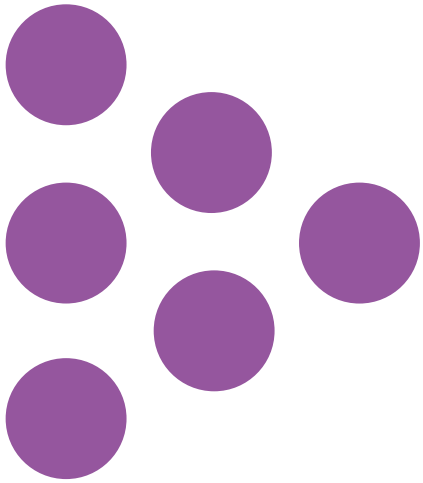
NFER
National Foundation for Educational Research

Report

What Impact Does Accountability Have On Curriculum, Standards and Engagement In Education?

A Literature Review

National Foundation for Educational Research (NFER)



What Impact Does Accountability Have On Curriculum, Standards and Engagement In Education?

A Literature Review

Frances Brill
Hilary Grayson
Lisa Kuhn
Sharon O'Donnell

Published 2018

By the National Foundation for Educational Research,

The Mere, Upton Park, Slough, Berkshire SL1 2DQ

www.nfer.ac.uk

ISBN: 978-1-911039-81-5

How to cite this publication:

Brill, F., Grayson, H., Kuhn, L. and O'Donnell, S. (2018). *What Impact Does Accountability Have On Curriculum, Standards and Engagement In Education? A Literature Review.* Slough: NFER.

NFER
National Foundation for Educational Research

Contents

Executive summary: what impact does accountability have on curriculum, standards and engagement in education?

Background In this review, we define accountability broadly as a government's mechanism for holding educational institutions to account for the delivery of high quality education. The idea that the practice of accountability can contribute *directly* to improvements in education is a powerful one that underpins policy. Paradoxically, though, some hold that accountability systems can also produce negative impacts on education, making it more difficult for schools to deliver the sought after quality. The question of what an optimal approach to accountability might look like is, therefore, intensely debated. The UK government's recent brief paper, *Principles for a clear and simple accountability system* (published 2018) foreshadows the launch of more detailed proposals for a government consultation scheduled for this autumn.

Aims NFER believes it is critical that research evidence should inform any rationales for policy change. The rapid literature review, reported here, aimed to evaluate a small body of international research evidence on the impact of accountability on three key areas: curriculum, standards and engagement.

Research question *What is the impact of different models of accountability in education on curriculum, standards, and teacher and pupil engagement and what factors affect this?*

Methodology We mapped the main features of accountability systems for primary education in 13 international jurisdictions. Six of the jurisdictions (Australia (New South Wales), England, Japan, New Zealand, Singapore and Wales) were finally selected to provide the focus for a small scale, rapid literature review of data studies and policy discussions. The initial searches retrieved 126 documents across the six selected jurisdictions; of these, a small set of 25 documents most relevant to the research question was selected for further appraisal. The literature identified was limited in scope and strength; it was dominated by small scale qualitative studies and reviews. The search identified few large scale investigations. Therefore, it was not possible to gauge quantitative impacts.

Thematic analysis of the 25 documents was undertaken in order to identify content related to the research question, although evidence was limited by the lack of quantitative research studies. The analysis yielded content relevant to relationships between accountability and the core topics of interest: curriculum, standards and teacher and pupil engagement. However, the severe limitations of the evidence base meant that it was not possible to fully determine impacts in these areas. Rather, it was the case that studies reported influences on curriculum, standards and pupils and teacher engagement that may be attributed, in part, to a jurisdiction's accountability system. The severe limitation of the evidence base must be taken into account in any interpretation of the findings.

Findings

Accountability and the curriculum

- Where pupil performance is used as a high stakes accountability measure, there is concern that certain parts of the curriculum become privileged above others at school delivery level, due to so-called 'teaching to the test'.
- Some pupils may receive an impoverished experience of the school curriculum as a result of targeted teaching where accountability systems focus on "borderline" or "cliff edge" measures.
- Jurisdictions may make deliberate system-level reforms to curriculum structure and documentation, typically in response to benchmarking the outcomes of international system comparisons.

Accountability and standards

- How accountability measures are carried out is important - the literature suggests three principles for a positive relationship between accountability and school effectiveness:
  - clarity over responsibilities
  - alignment of objectives at all levels of the system
  - transparency of criteria used for assessing performance.
- The application of accountability measures may increase the achievement gap (e.g. by focussing attention on the performance of 'borderline' pupils); or conversely they may be used to reduce the gap (e.g. by informing funding programmes for disadvantaged pupils).

Accountability and teacher and pupil engagement

- Teacher education can support teachers' engagement with assessment data to inform classroom teaching and learning.
- Pupils may become less engaged learners when undue emphasis is placed upon performance of some groups at the expense of others.

Evidence quality

- There is a paucity of data and robust, quantitative evidence about the impact of accountability on the curriculum, standards, and teacher and pupil engagement.
- In particular, there is little robust evidence about accountability on teacher workload, and teacher and pupil well-being.

Discussion and implications for policy

In our discussion, we focus on the two reported influences of accountability that were most strongly informed by the reviewed literature: curriculum narrowing and teachers' capacity to engage with data. It is important to note that the evidence base was limited; accountability operates within a specific context in each jurisdiction and that measures require translation to be applicable to alternative contexts.

The lack of impact identified does not necessarily imply that accountability does not have an impact on curriculum, standards and engagement. It is clearly difficult to draw out the impacts of accountability systems, for a number of different reasons. For example, there are many other factors that affect the quality of education, and it is difficult to isolate and tease apart the specific influences of accountability from those. In addition, it may take time for the effects of any accountability reforms to become apparent in a given education system.

1 Introduction

1.1 What is accountability?

Making sure that children and society receive maximum benefit from publicly-funded education is a high priority for governments worldwide. As governments are answerable for the use of public money, they, in turn, hold schools to account. In this way, a basic hierarchy of accountability - a familiar feature common across diverse national education systems - is constructed. Although there are many different definitions of accountability, in this review we define it broadly as a government's mechanism for holding education institutions to account for the delivery of high quality education. We understand the activity of school accountability in Stecher and Kirby's terms, as 'the practice of holding educational systems responsible for the quality of their products – students' knowledge, skills and behaviors' (Stecher and Kirby, 2004, p.1). Other forms of accountability (for example, financial integrity and individual school governance), though also important, are outside the scope of the present review.

Education accountability systems are complex, vary considerably across the world and are subject to change. As countries and jurisdictions strive to draw the most value from their education systems, it is not uncommon for governments to implement modest or extensive reforms to their accountability regimes. Motivation for change in national accountability systems – and education policy more widely – may, in some cases, be strongly related to a country's performance in international surveys such as the Programme for International Student Assessment (PISA), the Progress in International Reading Literacy Study (PIRLS) and the Trends in International Mathematics and Science Study (TIMSS). Indeed, the now-familiar reactive pattern of post-survey reform, often dubbed 'PISA-shock', has been well documented. For example, in an investigation of policy and media reactions to the 2009 and 2012 PISA results, Baird *et al.,* (2016) note, in respect to policy change in France, how an appetite for stronger accountability measures, partly driven by the PISA results, influenced the purpose and timing of school assessments:

> *The international and inexorable move towards greater outcomes-based accountability in education, a move reinforced by the OECD through PISA, saw France abandon a long-standing 'diagnostic' survey programme, in which testing took place at the beginning of key school years … in favour of the now familiar end-of-year model…*
>
> (Baird *et al.*, 2016, p.127)

Far from painting a picture of a static exercise in auditing education standards, conceptualisations of accountability tend to reflect the idea that the mechanism *itself* can be a dynamic agent of positive change. In 2016, in relation to proposed reforms to England's school accountability system, the Department for Education set out the belief that 'fair, robust, ambitious accountability is vital to monitor … standards, identify schools and areas that need extra support, and ensure children receive the education they deserve' (GB. Parliament. HoC, 2016). Implicit in this is the assumption that the accountability system plays an important role in bringing about improvements in education. Indeed, the whole purpose of accountability is widely accepted as one of strengthening the education system, rather than confirming the status-quo. As Ng observes, with reference to the literature on models of school accountability, 'Generally, it is assumed that the

goal of school accountability and its associated accountability-based interventions is to improve teaching and learning (Adams and Kirst, 1999; Darling-Hammond and Ascher, 1991; O'Day and Smith, 1993; O'Reilly, 1996) and (Ng, 2010, p.276).

Accountability approaches can take different forms. Burns and Köster (2016, p.25) describe how the term 'vertical accountability' is often used to describe 'top-down and hierarchical' accountability, which 'enforces compliance with laws and regulation and/or holds schools accountable for the quality of education they provide.' The question of which types of accountability approach are most likely to lead to successful outcomes is intensely debated. Some systems are underpinned by the idea that a high level of vertical accountability is necessary to deliver positive benefits. The wider education landscape is important here: in England, for example, the governmental proposals put forward to strengthen school-level accountability in 2016 (GB. Parliament. HoC, 2016, p.21) were in the context of policy decisions since 2010 that had given schools much greater autonomy. The rationale here is that when schools are granted more independence over the methods they use to achieve education outcomes, accountability increases in significance, as a 'more autonomous, school-led system depends even more heavily on a fair and effective accountability system, helping to identify any schools or areas that need extra challenge or support'.

However, there is also a perception that some forms of 'top-down' accountability can be counter-productive. Sometimes, accountability is deemed responsible for having negative influences on schools and education. For example, writing in the context of Australia's then 'new accountability regime', Lingard and Sellar (2013, p.634) argue that the use of national test results as a way of evaluating the performance of state education systems illustrates 'the wide scope for perverse incentives and effects to arise when funding and reputational capitals are tied to performance measures and comparisons' (ibid., p.651).

It is evident from cross-country comparisons that the type of accountability system adopted has far-reaching implications for schools themselves, and their relationships within the school system. For example, describing the accountability approach in Germany, Demski and Racherbäumer (2017, p.83) observe that 'Compared to, for instance, the United States, accountability is low in Germany, as there are no penalties for low-performing schools. Schools are neither placed on probation nor closed following poor test results. Furthermore, insufficient results in student testing do not lead to the replacement of school leaders or teachers'. This is cast in a positive light in terms of data use in schools, with the authors noting that, 'A low degree of accountability also has consequences for data-driven school improvement. In this regard, practitioners' willingness to use data seems to play an important role in trying to explain data use, as principals and teachers have room for manoeuvre'.

However, so-called 'lighter touch' accountability approaches may be subject to accusations of ineffectuality and are perhaps themselves responsible for a decline in education standards. In a large-scale quantitative comparison of school accountability practices in England and Wales, Burgess *et al.,* (2013, p.57) draw attention to the abolition of secondary school performance tables (known as 'league tables') in Wales in 2001, but not in England – thus removing 'a key element from the accountability system of two otherwise-identical education systems'. This study found 'systematic, significant and robust evidence that abolishing school league tables markedly reduced school effectiveness in Wales relative to England' (Burgess *et al.*, 2013, p.58).

Some have suggested that there has been a move away from strongly hierarchical 'top down' accountability models towards what, in Burns and Köster's (2016, p.25) terms, can be described as 'Horizontal accountability', which 'presupposes non-hierarchical relationships'. For example, (Robinson *et al.*, 2011, p.725), drew attention to 'a shift from more bureaucratic top-down forms to more emphasis on accountability to internalized professional norms, to peers and to parents and students.' Structures such as school-to-school or peer-to-peer partnerships and support systems may be regarded as examples of horizontal accountability, or as measures to support school improvement. How these measures relate to the overall concept of accountability is complex, since the relationships within them differ so markedly from hierarchical ones. Ehren and Perryman (2017, p.3) articulate and analyse the considerable tensions and challenges posed in situations where school 'networks and network governance' are introduced whilst there remains the legacy of 'existing accountability structures, most of which were developed to support hierarchical control of individual school quality' (ibid., p.1).

1.2 Accountability measures

Accountability measures may involve wide-ranging targets encompassing many aspects related to school and education governance and quality. These can include, for example, financial management, pupil well-being, behaviour and safety, as well as gauging standards of pupil attainment or progress against academic curricula. In this review, our interest lies, in particular, in two domains that are frequently used in accountability measures: (1) pupil assessment and (2) school evaluation (including school inspection).

Many countries and jurisdictions use students' test results from their national statutory assessment programmes as accountability measures. The challenges of doing this have been well-documented. Over a decade ago, concerns were raised in England that using national tests for a range of purposes including accountability may not be entirely satisfactory: The House of Commons Select Committee concluded that using national test results for the purpose of accountability "has resulted in some schools emphasising the maximisation of test results at the expense of a more rounded education for their pupils" (Stobart, 2009, p.173). Concerns notwithstanding, it is evident that gathering student performance data and making judgements about school effectiveness based upon it have long been central to many school accountability systems. In Levin's (1974) framework for accountability, the performance reporting process is one of four[1] accountability concepts and has been described as 'about reporting the performance of schools, usually based upon examination and other key student results, under the assumption that the information on such results enables stakeholders to appraise school effectiveness' (Ng, 2010, p.276). Of course, in addition to national tests, the reporting of performance at country level in the international surveys (PISA, PIRLS, TIMSS, etc.) is another assessment-based metric that can be used by governments as a means of assessing a country's performance against others, and as a starting point for introducing reforms into the system: Johansson (2016) points to a benefit of international large-scale assessments being their usefulness as a 'measure of the achievement trend *within* countries, particularly for countries with long-standing participation records'

[1] The other three are: a technical process, a political process and an institutional process (Levin, 1974).

(Johansson, 2016, p.145). Such data can also be used by governments to set school-based accountability targets.

In addition to using student assessments as a central component of a school accountability system, it is very common for a programme of external school inspection or evaluation to be used as a main accountability tool. This can be generally defined as an objective appraisal of the effectiveness of key aspects of a school's performance, including leadership, quality of teaching and pupil attainment and progress. It involves an accountable body outside of the school carrying out the evaluation. Evaluation procedures generally involve physical school inspection. However, there are also alternative, non-inspection based modes of evaluation: for example, evaluation of schools can involve the conducting of surveys of students, parents and teachers.

Relatively recently, the notion of school self-evaluation has gained currency and is considered to be an important component of some evaluation systems. As Vanhoof and Petegem (2007, p.261) observe, partly 'in response to recent trends with regard to decentralization and increasing autonomy for schools - evaluation methods have been developed in many countries which permit more participatory and self-directed forms of evaluation'. For example, New Zealand's Ministry of Education guidance document on 'How to do and use internal evaluation for improvement' (ERO, 2015, p.41) includes a quotation from Nusche *et al.,* (2012) to describe a consensual approach that integrates aspects of internal and external evaluation: 'both parties attempt to work together to agree on a rounded picture of the school in which there is mutual recognition of its strengths and consensus on areas for development' (Education Review Office, 2015, p.41). School support mechanisms, along with school-to-school collaborations (OECD, 2017) and peer-to-peer school evaluation systems are equally important areas within the domain of school evaluation.

The use of school-to-school and/or peer-to-peer evaluation approaches are likely to be concurrent with some form of external evaluation, although this is not necessarily the case. Ultimately, though, these forms of school evaluation differ in one important respect: in Burns and Köster's (2016) terms, they are 'horizontal' rather than 'vertical' systems. They may feed into the accountability hierarchy but do not replace it, as schools are usually ultimately accountable to governments (and other stakeholders) rather than each other. Here, it may be helpful to adopt Vanhoof and Petegem's distinction between two quality assurance perspectives: one focused on accountability and the other focused on school improvement:

> *The distinction between the two perspectives is based on different answers to the questions of (1) whether quality assurance is primarily concerned with monitoring and accountability or rather with development and improvement and (2) to the question of who determines "quality of education", in other words: the government or the school itself.*
>
> (Vanhoof and Petegem, 2007, p.264)

Elsewhere, the role of school-to-school collaboration is described somewhat differently. In a series of conceptual pieces about school-to-school relationships and school improvement in England, (Hargreaves, 2012, p.4) argued that 'clusters of schools working in partnership could potentially create a self-improving school system' and that the notion was supported by the government of the day: 'inter-school partnerships are flourishing in many different forms across thousands of

schools in England in response to the coalition government's policy of transferring the main responsibility for teacher development and school improvement away from local authorities and other providers and directly to schools themselves'.

As the discussion above indicates, a complex relationship exists between school accountability approaches on the one hand and school improvement and education outcomes, on the other. It is equally apparent that all systems have different advantages and disadvantages. However, it was hoped that by examining the features of different accountability models used in a range of different education settings, alongside documented evidence of impact on standards, curriculum and engagement, it may be possible to bring some insight into these issues and identify lessons that may help improve the effectiveness of the accountability system in England, whilst minimising any unintended consequences.

1.3 Purpose

Accountability is clearly an area of significant importance to education policy makers and other stakeholders in the education community. We believe it is critical for policy to be underpinned by research evidence. Against this background, the review aimed to identify and evaluate research evidence on the impact of different types of accountability systems. The potential breadth of 'accountability' as a concept has been discussed in the section above. We took, as our research focus, accountability of schools to government (either directly or indirectly, via more local levels in the hierarchy) for education standards. In restricting our scope, we acknowledge that we have not covered many other important aspects of accountability in education, including, for example, accountability for other types of performance metrics (e.g. the use of public funds).

A broad intention of the review was to offer some evidence-based insights into the best way forward for education policy in England and Wales. The timing of this study was designed to coincide with the National Association of Head Teacher's (NAHT's) review of accountability in England, and to provide some complementary objective evidence for policy makers.

The research question we sought to address was as follows:

What is the impact of different models of accountability on curriculum, standards, and pupil and teacher engagement, and what factors affect this?

For reasons of manageability, it was necessary to limit the scope of the inquiry to a single phase of education. The primary phase of education was selected as, for most countries, this represents the stage of the education system where pupils in a given country or jurisdiction are more likely to experience one common approach (as opposed to the various different school 'tracks', such as academic and vocational, that characterise the secondary school landscape in many settings). The review was defined further by a focus on a selected sample of countries or jurisdictions. These were: England, Wales, Australia (New South Wales), Japan, New Zealand and Singapore. These countries were selected as they represent a diverse range of systems (as explained in more detail below).

2 Methodology

2.1 Selection of relevant jurisdictions

Initial overview of 13 jurisdictions

The first step for selecting relevant countries and jurisdictions for this literature review was a systematic mapping of accountability features for a wider group of countries and jurisdictions, focusing on the domains of assessment and evaluation. Initially, there were 13 jurisdictions of interest. These were: England, Wales, Australia (New South Wales), Canada (Alberta), Estonia, Finland, Germany, Poland, Japan, New Zealand, Singapore, USA (Massachusetts) and Sweden. The rationale for selecting these countries and jurisdictions was as follows. Given our interest in education policy in England and Wales (noted in the previous section), England and Wales were given the role of 'reference countries'. The intention was to draw out comparisons with the accountability features of a geographically and culturally diverse selection of countries or jurisdictions. The systematic mapping involved interrogating information sources for answers to a detailed series of questions about assessment and evaluation for each country or jurisdiction. The sources we interrogated were primarily the websites of the relevant ministries of education, curriculum and assessment bodies, as well as the OECD. The findings from the mapping were then used to populate an overview table that presented, for each country or jurisdiction, information relevant to the following key questions for assessment and evaluation:

<u>Assessment questions</u>

Is there statutory national or jurisdiction-wide assessment in the primary phase (as defined by the jurisdiction)?

When during the primary phase does assessment take place?

Are assessment findings reported?

Are the assessments used to hold schools accountable for pupil attainment and progress?

<u>Evaluation questions</u>

Is there statutory external evaluation and does it include school inspection?

Are inspection outcomes reported?

Is there a requirement for self-evaluation?

Is there school-to-school or peer-to-peer support for school improvement?

The overview is presented in Appendix A. All of these 13 jurisdictions showed interesting patterns or accountability systems and presented a variety of assessment and evaluation strategies. The comparison showed that six out of these 13 jurisdictions carry out some form of statutory national assessment to hold schools accountable for pupil attainment and progress. All 13 jurisdictions have some form of external evaluation; of these, nine include inspection.

Selection of a subset of eight jurisdictions for initial literature search

On the basis of this overview, we narrowed down the initial jurisdictions to arrive at eight jurisdictions to search for more detail on: Australia (New South Wales), England, Canada (Alberta), Japan, New Zealand, Singapore, Sweden and Wales. The six countries/jurisdictions Australia (New South Wales), Canada (Alberta), Japan, New Zealand, Singapore and Sweden were selected as they appeared, from our initial mapping, to represent the most diverse and interesting subsample of assessment and evaluation approaches in comparison with England and Wales. Accordingly, an initial search of the literature was conducted for these eight jurisdictions.

2.2 Initial search

Search remit and parameters

The remit of the literature review was to draw out, in particular, any findings regarding the impact of accountability systems on:

- curriculum
- standards of attainment
- school improvement support mechanisms
- teacher workload
- teacher engagement/well-being/retention/recruitment
- pupil engagement/well-being
- school management decision-making
- perverse incentives and unintended consequences
- closing the gap between disadvantaged and non-disadvantaged pupils
- cost.

We employed the following search parameters/key inclusion criteria:

- Publications in the English language only, published between January 2010 and April 2018.
- The focus was on primary school years and the following study types were included: peer-reviewed and grey literature featuring statistical analyses of large datasets; nationally representative surveys; large-scale qualitative studies; we also searched for literature reviews, other qualitative and quantitative studies and evidence-based opinion pieces.
- A range of education bibliographic databases, websites and government reports served as search sources, in addition to which we harvested references from the reference lists of key reports. A full list of key words included in the search, together with the full search strategy, can be found in Appendix B.

Identification of six jurisdictions for full review

The evidence available for each jurisdiction varied in quantity as well as quality, resulting in a low number of papers for two jurisdictions (Canada (Alberta) and Sweden). This constrained the geographical scope of this work further to six jurisdictions in total with a reasonable amount of published evidence. These six were: England, Wales, Australia (New South Wales), Japan, New Zealand and Singapore. The searches retrieved 126 documents across all of the six selected jurisdictions. Overall, we found that studies that use data were primarily reporting on existing data rather than producing new analyses; also, the search identified few statistical analyses of large datasets, nationally representative surveys or large-scale qualitative studies.

2.3 Document selection

For these six selected jurisdictions, we identified, from the 126 documents, a small set of 25 documents for further appraisal. We selected those papers that were of most relevance and offered sufficient research quality (relative to the wider set of documents) for the key questions we sought to address – i.e. the impact of different models of accountability on curriculum, standards, and pupil and teacher engagement, and the factors that affect this. We also ensured that each jurisdiction was covered in two or more documents. The characteristics of the selected 25 pieces are detailed in a table in Appendix C of this report. As noted above, the search did not identify many large-scale studies; this is reflected in the table characteristics. It is evident from this table that the group of 25 pieces of most relevance to the search is dominated by small, qualitative studies, reviews and conceptual pieces. Although these studies all contribute research insights into the themes we addressed, the small scale pieces are clearly not able to offer indications of robust, statistically significant impacts. The severe limitation of the evidence base must, therefore, be taken into account in any interpretation of the findings. It should be noted that other research sources were also used to inform and contextualise this literature review. These sources are referenced separately at the end of the review.

The final distribution of 25 papers across jurisdictions for the critical appraisal and synthesis broke down as follows: England (11 documents), Wales (3 documents), Australia (New South Wales, 4 documents), Japan (4 documents), New Zealand (4 documents) and Singapore (3 documents). Some documents discussed more than one of the target jurisdictions, hence the individual quantities total to more than 25.

2.4 Literature appraisal

For the literature appraisal, we then extracted evidence for each of the themes of interest from each paper, together with an estimation of how relevant and how strong the evidence base was for each study. Following this appraisal, which identified the most salient and relevant findings, we identified three key themes related to our initial research questions: **curriculum**, **standards of attainment** (with closing the gap as a specific element), and **teacher and pupil engagement**. These core topics formed the basis for the thematic review, allowing a detailed discussion and evaluation of the target papers.

3 Findings and discussion

3.1 Background: country accountability thumbnails for the selected jurisdictions

As noted above, the literature review focused on six countries/jurisdictions: England, Wales, Australia (New South Wales), Japan, New Zealand, and Singapore. Below, as background to the thematic findings, we briefly outline, for each country, the very basic features of their school accountability systems in the domains of assessment and evaluation, derived from the accountability overview. It should be noted that these descriptions are broad-brush and do not give contextual information about the 'direction of travel' for a given country/jurisdiction (i.e. whether a country/jurisdiction is moving towards or away from a 'heavy' or 'light' accountability system).

England and Wales

In the United Kingdom, education policy is devolved to the home nations. The education systems in place for England and Wales are, therefore, different. However, there are commonalities in the overall accountability systems. In terms of assessment, both countries operate a statutory national assessment programme in the primary phase of education. The assessments are used to hold schools accountable for pupil attainment and progress, although in different ways. In both England and Wales, statutory external evaluation is carried out; in both cases, it involves school inspection. There is a requirement for school self-evaluation, and there are mechanisms for school-to-school support for school improvement.

Australia (New South Wales)

In Australia, there is statutory national assessment in the primary phase. The assessments are used to hold schools to account. In New South Wales, school evaluation involves a process of statutory annual self-assessment. In a five-year cycle, school self-assessments are validated by an external panel. There is a mechanism for peer-to-peer support for school improvement.

Japan

In the Japanese education system, there is statutory national assessment at the end of the primary phase. However, the assessments are not used to hold schools accountable for pupils' attainment and progress. As Nakayasu, (2016) explains, the aim is:

> […] *to check achievements and problems with national educational policies through collecting and analysing students' academic skills and learning conditions. This is done largely for the purpose of maintaining a uniform level of education and improving the level of education, rather than to check achievements in each school.*
>
> (Nakayasu, 2016, p.144-5)

Japan operates a system of statutory external evaluation which does include school inspection. In Japan, there is a statutory requirement for school self-evaluation and there is a mechanism for school-to-school support for school improvement.

New Zealand

New Zealand operates a national monitoring assessment in the primary phase: the National Monitoring Study of Student Achievement (NMSSA). This is based on nationally representative samples and does not test the cohort of pupils as a whole. These assessments are not used to hold schools accountable for pupil attainment and progress. The New Zealand government's information about the NMSSA states that the purpose 'is to get a broad picture of student achievement in New Zealand' and that 'The focus is on growth in educational achievement across time at a *national* level. National monitoring does not produce information about individual students, teachers or schools' (NMSSA, 2018). There is statutory evaluation of schools in New Zealand, and this does include inspection. School self-evaluation is a statutory requirement.

Singapore

Singapore operates a self-assessment model (the School Excellence Model, or SEM). Within this framework, schools self-evaluate and also undergo external validation. There is statutory national assessment of pupils at the end of the primary phase, for the main purpose of determining pupils' secondary school pathways (NFER and Arad Research, 2013).

3.2 What impact does accountability have on the curriculum?

One of the main aims of this review was to investigate the potential impact that accountability systems may have on the curriculum. Analysis of the literature suggests that influences on the curriculum attributed to accountability are as follows:

- accountability-driven school practices leading to a narrowing of the curriculum at school delivery level, as a consequence of so-called 'teaching to the test', particularly in educational settings where pupil performance in a limited range of subjects is used as a high stakes accountability measure
- an impoverished experience of the school curriculum for lower performing pupils, as a result of accountability-led resources being targeted elsewhere
- deliberate system-level reforms to curriculum structure and documentation, typically as a policy response based on the use of international survey data for broad benchmarking purposes.

Whilst the first two are largely perceived as unintended and negative effects, there are some indications that the third may have a positive impact on curriculum development, given the right conditions. The following section will discuss literature relating to these key aspects in greater detail, across a range of reviewed countries and jurisdictions.

3.2.1 Curriculum narrowing

Generally, the literature we reviewed perceived a troublesome relationship between high-stakes, accountability-driven assessment systems and the adoption of teaching practices that lead to over-emphasis on some parts of the curriculum at the expense of others. Characteristically, this involved analysis of school-based situations where certain curriculum areas become privileged, because teachers are conscious that these will count towards the school's publicly reported rating or ranking. Time and resource may then be eroded for subjects that do not receive such

accountability scrutiny; at the same time, targeted curriculum areas may be over-rehearsed as a result of classroom practices that are often termed 'teaching to the test'.

Concerns over curriculum narrowing as a perceived consequence of assessment has been reported in several of the reviewed documents across a range of jurisdictions (Bew, 2011 and Coldwell and Willis, 2017 (England); OECD, 2014 (Wales); Ng, 2010 (Singapore); Klenowski and Wyatt-Smith, 2012 and Lingard and Sellar, 2013 (Australia); Nusche *et al.*, 2012 (New Zealand). It is important to note, however, that this review did not identify large scale studies that provide quantifiable evidence of curriculum narrowing. Rather, the literature we reviewed typically offers documented reports of perceptions of curriculum narrowing and the attribution of this to assessment-related accountability requirements. For example, the UK government-commissioned review into national curriculum testing in England (Bew, 2011) was based on a 12-week call for evidence from interested parties. It notes concerns raised by some respondents:

> *We are aware that many teachers and head teachers feel that the current combination of statutory assessment and the school accountability system constrains schools, compelling them to over-focus on what is assessed. Many heads have told us in discussion that they 'need' to concentrate much of Year 6 teaching on preparation for National Curriculum Tests in order to prevent results dropping. We recognise that the accountability system to date may appear to have encouraged this kind of behaviour.*
>
> (Bew, 2011, p. 23–4).

Taking a historical perspective on the relationship between accountability and the curriculum in England, Coldwell and Willis (2017) discuss the idea that, although the introduction of a national curriculum for England and Wales in 1988 (Department of Education and Science, 1988) offered the provision of a broad curriculum, there may have been early signs of a narrowing of the curriculum (in terms of how the curriculum was delivered in the classroom), due to the use of national assessment for accountability purposes. The authors argue from previous literature that, from an early stage, 'the assessment of the curriculum at the end of Key Stage 2 as it was implemented, looking only at English, Mathematics and – at that time – Science, narrowed the focus' (Coldwell and Willis, 2017, p.580).They cite a longitudinal questionnaire survey analysis (the Monitoring Curriculum and Assessment project) by Boyle and Bragg (2006), which is also reported in Wyse and Torrance (2009). According to Wyse and Torrance, in Boyle and Bragg's study, a nationally representative sample of primary phase schools responded annually (1997 to 2004) to questions about the approximate percentage of teaching time spent on different subjects. As Wyse comments, the analysis leads Boyle and Bragg to 'suggest that their data point to a significant reduction in the broad and balanced curriculum as a result of central policy requirements' (Wyse and Torrance, 2009, p.219).

At this stage of education, it has also been reported that pressure associated with high stakes national curriculum testing may be a factor in reduced curriculum creativity (Troman, *et al.,* 2007), In their qualitative study that placed attention on more able pupils, Coldwell and Willis (2017) interviewed 80 school leaders and teachers across primary and secondary schools. One recurring finding from the thematic analysis was a perception expressed by secondary school teachers that the more able students had been taught to pass a test rather than to access a deeper level of

learning within a broad curriculum. For example, concern was expressed that the emphasis on mathematics and English at primary school, particularly for pupils entered for Level 6 tests in mathematics and reading, meant that pupils were arriving at secondary school with 'poor knowledge' (p.587) of other subject areas. Primary teachers reportedly felt that they faced a dilemma between preparing their pupils well for a specific test (i.e. a narrow experience) and offering pupils a balanced and broad curriculum.

Further indications of a relationship between standards-based accountability measures and curriculum narrowing come from a study of education in Wales. In a report about Welsh education practice, the OECD (2014) observes that Wales's policy focus on literacy and numeracy may create an unintended consequence of a greater concentration on tested subjects in the classroom. Whilst there is the intention to avoid this tendency, by demonstrating how literacy and numeracy can be taught across disciplines and the overall curriculum, the OECD report suggests a close monitoring of the potential shift in balance to higher stakes testing.

The phenomenon of schools implicitly or explicitly concentrating their efforts on outcomes relevant to the assessment system has also been reported in Singapore, despite the fact that assessment is not used primarily for school accountability purposes. For example, in a review of the literature, Ng (2010) draws attention to the suggestion that some schools overemphasize preparation for certain tests at the expense of acquiring more general skills. This was especially true if the test contributed to key performance indicators, such as physical education and the National Physical Fitness Test (Tan, 2005). An inclination to 'teach to the test' in order to satisfy performance targets has also been reported to heighten school competition and thereby, it is suggested, reduce cooperation amongst schools in Singapore (Ng and Chan, 2008).

The introduction of high-stakes testing in Australia is regarded, by some, as the harbinger of undesirable practices described as 'gaming the system', leading away from the provision of a broad curriculum. For example, Klenowski and Wyatt-Smith (2012) put forward such a viewpoint in their review of observations made by the Australian Primary Principals Association (APPA, 2010). Specifically, APPA proposed that the development of high-stakes national testing resulted in a narrowing of the curriculum and a lack of attention paid to curriculum areas that are not tested (APPA, 2010). A research study by Lingard and Sellar (2013), which included data from 30 research interviews of senior policy-makers, personnel in international organisations, researchers and academics across three Australian states (Victoria, New South Wales and Queensland), chimes with this interpretation. The study reported that the degree of 'gaming the system' to protect 'reputational capital' (p.634) varied between states, but Queensland, in particular, demonstrated signs of unintended consequences, such as 'teaching to the test' (p.647).

New Zealand provides an interesting example of reported curriculum narrowing in relation to the implementation of national standards. Officially, there is no full-cohort national testing system in place in New Zealand; however, Nusche *et al.,* (2012) report a negative perception of high-stakes accountability and national testing amongst professionals. In 2010, New Zealand saw the introduction of national standards in schools. Following an expert consultation, it was reported that there were concerns that 'the introduction of Standards increases the risk of a narrower focus on numeracy and literacy in primary schools' (Nusche *et al.*, 2012, p.53). From 2018 onwards, these

National Standards will be removed again, to be replaced with the National Monitoring of Student Achievement (https://www.education.govt.nz/news/national-standards-removed/).

It is clear that the reviewed literature identifies concerns about curriculum limitations that are regarded as a consequence of accountability. However, it is important to note the complexities of this issue, in terms of multiple perspectives. In some senses, 'curriculum consequences' may be regarded as entirely *intentional* effects by policy-makers. For example, there may be deliberate attempts, through an accountability mechanism, to bring about a curricular focus on core subjects (e.g. numeracy and literacy), if it is felt that insufficient attention is being paid to subjects which are key to engaging in other aspects of learning and accessing the rest of the curriculum.

3.2.2 Impoverished curriculum experience for pupils

There are suggestions in the literature that some pupils, typically the less able, may have their exposure to the school curriculum limited by a concentration of resources on other pupil groups who become the focus of accountability measures. The phrase 'educational triage' (Marks, 2014, p.38) is used to describe 'a process of goods distribution whereby a number of linked practices are enacted to achieve a specified aim, usually related to maximising attainment outcomes' (Marks, 2014, p.38). In her small case study (88 pupils in one year 6 (pupil age 10 -11) group in a single Greater London primary school), the author describes the unintended negative consequences that can arise. In this scenario, certain groups of pupils, such as 'borderline' students who are below but within reach of a grade that is significant in the accountability system, are specifically targeted with additional preparation and coaching. It was found that within this particular school, 'triaging' took place, in response to accountability measures, with the specific aim of increasing the number of pupils meeting government targets. Low-performing students not in this targeted group appeared to have reduced mathematical learning experiences which, in turn, increased the attainment gap. As Layard and Dunn (2009) observe elsewhere, in these types of situations there is little extrinsic incentive for schools to focus on improving low performers' scores, and this pattern was also evident in secondary schools during GCSE preparation (Layard and Dunn, 2009).

Similar concerns have also been raised by an OECD report on the Welsh education system (OECD 2014). Literature within the OECD review suggests a danger of focusing on pupils who are just below the proficiency threshold, in order to meet school targets (Hargreaves and Shirley, 2009). This would have the negative consequence of focusing on pupils whose performance is most likely to improve sufficiently to attain the target, but placing less emphasis on other lower (or indeed higher) performers. In turn, this may work against the government's target to close the gap between advantaged and disadvantaged students (OECD, 2014).

In Australia, there are, similarly, indications of a belief that high-stakes testing may have a negative impact on the curricular experiences of low-performing students, as the focus of resources is on pupils close to the academic threshold. It has even been suggested that parents may have been encouraged to not let their low-performing child take part in the test (APPA, 2010).

Whilst the availability and strength of evidence varies greatly between reviewed countries, several papers identify concerns that where curriculum support has led to a focus on students that can positively contribute to the school targets, the incentive can become somewhat lower to support the curriculum experience for low-performers who are unlikely to reach the target threshold. In

England, the emphasis in the accountability system has shifted towards measures of "progress" rather than threshold measures such as the "C/D" borderline in order to reduce these negative incentive effects.

3.2.3 System-level curricular reforms

International benchmarking as part of system-wide "accountability" can have a marked effect on curriculum policy. The publication of findings from international comparisons of pupil performance, such as PISA, has galvanised system-level curriculum reform. The process of curriculum revision is spurred on by the belief that curriculum improvements can lead to improved pupil performance. For example, Nakayasu (2016) provides insight into this via a case study from Japan, noting 'that the Japanese government was really concerned about the decline of the ranking in PISA in the early twenty-first century' (p.146). As a direct result, Japan revised its national curriculum 'in response to children's skills which the PISA aimed to measure (key competencies)' (p.134). In Japan, whilst national curriculum standards are highly centralised, individual teachers are responsible for how the curriculum is taught and teachers have considerable authority over classroom practice (OECD 2012; Miki *et al.*, 2015). Apart from key competency areas, the Japanese system adopted a model which places emphasis on cross-disciplinary skills such as independent thinking and problem solving. To improve PISA performance, a curriculum revision in 2008-2009 included a new main goal, which was to teach the 'Zest for Life': a combination of academic skills, morality and physical health, in order to encourage proactive learning.

The notion that curriculum changes can lead to increased performance is also evident in relation to Singapore. In order to be able to achieve and maintain high standards, Singapore pays 'Serious attention to curriculum development' (OECD, 2012, p.124) whilst also making sure that teachers are well equipped to teach and deliver this curriculum, reportedly resulting in strong programmes across a range of core subjects. In terms of continuous development, despite being amongst the highest scoring countries in the PISA studies, Singapore developed Curriculum 2015 (Singapore Ministry of Education, 2010) which included socio-emotional skills and 21st century skills (e.g. critical and independent thinking, civic responsibilities, communication and forming relationships). The aim was to develop a more active learning experience, following the concept of 'teach less, learn more', in which the curriculum content was reduced by 10 to 20 per cent in order to allow for a wider range of teaching approaches whilst reducing content-overload (OECD, 2012, p. 126).

The literature reports some other findings that suggest an impact on the curriculum stemming from the local use of international survey data. The OECD (2012) review suggested that student achievement data from international studies such as PISA can help to identify areas of the curriculum that could be enhanced. For example, PISA data from 2009 showed that 83 per cent of students in Japan were in schools that use achievement data to identify aspects of instruction or the curriculum for improvement (OECD, 2012, p.84).

3.3 What impact does accountability have on school standards?

As noted in the introduction, governments readily use measures of pupil achievement and school evaluation as a way of measuring education quality. The literature reflects a particularly close relationship between accountability and standards of attainment. Jurisdictions frequently express

their aims for the accountability systems they establish in terms of ensuring that schools or students meet defined performance standards. A number of themes emerged from the literature, although the limited nature of the evidence base means that findings must be interpreted with extreme caution. The literature identified influences related to:

- the operation of accountability
- the public reporting of accountability outcomes, and
- the influence of accountability on the widening or lessening of the achievement gap between particular groups of students.

3.3.1 The operation of accountability

The literature suggests that how accountability measures are implemented can affect the extent to which there is confidence in standards.

Clarity in terms of responsibilities appears important here. For example, Robinson *et al.,* (2011) identify, from international comparisons, that clearly assigned responsibility for system progress and performance is one of the features of high-performing education systems:

> […] *high performing systems have systematic institutional routines for the improvement of practice at both system and school level. At system level, there is a clearly identified agency or agencies that are responsible for the progress and performance of the system as a whole.*
>
> (Robinson *et al.*, 2011, p.726)

Alignment of objectives at all levels of the education system is also needed. Nusche *et al.,* (2012) OECD expert review of evaluation and assessment in New Zealand education offers a detailed analysis of the country's assessment and evaluation frameworks, and an exploration of how these can be used to improve student outcomes. The review recommended that standards, curriculum and assessment should be better aligned to provide a more coherent national evaluation and assessment agenda, and to prevent inconsistencies in evaluation practice, because current variations in practice across New Zealand put the degree of consistency in doubt.

Another OECD review (2014) proposed a comprehensive strategy to support equity and quality in Wales' school system. This study provides analysis, using PISA data, of the identified strengths and challenges of the school system in Wales, and proposes recommendations for improvement. It makes a related point, that while Wales is increasingly focusing on evaluation and assessment, its arrangements are lacking in synergy. In particular, the review noted a lack of coherence between the school evaluations conducted by the inspectorate, Estyn, and the school banding system that assigned schools to one of four categories depending on the level of support they were deemed to need to make improvements.

There is also a suggestion in the literature that criteria used for the assessment of performance need to be clearly discernible. OECD's review of Wales makes an important recommendation in this regard: that the criteria or calculation methods used for assigning schools to certain categories (in this case to bands designating the level of support judged appropriate) should be transparent.

Additionally, there needs to be alignment between criteria for performance assessment and what is targeted for improvement. Lingard and Sellar's (2013) qualitative research and policy review studied actions taken by Australian educators (teachers and administrators) in response to the national reform agenda, particularly the country's national testing programme (NAPLAN). The study is based on data from 30 interviews with policy makers, personnel from international organisations, and researchers and academics. It provides a case study with two parts, which analyses effects related to the use of national test data for accountability purposes. The authors report how, in one scenario, performance targets were based on what was easy to measure rather than, arguably, on aspects that were more important. The paper refers to a potential for 'goal displacement' (p.649) – i.e. a situation where the aspects measured by internal state-wide indicators did not align with the domains targeted by the national testing programme (NAPLAN). Poor performance by one state in NAPLAN 2008 led to the introduction of Teaching and Learning Audits and state-wide targets for improvement in 2010; however, the authors found no direct relationship between a school improving its standing against the audit domains and its pupils achieving improved performance on NAPLAN. They considered that there was a risk of pressure to improve areas that scored low in the audit, even though that may not result in the desired performance in NAPLAN.

3.3.2 The public reporting of accountability outcomes

The literature indicates that transparency is important when it comes to reporting on the aspects a given jurisdiction chooses to include in its accountability system. In a conceptual piece, setting out a re-balanced system of accountability, Gilbert (2012) considers the history of accountability in England and acknowledges that, in the 1990s, 'Publishing information on all schools had a profound impact on the national debate around education. In particular, it shone a light on poor performance and poor service' (Gilbert, 2012, p.7). As mentioned in the Introduction, in the context of secondary school pupil attainment in Wales, ceasing to publish performance tables has been shown quantitatively to have a negative effect on standards: 'policy reform in Wales reduced average performance and raised educational inequality relative to England' (Burgess *et al.*, 2013, p.66).

The phenomenon of results publication being associated with increased school effectiveness is also alluded to in another form in a quantitative study by Hutchinson and Dunford (2016), pertaining to multiple stages of schooling (early years, primary and secondary) in England. This report explores the growth of 'progress gaps' (p.17). In the study, these are defined 'as the number of months of *additional* academic development experienced by non-disadvantaged pupils, compared with the progress made by disadvantaged pupils' (Hutchinson and Dunford, 2016, p.17). This is to enable identification of *when* disadvantaged pupils are falling further behind, and thus when and where additional efforts are most needed to improve outcomes for disadvantaged pupils. The study suggests increasing the prominence of three-year aggregated measures of progress and attainment in school performance tables, especially for primary schools where numbers of disadvantaged pupils in a single year group may be too low for the statistics to be published.

The issue of reporting on the performance outcomes of specific pupil groups by providing a more detailed breakdown of the results by different characteristics also arises in Australia, with regard to this country's statutory national assessments in literacy and numeracy. Bien (2016), analysing

NAPLAN results for the period 2008-13, identified that these are disaggregated only for indigenous students and for students whose family language background is other than English, and not for all disadvantaged groups. As parents may withdraw their children from the test, there is an incomplete picture as to who has participated and to whom the results apply. It is thus more difficult to determine the impact of tests if the breakdown of results is not sufficiently granular.

Decisions, and changes in the decision-making about the public reporting of assessments are inevitably deeply intertwined with a nation's education history, reminding us that context is ever-important in the interpretation of different approaches to accountability. Nakayasu's (2016) article on *School Curriculum in Japan* explains that 'Japanese principles and methods of accountability for school education are different from other developed countries because of its own system of regional education administration' (p.144).

An interesting aspect of this system is that the Japanese Ministry of Education, Culture, Sports Science and Technology (MEXT) does conduct a national assessment – the National Survey on Academic Skills and Learning Conditions. How the results of the survey have been published has changed over time. Before 2013, MEXT 'published the results at national and prefectural level, but not at municipal and school level, and it had prohibited each prefecture and municipal from releasing the survey results by school' (Nakayasu, 2016, p.145). After this time, there was a policy change 'and it was proposed that prefectures and municipalities should be allowed to release the survey results for each jurisdiction because schools should be made accountable' (p.145). However, according to Nakayasu, 'there are only a few municipalities that show willingness to release the results by school, only 32 municipalities out of 1736 released the percentage of questions answered correctly by school in 2014' (p.145). Thus, even though Japan has moved from a situation where there was *prohibition* of the publication of national assessment results by school, it is still distinct from other centralised accountability systems because the decision of whether or not to release results resides *locally* in the system. It is important to understand this approach in the historical context of the decentralisation of the Japanese education system. It would be difficult to imagine a country or jurisdiction with a more centralised education system proposing to devolve the decision-making about the publication of results down to localised education constituencies in quite the same way.

3.3.3 Closing the gap

The *achievement gap* is a term commonly used to describe the

> […] *differences in pupil attainment associated with social class, ethnicity and gender…The achievement gap …is a shared concern for the education policies of many countries around the world with governments frequently introducing globalised new approaches to schooling, aimed at narrowing this gap*
>
> (Goodman and Burton, 2012, p.500)

There are different views about whether accountability measures have a positive or negative impact on the achievement gap. In the literature we reviewed, there were some suggestions that accountability data can be used productively as a starting point in efforts to reduce the gap.

However, in other studies, there were indications that some accountability approaches can contribute to a widening of the gap, where these encourage a focus on other groups of pupils.

The use of accountability data to reduce the achievement gap

Whilst most education systems will battle with achievement gaps to some degree, some literature suggested ways of using information derived from accountability data (such as performance data) to optimise funded support for disadvantaged students, in targeted attempts to address the negative relationship between achievement and disadvantage (e.g. Macleod *et al.*, 2015). A study from England (National College for School Leadership, 2011) focusing on disadvantaged pupils draws attention to the use of attainment data in the context of school-to-school support. There are suggestions that primary and secondary schools that were supported by the initiative of 'National Support Schools (NSSs)' (p.5) contributed to the closing of an attainment gap between pupils eligible for free school meals (FSM) and those who were not. This study reports that the schools which were 'supported by an NSS for more than one year showed that the attainment of pupils eligible for FSM in these supported schools improved at a faster rate than national averages between 2008 and 2010' (p.5). Japan offers another example of performance data being used to address efforts to close the gap between disadvantaged and non-disadvantaged students. One specific way disadvantage is tackled is by allocating more teachers to low-performing schools with more disadvantaged students, therefore having a more favourable student-teacher ratio in disadvantaged schools and spending more per disadvantaged student than the Japanese average (OECD, 2012).

The increasing achievement gap as consequence of high-stakes accountability

Whilst national testing and publication of school league tables are expected to raise school effectiveness, the literature suggests that these high-stakes outcomes of assessments also raise the pressure on schools to perform at certain standards. Consequently, Goodman and Burton (2012) contend that this may result in a tendency to support pupils who can positively contribute to the expected school standards outcomes more than pupils who may not be able to contribute to the expected standard outcome. This may include disadvantaged pupils, or pupils with English as an additional language (EAL), who work well below the expected achievement threshold, where accountability systems focus on achieving absolute attainment standards. The Bew Review (2011) received consultation responses suggesting that, at the time, disadvantaged pupils, such as those with special educational needs (SEN), might even be actively discouraged from taking part in official assessment practices.

Evidence for an achievement gap has also previously been identified for Wales. Generally, data from PISA 2012 suggested that Wales has a relatively equitable education system, with 10.4 per cent of variance in Welsh student performance in mathematics explained by the students' socioeconomic background, compared to the OECD average of 20.8 per cent (OECD 2014, p.21). However, the OECD (2014) suggested that whilst Wales has moved towards more higher stakes and summative full-cohort assessments in recent years, this has also increased the pressure to focus on children just below the system's threshold for proficiency so that they can meet school-level targets, leaving other groups out (Hargreaves and Shirley, 2009). This may result in

attainment levels rising only for a specific sub-group of pupils, while widening the achievement gap.

Australia's government aimed to use accountability measures in order to raise equality and opportunities for all students by signing the Melbourne Declaration on Educational Goals for Young Australians in 2008. Bien (2016) analysed national data from Australia's NAPLAN tests for Year 3 and 5 students to identify changes in the achievement gap from 2009 to 2013. Whilst the data suggested a trend towards higher attainment scores overall, Bien's analysis found that

> *Despite some encouraging gains, categories indicative of a less-advantaged school were consistently associated with lower achievement in reading and numeracy across both year levels* […] *Unfortunately, the proficiency gaps between the two most disadvantaged groups* […] *and their advantaged counterparts have significantly widened or remained unchanged.*
>
> (Bien, 2016, p.225)

In all studies into achievement gaps, it is important to note, however, that reported differences (positive or negative) cannot be attributed directly to accountability systems: as discussed earlier in this report, accountability systems are themselves part of complex interrelated educational systems and there are many other factors that influence the attainment of pupil groups.

3.4 What impact does accountability have on teacher and pupil engagement?

In this section, we focus on how the reviewed literature addressed the impacts of accountability on teacher and pupil engagement. In terms of teachers, the literature made a distinction between two engagement-related themes. The first of these involved reports of how teachers engage with the requirements of accountability, both in terms of assessing pupils effectively and interpreting and making best use of data generated by accountability measures. A second grouping involves issues of workload, tensions and pressures which have an effect on teachers' levels of motivation and engagement in their role as teachers. In the literature we reviewed, findings relating to pupils largely centre on test anxiety.

3.4.1 Teacher engagement

Assessment literacy and data literacy

One of the findings to emerge from the literature was the reported need among teachers for preparation to engage effectively with assessment data, and the difficulties that are encountered when teachers do not feel adequately prepared. For example, Bien (2016) reports from earlier literature (Pierce & Chick, 2011; Wayman *et al.*, 2007 and Young and Kim, 2010) that the use of assessment data for internal accountability is challenged when schools lack the technological capability to operationalise data or do not have a culture of data use, and when teachers lack the necessary analytical skills. In some education settings it is the *requirement* for accountability that has highlighted the importance of supporting teachers to develop a stronger understanding of education data. Formative assessment – to identify children's learning needs and inform the next

steps in teaching – is a crucial element of effective pedagogy. In some senses, then, accountability could be regarded as having a positive impact here, as it has exposed a weakness in teacher education and provided impetus for its redress. Certainly, more generally, there is some evidence to support the education benefits of teacher engagement with data and the importance of providing teachers with professional learning opportunities (see, e.g. Bien, 2016).

Elsewhere, a rapid evidence assessment undertaken by NFER and Arad Research (2013) identified that in high performing systems, assessment data is used for summative purposes (monitoring, accountability and reporting) and for formative ones (identifying learning needs, tailoring teaching and learning approaches, and ensuring that pedagogical approaches meet the needs of the individual child). They noted that in many such countries, initial teacher education and continuing professional development opportunities emphasise the importance of building teacher capacity to undertake assessment and engage with the resulting data. Likewise, Nusche *et al.,* (2012), in their OECD review of New Zealand, note that teachers need to develop valid and reliable assessment tools so that they can independently collect and interpret school-wide assessment data.

Public accountability pressures and teacher motivation and workload

The accountability system has been claimed as a source of frustration for teachers, impacting on their morale and levels of satisfaction with their jobs (see, for example, Lynch *et al.,* 2016) for an investigation of why some teachers leave the profession). Some of the reviewed studies investigated how pressures associated with public accountability measures can have negative impacts on teachers' professional experiences and their sense of engagement with their work. Adoniou's (2016) small-scale study of newly qualified teachers (n=14) in Australia explored the impact of education reform agendas, specifically national standardised literacy testing and public reporting of results, on teachers and their intentions to remain in the profession. Participants expressed frustrations with aspects such as workload (judged to have increased due to accountability requirements), lack of autonomy and how the reform agendas were operationalised. It is important to note that other factors influenced these early career teachers' motivations, including their own skill sets, school contexts and the levels of support they received (e.g. the extent to which they felt supported by the structure of the school's reporting and assessment arrangements). Bien's (2016) study, also based partly in Australia, concluded that accountability pressure 'can stifle teacher creativity' (Bien, 2016, p.320). One principal gave the example that, although his teachers would like to focus on wider student development, 'they would not do it at the expense of redirecting attention from the core subjects.' (Bien, 2016, p.320). In England, Marks' (2014) case study of the organisation of mathematics teaching in the final year of one English primary school identified that the teachers were keenly aware of being accountable to a range of stakeholders, including parents, the local authority, the school inspection body Ofsted and government, and of the pressure to maximise outputs. The deputy head of the case study school felt compelled to focus on end of primary school national test results, viewing them as 'the currency that the Government [uses to] check that schools are working properly' (Marks, 2014, p.50).

Like Adoniou (2016), Wilkins (2011) undertook a small-scale study of teachers (N=18), this time in England. These individuals, all graduates of a single university's primary PGCE programme and working as newly qualified teachers (NQTs), reflected on their experiences of autonomy and

accountability. Wilkins' background literature review identified negative effects of the growth of a performance management culture within schools, including 'low morale, job dissatisfaction and a sense of diminished autonomy' (Wilkins, 2011, p.390), in contrast to school inspection (Ofsted) and pupil attainment data that suggested to the author largely positive outcomes. Conversely, in Wilkins' own study, the majority of the participants appeared 'sanguine' (p.400) about the demands of bureaucracy involved in monitoring pupil progress and taking on extra-curricular responsibilities. A number of them noted, unprompted, a generational shift in attitudes to recorded accountability, feeling that, at the time of the study, new and early career teachers regarded it as part of the job, in contrast to colleagues who had been teaching longer. Wilkins notes that these individuals' own experience as pupils was of an increasingly performative schooling system, which perhaps made them generally more comfortable with the balance between accountability demands and the desire for autonomy.

3.4.2 Pupil engagement

Test anxiety is an impact of assessment often featuring in public discourse and is discussed in several pieces of reviewed literature. Goodman and Burton's (2012) review article investigating the complexity of the achievement gap in England suggests that anxiety and pressure may be a particular issue for disadvantaged pupils. At the other end of the ability distribution, Coldwell and Willis's (2017) qualitative research into primary schools' use of optional tests for high performing pupils at the end of primary education (now discontinued) found some schools suggesting that the pressure of testing could have a negative impact on pupils' well-being and confidence, which affected the schools' decision-making over whether to enter able pupils for these tests. Kuramoto and Koizumi (2016), reviewing national tests against the culture and principles governing Japanese education, identify 'test aversion' among pupils in Japan, where they argue that tests are perceived as leading to a competitive classroom atmosphere and to anxiety (p.1). Ng's (2010) Singapore study identified a stakeholder view that another form of accountability measure, school ranking, placed great pressure on pupils. A further side-effect of assessment noted by Ng (2008) is the possibility to render pupils 'passive learners, driven externally to perform but not necessarily engaged or inspired' (Ng, 2010, p.279). Ng judges that there was an encouraging change following the introduction of Teach Less, Learn More (TLLM) in 2005. TLLM aims to develop 'engaged learners [through] curriculum and pedagogical reform […and] to achieve among students [outcomes including] self-directed learning, deep understanding of concepts, appreciation of subjects, and knowledge construction and sharing' (Ng, 2010, p.281).

Literature suggests that the relationships between assessment and accountability are complex; not least because there are many other factors and influences that affect pupils' experiences of assessment. For example, assessments often have implications for pupils that are distinct from the accountability system. It is, therefore, not easy to differentiate the extent to which it is testing that is carried out for the purposes of *accountability* that generates anxiety. From the limited scope of this review, it is not possible to comment on this and other relevant matters such as, for example, the levels of test anxiety, the extent to which pupils and/or schools considered the tests to be high stakes, and changes in experiences over time. Other important areas for exploration would include insights into the approaches taken by schools to minimise test anxiety.

3.5 Wider discussion and policy implications

As discussed above, the literature we reviewed did enable insight into the influences that accountability can have on important aspects of education. However, this was a small scale, rapid review; the limited evidence base means that it is not possible to draw substantial conclusions about impact. Instead, we identify some points and issues that those involved in developing more effective accountability approaches may wish to consider.

Given that, as recognised in the introduction, the underlying aims of accountability policies are to improve education, it is surely of interest and concern that many of the impacts identified in the literature appear to be – or at least are regarded as – negative ones. However, teasing out the potential performance benefits of different accountability systems in a robust way is more complex, and perhaps why so few studies have attempted to quantify the overall impact of accountability on education standards.

It must also be acknowledged that public debates about these impacts can pave the way for policy discussions and reforms that seek to address some of the issues and reconceptualise accountability as a result. The review of international jurisdictions does suggest that all accountability systems have benefits and limitations. Whilst the functioning of accountability systems is obviously context dependent, there is value in learning from the identified strengths and weaknesses perceived in different accountability systems. The phenomenon of **curriculum narrowing**, and the **professional capacity for teachers to engage with data** were two areas that emerged particularly strongly in terms of reported accountability impacts.

In terms of curriculum narrowing, we explored reports of this phenomenon as a consequence of so-called 'teaching to the test'. Pieces of literature across different jurisdictions identified this as a practice associated with accountability measures based on pupil attainment. These findings are further supported by research which highlighted the prioritisation of English Baccalaureate subjects following their inclusion as performance measures, alongside teacher supply and school funding as a major influence on curriculum time (Worth, 2017). In the context of Wales, it is interesting that *Successful Futures*, the Independent Review of Curriculum and Assessment Arrangements in Wales (Donaldson, 2015) also found evidence that 'assessment and accountability had become unhelpfully intertwined in ways that weaken the power of both to serve children's learning' (p.105). The report concluded that 'Overemphasis on a small range of outcomes (especially when they are linked to high-stakes assessment) risks narrowing the curriculum and there is evidence that this is the case in significant numbers of schools in Wales' (Donaldson, 2015, p.75).

An important question here is whether there is a way of breaking the reported link between schools feeling compelled to focus on curriculum areas that are most salient for accountability purposes at the expense of other areas that do not have accountability consequences. Clarity over what is expected through the inspection regime is a key driver of school behaviour, and whether schools focus on those subjects which are assessed, or take a wider view of the curriculum. Of particular interest here is a preliminary commentary on the emerging findings of a research programme into curriculum implementation, conducted by England's school inspection body, Ofsted (Spielman, 2017). The section on 'Narrowing of the primary curriculum' describes an emphasis on test preparation in some primary schools and a perception by some parents that 'test preparation had

reduced the teaching time available for the other foundation subjects or for reading for pleasure.' Spielman notes that 'This is not the first time we have seen evidence of a narrowing curriculum in primary schools. As far back as 2001, we reported that the National Literacy and Numeracy Strategies, along with increasingly demanding performance targets, had adversely affected the breadth of the primary curriculum.' In this piece, there is also recognition that accountability, in the form of school inspection, may have played its part in creating conditions where it may be easy for school leaders 'to focus on the performance of the school and lose sight of the pupil.' There are indications that this recognition will help to inform the nature of future inspection frameworks.

Another recurring theme in the literature was the complexity of accountability and the suggestion that the training teachers receive may not align with the requirements their jurisdictions' accountability systems place on them. Literature from four of the jurisdictions – Australia, New Zealand, Singapore and Wales – suggested that teachers' initial training might not adequately prepare them to be fully assessment literate and data literate – to have a comprehensive understanding of how to implement assessment or of how to interpret assessment or other outcome data. This chimes with the findings of two recent explorations of aspects of the education system in England, the *Carter Review of Initial Teacher Training* and the *Commission on Assessment without Levels* (both 2015). As noted earlier, this might be seen in some ways as a positive impact of accountability, in that it has exposed an area of weakness in professional development that should be addressed anyway, irrespective of the accountability system in place. In other words, support for teachers to understand how to use assessment data to support their teaching and learning should be part and parcel of any professional set of teaching skills.

In England, the Carter Review identified assessment as the area of initial teacher training (ITT) programmes in need of most improvement, highlighting significant gaps in both schools' and ITT providers' capacity in the theoretical and technical aspects of assessment. It recommended that 'Assessment, including the theories of assessment and technical aspects of assessment, should be part of a framework for ITT content'; and that 'Alongside a central portal on evidence-based practice, a central repository of resources and guidance on assessment should be developed' (Carter, 2015, p.9). The Commission on Assessment without Levels agreed that the quality of assessment training was currently too weak and reiterated the importance of schools taking up opportunities to train staff in assessment, beyond their ITT: '[E]very teacher should have the opportunity to become skilled and confident at assessing pupils' learning. Furthermore, there should be an explicit expectation that school leaders and the Ofsted inspectorate develop a rigorous and shared understanding of all aspects of assessment' (p.8). *Successful Futures* (the Donaldson Review, 2015) likewise recommended that the Welsh Government should develop teacher capacity to practice assessment effectively, noting that the then ongoing Furlong review of initial teacher training (Furlong, 2015) provided opportunities for exploring this.

More generally, recent reforms in England have aimed to address some of the unintended consequences this review has discussed, with the removal of assessment levels and refocusing of the accountability system onto progress measures rather than absolute standards (STA, 2017). It will be interesting to see how far this approach reduces perverse incentives in the system in the future. Meanwhile, the current policy focus appears to be on developing an accountability system that is 'clear and simple' (DfE, 2018), with publication of the *Principles for a Clear and Simple*

Accountability System (2018), ahead of more detailed proposals and a full consultation in the Autumn.

What, then, might different, more effective approaches to accountability look like? Schleicher (2018, p.115) suggests that they may involve a move in emphasis towards 'professional accountability'. Within 'administrative accountability' lies the use of pupil performance and other school evaluation data to make decisions about school quality. Professional accountability, though,

> […] *refers to systems in which teachers are accountable not so much to administrative authorities but primarily to their fellow teachers and school principals. Professionals in most fields feel themselves accountable to other members of their profession. In the case of education, professional accountability also includes the kind of personal responsibility that teachers feel towards their peers, their students and their students' parents.*
>
> (Schleicher, 2018, p.116).

Interestingly, Schleicher gives Canada (Ontario), Finland, Japan and New Zealand as examples of jurisdictions that exemplify aspects of this collaborative, less hierarchical approach; moreover Singapore is described as a system where 'administrative and professional accountability are combined.' (p.117). It is pertinent here to return to the conceptual distinction between vertical and horizontal accountability (Burns and Köster (2016)), discussed at the beginning of this review. Since, in most cases, schools are ultimately accountable to governments for the use of public money, a degree of hybridisation is likely to be involved, such that horizontal structures (such as school-to-school networks) can complement, or feed into the vertical system.

Limitations

As noted earlier, a characteristic of this review was the limited nature of the evidence base. Although the studies identified by the review are all of interest and value, it was not possible to draw firm conclusions about the impact of accountability as we lacked a body of large scale, robustly-conducted quantitative studies. One possible reason for this is that investigations of this nature are costly and complex: however, given that accountability systems are expensive entities funded by public money, we argue that such studies will be worthwhile investments. Well-conducted quantitative studies into the effects of accountability may provide powerful and convincing evidence for reform. In particular, there remain many unanswered questions about effective ways of integrating so-called 'horizontal' forms of accountability (such as school-to-school and peer-to-peer partnerships) into national accountability systems where, ultimately, the aim of accountability is to demonstrate to the public that state-funded schools are delivering high standards of education for their children.

Directions for future research

The methodological challenges involved in investigating the impact of accountability in a scientifically robust way must not be underestimated. However, where circumstances permit and change in the accountability regime takes place without other simultaneous reforms, it can be possible to explore subsequent impacts by making comparisons with previous performance or with the performance of matched countries who have not experienced a change. The work of Burgess

et al., (2013, p.57) referred to in this review provides an example of such a 'natural experiment' with school accountability. In that study, the researchers were able to isolate a change in the accountability system and measure its impact.

Our review points to a need for greater clarity over roles and responsibilities within accountability systems. We would recommend these are further explored, with work to understand the system and wider set of actors who influence school effectiveness. For example, we need greater understanding about the entire system within which schools sit to be able to allocate responsibility and thus accountability for children's attainment and subsequent outcomes. An examination of place-based education and associated accountability mechanisms could provide helpful insight particularly for horizontal accountability.

It would also be beneficial to investigate how horizontal models of accountability work most effectively in other countries, and how they are best integrated into vertical systems. Closer examination of the relationship between vertical and horizontal accountability and school improvement would also give us insight into the wider ramifications of accountability and the experience of teachers and students.

Internationally, a range of approaches are taken to facilitating and supporting change for school improvement. We recommend further exploration and evaluation of the support for school and teacher development offered in different accountability systems, including from key national bodies as well as "peers".

The influence of accountability on children and young peoples' experience of school is an important finding from the review. Research about effective ways of attenuating negative impacts of accountability on curricula should be a priority and include the important role of evaluation and the agencies involved in the accountability system. Equally important is further exploration of test anxiety in children and young people to identify effective ways to mitigate any negative influence of high stakes assessment.

The review highlights the value of publicly available data as an aid to accountability and to the maintenance or improvement of school standards. However, we have also found that the confidence and competence of various actors in the education system, including teachers, in using data can be a limiting factor. We would, therefore, recommend that further research be undertaken to explore the options around what data, presented with what contextual and other information is most effective in identifying school performance and guiding decision making by stakeholders.

Concluding comments

That accountability operates within a particular cultural and organisational context in each jurisdiction is, of course, a given: the cultural outlook on education and its principles that pertains in a jurisdiction informs how stakeholders in that system view its accountability arrangements. This has implications for how far a practice that is more successful in one national context may or may not be transferable internationally - it is right to be cautious about unquestioning 'policy tourism'. It must be borne in mind that findings from the literature from one jurisdiction may only translate successfully to another if their context is recognised. As Burdett and O'Donnell (2016) indicate: 'to

be used effectively in policy-making, the evidence on educational performance needs to be correctly and thoroughly interpreted in context.' (Burdett and O'Donnell, 2016, p.113).

References

Literature sources identified by the search process

Adoniou, M. (2016). 'Beginning teachers' responses to education reform agendas', *School Leadership and Management*, **36**, 1, 80–95.

Australian Primary Principals Association (APPA). 2010. *The Reporting and Use Of NAPLAN*. Hindmarsh, S. Australia: APPA. Cited in: 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.

Bew Review (2011). *Independent Review of Key Stage 2 Testing, Assessment and Accountability*. London: DfE [online]. Available: https://www.gov.uk/government/publications/independent-review-of-key-stage-2-testing-assessment-and-accountability-final-report [5 September, 2018].

Bien, N.M. (2016). *Primary School Achievement Gaps and School Decisions to Support the Academic Achievement of Disadvantaged Students with Data: a Cross-Country Comparative Study*. PhD Thesis, University of Sydney [online]. Available: https://ses.library.usyd.edu.au/bitstream/2123/15742/1/bien_nm_thesis.pdf [5 September, 2018].

Burgess, S., Wilson, D. and Worth, J. (2013). 'A natural experiment in school accountability: The impact of school performance information on pupil progress', *Journal of Public Economics,* **106**, 2013, 57–67.

Coldwell, M. and Willis, B. (2017). 'Tests as boundary signifiers: level 6 tests and the primary secondary divide', *Curriculum Journal*, **28**, 4, 578–97.

Education Review Office (2015). *Effective School Evaluation*. Wellington: ERO [online]. Available: http://www.ero.govt.nz/assets/Uploads/Effective-School-Evaluation-web.pdf [5 September, 2018].

Gilbert, C. (2012). *Towards a Self-Improving System: the Role of School Accountability*. Nottingham: NCSL [online]. Available: http://dera.ioe.ac.uk/14919/1/towards-a-self-improving-system-school-accountability-thinkpiece%5B1%5D.pdf [5 September, 2018].

Goodman, R. and Burton, D. (2012). 'What is the nature of the achievement gap, why does it persist and are government goals sufficient to create social justice in the education system?' *Education 3-13*, **40**, 5, 500–14.

Hargreaves, A. and Shirley, D. (2009). *The Fourth Way: The Inspiring Future for Educational Change.* Thousand Oaks, CA: Corwin Press. Cited in: Organisation for Economic Co-operation and Development (OECD) (2014). *Improving Schools in Wales: An OECD Perspective*. Paris: OECD [online]. Available: http://www.oecd.org/education/Improving-schools-in-Wales.pdf [5 September, 2018].

Hargreaves, D. H. (2012). *A Self-Improving School System: Towards Maturity*. Nottingham: NCSL [online]. Available: http://dera.ioe.ac.uk/15804/1/a-self-improving-school-system-towards-maturity.pdf [5 September, 2018].

Hutchinson, J. and Dunford, J. (2016). *Divergent Pathways: the Disadvantage Gap, Accountability and the Pupil Premium*. London: EPI [online]. Available: https://epi.org.uk/publications-and-research/divergent-pathways-disadvantage-gap-accountability-pupil-premium/ [5 September, 2018].

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.

Kuramoto, N. and Koizumi, R. (2016). 'Current issues in large-scale educational assessment in Japan: focus on national assessment of academic ability and university entrance examinations', *Assessment in Education: Principles, Policy & Practice* [online]. Available: https://www.adrec.ihe.tohoku.ac.jp/wp/wp-content/uploads/2016/11/KuramotoKoizumiin-press.pdf [5 September, 2018].

Layard, R. and J.A. Dunn. (2009). *Good Childhood: Searching For Values In A Competitive Age*. London: Penguin. Cited in: Goodman, R. and Burton, D. (2012). 'What is the nature of the achievement gap, why does it persist and are government goals sufficient to create social justice in the education system?' *Education 3-13*, **40**, 5, 500–14.

Levin, H. (1974). 'A conceptual framework for accountability in education', *School Review*, **82**, 3, 363–91. Cited in: Ng, P.T. (2010). 'The evolution and nature of school accountability in the Singapore education system', *Educational Assessment, Evaluation and Accountability*, **22**, 4, 275–92.

Lingard, B. and Sellar, S. (2013). '"Catalyst data": perverse systemic effects of audit and accountability in Australian schooling', *Journal of Education Policy*, **28**, 5, 634–56.

Macleod, S., Sharp, C., Bernardinelli, D., Skipp, A. and Higgins, S. (2015). *Supporting the Attainment of Disadvantaged Pupils: Articulating Success and Good Practice*. London: DfE [online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/473974/DFE-RR411_Supporting_the_attainment_of_disadvantaged_pupils.pdf [5 September, 2018].

Marks, R. (2014). 'Educational triage and ability-grouping in primary mathematics: a case-study of the impacts on low-attaining pupils', *Research in Mathematics Education*, **16**, 1, 38–53 [online]. Available: [29 June, 2018].

Miki, T., Toledo Figueroa, D., Peterka, J., Fraccola, S. and Ikesako, H. (2015). *Education Policy Outlook:* Japan. Paris: OECD [online]. Available: https://docslide.us/documents/education-policy-outlook-japan-oecd-policy-outlook-japan-oecd-2015-6-equity.html [September, 2018].

Nakayasu, C. (2016). 'School curriculum in Japan', *Curriculum Journal*, **27**, 1, 134–50.

National College for School Leadership (2011). *System Leadership: Does School-to-School Support Close the Gap?* Nottingham: NCSL [online]. Available: http://www.isospartnership.com/uploads/files/system-leadership-does-school-to-school-support-close-the-gap.pdf [5 September, 2018].

NFER and Arad Research (2013). *A Rapid Evidence Assessment on the Impact of Curriculum and Assessment Arrangements within High Performing Countries*. Cardiff: Welsh Government [online]. Available: Available: https://gov.wales/docs/caecd/research/131022-rapid-evidence-assessment-impact-curriculum-assessment-arrangements-high-performing-countries-en.pdf [5 September, 2018].

Ng, P. T. (2008). 'Educational reform in Singapore: from quantity to quality', *Educational Research for Policy and Practice*, **7**, 1, 5–15. Cited in: Ng, P.T. (2010). 'The evolution and nature of school accountability in the Singapore education system', *Educational Assessment, Evaluation and Accountability*, **22**, 4, 275–92.

Ng, P.T. (2010). 'The evolution and nature of school accountability in the Singapore education system', *Educational Assessment, Evaluation and Accountability*, **22**, 4, 275–92.

Ng, P. T. and Chan, D. (2008). 'A comparative study of Singapore's school excellence model with Hong Kong's school-based management', *International Journal of Educational Management,* **22**, 6, 488–505. Cited in: Ng, P.T. (2010). 'The evolution and nature of school accountability in the Singapore education system', *Educational Assessment, Evaluation and Accountability*, **22**, 4, 275–92.

Nusche, D., Laveault, D., MacBeath, J. and Santiago, P. (2012). *OECD Reviews of Evaluation and Assessment in Education: New Zealand 2011.* Paris: OECD [online]. Available: http://img.scoop.co.nz/media/pdfs/1202/49681441.pdf [5 September, 2018].

Organisation for Economic Co-operation and Development (OECD) (2012). *Lessons from PISA for Japan: Strong Performers and Successful Reformers in Education*. Paris: OECD [online]. Available: http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/49802616.pdf [5 September, 2018].

Organisation for Economic Co-operation and Development (OECD) (2014). *Improving Schools in Wales: An OECD Perspective*. Paris: OECD [online]. Available: http://www.oecd.org/education/Improving-schools-in-Wales.pdf [5 September, 2018].

Pierce, R. and Chick, H. (2011). 'Teachers' intentions to use national literacy and numeracy assessment data: A pilot study', *Australian Educational Researcher,* **38,** 4, 433–447. Cited in: Bien, N.M. (2016). *Primary School Achievement Gaps and School Decisions to Support the Academic Achievement of Disadvantaged Students with Data: A Cross-Country Comparative Study*. PhD thesis, University of Sydney [online]. Available: https://ses.library.usyd.edu.au/bitstream/2123/15742/1/bien_nm_thesis.pdf [5 September, 2018].

Robinson, V., McNaughton, S. and Timperley, H. (2011). 'Building capacity in a self-managing schooling system: The New Zealand experience', *Journal of Educational Administration*, **49**, 6, 720–38 OECD [online]. Available: https://cdn.auckland.ac.nz/assets/education/about/research/LRG/Robinson,%20McNaughton%20&%20Timperley%20(2011)%20Building%20capacity%20in%20a%20self-managing%20schooling%20system-%20The%20New%20Zealand%20experience.pdf [5 September, 2018].

Singapore Ministry of Education (2010). *Nurturing our Young for the Future: Competencies for the 21st Century*. Ministry of Education, Singapore. Cited in: Organisation for Economic Co-operation and Development (OECD). (2012). *Lessons from PISA for Japan: Strong Performers and Successful Reformers in Education*. Paris: OECD [online]. Available: http://www.oecd.org/education/school/programmeforinternationalstudentassessmentpisa/49802616.pdf [5 September, 2018].

Tan, E. T. J. (2005). 'The marketization of education in Singapore: What does this mean for thinking schools, learning nation?' In: Tan, E. T. J. and Ng, P. T. (Eds) *Shaping Singapore's Future: Thinking Schools, Learning Nation*. Singapore: Prentice Hall. Cited in: Ng, P.T. (2010). 'The evolution and nature of school accountability in the Singapore education system', *Educational Assessment, Evaluation and Accountability*, **22**, 4, 275–92.

Troman, G., Jeffrey, B. and Raggl, A. (2007). 'Creativity and performativity policies in primary school cultures', *Journal of Education Policy*, **22**, 5, 549–572. Cited in: Coldwell, M. and Willis, B. (2017). 'Tests as boundary signifiers: Level 6 tests and the primary secondary divide', *Curriculum Journal*, **28**, 4, 578–97 (DOI: 10.1080/09585176.2017.1330220) [online]. Available: https://www.tandfonline.com/doi/full/10.1080/09585176.2017.1330220 [5 September, 2018].

Wayman, J. C., Cho, V. and Johnston, M. T. (2007). *The Data-Informed District: A District-Wide Evaluation of Data use in the Natrona County School District.* Austin: The University of Texas. Cited in: Bien, N.M. (2016). Primary School Achievement Gaps and School Decisions to Support the Academic Achievement of Disadvantaged Students with Data: a Cross-Country Comparative Study. PhD Thesis, University of Sydney [online]. Available: https://ses.library.usyd.edu.au/bitstream/2123/15742/1/bien_nm_thesis.pdf [5 September, 2018].

Wilkins, C. (2011). 'Professionalism and the post-performative teacher: new teachers reflect on autonomy and accountability in the English school system', *Professional Development in Education*, **37**, 3, 389–409 [online]. Available: https://lra.le.ac.uk/bitstream/2381/21565/1/professionalism%20and%20the%20post-performative%20teacher%20final%20PDE%20version%202011.pdf [5 September, 2018].

Worth, J. (2017). 'How subject teaching time has changed since 2011'. *SecEd*, 14 June [online]. Available: http://www.sec-ed.co.uk/best-practice/how-subject-teaching-time-has-changed-since-2011/ [3 September, 2018].

Young, V. M. and Kim, D. H. (2010). 'Using assessments for instructional improvement: A literature review', *Education Policy Analysis Archives,* **18**, 19. Cited in: Bien, N.M. (2016). *Primary School Achievement Gaps and School Decisions to Support the Academic Achievement of Disadvantaged Students with Data: a Cross-country Comparative Study*. PhD Thesis, University of Sydney [online]. Available: https://ses.library.usyd.edu.au/bitstream/2123/15742/1/bien_nm_thesis.pdf [5 September, 2018].

Additional references

Adams, J. E. and Kirst, M. (1999). 'New demands for educational accountability: striving for results in an era of excellence'. Cited in: Murphy, J and Louis, K. S. (Eds). *Handbook of Research In Educational Administration,* 463–489, 2nd ed. San Francisco, CA: Jossey-Bass.

Baird, J., Johnson, S., Hopfenbeck, T. N., Isaacs, T., Sprague, T., Stobart, G. and Yu, G. (2016). 'On the supranational spell of PISA in policy', *Educational Research*, **58,** 2, 121–138. Special Issue: International Policy Borrowing and Evidence-based Educational Policy Making: Relationships and Tensions.

Burdett, N. and O'Donnell, S. (2016). 'Lost in translation? The challenges of educational policy borrowing', *Educational Research*, **58**, 2, 113–120, (DOI: 10.1080/00131881.2016.1168678) [online]. Available: https://www.tandfonline.com/doi/full/10.1080/00131881.2016.1168678 [5 September, 2018].

Boyle, B. and Bragg, J. (2006). 'A curriculum without foundation', *British Educational Research Journal*, **32**, 4, 569-582. Cited in: Darling-Hammond, L. and Ascher, C. (1991). *Creating Accountability In Big City Systems*. New York, NY: ERIC Clearinghouse on Urban Education.

Burns, T. and Köster, F. (Eds) (2016). *Governing Education in a Complex World. Educational Research and Innovation.* Paris: OECD Publishing.

Carter, A. (2015). *Carter Review of Initial Teacher Training (ITT).* London: DfE [online]. Available: https://www.gov.uk/government/publications/carter-review-of-initial-teacher-training [5 September, 2018].

Demski. D. and Racherbäumer, K. (2017). 'What data do practitioners use and why? Evidence from Germany comparing schools in different contexts', *Nordic Journal of Studies in Educational Policy,* **3**, 1, 82–94 (DOI: 10.1080/20020317.2017.1320934) [online]. Available: https://www.tandfonline.com/doi/full/10.1080/20020317.2017.1320934 [5 September, 2018].

Department for Education (2018). *Principles for A Clear And Simple Accountability System.* [online]. Available: https://www.gov.uk/government/publications/principles-for-a-clear-and-simple-school-accountability-system [5 September, 2018].

Department of Education and Science (DES). (1988). *Education Reform Act*. London: HMSO. Cited in: Coldwell, M. and Willis, B. (2017). 'Tests as boundary signifiers: level 6 tests and the primary secondary divide', *Curriculum Journal*, **28**, 4, 578–97 (DOI: 10.1080/09585176.2017.1330220) [online] Available: https://www.tandfonline.com/doi/abs/10.1080/09585176.2017.1330220?journalCode=rcjo20 [5 September, 2018].

Great Britain. Parliament. House of Commons (2016). *Educational Excellence Everywhere* (Cm. 9230). London: TSO [online]. Available https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/508447/Educational_Excellence_Everywhere.pdf [3 September, 2018].

Donaldson, G. (2015). *Successful Futures. Independent Review of Curriculum and Assessment Arrangements in Wales* [online]. Available: https://www.nasuwt.org.uk/asset/A788604C-3046-4005-A1EA0EAFF023E0DD/ [5 September, 2018].

Ehren, M. and Perryman, J. (2017). 'Accountability of school networks: who is accountable to whom and for what?', *Educational Management Administration and Leadership*, 1–18.

Furlong, J. (2015). *Teaching Tomorrow's Teachers: Options for the Future of Initial Teacher Education in Wales.* Oxford: University of Oxford.

Johansson, S. (2016). 'International large-scale assessments: what uses, what consequences?' *Educational Research*, **58**, 2, 139–148.

Lynch, S., Worth, J., Bamford, S. and Wespieser, K. (2016). *Engaging Teachers: NFER Analysis of Teacher Retention.* Slough: NFER [online]. Available: https://www.nfer.ac.uk/engaging-teachers-nfer-analysis-of-teacher-retention/ [5 September, 2018].

National Monitoring Study of Student Achievement (NMSSA) (2018). *NMSSA 2018 Information for Schools, Parents, Whānau and Caregivers.* New Zealand: New Zealand Government [online]. Available: http://nmssa.otago.ac.nz/files/NMSSA_2018.pdf [5 September, 2018].

O'Day, J. A. and Smith, M. S. (1993). 'Systemic reform and educational opportunity'. In S. H. Fuhrman (Ed.) *Designing Coherent Education Policy: Improving The System*. San Francisco, CA: Jossey-Bass.

O'Reilly, F. E. (1996). *Educational Accountability: Current Practices and Theories In Use*. Cambridge, MA: Harvard University, Consortium for Policy Research in Education. Cited in: Ng, P.T. (2010). 'The evolution and nature of school accountability in the Singapore education system', *Educational Assessment, Evaluation and Accountability*, **22**, 4, 275–92.

Organisation for Economic Co-operation and Development (OECD) (2017). *The Welsh Education Reform Journey. A Rapid Policy Assessment*. Paris: OECD [online]. Available: http://www.oecd.org/education/The-Welsh-Education-Reform-Journey.pdf [5 September, 2018].

Schleicher, A. (2018). *World Class: How To Build A 21st-Century School System*. Paris: OECD Publishing [online]. Available: http://www.oecd.org/education/world-class-9789264300002-en.htm [5 September, 2018].

Standards and Testing Agency (2017). *Information for Schools and Parents: Primary School Progress Measures. Coventry:* STA [online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/596097/STA177916e_Primary_school_progress_measures_leaflet_V6_FINAL.pdf [5 September, 2018].

Spielman, A. (2017). *HMCI's Commentary: Recent Primary and Secondary Curriculum Research*. London: Ofsted [online]. Available: https://www.gov.uk/government/speeches/hmcis-commentary-october-2017 [5 September, 2018].

Stecher, B. and Nataraj Kirby, S. (Eds) (2004). *Organizational Improvement and Accountability: Lessons for Education from Other Sectors.* Santa Monica, CA: RAND Corporation.

Stobart, G. (2009). 'Determining validity in national curriculum assessments', *Educational Research*, **51**, 2, 161–179, (DOI: 10.1080/00131880902891305) [online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00131880902891305 [5 September, 2018].

The Commission on Assessment Without Levels (2015). *Final Report of the Commission on Assessment without Levels.* London: DfE [online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/483058/Commission_on_Assessment_Without_Levels_-_report.pdf [3 September, 2018].

Vanhoof, J. and Van Petegem, P. (2007). 'How to match internal and external evaluation? Conceptual reflections from a Flemish perspective.' Presentation at the 20th International Congress for Efficiency and Improvement 'Professional Challenges for School Effectiveness and Improvement in the Era of Accountability', Portorož, Slovenia, 3-6 January.

Wyse, D. and Torrance, H. (2009). 'The development and consequences of national curriculum assessment for primary education in England', *Educational Research*, **51**, 2, 213–228.

Appendix A An overview of the key features of selected accountability systems in primary education

Countries / jurisdictions	Assessment				Evaluation			
	Is there statutory national or jurisdiction-wide assessment in the primary phase as defined by the jurisdiction?	When during the primary phase does assessment take place?	Are assessment findings reported?	Is assessment used to hold schools to account for pupil performance?[2]	Is there statutory external evaluation and does it include school inspection?	Are inspection outcomes reported?	Is there a requirement for self-evaluation?	Is there school-to-school or peer-to-peer support for school improvement?
England	●	P	●	●	● I	●	●	●
Wales	●	P	●	●	● I	●	●	●
Australia – New South Wales	●	P	●	●	●	n/a	●	●
Canada – Alberta	●	E	●	●	●	n/a	●	●
Estonia	●	P	●	■	● I	●	●	●
Finland	●◆	E	●	■	●	n/a	●	●
Germany	●	P	●	■	● (I)	(●)	(●)	●
Japan	●	E	●	■	● I	●	●	●
New Zealand	●◆	P	●	■	● I	●	●	●
Poland	(●)	E	●	■	● I	●	●	●
Singapore	●	E	●	■	●	n/a	●	●
Sweden	●	P	●	●	● I	●	●	●
USA – Massachusetts	●	P	●	●	● I	●	●	●

About this overview

This overview of primary phase accountability systems in 13 countries / jurisdictions informed the development of a literature review of the impact of different accountability systems on standards, curriculum, and teacher and pupil engagement.

According to the overview[1]:

• All 13 countries / jurisdictions carry out some form of jurisdiction-wide assessment in the primary phase but it is only in six countries / jurisdictions (including England and Wales) that the assessments are used to hold schools accountable for pupil attainment and progress.

• All 13 countries / jurisdictions carry out statutory external evaluation in the primary phase. Of these, nine (including England and Wales) include inspection.

Key

●	This feature is present
(●)	This feature is present, with qualifications
■	Feature not apparent in sources examined
E	Only end-of-primary phase assessments
P	Assessments at one or more points during primary phase
I	Inspection is a feature of external evaluation
n/a	Not applicable
◆	Sample-based assessment, not whole cohort.

Notes

[1] The information derives from publicly available official sources such as ministry and inspectorate websites and draws on public information about assessment and evaluation systems. Whilst every effort has been made to reflect the categorisations with accuracy, we recognise that different interpretations are possible.

[2] The extent to which this may be through intermediate levels of authority, e.g. local administrative arrangements, varies.

Appendix B Search strategy

The aim of this search was to provide evidence on the impact of different models of accountability on standards, curriculum, and pupil and teacher engagement, and on what factors affect this. The particular focus areas were:

- curriculum
- standards of attainment
- school improvement support mechanisms
- teacher workload
- teacher engagement, well-being, retention and recruitment
- pupil engagement and well-being
- school management decision-making
- perverse incentives and unintended consequences
- closing the gap between disadvantaged and non-disadvantaged pupils
- cost.

The date range was 2010 to the present and only literature published in the English language was selected. The age range was the primary phase and the geographical scope covered Australia – New South Wales, Canada – Alberta, England, Japan, New Zealand, Singapore, Sweden and Wales. Study types included peer-reviewed and grey literature featuring statistical analyses of large datasets, nationally representative surveys, large-scale qualitative studies, literature reviews and evidence-based opinion pieces.

Search sources included key education bibliographic databases, relevant websites including those of the ministries of education, curriculum and assessment bodies of the selected jurisdictions, and reference harvesting from reference lists of key reports identified.

Bibliographic databases

The Australian Education Database (AEI), British Education Index (BEI) and the Education Resources Information Center (ERIC) databases were searched using the keywords listed below, tailored to the specific search capability of each database. The keywords used are based on the thesaurus terms used in each database. An additional set of “country/jurisdiction” keywords was added to the search strategy for ERIC to narrow the results.

Throughout the abbreviation “FT” denotes that a free text term was used and * denotes truncation of terms.

Australian Education Index (searched via ProQuest 29/4/18) 33 hits

#1 Accountability

#2 Educational accountability FT

#3 Educational evaluation FT

#4 Education* standards FT

#5 Peer support FT

#6 School accountability FT

#7 School autonomy FT

#8 School to school support FT

#9 Self-evaluation

#10 System efficiency FT

#11 #1 OR #2…OR #10

#12 Primary education

#13 Elementary education

#14 #12 OR #13

#15 Curriculum

#16 Curriculum breadth FT

#17 Curriculum implementation

#18 Curriculum narrowing FT

#19 #15 OR #16…OR #18

#20 #19 AND #11 AND #14

#21 Academic achievement

#22 Academic improvement FT

#23 Educational attainment

#24 Educational improvement

#25 Performance

#26 School performance FT

#27 #21 OR #22……OR #26

#28 #27 AND #11 AND #14

#29 Educational facilities improvement

#30 School support teams FT#

#31 #29 OR #30

#32 #31 AND #11 AND #14

#33 Teacher workload

#34 Teaching load FT

#35 Teacher welfare

#36 #33 OR #34 OR #35

#37 #36 AND #11 AND #14

#38 Teacher attitudes

#39 Teacher burnout

#40 Teacher commitment FT

#41 Teacher engagement FT

#42 Teacher morale

#43 Teacher motivation

#44 Teacher participation

#45 Teacher persistence

#46 Teacher recruitment

#47 Teacher retention FT

#48 Teacher supply and demand

#49 Teacher welfare

#50 #38 OR #39….OR #49

#51 #50 AND #11 AND #14

#52 Student attitudes

#53 Student commitment FT

#54 Student engagement FT

#55 Student morale FT

#56 Student motivation FT

#57 Student participation

#58 Student persistence FT

#59 Student well-being FT

#60 #52 OR #53….OR #59

#61 #60 AND #11 AND #14

#62 School administration

#63 School decision-making FT

#64 School leadership FT

#65 School based management

#66 #62 OR #63….OR #65

#67 #66 AND #11 AND #14

#68 Achievement gap FT

#69 Closing the gap FT

#70 Disadvantaged

#71 Educationally disadvantaged

#72 Low achievement

#73 Underachievement

#74 #68 OR #69….OR #73

#75 #74 AND #11 AND #14

#76 Costs

#77 Educational finance

#78 #76 OR #77

#79 #78 AND #11 AND #14

British Education Index (searched via EBSCO Host 26/4/18) – 107 hits

#1 Accountability FT

#2 Educational accountability

#3 Educational evaluation

#4 Educational standards

#5 Peer support FT

#6 School accountability FT

#7 School autonomy FT

#8 School to school support FT

#9 Self-evaluation

#10 System efficiency FT

#11 #1 OR #2…OR #10

#12 Primary education

#13 Elementary education

#14 #12 OR #13

#15 Curriculum FT

#16 Curriculum breadth FT

#17 Curriculum implementation

#18 Curriculum narrowing FT

#19 #15 OR #16…OR #18

#20 #19 AND #11 AND #14

#21 Academic achievement

#22 Academic improvement FT

#23 Educational attainment

#24 Educational improvement FT

#25 Performance

#26 School performance FT

#27 #21 OR #22……OR #26

#28 #27 AND #11 AND #14

#29 School improvement FT

#30 School support teams FT#

#31 #29 OR #30

#32 #31 AND #11 AND #14

#33 Teacher* workload FT

#34 Teaching load

#35 Teacher welfare

#36 #33 OR #34 or #35

#37 #36 AND #11 AND #14

#38 Teacher attitudes

#39 Teacher burnout

#40 Teacher commitment FT

#41 Teacher engagement FT

#42 Teacher morale

#43 Teacher motivation FT

#44 Teacher participation FT

#45 Teacher persistence

#46 Teacher recruitment

#47 Teacher retention

#48 Teacher supply and demand FT

#49 Teacher well-being FT

#50 #38 OR #39….OR #49

#51 #50 AND #11 AND #14

#52 Student attitudes

#53 Pupil commitment FT

#54 Student engagement

#55 Pupil morale FT

#56 Pupil motivation FT

#57 Student participation

#58 Pupil persistence FT

#59 Student well-being

#60 #52 OR #53….OR #59

#61 #60 AND #11 AND #14

#62 School administration

#63 School decision-making FT

#64 School leadership FT

#65 School management FT

#66 #62 OR #63….OR #65

#67 #66 AND #11 AND #14

#68 Achievement gap

#69 Closing the gap FT

#70 Disadvantaged FT

#71 Educationally disadvantaged students

#72 Low achievement FT

#73 Underachievement

#74 #68 OR #69….OR #73

#75 #74 AND #11 AND #14

#76 Cost* FT

#77 Educational finance

#78 #76 OR #77

#79 #78 AND #11 AND #14

Educational Resources Information Center (searched via EBSCO Host 28/4/18) 970 hits

#1 Accountability

#2 Educational accountability FT

#3 Educational evaluation FT

#4 Education* standards FT

#5 Peer support FT

#6 School accountability FT

#7 School autonomy FT

#8 School to school support FT

#9 Self-evaluation

#10 System efficiency FT

#11 #1 OR #2…OR #10

#12 Primary education

#13 Elementary education

#14 #12 OR #13

#15 Curriculum

#16 Curriculum breadth FT

#17 Curriculum implementation

#18 Curriculum narrowing FT

#19 #15 OR #16…OR #18

#20 #19 AND #11 AND #14

#21 Academic achievement

#22 Academic improvement

#23 Educational attainment

#24 Educational improvement FT

#25 Performance

#26 School performance FT

#27 #21 OR #22……OR #26

#28 #27 AND #11 AND #14

#29 School improvement OR Educational facilities improvement

#30 School support teams FT

#31 #29 OR #30

#32 #31 AND #11 AND #14

#33 Teacher* workload FT

#34 Teaching load

#35 Teacher welfare

#36 #33 OR #34 OR #35

#37 #36 AND #11 AND #14

#38 Teacher attitudes

#39 Teacher burnout

#40 Teacher commitment FT

#41 Teacher engagement FT

#42 Teacher morale

#43 Teacher motivation

#44 Teacher participation

#45 Teacher persistence

#46 Teacher recruitment

#47 Teacher retention FT

#48 Teacher supply and demand FT

#49 Teacher well-being FT

#50 #38 OR #39….OR #49

#51 #50 AND #11 AND #14

#52 Student attitudes

#53 Student commitment FT

#54 Student engagement

#55 Student morale FT

#56 Student motivation

#57 Student participation FT

#58 Student persistence FT

#59 Student well-being FT

#60 #52 OR #53….OR #59

#61 #60 AND #11 AND #14

#62 School administration

#63 School decision-making FT

#64 School leadership FT

#65 School management FT

#66 #62 OR #63….OR #65

#67 #66 AND #11 AND #14

#68 Achievement gap

#69 Closing the gap FT

#70 Disadvantaged FT

#71 Educationally disadvantaged

#72 Low achievement

#73 Underachievement

#74 #68 OR #69….OR #73

#75 #74 AND #11 AND #14

#76 Costs

#77 Educational finance

#78 #76 OR #77

#79 #78 AND #11 AND #14

#80 Australia

#81 Canada

#82 England

#83 Japan

#84 New Zealand

#85 Singapore

#86 Sweden

#87 Wales

#88 #80 OR #81… #87

#89 #20 OR #28 OR #32 OR #37 OR #51 OR #61 OR #67 OR #75 OR #79

#90 #88 AND #89

Website searches

The following websites, including those of the ministries of education, curriculum and assessment bodies of the selected jurisdictions, were searched on main keywords and/or the publications/research/policy sections were browsed as appropriate:

Alberta Education

Association of School and College Leaders

Australian Council for Educational Research

Australian Curriculum, Assessment and Reporting Authority

Department for Education (including National College for Teaching and Leadership)

Education Committee

Education Policy Institute

Education Review Office (New Zealand)

Estyn

House of Commons Library

Institute for Fiscal Studies

MEXT: Ministry of Education, Culture, Sports, Science and Technology (Japan)

Ministry of Education and Research (Sweden)

Ministry of Education New Zealand

Ministry of Education Singapore

National Association of Head Teachers

National Audit Office

National Education Union

National Foundation for Educational Research

New South Wales Department of Education

OECD

Ofsted

Social Mobility Commission

Standards and Testing Agency

Sutton Trust

Swedish National Agency for Education

Swedish Schools Inspectorate

UNESCO

Welsh Government

World Bank

Appendix C Details of the 25 pieces of literature identified by the search proces

Documents relate to: **England** (coverage in 11 documents); **Wales** (coverage in 3 documents)**; Australia** (New South Wales and national) (coverage in 4 documents); **Japan** (coverage in 4 documents); **New Zealand** (coverage in 4 documents); **Singapore** (coverage in 3 documents).

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
1	Adoniou, M. (2016). 'Beginning teachers' responses to education reform agendas', *School Leadership and Management*, **36**, 1, 80–95.	Australia	Primary education	Qualitative data study	• 14 teachers in their first 16 months of primary school teaching • 8 semi-structured interviews; also classroom observations, field notes, online surveys • Thematic analysis
2	Bew Review (2011). *Independent Review of Key Stage 2 Testing, Assessment and Accountability*. London: DfE [online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/176180/Review-KS2-Testing_final-report.pdf [29 June, 2018].	England	Primary education	Evidence-based review commissioned by UK government	• 12-week call for evidence leading to receipt of 3940 online responses, oral evidence from 50 stakeholders, written submissions • Report also discusses published research material

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
3	Bien, N.M. (2016). *Primary School Achievement Gaps and School Decisions to Support the Academic Achievement of Disadvantaged Students with Data : a Cross-country Comparative Study*. PhD thesis, University of Sydney [online]. Available: https://ses.library.usyd.edu.au/bitstream/2123/15742/1/bien_nm_thesis.pdf [29 June, 2018].	Australia (NSW)	Primary education	Qualitative and quantitative data study	• Six primary school case studies across New South Wales, California and Hawaii: 50 educators (43 interviewed; 8 observed) • Analysis of standardised assessment results from 2008-2013 from Australia and two counties in California
4	Burgess, S., Wilson, D. and Worth, J. (2013). 'A natural experiment in school accountability: The impact of school performance information on pupil progress', *Journal of Public Economics,* **106**, 2013, 57–67.	England and Wales	Secondary education	Quantitative data study	• Data from school censuses in England and Wales • Quantitative comparison of attainment outcomes (secondary schools) • Quasi-experimental evaluation design and/or "natural experiment" • Analysis of attainment and pupil characteristic data from a population of schools in England and Wales (around 3,500 schools)

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
					over an eight-year period (2001-2008).
5	Coldwell, M., Willis, B. (2017). 'Tests as boundary signifiers: level 6 tests and the primary secondary divide', *Curriculum Journal*, **28**, 4, 578–97.	England	Primary education	Qualitative data study	• Interviews with: teachers and school leaders in 20 primary schools participating in the test; telephone interviews with 40 school leaders who chose not to participate in the test; 20 secondary-school leaders • Thematic analysis
6	Education Review Office (2015). *Effective School Evaluation*. Wellington: ERO [online]. Available: http://www.ero.govt.nz/assets/Uploads/Effective-School-Evaluation-web.pdf [29 June, 2018].	New Zealand	Primary and secondary education	Government guidance documentation including illustrative school case studies	• Typology of internal school evaluation types • Outline of 5 step school improvement journey • Case studies of schools using internal evaluation to inform strategies for raising achievement
7	Gilbert, C. (2012). *Towards a Self-Improving System: the Role of School Accountability*. Nottingham: NCSL [online]. Available:	England	Primary and secondary education	Conceptual discussion	• Argument for re-balancing current school accountability system in England

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
	http://dera.ioe.ac.uk/14919/1/towards-a-self-improving-system-school-accountability-thinkpiece%5B1%5D.pdf [29 June, 2018].				
8	Goodman, R. and Burton, D. (2012). 'What is the nature of the achievement gap, why does it persist and are government goals sufficient to create social justice in the education system?' *Education 3-13*, **40**, 5, 500–14.	England	Primary and secondary education	Literature review	• Review of literature (mainly from England – also from elsewhere in UK, and US) • Argument that government focus on accountability and assessment does not help address fundamentals of inequality
9	Hargreaves, D.H. (2012). *A Self-Improving School System: Towards Maturity*. Nottingham: NCSL [online]. Available: http://dera.ioe.ac.uk/15804/1/a-self-improving-school-system-towards-maturity.pdf [29 June, 2018].	England	Primary and secondary education	Conceptual discussion including illustrative school case studies	• Fourth in a series of conceptual pieces about the development of a self-improving school system in England • Proposes a model with mature inter-school partnerships central to a self-improving school system

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
10	Hutchinson, J. and Dunford, J. (2016). *Divergent Pathways: the Disadvantage Gap, Accountability and the Pupil Premium*. London: EPI [online]. Available: https://epi.org.uk/publications-and-research/divergent-pathways-disadvantage-gap-accountability-pupil-premium/ [29 June, 2018].	England	Primary and secondary education	Quantitative data study	• Quantitative analysis of national pupil performance data in England between 2006 and 2015, split by subsets of schools with different proportions of disadvantaged pupils
11	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.	Australia	Primary education	Discussion	• Discussion of impact of national testing in Australia, with reference to reported observations from the Australian Primary Principals' Association (APPA) 2009-10, and published literature. • Consideration of alternative approaches for accountability purposes
12	Kuramoto, N. and Koizumi, R. (2016). 'Current issues in large-	Japan	Primary, secondary	Discussion	• Profile of Japanese education assessment system; discussion of

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
	scale educational assessment in Japan: focus on national assessment of academic ability and university entrance examinations', *Assessment in Education: Principles, Policy & Practice* [online]. Available: https://www.adrec.ihe.tohoku.ac.jp/wp/wp-content/uploads/2016/11/KuramotoKoizumiin-press.pdf [29 June, 2018].		and university education		Japan's large-scale educational assessments • Argument that there are conflicts evident between the principle of education and the principle of measurement
13	Lingard, B. and Sellar, S. (2013). '"Catalyst data": perverse systemic effects of audit and accountability in Australian schooling', *Journal of Education Policy*, **28**, 5, 634–56.	Australia	Primary and secondary education	Qualitative data study	• 30 interviews with policy makers, personnel from international organisations, and researchers and academics • Case study with two parts, analysing effects related to the use of national test data for accountability purposes

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
14	Macleod, S., Sharp, C., Bernardinelli, D., Skipp, A. and Higgins, S. (2015). *Supporting the Attainment of Disadvantaged Pupils: Articulating Success and Good Practice*. London: DfE [online]. Available: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/473974/DFE-RR411_Supporting_the_attainment_of_disadvantaged_pupils.pdf [29 June, 2018].	England	Primary and secondary education	Quantitative and qualitative data study	• Quantitative analysis of school-level data from national school performance tables • Survey of 759 primary and 570 secondary schools in England • Interviews with senior leaders in 49 schools
15	Marks, R. (2014). 'Educational triage and ability-grouping in primary mathematics: a case-study of the impacts on low-attaining pupils', *Research in Mathematics Education*, **16**, 1, 38–53 [online]. 2018].	England	Primary education	Qualitative case study (drawn from a mixed-methods main study)	• Single-school longitudinal case study conducted over one year • 88 primary school pupils in Year 6 (11-12 year-olds), with a focus on nine pupils in the lowest ability set for mathematics • Thematic analysis of classroom observation and pupil interviews

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
16	Miki, T., Toledo Figueroa, D., Peterka, J., Fraccola, S. and Ikesako, H. (2015). *Education Policy Outlook:* Japan. Paris: OECD [online]. Available: [29 June, 2018].	Japan	Education system – all stages	Policy review	• Comparative analysis of Japan's education policy, drawing on OECD indicators
17	Nakayasu, C. (2016). 'School curriculum in Japan', *Curriculum Journal*, **27**, 1, 134–50.	Japan	Primary and secondary education	Discussion	• Discussion of the history of the Japanese education system relevant to the school curriculum • Examination of the relationship between curriculum reform and Japan's performance in PISA assessments
18	National College for School Leadership (2011). *System Leadership: Does School-to-School Support Close the Gap?* Nottingham: NCSL [online]. Available: http://www.isospartnership.com/uploads/files/system-leadership-does-	England	Primary and secondary education	Quantitative and qualitative data study, including case studies	• Samples of: 87 primary and 105 secondary 'National Support Schools' (NSSs); and 164 primary and 130 secondary NSS-supported schools

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
	school-to-school-support-close-the-gap.pdf [29 June, 2018].				• Analysis of pupil performance data, split by eligibility for free school meals (FSM) • Data compared with national pupil performance (2008 - 2010) • Interviews with 30 'National Leaders of Education' (NLEs) ; 10 follow up visits; workshops with another 20 NLEs
19	NFER and Arad Research (2013). *A Rapid Evidence Assessment on the Impact of Curriculum and Assessment Arrangements within High Performing Countries*. Cardiff: Welsh Government [online]. Available: https://gov.wales/docs/caecd/research/131022-rapid-evidence-assessment-impact-curriculum-assessment-arrangements-high-performing-countries-en.pdf [29 June, 2018].	Wales, New Zealand, Singapore	Primary and secondary education	Literature review	• Literature review of 20 studies • Analysis of literature to explore the impact of curriculum and assessment arrangements in Canada, Finland, Korea, New Zealand and Singapore (5 high performing countries according to PISA results) and consideration of findings in comparison with Wales

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
20	Ng, P.T. (2010). 'The evolution and nature of school accountability in the Singapore education system', *Educational Assessment, Evaluation and Accountability*, **22**, 4, 275–92.	Singapore	Primary and secondary education	Discussion	• Analysis of the development of school accountability in Singapore, using a theoretical framework • Exploration of how schools respond to school accountability requirements
21	Nusche, D., Laveault, D., MacBeath, J. and Santiago, P. (2012). *OECD Reviews of Evaluation and Assessment in Education: New Zealand 2011.* Paris: OECD [online]. Available: http://img.scoop.co.nz/media/pdfs/1202/49681441.pdf [29 June, 2018].	New Zealand	Primary and secondary education	Policy review	• Analysis of New Zealand's assessment and evaluation frameworks • Exploration of how these can be used to improve student outcomes
22	Organisation for Economic Co-operation and Development (OECD) (2012). *Lessons from PISA for Japan: Strong Performers and Successful Reformers in Education*. Paris: OECD [online]. Available: http://www.oecd.org/education/school/programmeforinternationalstude	Japan, Singapore	Primary and secondary education	Policy review	• Analysis, using PISA data, of how Japan and other strongly performing countries are reforming their education systems to prepare students for the 21st century work place

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
	ntassessmentpisa/49802616.pdf [29 June, 2018].				
23	Organisation for Economic Co-operation and Development (OECD) (2014). *Improving Schools in Wales: an OECD Perspective*. Paris: OECD [online]. Available: http://www.oecd.org/education/Improving-schools-in-Wales.pdf [29 June, 2018].	Wales	Primary and secondary education	Policy review	• Analysis, using PISA data, of the identified strengths and challenges of the school system in Wales • Proposal of recommendations for improvement
24	Robinson, V., McNaughton, S. and Timperley, H. (2011). 'Building capacity in a self-managing schooling system: the New Zealand experience', *Journal of Educational Administration*, **49**, 6, 720–38 OECD [online]. Available: https://cdn.auckland.ac.nz/assets/education/about/research/LRG/Robinson,%20McNaughton%20&%20Timperley%20(2011)%20Building%20capacity%20in%20a%20self-managing%20schooling%20system	New Zealand	Primary and secondary education	Discussion	• Analysis of policy-related documentation in order to evaluate two examples of the Ministry's approach to addressing achievement gaps

Doc ref no.	Reference	Relevant Jurisdiction(s)	Phase of education focused on in study	Broad description of study type	Details of study – to show the nature of the evidence
	- %20The%20New%20Zealand%20 experience.pdf [29 June, 2018].				
25	Wilkins, C. (2011). 'Professionalism and the post-performative teacher: new teachers reflect on autonomy and accountability in the English school system', *Professional Development in Education*, **37**, 3, 389–409 [online].	England	Primary education	Qualitative data study	• 18 newly-qualified teachers interviewed at end of first year of teaching • Thematic analysis

National Foundation for Educational Research

Evidence for excellence in education

Public

The Mere, Upton Park, Slough, Berks SL1 2DQ
T: +44 (0)1753 574123 • **F:** +44 (0)1753 691632 • enquiries@nfer.ac.uk

www.nfer.ac.uk

NFER ref. ACCT