Group scraps A Level subject guidance." Accessed 22 July 2019. https://www.artsprofessional.co.uk/news/russell-group-scraps-level-subject-guidance.

Smith, L. 2015. *Understanding and Using Progress 8 and Other Features of the New GCSE English Examinations*. London: King's College London. Accessed 19 March 2019. https://www.letsthinkinenglish.org/wp-content/uploads/2013/08/Understanding-and-using-Progress-8-etc-full-version.pdf

Stuart Wells, A. 2019. "An Inconvenient Truth About the New Jim Crow of Education." Presidential Address, American Educational Research Association Annual Meeting, Toronto, 7th April.

Teacher Tapp. 2018. "A startling story on SATs cheating." Accessed 14 July 2019. https://teachertapp.co.uk/startling-story-sats-cheating/

Thompson, G. & Cook, I. 2014. "Manipulating the data: Teaching and NAPLAN in the control society." *Discourse: Studies in the Cultural Politics of Education* 35 (1): 129- 142.

Thrupp, M. and S. Tomlinson. 2005. "Introduction: Education Policy, Social Justice and 'Complex Hope'." *British Educational Research Journal* 31 (5): 549-556.

Tickle, L. 2017. "Disappeared: The Headteachers Sacked and Gagged by Academy Trusts." Accessed 9 February 2019. https://www.theguardian.com/education/2017/oct/24/disappeared-headteacher-sacked-academy-dismissal.

West, A. and H. Pennell. 2002. "How New is New Labour? The Quasi-Market and English Schools 1997 to 2001." *British Journal of Educational Studies* 50 (2): 206-224.

Wilson, D., B. Croxson, and A. Atkinson. 2006. “What Gets Measured Gets Done: Headteachers’ Responses to the English Secondary School Performance Management System.” *Policy Studies* 27 (2): 153–171.

Appendix – Open ended questions from the survey (with numbers of responses)

If you wish, please add any further comments you may have about the impact of Ebacc in your school. (463 responses)

If you wish, please add any further comments you may have about the impact of the new GCSEs on your subject or your school. (421 responses)

Have any other areas of teaching and learning or organisational arrangements been affected by P8? If you wish, please add any further comments on this. (277 responses) [Previous structured questions had asked whether in response to Progress 8 there had been an increase in targeting of extra provision (1:1 lessons, boosters, exam preps) and ability setting, a more rigorous approach to data gathering, assessing and tracking students, whether data on individual pupils had been used to inform lesson planning and whether teacher appraisal has become increasingly data-focussed.)

If you selected yes [to the question 'Are there any other groups which are likely to receive improved support because of Progress 8], which one(s)? (194 responses). [Previous structured questions had asked whether students with special needs, EAL, boys, girls, those high and low attaining in core subjects and those eligible for FSM are likely to receive improved support as a result of Progress 8.)

Which groups of students are most likely to benefit from Progress 8? (801 responses)

In what ways? (574 responses)

Which groups of students are most likely to be disadvantaged by Progress 8? (846 responses)

In what ways? (642 responses)

If you wish, please add any further comments you may have about the impact of Progress 8 and Attainment 8 in your school. (222 responses)

If you selected yes [in response to the question 'Is your school considering any form of workforce restructuring or re-training in response to the EBacc and the recent GCSE and

school accountability reforms (Progress 8 and Attainment 8)?’] please specify. (324 responses)

Is there anything else you would like to add about any of the questions in this survey? (162 responses)