Special educational needs, disability and school accountability: an international perspective

Emma Smith and Graeme Douglas

Accepted for publication in the International Journal of Inclusive Education

Emma Smith (corresponding author)

University of Leicester

es228@le.ac.uk

Graeme Douglas

University of Birmingham

g.g.a.douglas@bham.ac.uk

Abstract

Standards-based accountability approaches in education are increasingly being adopted in many industrialised countries. This tends to involve the scrutiny of student performance in national academic tests and then holding schools accountable for subsequent progress. International assessments such as PISA are not accountability measures in themselves, but are often utilised in a similar way and national policies are judged against performance in these tests. Drawing upon the examples of PISA and national assessments in the US, this paper explores how students with special educational needs and disability (SEND) are included in, and served by, these assessment policies. The analysis suggests that while these approaches are inclusive in their intent (most obviously through efforts to include students with SEND in testing) they may poorly represent the progress of these students. It is argued that more equitable inclusion of students with SEND must find a balance between accountability measures which measure performance in traditional attainment-focused tests and other relevant assessments which are meaningful to the given student and / or SEND sub-group.

Keywords: special education, accountability, high stakes testing, social justice

Introduction

Standards-driven accountability is now the norm in public schools in many industrialised countries (e.g. Cabinet Office 2010, Klenowski and Wyatt-Smith 2012). It involves holding schools accountable for the academic attainment of their students and, according to Hursh (2005, 609), is driven by ‘a three-fold need to raise educational and economic productivity in an increasingly globalized economy, to decrease educational inequality and to improve the quality and objectivity of school-based assessment’. Arguably the key questions that increased levels of accountability have to answer are not only whether it leads to improved student performance but if it does, who benefits? It is well established that during compulsory education, academic success is predicated on success at the previous educational stage and that young people from less affluent social groups achieve at lower levels throughout schooling (e.g. Gorard *et al.* 2007, Smith 2012). Therefore the main challenges that face those concerned with both educational justice and school improvement involve unpacking the educational determinants of success in school (such as effective teaching) from the wider social and cultural influences; as well as understanding the ways in which accountability reform might mitigate against prior academic disadvantage.

This paper will consider the relationship between test-based accountability reform and the education of students with special educational needs and disability (SEND) (see Appendix 1), a group who may often be the least academically successful in school. It will begin by looking at the role that international comparative tests of student attainment have in informing accountability policies at the national level and the extent to which these assessments can include all students, particularly those with SEND. The second part of the paper looks at the increase of accountability based reform in the USA and the extent to which high-stakes testing at the school level is able to effectively address the needs of SEND students. More specifically this paper has the following aims:

- To consider the relationship between test-based accountability policies and the education of students with SEND.
- To examine the inclusion of students with SEND in international comparative assessments of student performance (specifically PISA).
- To illustrate the issues around inclusion of students with SEND in national accountability based reform through a consideration of the No Child Left Behind Act in the United States.

We begin by providing a brief introduction to standards-based accountability policies and high stakes testing in general, before considering some of their most widely rehearsed advantages and disadvantages.

Standards-based accountability policies

The aim of test- or standards-based accountability policies is to identify a set of measurable performance standards across a number of core subject areas, to align the curriculum to these standards, and to assess and scrutinise subsequent student attainment (Figlio and Loeb 2011). This evaluation of learning against a common set of test scores will ostensibly provide policy makers, school administrators and members of the general public with an indicator of how well schools are performing relative to each other. Schools that appear to be under-performing on this indicator might be subject to further sanctions and scrutiny that, in time, might lead to improvement. High-stakes tests are those in which the results have serious consequences for stakeholders: such as a student who achieves poorer results than expected in college entrance examinations. Or, where tests are used for accountability purposes (and might be linked to school funding or teacher tenure), then they also become high-stakes for the institution involved. Additionally, tests that might otherwise be low-stakes for the individuals who take them and their schools (such as the international comparative tests that we discuss here) can be high-stakes for policy makers and governments (Stobbart and Eggen 2012).

This emphasis on academic attainment as the most tangible outcome of schooling makes the very premise of standards-based reform open to scrutiny, particularly because of concerns that high-stakes testing may result in a side-lining of other aspects of education, such as citizenship, well-being and critical thinking (e.g. West et al. 2011). Perhaps it is unsurprising then that concerns about accountability driven reform are well-rehearsed. Its critics argue that they result in a narrowing of the curriculum and an over-emphasis on the subjects that are tested (such as science and maths) at the expense of those that are not (such as music and social studies); they promote teaching to the test and emphasise low quality assessment such as multiple-choice responses at the expense of richer and more nuanced tasks; and they can also harm poor, disabled and minority-group students and increase the dropout rate (e.g. Darling-Hammond 2007, Haney 2000, Hout and Elliot 2011).

However, alongside the apparent negative consequences of test-based accountability, there is also a body of evidence that suggests that there can be a positive relationship between such policies and student achievement (e.g. Dee and Jacob 2009, Rockoff and Turner 2008, Hanushek 2005). But while there may be an 'emerging consensus that students whose scores are the most consequential for school accountability are those who gain the most' (Figlio and Loeb 2011, 411); evidence also suggests that not all student groups benefit equally from accountability based reform. Hanushek and Raymond (2005), for example, find that while accountability policies in the US have had some success in narrowing the achievement gap between Hispanic and White students, this is not the case for the Black-White achievement gap. In fact, one of the key issues for those who design accountability measures is in deciding which students should be included. Inclusively named policies such as No Child Left Behind in the USA and Every Child Matters in England and Wales, would

suggest that measures to improve schools ought to include everyone. But, as Figlio and Loeb (2011, 394) argue ‘universal inclusion raises important questions about fairness and attribution’ and, most pertinently for this paper, ‘should schools be held accountable for students for whom testing is more challenging, or potentially less reliable, such as students with disabilities?’

Whatever the arguments for and against school accountability policies, it is nevertheless the case that they are an integral part of the educational landscape in many nations. Including students with SEND in accountability linked assessments is perhaps a more contentious matter and three key issues arise: first about whether all students ought to participate, second about the nature and effectiveness of the assessment tools used, and third whether the assessments are in fact relevant and comprehensive for these students’ needs. In the next section we look at the issues that arise when students with SEND are included in international assessments of student performance. While low stakes for the actual students, these tests are anything but low stakes for policy makers and the architects of national education systems.

International perspectives on school accountability: the role of PISA

One of the key drivers of school reform in many developed countries is their performance in international comparative tests such as the Programme for International Student Assessment (PISA) (see appendix 2) and the Trends in International Maths and Science Study (TIMSS). While not an accountability measure in themselves, it is difficult to underestimate the power of these assessments to influence national education policy (e.g. Dolin 2007, Rutkowski and Rutkowski 2010, Baird et al. 2011). Indeed, the tests occupy a distinctive position in being low-stakes for the test-takers and their schools (participants are not identified) and yet high-stakes for politicians, whose education policies will stand or fall on their outcomes (Stobbart and Eggen 2012). The increasing popularity and sophistication of these tests has enabled nations to look critically at the performance of their schools in comparison with their international neighbours. This has led to many nations re-examining their education systems in light of perceived failings in these assessments (Ertl 2006). In several countries this has been used to further justify dissatisfaction with the domestic school system and has led to accusations of falling academic standards and failing students, as well as calls for policy reform and increased levels of school accountability (Young 2010). A recent example from the UK serves to highlight this point. The results from the 2009 round of PISA were released just weeks after the UK Coalition government published its first White Paper on education; they could hardly have been more timely:

So much of the education debate in this country is backward looking: have standards fallen? Have exams got easier? These debates will continue, but *what really matters* is how we’re doing compared with our international competitors. That is what will define our economic growth and our country’s

future. The truth is, at the moment we are standing still while others race past. (DfE 2010, 1, emphasis added).

Through the opening paragraph of this report, the new government reiterated the imperative that education is the cornerstone of an effective economy, or in other words that 'education is the best economic policy' (Brown 2007). The relatively poor performance of British 15 year olds in PISA 2009 (OECD 2010) only served to reaffirm their dissatisfaction with the current domestic school system.

According to Gary Orfield (2000), policy-makers' preoccupations with perceived mediocre performance in international tests such as PISA, have contributed to a refocusing of educational priorities in many Western countries. This has resulted in a move away from policies that have attempted to resolve the inequities faced by poor and other disadvantaged children – for example by increasing access to high-quality education, desegregating schools and reducing achievement gaps – and towards ones concerned with more testing, more accountability and increasingly market-driven systems of school choice. In the US context this is reflected in systems of high-stakes testing and strict school accountability sanctions, epitomised by the No Child Left Behind legislation. In England, an already highly regulated and highly accountable national school system has experienced further moves towards increased diversity of provision driven by choice and market-based policies, all of which are set to continue under the current Coalition government.

International assessments such as PISA are important processes for collecting comparative data on the academic attainment of different groups of students. However, while the physical and academic inclusion of students with SEND in schools and classrooms has steadily increased, their inclusion in this sort of standardised assessment has not kept pace (OECD 2007). However, in terms of international comparative tests, PISA arguably endeavours to provide the most comprehensive coverage of students with SEND and is the one considered here.

The exclusion of young people with SEND, as well as non-native speakers and low-achievers, from international comparative tests has long been controversial (Hormann 2007, Wuttke 2007). In an effort to address this, a number of amendments were introduced from PISA 2003. Data is now collected on a student's SEND status, instruction manuals are explicit about the need to include these students in the assessments (e.g. OECD 2008:24); and an optional test instrument (Une Heure) which provides an abridged version of the main assessment was introduced. While PISA aims to be as inclusive as possible, there will be some students who are not able to participate for various reasons and there are five categories of student-based exclusion. The first four categories are also used as SEND categories for those who are included in PISA (see OECD 2008, OECD 2010, and Table 1 below):

- i. students with a cognitive, behavioural or emotional disability. The manual for PISA 2003 (OECD 2002, 16) used a slightly different label of 'Intellectual Disability' and this same term is used as an exclusion category term.
- ii. students with a functional disability – student has a moderate to severe permanent physical disability
- iii. students with limited proficiency in the assessment language
- iv. other (as defined by the national centres and approved by the international centre)
- v. students taught in a language of instruction for the main domain for which no materials were available (category for students who are excluded only).

Students can be excluded from participation in PISA on the basis of SEND or unfamiliarity with the language of instruction but cannot be excluded because of low proficiency or common discipline problems. School-level exclusions for students in the first three SEND categories (listed above) should not exceed 2% and within-school exclusions should not exceed 2.5% of students (OECD 2007). The decision to use any of the above categories to exclude students from PISA is taken by the school PISA-coordinator following a set of OECD-approved guidelines and the details of all excluded students are recorded. Sampling for PISA is intended to ensure that there is a representative national sample of students included in the assessment. However, there is flexibility within the sampling frame to exclude certain types of schools including geographically remote schools, extremely small schools and those where the administration of PISA would be not feasible. This final criterion might include special schools. In England in PISA 2009, special schools and pupil referral units were excluded from the sampling frame (Bradshaw *et al.* 2010) and special schools were also excluded from the country sampling frame in Ireland (La Roche *et al.* 2010). Table 1 shows the number of students excluded and included, the category of SEND and the within-school exclusion/inclusion rate for a selection of the countries that participated in PISA 2009.

Table 1: Numbers of SEND students according to category of inclusion/exclusion in PISA 2009, selected countries

Country	All SEND		Cat. 1 Functional disability		Cat. 2 Intellectual disability		Cat. 3 Limited lang. proficiency		Cat. 4 Other SEND		Within school exclusion rate**	Within school inclusion rate**	Proportion of sample with SEND**
	Inc.	Exc.	Inc.	Exc.	Inc.	Exc.	Inc.	Exc.	Inc.	Exc.			
Australia	377	313	35	24	175	210	167	79	0	0	1.79	2.39	4.2
Belgium	607	30	40	3	486	17	81	10	0	0	0.24	7.61	7.9
Canada	118	1607	6	49	82	1458	30	100	0	0	5.47	0.66	6.1
Finland	312	75	5	4	143	48	53	12	111	11	1.15	4.46	5.6
France	142	1	49	1	59	0	34	0	0	0	0.04	3.53	3.6
Germany	78	28	18	6	57	20	3	2	0	0	0.47	1.72	2.2
Hong Kong-China	35	9	8	0	21	9	6	0	0	0	0.16	0.70	0.9
Iceland	252	183	5	3	75	78	25	64	147	38	4.10	6.52	10.6
Ireland	163	136	4	4	77	72	28	25	54	35	2.75	3.85	6.6
Japan	0	0	0	0	0	0	0	0	0	0	0.00	0	0
Korea	18	16	2	7	15	9	1	0	0	0	0.28	0.49	0.8
Mexico	24	52	6	25	15	25	3	2	0	0	0.15	0.14	0.3
Netherlands	42	19	5	6	33	13	4	0	0	0	0.35	1.10	1.5
New Zealand	102	181	8	19	42	84	52	78	0	0	3.15	2.26	5.4
Norway	95	207	4	8	59	160	32	39	0	0	3.79	1.87	5.7
Shanghai-China	27	7	2	1	14	6	11	0	0	0	0.13	0.52	0.6
Singapore	71	48	7	2	15	22	49	24	0	0	0.80	1.13	1.9
Sweden	254	146	160	115	0	0	94	31	0	0	2.89	5.59	8.5
United Kingdom*	-	318	-	40	-	247	-	31	-	0	2.44	-	-
United States	265	305	10	29	224	236	29	40	2	10	4.81	4.77	9.6

*No data available for SEND inclusions

** calculations made using weighted data

Source: OECD 2010 tables A2.1 and A2.2 plus authors' own analysis

The exclusion and inclusion rates vary considerably. It is notable that some of the highest performing PISA nations (Finland, Shanghai-China and Australia) appear to exclude relatively small proportions of students with SEND within sampled schools. As might be expected, some countries have a high exclusion rate and a low inclusion rate (or vice versa), e.g. Canada, Finland. However, some countries have a different profile suggesting different levels of prevalence of SEND in the education system which appears not to be explained by school-level sampling (e.g. US has relatively high exclusion and inclusion figures, while Shanghai-China and Japan have low). The zero score of Japan is particularly noteworthy and will be the focus of future research. Additionally, the categorisation of students with SEND is variable, e.g. Iceland and Finland making higher use of the 'other' category. On the one hand, it might be argued that there is little evidence to show that students with SEND are being routinely excluded from PISA in order to influence scores at the national level. However the absence of large scale observational data on the administration and exclusion of young people with SEND from PISA means that it is difficult to draw any firm conclusions on this matter. It is important to note that issues surrounding the inclusion of young people with SEND and PISA are relatively under-researched and there is a need for a more detailed examination of the area.

The issue of cross-national comparisons of the educational participation and attainment of young people with SEND is a complex one (Douglas et al. 2012). It is certainly possible that some countries/economies could exclude students with a particular need while others would include them. Another issue is that students taught in discrete special schools might be excluded from the process altogether (as happened in England and Ireland in 2009). In addition, despite the efforts of organisations such as the OECD to develop common categories of SEND, it is still the case that practice and categorisation of needs can vary widely. Given the perhaps inevitable challenges of including and comparing students with SEND on the international level, it is worth examining whether the process is any easier within a national system; in this case, the USA.

National perspectives on school accountability: the United States

Dissatisfaction with American performance in both domestic and international assessments came to the fore in 1983 when President Ronald Regan's administration published the document *A Nation at Risk?* (NCEE 1983). This searing indictment of educational standards in the US signalled a shift in focus towards accountability and testing. The invective used in this document is strong and condemns the 'rising tide of mediocrity' eroding the American public school system. The rhetoric is little different today. American students still occupy a mid-table position in the international achievement rankings and commentators and politicians still despair over the apparent 'underachievement' of American schools:

We have a genuine national crisis. More and more, we are divided into two nations. One that reads and one that doesn't. One that dreams and one that doesn't (George Bush, US Department of Education 2002).

Systems of high-stakes testing, epitomised by the 'No Child Left Behind' (NCLB) Act, have been put in place in order to raise standards in schools and to help propel the United States up the international school league tables. The NCLB legislation was passed into law by former President Bush early in 2002. It required that schools and school districts put into place a set of standards for improving student achievement, together with detailed plans charting how these standards will be monitored and met. A major consequence of this is that schools have been required to set targets and monitor the progress of *all* students in order to ensure that they reach minimum proficiency levels by 2014. Failure to achieve proficiency could lead to school sanctions, which, at their most extreme, would result in school closure.

NCLB requires that states assess performance annually in grades 3 to 8 in Literacy, Mathematics, and Science, with additional tests administered in high school. Under NCLB, states had to develop plans detailing how both schools and school districts would demonstrate Adequate Yearly Progress (AYP) towards full proficiency and make public their test results. States determine their own timelines for achieving proficiency and must ensure that the tests are implemented (although they decide which tests to use) and that curriculum standards are in line with federal guidelines. School performance data is disaggregated according to different student sub-groups. A sub-group is a group of students who share similar characteristics and they are grouped this way for reporting purposes. So a sub-group might be students' minority group, disability or level of economic disadvantage. States do vary in their definition of what constitutes a sub-group. For example in Arizona sub-group size is 40 students, in Maryland it is five and in Texas it can range from 50 to 200 students depending on school size (State of Arizona Department of Education 2010, Figlio and Loeb 2011). This is of particular importance when monitoring the performance of students with disabilities, as group sizes tend to be small. So in Arizona, if there are fewer than 40 students with disabilities in one school grade then the performance of that subgroup will not count towards AYP under NCLB – if there are 42 students in the sub-group, then their performance will count. This is crucial because according to Figlio and Loeb (2011) there is a strong relationship between sub-group size and the likelihood that the school will make AYP, and they attribute between one-fifth and one-third of the failure to achieve AYP to variations in sub-group size requirements.

State-wide testing and students with disabilities

The 1975 Individuals with Disabilities Education Act (IDEA) mandates that all children and young people aged between 3 and 21 who have disabilities are entitled to a Free and Appropriate Public Education (FAPE). Under American law no student is too disabled to

qualify and it is up to school districts to ensure that all children and young people have access to FAPE. Students with a disability have their development monitored and assessed through an Individual Education Plan (IEP). By law, the IEP must include a statement of the child's present level of performance, annual goals and short-term objectives plus detail of all additional support that will be provided (McLaughlin and Thurlow 2003). In 2008–09, 6.483 million children and young people were receiving support under IDEA; this corresponds to about 13% of all public school enrolment (NCES 2011).

Since being introduced IDEA has undergone numerous revisions, the most recent in 2004 aligned the Act with NCLB. The 2004 reauthorisation of IDEA mandated that *all* students with disabilities be involved in state-wide testing. The decision to include students with disabilities in all forms of accountability assessment is supported by concerns that this group of young people were being overlooked in school:

“By including all students in state accountability systems, schools pay attention to the performance and progress of all students; educating students with disabilities becomes a shared responsibility of both general and special education teachers” (US Department of Education 2007, 11).

Most students with disabilities are expected to participate in general grade-level state assessments with or without accommodations. Accommodations are modifications made to the presentation of the assessment without changing the assessment criteria (Lazarus *et al.* 2009 provides a summary of practice across the US). If students could not participate in state grade-level assessments then states must make alternative provision through ‘alternate’ assessments. The nature of these assessments varies from state to state (Cameto *et al.* 2009). Some states use a portfolio based system where students are assessed by their classroom teacher over an extended period while others may use externally moderated pencil and paper tests. The decision about whether or not a student will participate in alternate assessments is made by their IEP team following state-approved criteria (US Department of Education 2007). There are therefore a number of ways in which students with disabilities might be included in state-wide assessments (with examples from three states):

- General grade-level assessment (sometimes with accommodations such as additional time) – see Lazarus *et al.* 2009. Examples: Arizona Instrument to Measure Standards (AIMS); California Standards Tests (CSTs); New Jersey Assessment of Skills and Knowledge (NJASK) or the High School Proficiency Assessment (HSPA)
- Alternate assessment based on modified grade-level academic achievement standards – see Lazarus *et al.* (2011). Examples: California Modified Assessment (CMA); neither Arizona nor New Jersey operate this type of assessment (students are included through one of the other methods).
- Alternate assessment based on alternative academic achievement standards – see Cameto *et al.* (2009). Examples: Arizona Instrument to Measure Standards (A) (AIMS-A);

California Modified Assessment (CMA); New Jersey Alternate Proficiency Assessment (APA).

The outcomes of these assessments, especially for students with disabilities, are mixed. For example, nationwide in 2011, only 51% of schools made AYP, this figure varied from just 11% of schools in Florida to 89% in Wisconsin (Usher 2012). In Iowa, just 29% of students in disability sub-groups made AYP in grade 8 Reading; the figure for all students was 75% (Iowa Department of Education 2011). The alignment of IDEA and NCLB meant that for the first time, students with SEND were to be subject to the same standards and scrutiny in relation to academic attainment as their non-disabled peers. For some commentators this puts the individualised approach to student achievement (as advocated by IDEA) into conflict with the standardised approach to achievement advocated by NCLB (Kraus 2012). Furthermore it requires that schools juggle meeting AYP with making progress required by the IEP through which the school is already held accountable under IDEA. The general issue of whether including these students in accountability-linked tests of this nature is in fact equitable, is considered next.

The role of accountability policies in the education of students with SEND

This paper has described how high-stakes tests that are linked, either directly or indirectly, to school accountability measures can affect students with SEND both internationally and in the USA. In this final section we consider the arguments for and against the inclusion of children and young people with SEND in such assessments.

Ensuring that schools are held accountable for the educational progress of students with SEND has generally been viewed positively by commentators and has, in part, been welcomed because of the disability community's concerns over the often poor educational performance of this group of young people (e.g. Cole 2006, McLaughlin and Thurlow 2003, Centre on Education Policy 2007). For example, according to the National Council on Disability, the reauthorisation of IDEA in 2004 has led to a "palpable and positive change" (2008, 55) in the approach to including young people with disabilities in American schools and in state-wide testing. Indeed, one of the fundamental arguments in support of their inclusion is that because students with SEND now count in accountability-based assessments, due regard has to be paid to their progress:

"People teach what is tested and who is tested—so now that students with disabilities are included in the accountability system, they are being taught" (National Council on Disability 2008, 55).

Perhaps unsurprisingly however, there are concerns about the negative consequences of an increased emphasis on testing and monitoring on the educational experiences of students with SEND. Many of these concerns stem from issues with the practicalities of assessing

such a diverse group of learners, rather than an ideological objection to testing in its own right (e.g. technical analyses of state implementation of the national assessments are offered by Lazarus *et al.* 2009, 2011, Cameto *et al.* 2009). Cole (2006) provides a number of examples of unintended consequences of standards-based accountability measures that are specific to students with SEND. First students with SEND might be treated as scapegoats because if schools fail to meet accountability targets it is often (although not exclusively) because of the performance of students with SEND. Secondly there is a potentially negative impact on inclusion and integration because if students with SEND are considered to be one reason for a school's failure to meet their targets, then this might slow progress on integration. Finally the need for schools to meet targets might lead to increased drop-out and poor retention rates among students with SEND. Indeed, schools with large proportions of students with SEND might legitimately argue that holding them accountable for the performance of these students puts them at a disadvantage (Figlio and Loeb 2011).

Additional areas of concern lie around the practice of grouping together different types of SEND rather than reporting on subgroups of disabilities. With reference to NCLB, it can be argued that:

There is so much lumping together of disabilities, and we need to really differentiate them. NCLB should have more varied testing and accountability standards for students with disabilities given the differences in disabilities. NCLB should be more sophisticated in its requirements for proficiency, not just one standard (National Council on Disability study, 2008, 66).

Indeed in 2009, fewer than half of the US states disaggregated results by disability category (primary disability) (Altman *et al.* 2010) and under NCLB, there is no requirement for states to report separate disability categories rather all students with disability are collapsed into a single category. Furthermore, the concept of primary disability can be problematic because many young people have a complex set of multiple conditions.

Unsurprisingly, the same issues about how to group and report test scores for students with disabilities can be seen when comparisons are made on an international level. The OECD (2007) present a re-classification of countries own schemes and data into a proposed new tri-partite cross-national classification system. The exercise has served to confirm that definitions of disability categories vary widely.

But possibly the most widespread area of concern is the increased frequency of testing that these accountability-linked assessments require. The increased emphasis on testing leads to a narrowly defined curriculum where, arguably, too much attention is paid to curriculum content that will feature in the tests. This, of course, is not a concern that applies only to students with SEND but is relevant for all young people in school. But the consequences for those with SEND are perhaps more profound. Focus upon the narrower and academic aspects of the curriculum may undermine those with learning disabilities for whom other

areas of a balanced curriculum are of greater enjoyment, accomplishment or value (e.g. vocational topics, music, art, sport). Further, it may also detract from more nuanced (even disability-specific) aspects of the curriculum which are of particular value to students with SEND (e.g. mobility, access technology, independence skills, social communication). Such tests may result in them being held back an academic year, subjected to ‘humiliating’ tests which they are unable to answer or being denied entry to a particular school in the first place:

Everything revolves around testing and the punitive nature of the system. It pervades everything, and kids pick up on it. And then you have the stress of the IEP. Teachers don’t feel like they can just try something creative or different to help meet the needs of students with disabilities. There is no time to be creative—teachers are always planning for tests (National Council on Disability study, 2008, 75).

Lowrey *et al.* (2007, 251) highlight this concern in relation the use of alternate assessments in the USA for students with severe learning disabilities noting that they may be at the expense of other ‘meaningful targets that will improve a student’s quality of life after leaving the public school system’.

Another argument against test-based accountability systems is that they may encourage schools to exclude low performing students from assessments or even, in the case of some Charter schools, from the school itself (McLaughlin and Rhim 2007, Wolf *et al.* 2012). Arguably, the exclusion of students with SEND from assessments is a particular issue for tests that are low-stakes at the school/student level, such as PISA, making it perhaps easier to exclude these students from the assessments. With NCLB, because it requires that 95% of all students in a school are assessed, a concern is that rather than sitting the general assessment, schools will reclassify students into SEND categories and they will then sit modified versions of the assessments, often with accommodations (such as extra time) that might potentially improve their performance (e.g. Figlio and Loeb 2011). For example, schools might improve their likelihood of raising test scores by selectively assigning students to SEND sub-groups, perhaps by placing relatively high-achieving students with mild dyslexia into the disabilities subgroup. This practice of ‘gaming’ the system has been explored in Florida by Figlio and Getzler (2002) who found that following the introduction of an accountability testing program, low-performing students and students from low socioeconomic backgrounds were substantially more likely to be reclassified into SEND categories and exempted from the accountability process. They also found that high-poverty schools were significantly more likely than more affluent schools to reclassify low-achieving students.

Discussion

Writing about the UK context in 1985, Sally Tomlinson argues that the primary purpose of an expanding SEND sector was to respond to the dilemma of how to restructure the

education-training system to deal with the 'increasing number of young people who are defined as being unable or unwilling to participate satisfactorily in a system primarily directed towards producing academic and technical elites' (Tomlinson 1985, 157). While much may have changed in the intervening years (with special education, and inclusive educational placement in particular, being now a key part of the education systems in many countries), it is still the case that the dilemma of how to prepare students with SEND for full and active citizenship in a high skilled, globalised economy does persist. As Tomlinson goes on to argue more recently:

'the ideology and practice of inclusive education has brought into mainstream schools and colleges large numbers of young people who would previously have been segregated in special schools or classes, joining their fellow students who were previously labelled slow, dull, retarded or remedial. The numbers of those variously regarded as having learning or behavioural difficulties and disabilities and unable to achieve required levels has increased, and pressures on public comprehensive schools and lower level schools in selective systems are intense' (2012, 275).

The economic imperative of higher levels of skills and training for all young people, including those with SEND, is a compelling driver of school reform in many industrialised countries. However, the question for governments is what to do with some young people with SEND in an economy where even low-skilled employment requires some level of qualification (Tomlinson 2012). One response to this has been to increase levels of engagement in education: in England, for example, the school leaving age will rise from 2015 and all young people will have to remain in some form of education or training until they are 18. Another has been to urge school improvement for all students through systems of high-stakes testing that are linked to strict accountability policies that are aimed at raising standards. It is this approach and its implications for students with SEND that we have considered in this paper.

While most will agree that some form of accountability for those teaching and administering special education programmes is desirable, the issue of what form that accountability should take and the impact of accountability policies on school reform are uncertain. While some evidence points to instances of improvement, there are concerns that not all students, in particular the most vulnerable, have benefited. For instance, research by the Civil Rights Project in the USA shows that rather than encouraging schools to become better integrated, recent accountability policies, are in fact moving schools towards greater inequalities and increased levels of segregation (Sunderman 2008). It is beyond the scope of this paper to address all these issues. Instead we have sought to examine the types of high-stakes accountability-linked assessments that are being administered to students with SEND. Our discussion shows that the prospects for including students with SEND in these assessments are mixed. For instance, in spite of PISA aiming to include students with SEND, variable definitions of SEND and sampling approaches in participating countries raise questions about their meaningful inclusion and is worthy of closer analysis. While the participation of

students with SEND in national assessments in the USA appears to be more successful, potential unintended consequences appear to persist: re-categorising low achieving students as having a SEND, holding these students responsible for schools not meeting targets, as well as inappropriately designed curricula. This comes alongside more fundamental concerns that students with SEND can be 'lumped together' in single categories when their circumstances and educational needs may vary enormously.

While accountability policies, and NCLB in particular, may well have changed the character of debate from excusing bad results because students were disadvantaged, to focusing attention on the education of a previously neglected group (Hanushek, 2005), the issue of 'what form of accountability for what type of student?' remains. Including the most vulnerable students in stressful tests that may, or may not, measure the educational outcomes that are the most important for them has important implications for the fair treatment of these students in school; but arguably so does leaving them out of these assessments altogether. One way forward according to Shippen et al. (2006, 327) is for schools to acknowledge that most students with SEND will never achieve at the 'average' levels demanded by most high stakes tests and that instead appropriate standards and curriculum should be established for these students. This does not mean that there should not be 'effective and efficient instruction' for these students but merely recognises that the 'contention that *all* students with disabilities will attain average achievement is unrealistic'. However, such a proposal is likely to sit uncomfortably with those anxious to avoid a segregated 'special education' system in which different students experience different curricula. The inclusion of all children and young people with SEND in assessments related to accountability is an important aspiration. Considerable effort is required if such assessments are accessible, reliable and valid. Accommodated and alternative assessment approaches are important strategies in this regard and the considerable work in the USA offers many lessons in how this can be done. Nevertheless, young people with SEND vary enormously in terms of the challenges and dangers they face in relation to inclusion in assessment – 'lumping together' these groups and looking at averages across SEND as a whole has limited (if any) value when evaluating the efficacy of education. Therefore the collection of reliable and more detailed data in relation to student SEND characteristics would allow for more meaningful data analysis and disaggregation. Overall, balance seems crucial – the system must value a balance between performance in traditional attainment-focussed accountability tests and other relevant assessments which are meaningful to the given student and / or SEND sub-group. The design of such inclusive assessment procedures is undoubtedly challenging and, in addition to accommodated and alternative assessment approaches described, may also require additional sources of evidence in relation to student progress. For example, this might include evidence gathered through school inspection procedures or bespoke focussed studies (for example the US-based National Longitudinal Transition Study-2 [NLTS2] gathers a broad range of progress and outcome data from over 8000 students with disability). This requires considerable thought, but the imperative

remains that systems for assessing student progress (with their associated use for accountability purposes) should seek to include and serve *all* students.

Appendices

Appendix 1: Different countries use different terms to refer to students with 'special educational needs' (e.g. in USA 'disability' is used, while Scotland adopts the broader 'Additional Support Needs'). In this article we use the term 'special educational needs and disability' (SEND) to avoid ambiguity. An exception is when we discuss the USA specifically where we use the term most commonly used in that country: students with disability.

Appendix 2: PISA is a series of assessments that are taken by 15 year old students and that are designed to 'assess student performance and collect data on the student, family and institutional factors that can help explain differences in performance' (OECD 2001:4). Students are assessed in three yearly waves in three areas: Language, Mathematics and Science. The first wave, in 2000 focused mainly on Language, the second on Mathematics and the third on Science. The most recent wave of PISA was in 2009 and focused on Language and included around 470,000 students from 65 countries or economies.

Acknowledgements

The research upon which this paper is based was partly funded by the National Council for Special Education, Ireland.

References

Altman, J. R., Lazarus, S. S., Quenemoen, R. F., Kearns, J., Quenemoen, M., & Thurlow, M. L. 2010. 2009 survey of states: Accomplishments and new issues at the end of a decade of change. Minneapolis: National Center on Educational Outcomes. www.cehd.umn.edu/NCEO/OnlinePubs/2009StateSurvey.pdf

Baird, J.A., Isaacs, T., Johnson, S., Stobart, G., Yu, G., Sprague, T., Daugherty, R. 2011. Policy effects of PISA. Oxford University Centre for Educational Assessment. <http://oucea.education.ox.ac.uk/wordpress/wp-content/uploads/2011/10/Policy-Effects-of-PISA-OUCEA.pdf>

Bradshaw, J., Ager, R., Burge, B., Wheeler, R. 2010. *PISA 2009: Achievement of 15-Year-Olds in England*. Slough: NFER.

Brown, G. 2009. The Romanes Lecture given at the Sheldonian Theatre Oxford. 27th February 2009. www.number10.gov/Page18472.

Cabinet Office 2010. *The Coalition: Our Programme for Government*. London: Cabinet Office.

Cameto, R., Knokey, A.-M. et al. 2009. State Profiles on Alternate Assessments Based on Alternate Achievement Standards: A Report from the National Study on Alternate Assessments, NCSER 2009-3013. National Center for Special Education Research. <http://ies.ed.gov/ncser/>.

Centre on Education Policy 2007. NCLB's Accountability Provisions for Students with Disabilities. Center on Education Policy Roundtable Discussion. www.cep-dc.org

Cole, C. 2006. Closing the Achievement Gap Series: Part III What is the Impact of NCLB on the Inclusion of Students with Disabilities?. Center for Evaluation and Education Policy. www.ceep.indiana.edu/projects/PDF/PB_V4N11_Fall_2006_NCLB_dis.pdf

Darling-Hammond, L. 2007. Race, inequality and educational accountability: the irony of "No Child Left Behind". *Race, Ethnicity and Education*, 10, no.3:245–60.

Dee, T., Jacob, B. 2009. The impact of No Child Left Behind on student achievement. National Bureau of Economic Research. NBER Working Paper No. 15531.

DfE. 2010. *The Importance of Teaching*, The Schools White Paper 2010. <http://www.education.gov.uk/publications/eOrderingDownload/CM-7980.pdf>.

Dolin, J. 2007. PISA: An Example of the Use and Misuse of Large-Scale Comparative Tests. In *PISA According to PISA: Does PISA Keep What It Promises?*, ed. Hopmann, S.T., Brinek, G., Retzl, M. Berlin and Vienna: LIT Verlag.

Douglas, G., Travers, J., McLinden, M., Robertson, C., Smith, E., Macnab, N., Powers, S., Guldborg, K., McGough, A., O'Donnell, M., Lacey, P. 2012. Measuring Educational Engagement, Progress and Outcomes for Children with Special Educational Needs: A Review. Ireland: National Council for Special Education.

Ertl, H. 2006. 'Educational standards and the changing discourse on education: the reception and consequences of the PISA Study in Germany'. *Oxford Review of Education*, 32, no. 5: 619–34.

Figlio, D., N., Getzler, L., S. 2002. Accountability, ability and disability: Gaming the system, Working Paper 9307. National Bureau of Economic Research. <http://www.nber.org/papers/w9307>

Figlio, D., Loeb, S. 2011. School Accountability. In *Economics of Education Vol. 3*. Ed. Hanushek, E.A., Machin, S and Woessmann, L., 383-421. The Netherlands: North-Holland.

Gorard, S., Adnett, N., May, H., Slack, K., Smith, E. and Thomas, L. 2007, *Overcoming the Barriers to Higher Education*. Stoke-on-Trent: Trentham Books.

Hautamäki, J., Harjunen, E., Hautamäki, A., Karjalainen, T., Kupiainen, S., Laaksonen, S., Lavonen, J., Pehkonen, E., Rantanen, P., Scheinin, P. 2008. PISA 2006 Finland: Analyses,

Reflections, Explanations.
www.minedu.fi/export/sites/default/OPM/Julkaisut/2008/liitteet/opm44.pdf?lang=en

Haney, W. 2000. The myth of the Texas miracle in education. *Education Policy Analysis Archives*. 8, no. 41, <http://epaa.asu.edu/ojs/article/view/432/828>.

Hanushek, E.A., Raymond, M.E. 2005. Does School Accountability Lead to Improved Student Performance? *Journal of Policy Analysis and Management*. 24, no. 2:297–327.

Hanushek, E.A. 2005. Why the Federal Government Should Be Involved in School Accountability. *Journal of Policy Analysis and Management* 24, no. 1:167–178.

Hörmann, B. 2007. Disappearing Students PISA and Students With Disabilities, In *PISA According to PISA: Does PISA Keep What It Promises?* Ed. Hopmann, S.T., Brinek, G., Retzl, M. Berlin and Vienna: LIT Verlag.

Hout, M., Elliot, S.W. 2011. Incentives and test-based accountability in education. National Academies Press. http://www.nap.edu/catalog.php?record_id=12521

Hursh D. 2005. The growth of high-stakes testing in the USA: accountability, markets and the decline in educational equality. *British Educational Research Journal*, 31, no.5:605-622.

Klenowski, V., Wyatt-Smith, C. 2012. The impact of high stakes testing: the Australian story. *Assessment in Education* 19, no.1:65-79.

Kraus, L. 2012. The Impact of No Child Left Behind on IDEA's Guarantee of Free, Appropriate Public Education for Students with Disabilities: A Critical Review of Recent Case Law. www.luc.edu/law/academics/special/center/child/childed_forum/pdfs/2009_student_papers/kraus_impact_no_child.pdf

LaRoche, S., Cartwright, F. 2010. Independent Review of the 2009 PISA Results for Ireland. Dublin: Educational Research Centre.

Lazarus, S. S., Thurlow, M. L., Lail, K. E. & Christensen, L. 2009. A Longitudinal Analysis of State Accommodations Policies: Twelve Years of Change, 1993–2005. *Journal of Special Education*, 43: 67–80.

Lazarus, S., Cormier, D. & Thurlow, M. 2011. States' Accommodations Policies and Development of Alternate Assessments Based on Modified Achievement Standards: A Discriminant Analysis. *Remedial and Special Education*, 32, no.4: 301–308.

Lowrey, K. A., Drasgow, E. et al. 2007. Impact of Alternate Assessment on Curricula for Students with Severe Disabilities: Purpose Driven or Process Driven? *Assessment for Effective Intervention* 32, no.4:244–253

McLaughlin, M.J., Thurlow, M. 2003. Educational Accountability and Students with Disabilities: Issues and Challenges. *Educational Policy*, 17:43.

McLaughlin, M.J., Rhim, L.M. 2007. Accountability Frameworks and Children with Disabilities: A test of assumptions about improving public education for all students. *International Journal of Disability, Development and Education*, 54, no.1:25-49.

National Council on Disability. 2008. The No Child Left Behind Act and the Individuals with Disabilities Education Act: A Progress Report. National Council on Disability. www.educationalpolicy.org/publications/pubpdf/NCD_2008.pdf

NCEE. 1983. *A Nation at Risk?* www.njafter3.org/edu/docs/Reports_A-Nation-At-Risk.pdf.

NCES. 2011. The Condition of Education: Children and Youth with Disabilities. National Centre for Education Statistics. http://nces.ed.gov/programs/coe/indicator_cwd.asp

OECD. 2001. Knowledge and Skills for Life: First results from PISA 2000. <http://browse.oecdbookshop.org/oecd/pdfs/free/9601141e.pdf>

OECD. 2002. PISA 2003 Main Study, School Co-ordinator's Manual. http://www.namsmat.is/vefur/rannsoknir/PISA_skyrslur_almennt/3_Handbakur/PISA_2003_school_coordinator_manual.pdf

OECD. 2007. *Students with Disabilities, Learning Difficulties and Disadvantages: Policies, Statistics and Indicators*. Paris: OECD.

OECD. 2008. PISA 2009 Main Survey, School Co-ordinator's Manual. https://mypisa.acer.edu.au/images/mypisadoc/sc_manual_ms09_v2_1.pdf

OECD. 2010. PISA 2009 Results: What Students Know and Can Do: Student Performance in Reading, Mathematics and Science (Volume I). <http://dx.doi.org/10.1787/9789264091450-en>

Orfield, G. 2000. Policy and equity: lessons of a third of a century of educational reforms in the United States. In *Unequal Schools, Unequal Chances: The Challenges to Equal Opportunity in the Americas*, ed. F. Reimers. Cambridge, MA: Harvard University Press.

Rockoff, J., Turner, L. 2008. *Short run impacts of accountability on school quality*. National Bureau of Economic Research. NBER Working Paper No. 14564.

Rutkowski, L., Rutkowski, D. 2010. Getting it 'better': the importance of improving background questionnaires in international large-scale assessment, *Journal of Curriculum Studies*, 42, no.3:411-430.

Shippen, M., Houchins, D.E., Calhoon, M., Furlow, C., Sartor, D. 2006. The Effects of Comprehensive School Reform Models in Reading for Urban Middle School Students With Disabilities. *Remedial and Special Education*, 27, no.6:322-328.

Smith, E. 2012. *Key Issues in Education and Social Justice*. London: Sage.

State of Arizona Department of Education. 2010. State of Arizona Consolidated State Application Accountability Workbook. Arizona Department of Education, http://www.azed.gov/research-evaluation/files/2011/07/conappaypwb-10_07_06.pdf

Stobart, G., Eggen, T. 2012. High-stakes testing – value, fairness and consequences, *Assessment in Education*, 19, no.1:1-6.

Sunderman, G. 2008. *Holding NCLB Accountable: Achieving Accountability, Equity and School Reform*. New York: Corwin Press.

Tomlinson, S. 1985. The expansion of special education. *Oxford Review of Education*, 2, no.2:195–205.

Tomlinson, S. 2012. The irresistible rise of the SEN industry. *Oxford Review of Education*, 38, no.3:267-286.

US Department of Education. 2002. No Child Left Behind: Executive Summary. www.ed.gov/nclb/overview/intro/presidentplan/page_pg3.html.

US Department of Education. 2007. *Modified Academic Achievement Standards: Non-Regulatory Guidance*. US Department of Education.

Usher, A. 2012. *AYP Results for 2010-11 — May 2012 Update*. Washington DC: Centre on Education Policy.

West, A., Mattei, P., Roberts, J. 2011. Accountability and Sanctions in English Schools. *British Journal of Educational Studies*, 59, no.1:41-62.

Wolf, P.J., Witte, J., Fleming, D.J. 2012. Special Education and the Milwaukee Parental Choice Program. SCDP Milwaukee Evaluation Report #35. http://www.uaedreform.org/SCDP/Milwaukee_Eval/Report_35.pdf

Wuttke, J. 2007. Uncertainties and Bias in PISA. In *PISA According to PISA: Does PISA Keep What It Promises?* Ed. Hopmann, S.T., Brinek, G., Retzl, M. Berlin and Vienna: LIT Verlag.

Young, T. 2010. British schoolchildren now ranked 23rd in the world, down from 12th in 2000. *The Daily Telegraph*, 7th December 2010.