

2017-08-16

Education governance and standardised tests in Denmark and England

Kelly, P

<http://hdl.handle.net/10026.1/9910>

10.1080/02680939.2017.1360518

Journal of Education Policy

Informa UK Limited

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

Education governance and standardised tests in Denmark and England¹

Peter Kelly, Karen Andreasen, Kristine Kousholt, Elizabeth McNess & Christian Ydesen

Abstract: *In this study we identify and compare the impact of standardised student assessment in England, an established neoliberal context, and in Denmark where a neoliberal education reform agenda is emerging in response to both national concerns and international governance. National reading tests for students aged 11-12 years, long established in England, were introduced in Denmark in 2010. The form they take differs considerably, being primarily formative in Denmark and largely summative in England. Culturally sensitive extended semi-structured interviews were conducted with both teachers and students and analysed to identify the extent to which neoliberal reform is mobilised through testing in each context and how testing shapes curriculum and pedagogy. Significantly, we find that in Denmark, where professional judgement still dominates, teachers often deploy pedagogical approaches to service what they believe to be their students' best interests. In England, however, teachers try to accommodate a concern for both their students' and their own interests, and the pedagogy they enact is more often controlling, instrumental and reductionist; their wish to be proactive is compromised by their need to be responsive. Hence we show how policy technologies shape practice to undermine a deliberate pedagogy rooted in ideas legitimated through scholarship and experience.*

Keywords: Comparative, sociology, global and national educational governance, policy technologies, national standardised testing, reading tests

Corresponding author: Dr Peter Kelly, Institute of Education, Plymouth University, Plymouth PL4 8AA, UK, email: peter.kelly@plymouth.ac.uk

Short Biographies

Dr Peter Kelly is Reader in Comparative Education at Plymouth University, UK, where he has worked since 2003. Main focus: the relationship between education policy and pedagogy, and their effect on educational inequality in Northern European countries.

Plymouth University, UK, email: peter.kelly@plymouth.ac.uk

Dr Karen Andreasen is an Associate Professor at Aalborg University where she has worked since 2012. Main focus: Education, educational assessment, history of education, educational sociology, in- and exclusion.

Aalborg University, Denmark, email: karena@learning.aau.dk

Dr Kristine Kousholt is an Associate Professor at Aarhus University where she has worked since 2015. Main focus: assessment practice in primary and lower secondary school and teachers' and pupils' perspectives on and participation in the Danish national standardized tests.

Aarhus University, Denmark, email: krko@edu.au.dk

Dr Elizabeth McNess has been a Senior Lecturer in Education at the Graduate School of Education, University of Bristol, UK, since 2005. Main focus: comparative analyses of national education policy and practice and its impact on teachers' work and pupil experience.

University of Bristol, UK, email: Elizabeth.McNess@bristol.ac.uk

Dr Christian Ydesen is an Associate Professor at Aalborg University where he has worked since 2012. Main focus: history of educational testing, accountability, ethnicity and education and international organisations.

Aalborg University, Denmark, email: cy@learning.aau.dk

¹ Kelly, P., Andreasen, K., Kousholt, K., McNess, E. & Ydesen, C. (2017) Education governance and standardised tests in Denmark and England, *Journal of Education Policy* [Published online September 2017].

1. Neoliberal education governance, policy technologies and standardised national tests

For neoliberals, markets are the social arrangements best placed to maximise human flourishing (Harvey, 2007). As such, the marketization of education provision coupled with an emphasis on competitive student evaluation through high stakes testing dominates both international and national neoliberal education reform agendas (Au, 2008), and in this context international comparative student tests have flourished. The Organisation for Economic Co-operation and Development's (OECD) Programme for International Student Assessment (PISA) surveys, for example, are highly influential (Mundy et al., 2016), and in recent years the OECD has sought to increase their impact by linking explanations for differences in student scores to policy recommendations (Martens 2007; Sellar & Lingard, 2014). This has increased pressure on policymakers for neoliberal reform, leading Lingard and his colleagues (2013) to add a level of global governance to Ball and Bowe's (1992) policy cycle. Administrations have become preoccupied with governance by numbers (Ball, 2015; Ozga, 2008) as international comparative survey data complements national testing regimes, with the global and the national providing commensurate spaces for measurement (Lingard, 2011). As a consequence, national education reform agendas often share performative similarities (Ball, 2013); those for whom the worth of education extends beyond that provided by neoliberal accounts fear growing global influences (Ball, 2012) are helping to homogenise educational practice (Breakspear, 2012).

Our concern here is to elaborate two particular cases of extended policy cycles which differ by degree; in England where an established neoliberal discourse dominates education, and in Denmark which we characterise as representing an emerging neoliberal education discourse. In his analysis of neoliberal education governance, Ball (2013) identifies marketization, managerialism and performativity together as policy technologies; intersecting practices and related artefacts, deliberately deployed to shape and regulate human activity. Together, these serve as a framework for us to ascertain the extent of neoliberal governance in each context.

In so doing, we consider standardised national reading tests in England to be an element of each of these policy technologies, whilst their counterparts in Denmark have been partially mobilised for similar purposes. Their activity is partially deliberate but also brings unintended consequences. We employ Bernstein's account of the pedagogic device (1990; 1996) to analyse how assessment shapes curriculum and pedagogy and the consequences of this process.

In contrast to Lingard and his colleagues (2013), the originality of this study lies in comparing the role standardised assessment plays in an established and an emergent neoliberal context by focussing on policy enactment in schools and classrooms, something Ball and his colleagues (2012) assert is a complex but significant focus for research. This allows us to identify commonalities and variations across two regimes differing in the extent to which neoliberal discourse is implicated in the enactment of standardised assessment and practices associated with it, and thereby elaborate a fuller understanding of neoliberal governance and its effects.

2. The wider policy context

Education has become the focus of reform because it is believed to assist economic growth. With its emphasis on services, England's liberal market economy allows lightly regulated enterprises to respond quickly to changing markets (Hall and Soskice, 2001). But this also brings inequalities in pay, job security and working conditions between high and low skilled workers. Of particular concern are the increasingly precarious lives of unskilled workers in short-term employment (Standing, 2011). In wishing to cultivate a generically skilled, flexible and socially mobile workforce, successive

governments have considered education to be at the heart of economic development. These same governments have also tightened welfare provision by policing eligibility and reducing payments. Their intention; to discourage welfare dependency by reducing opportunities for undeserved gain and ensuring people who work are rewarded for doing so. As a result, the provision of in-work benefits has increased, leading inadvertently to a further rise in low paid and temporary employment.

Over time, Denmark's social market, which combines manufacturing and the exploitation of natural resources with service provision, has delivered generally high wages and relatively low inequality (Hall and Soskice, 2001). Workplaces are highly regulated and emphasise community, consensus and trust. Strong vocational education and training provides large numbers of specialised skilled workers who then enjoy relatively stable employment. However, increased unemployment since the mid-2000s has provoked calls to reduce regulation and taxes, which are thought to stifle job creation, and to reform expensive social welfare provision. Based on the principle of universal access, the welfare benefits and services allowed a relatively high degree of citizen autonomy, limiting their reliance on family and market (Esping-Andersen, 1990). But recently various benefits have been reduced inciting widespread political debate. This liberalising influence has also touched education, where there has been a shift from the traditional emphasis on *Bildung*, the process of personal formation that brings about the inner development, to a more utilitarian focus on competencies.

3. Assessment within an emerging Danish and established English neoliberal education discourse

In England, children attend primary schools from age 5 until 11 when they transfer to secondary schools, whereas most Danish children go to *folkeskoler* which combine primary and lower secondary education between the ages of 7 and 16. The form of the statutory curriculum in both Denmark and England is similar, although, in general, curriculum goals in Denmark are broader than in England. Partly because of the professed economic value of education, English primary schools have been subject to nearly three decades of reform with teachers' work subject to considerable surveillance. Yet the traditional focus on individualised teaching remains (Goodson & Lindblad, 2011). Over the past fifteen years a significant number of school policy reforms have been implemented in Denmark beginning with the introduction of common objectives for *folkeskole* students in 2003. Nevertheless, Danish education continues to place high importance on the group rather than the individual, and values greatly a close relationship between a class teacher and one group of pupils (Osborn, 2004), and there remains greater trust of teachers and schools in Denmark than in England.

Lingard and his colleagues (2013) suggest international assessment regimes are in dialogue with vernacular ones; those administered at a national or regional level, each contingent on their particular social, historical and cultural circumstances. We have written in detail about the history and context of national testing in Denmark and England elsewhere (Andreasen et al., 2015). In summary, whilst standardised assessment is now established for students of all ages in English schools, it is a relatively recent addition in Denmark, having been introduced in 2010. Whilst traditionally liberal and individualist, the turn to techno-rationalism and preoccupation with accountability began in England more than twenty years ago (Goodson & Lindblad, 2011), and standardised tests, particularly in English and mathematics, have been used to monitor and compare student and thereby teacher and school performance since the early 1990s indeed, this has been central to education reform. Because of this, the 'shock' (Wiseman, 2013) of the PISA 2000 results (OECD, 2001) provoked little reaction in England (Knodel & Walkenhorst, 2010), whereas in Denmark and elsewhere, the policy response included an increased focus on standardised national testing and

evaluation strategies (Andreasen et al., 2015; Egelund, 2008). Behind these changes we can identify dominant Danish education discourses since 2002 as having partly moved from humanist and socially-orientated (Hermann, 2007) towards skills-competence and individualist.

With policy processes subject to the mediating influences of national cultures, traditions and dominant education discourses (Ball, 2013), the vernacular forms of standardised testing in these two countries differ considerably. Notionally it is possible for subject assessment in schools to take many forms. It can be formalised in tests or tasks which students carry out, designed to allow performances to be evaluated against set criteria; such activities and their outcomes are sometimes standardised to allow comparison across groups and contexts. Or it can be less formal, where various people including students themselves appraise subject development more broadly, sometimes using agreed knowledge hierarchies. The outcomes of assessments can be reported as a summation of what students have accomplished, used to feedback more formatively how they might improve, or serve a combination of these purposes. Our focus here is standardised assessment and, in outline, English children sit common paper and pencil tests for reading, writing, mathematics and science together, in one week in May in the school year in which they are 11 years old, whereas Danish children sit individualised, adaptive computer-based tests, with different subject areas tested at different times over their school career; children are tested in Danish in the school years in which they are aged 8 and 11 and in mathematics at ages 9 and 12. Much has been written about testing in each of Denmark and England (for example, in Denmark, Andreasen & Hjørne, 2013; Ekholm et al., 2004; in England, Adams, 2008; Tomlinson, 2005), but only a small number of studies have compared assessment practices across these countries (for example, McNess et al., 2003) and none have done so since the introduction of national standardised tests in Denmark in 2010 or linked this to a broader analysis of the extent of neoliberal education reform. As reading is tested in both Denmark and England in the school year in which students become 11 years old, this provides our comparative focus. But the form of these tests differs considerably and whilst test results in Denmark at the time of this study had primarily a formative purpose, those in England remained largely summative.

Despite clear differences in policy, it is possible that the wider effects of tests in Denmark and England will partially converge because of the relative significance student assessment carries. In England in particular, the consequences for teachers and schools of poorer than expected test performances by students are significant, damaging reputations and triggering increased scrutiny in the form of Ofsted inspections. This can result in fewer pupil enrolments and lost income for the school. Students' views of themselves as learners can suffer, as can teachers' expectations of them as they move into secondary education. These consequences heighten the stakes of standardised tests, with negative effect (Stevenson and Wood, 2013; West, 2010); one British Parliament select committee report (House of Commons, 2008) identifies tests as narrowing the curriculum in favour of those subjects tested, increasing the amount of teaching to the test, promoting shallow learning and short-term retention of knowledge and increasing pupil stress and demotivation. Later we consider whether the formative intentions of the Danish tests lower the stakes and serve to moderate these effects.

4. The value of cross-national comparisons in education policy research

Alexander argues for a comparative approach which, 'maps the key elements in the act of teaching and links them with the processes of curriculum transformation from state to classroom' (Alexander, 2001: 507). This includes exploring how social and political values translate via situational circumstances into acts of teaching. The neoliberal education reform agenda in England is long

established, whilst a similar programme is emerging in Denmark. Comparison in each of these circumstances allows an improved understanding of the broad relations between teachers, practices and pupil experiences. Together these can inform an understanding of policy implementation and the policy cycle (Ball & Bowe, 1992); concluding an extensive review of comparative education policy research, Busemeyer and Trampusch suggest this is much needed: 'this domain would benefit from theoretical work on the micro–macro problem in understanding outcomes of education policy' (2011: 434).

The comparison provided in this study was designed to allow insight into how the contrasting approaches to standardised assessment used in two neoliberal policy contexts differing by degree shape curriculum and pedagogy. Methodological rigour was provided in two ways. First, the insights of insider researchers were combined with the perspectives gained from teachers and students; for Kelly (2014: 2), 'insiders bring potential insights into nuanced cultural signifiers, but their familiarity may lead to the recycling of dominant assumptions; outsiders bring a freshness of perspective, but may impose their own worldviews uncritically'. The interviews and early analyses of these were conducted by cultural insiders who appreciated the values and expectations of participants. Later, cultural outsiders looked for similarities and differences across countries, and together the research team considered the wider analysis and implications. Second, we frame our analysis of each country using Bernstein's (1990; 1996) account of the formation of curriculum and pedagogy under the influence of assessment. To enhance the veracity of our comparisons we have tried to match the two contexts in terms of regional and local characteristics, school size and catchments and the age and relative attainment of students. Further, we only engaged with schools which external advisors considered to be successful and with classes whose teachers were experienced and considered competent.

5. Analysing the influence of testing on curriculum, pedagogy and students

In this study we adopt Au's (2008) theoretical frame which is based in Bernstein's account of the pedagogic device (1990; 1996) to analyse how testing helps form and shape curriculum and pedagogy. Bernstein does not regard this mechanistically; rather, the pedagogic device describes the process by which knowledge is transformed into curricula, classroom resources and forms of organisation, teaching materials and approaches, and ultimately interactions between teachers and students (Singh, 2002). In this Bernstein proposes distributive rules which describe the regulation of knowledge distribution; how the knowledge and consciousness of some social groups are privileged as common sense, thus delimiting the thinkable from the unthinkable and marginalising those whose norms lie outside these parameters. Bernstein asserts that ways of knowing thus selected are translated into curricula and pedagogies through processes of recontextualisation, and he invokes evaluative rules to explain how success is regulated within this through the identification and valorisation of particular knowledge and classroom practices - and thereby teachers and students - over others. Rules here can be regarded as regulatory processes which privilege particular outcomes. For Au, high-stakes standardised tests thus express distributive rules, which account for the spread of different forms of knowledge and consciousness to diverse social groups (Bernstein 1990; 1996), by: (a) defining what counts as legitimate school knowledge; (b) exerting considerable control over the form in which teachers present content knowledge in the classroom, often mirroring that used in the tests; and (c) leveraging control over teacher pedagogies towards teacher-centred pedagogies in an effort to keep up with the content and knowledge forms required by the tests. Evaluative rules, which condense the whole meaning of the pedagogic device (Bernstein 1990; 1996), are expressed in how, by structuring knowledge and its transmission, high-stakes standardised tests 'actively select

and regulate ... students' educational success' (Au, 2008: 639). To ascertain the influence of standardised assessment in the present study, we compared the distributive and evaluative rules in Danish and English tests to discern parallels and variations in how they formed and shaped curriculum and pedagogy and their impact on students.

We focused on the experiences of six students aged 10-11 and their class teachers in each of three public schools or *folkeskoler* in Denmark and three primary schools in England; that is, eighteen students and three teachers from Denmark and the same number from England. Schools were chosen to represent the geographical, sociocultural and economic diversity of each country. Schools included one rural, one small town and one inner city, and all were of mixed catchment (see Table 1). Students were divided into three boy-girl student-pairs in each school, chosen as high, middle and low-attainers by their teachers (where attainment is their anticipated test score level). Teachers and student-pairs participated in one semi structured interview each led by a fluent national language speaker with experience of similar settings; the teacher interviews lasted about one hour and student-pair interviews about 30 minutes. These took place in quiet areas close to the students' classrooms during the school day. Interviews took place prior to recent policy changes in assessment in England (STA, 2015), the significance of which we will discuss at the end of this paper.

Interviews explored teachers' and students' experiences of the national reading tests and were translated into English and then analysed using Au's (2008) theoretical frame to identify and compare the distributive and evaluative rules in each country (see Table 2). For the distributive rules we looked at: (a) the knowledge and processes tested in the national reading tests; (b) how these were processed to make them testable; (c) the ways of teaching privileged by the national reading tests; and, in consequence (d) the students who were advantaged by the national reading tests and those who are disadvantaged. For the evaluative rules we considered: (e) how the national reading tests acted as a set of rules for evaluating and regulating classroom practices and teacher and student identities; and (f) how the national reading tests thereby served to privilege and marginalise aspects of curriculum and pedagogy. Thus variations across the two countries were identified.

Table 1: Descriptions of schools and teachers

	Denmark	England
Teacher A (Denmark) D (England)	This is a large school in a major provincial town with some 600 pupils on roll; the teacher is male with 30 years teaching experience	The school is situated in a large seaside town and has a socially mixed catchment with 200 pupils on roll; the teacher is female with 25 years teaching experience
Teacher B (Denmark) E (England)	The school is situated in the outskirts of a major provincial town and has some 200 pupils on roll; the teacher is female with 9 years teaching experience	This is a medium sized village school serving a socially mixed catchment and has 350 pupils on roll; the teacher is female with 20 years teaching experience
Teacher C (Denmark) F (England)	This is a medium sized village school with some 250 pupils on roll; the teacher is female with 15 years teaching experience	This is a medium sized village school serving a socially mixed catchment and has 300 pupils on roll; the teacher is male with 17 years teaching experience who is also deputy head of the school

We now consider, in turn, the evaluative and distributive rules which regulate pedagogic discourse in each country, overlaying this with how standardised national tests mobilise elements of

each of the policy technologies identified by Ball (2013). We begin by considering the extent to which student test results in each country were performatively regarded as exclusively capturing the worth of the organisation to which they pertained and the various actors within. Next we consider how, in working to improve test performances, teachers and students together were the subjects of school management. And finally we consider the extent to which both teachers and students were thereby constructed in each country as operating within marketplaces where students exchanged their knowledge and skills for grades.

Table 2: Linking the analysis to the interviews

Analysis	Interviews
<p>To explore the evaluative rules regulating curriculum and pedagogy through the tests we asked of the data:</p> <ul style="list-style-type: none"> – What are the tests for? – How do the tests shape cross-school priorities and practice? – How do the tests shape classroom priorities and practice? 	<ul style="list-style-type: none"> – Teachers were asked about what they felt were the aims of the tests and the expectations, broadly and with regard to the tests, of school inspectors and advisors, senior managers, colleagues, parents and students. – Teachers were asked how significant or important they thought the tests were for themselves. – Students were asked how significant or important they thought the tests were for their schools and teachers. – Teachers were asked about how they had responded to the tests. – Students were asked about their expectations of the tests and how they had prepared for them.
<p>To explore the distributive rules in relation to the national tests we asked of the data:</p> <ul style="list-style-type: none"> – Which students are most and least comfortable with the form and content of the tests? – Who are advantaged and legitimated by the tests and who are disadvantaged? 	<ul style="list-style-type: none"> – Teachers were asked about the advantages, disadvantages, strengths and weaknesses of the tests, and what students needed to be able to do to be successful at each level. – Students were asked what they needed to be able to do to be successful.

6. Evaluative rules for the national tests: their purpose and how this was realised

6.1 Performativity and purpose

Despite their stated formative purpose, standardised national testing was received cautiously by Danish teachers, with some suspicious of the motives behind them. All teachers interviewed believed the long-term aim of the tests was more than providing information on students; it was to create a foundation for the comparative evaluation of schools and teachers. However, as Teacher C reported, initially the tests suffered numerous problems and were heavily criticised, leading the Ministry of Education to approve wide timeframes during which the tests could be conducted to ensure schools and teachers cooperated; this compromised their standardisation and thus the possibility of comparing results. In any case, all agreed there was no evidence that tests were being used to evaluate schools or, indeed, individual teachers, although this would be possible in a limited way as things stood. Teacher A described the municipal quality report covering education, a public document; whilst national test results were only one aspect of the report, there was concern that this information could be abused, although, according to Teacher C, civil servants only had access to

the collective test results from schools within their jurisdiction and not results of individual pupils, and it was these alone which were reported. Together these teachers signalled a picture of possible performative intent, as yet unrealised.

In contrast, standardised national tests in England were high-stakes precisely because their intent was both summative and comparative as they were used for school evaluation. This represented an established neoliberal position. So much rested on the test results for schools that, according to Teacher D, 'they are everything that the school works for, more than anything'. The results were used to categorise and compare schools and discern their year on year progress, with implications for the frequency and intensity of Ofsted inspection and intervention and ultimately continuation or change in the school and its staff. Not only that, but, as Teacher F suggested, governors, colleagues, parents and prospective parents made judgements about schools and teachers on the basis of their published results and associated commentaries in the local media, with implications for school popularity and recruitment and for the reputation of individual teachers. Hence school leadership also put considerable energy into anticipating what Ofsted would make of the results, their biggest concern given the importance of Ofsted evaluations both for schools and the individual teachers who work there.

As such, the worth of schools and all of those concerned with them was reduced almost entirely to test results. Teachers were openly judged on test results. Year 6 (the final year of primary school for pupils in England) teachers especially benefitted from the status afforded by their importance in preparing students well for the tests, but, according to Teacher D, were also in a risky position as their reputation could be lost in one poor year; as Teacher F said, 'I feel a lot of pressure on me as Year 6 teacher because I cannot afford to let it go - with parents' expectations and the school.' However, so as not to be overly dependent on one individual teacher, in all three of the English schools there was some spreading of the responsibility for preparing the students and administering the tests across a team of teachers and teaching assistants; as one teacher indicated, 'the load is spread. The boosters [additional classes for targeted groups of students, aimed to 'boost' their test scores] are delivered by six members of staff ... including the deputy head and special educational need coordinator'.

Within Denmark, then, the extent to which testing played a performative role indicated an emerging neoliberal discourse. Some teachers feared that head teachers might use the test results to compare teachers in the future with one teacher concerned that the school management was very interested in the test results. He cautioned that the composition of children in each class be taken into account when evaluating the test results, not least the number of children with Danish as a second language. He hoped that conclusions about teacher performance would only be drawn cautiously and with common sense, adding that, so far, there had not been a dialogue with the school management about the use of the test results for evaluative purposes. But another teacher saw less of a problem with the school management having access to test results and confirmed that it might be possible to use the tests as a governing tool, especially if a teacher had long had below average results.

With such fears in mind, some in Denmark also voiced the need for a shared approach. Teacher C insisted teachers speak together and school management take responsibility for coordinating how to handle the national tests to avoid exposing individual teachers to any criticism resulting from poor results. Further, in terms of weakening their influence, Teachers A and B insisted national tests be used in parallel with other evaluative tools to inform their judgement in coming to a holistic appraisal of students. Like their Danish equivalents, English teachers felt test data could be better used in the evaluation to support individual children, but their timing at the end of their primary

school careers precluded this. However, the Danish teachers felt that the tests as they stood were primarily for teachers, parents and children by providing information which could contribute to an evaluation of the individual child.

6.2 National tests and the management of schools, curriculum and pedagogy

Because in England tests were assumed to capture the value of schools and actors within, the work of management was to use both standardised and their own test data to follow closely student development and inform changes in teaching, the success of which were subsequently judged by their effect on test results. Hence everything regarded as worthwhile in schools was seen to be embodied in these measures which all activity sought to increase. As such, tests had a direct governance effect by strongly framing (Bernstein, 1996) or shaping the content and form of teaching in ways whereby teachers acted in advance to ensure results were acceptable to Ofsted. This framing positioned principals and teachers as middle managers and implementers, without ultimate authority, and involved them constantly monitoring students' performance whilst coaching them to improve. Hence data drove everything. Schools used pupil tracking software to follow the progress of individuals and selected groups of students. To allow greater precision, levels were partitioned to identify higher, middle and low achievement of them. This software was quite sophisticated and allowed children to be classified into groups (Bernstein, 1996) and then compared with both the national mean performance of similar groups and that of similar groups in similar schools on the basis of gender, ethnicity, learning needs and their relative socio-economic status (using their eligibility to free school meals as an indicator). Through this process, students who might underperform were identified, allowing the school to intervene.

Test data in Denmark could not be deployed in ways like this because it was not measured against standardised descriptions of pupil progress. Nevertheless, according to Teacher C, the publication of aggregate school results in municipal reports had had a less direct effect by engendering a sense of defensiveness in a number of school principals, spurring some to wish for improvement. This had led to the introduction across schools of a new strategy with guidelines for national test preparation and administration. However, the Danish strategy had had a less pervasive influence on curriculum and pedagogy and, unlike English data-led approaches, had not targeted individual students. Indeed, there were generally mixed feelings amongst Danish teachers about the usefulness of tests in supporting students. All teachers interviewed were broadly positive, agreeing that the tests provided a reasonable indication of students' proficiency level in reading, and each, to a greater or lesser extent, took account of students' test results in their subsequent planning. However, Teacher B emphasised that tests provided a momentary picture reliant on how students felt at the time of the test, adding that the interpretation of test results called for qualitative knowledge of students. For example, Teacher C described how some students rushed in a cavalier manner through the test, whereas others were less secure and read questions carefully several times before answering. Nevertheless, all teachers interviewed complained some features of the Danish tests reduced their reliability, including some test items and how the tests adapted as students proceeded through them. The adaptive principle worked by placing more weight on the first questions compared to later questions. As a result, many teachers stressed with students the need for them to attend carefully to these first questions, and not to move on too hastily if, on first impression, they found the questions hard. Still, Teacher B felt that broadly the adaptive principle was a significant strength of the tests, and she contrasted them with pencil and paper tests where adaptation would not be possible. Generally, though, Danish teachers had less confidence in the tests as a unique measure of worth, regarding them instead as helpful in supporting learning and

informing teaching, even whilst their managers began to shape classroom practice to try to improve results, despite the low stakes attached.

Notwithstanding teacher reservations about their veracity, the governing effects of the tests at the classroom level in Denmark were to increasingly frame the content, sequencing and pace of teaching and the approach to and focus of assessment, what Bernstein (1990) calls the instructional discourse, to the extent that Teacher C felt that they had a negative effect on some children's schooling. According to all three Danish teachers, they and many of their colleagues planned to cover what would be tested in their teaching; indeed, many teachers used the test results to decide what and whom to focus on, sometimes referring to the national tests in their teaching and identifying for students the things which were likely to be included. For example, all three teachers reported that, since their inclusion in the national tests, they and their colleagues increasingly incorporated work on proverbs and idioms in lessons. They also taught specific lessons on letter sequence, which some test items required to be decoded and trained students in techniques they could use to show they had a good understanding of texts along with tips on what to look out for. So, with word chains (a category of test item where three words were written without separation, such as *informedilluminatedinvaded*, and students had to identify the words and divide the word chain correctly) pupils were told that often the words to be identified all began with the same letter. Danish students confirmed these activities, adding that they also took a web-based mock test in the autumn which described test items and provided instruction on how to complete them. All of the Danish teachers were positive about the mock test, which served to reassure students, familiarising them with the test situation and how the test worked; as a result, teachers believed, testing was implemented as a natural part of schooling. Teacher B said that the mock test was also used as an occasion for talking about reading strategies. Many of these approaches were included in national guidelines for national test preparation and administration, and Teacher C reported that following these helped raise her school's results considerably. These contained notes on aids which could be used to help pupils, the optimal work speed, how to work slowly and carefully through the mock test items with the children and how to work around difficult test items. But otherwise preparations were limited, since not even the teachers knew much about the content of the test other than what was made public on the national tests' homepage and student and teacher recollections of previous tests. Indeed, in general terms the classification and distinction between subject teaching and teaching for the tests was somewhat porous. Further, despite the variety of approaches used, all of the students interviewed indicated that test preparations were kept low-key, and they remained unaware of any teachers identifying individual students in Denmark through their predicted test performance. And in terms of the division of labour between teachers and students, an aspect of what Bernstein (1990) calls the regulatory discourse, both largely continued to share responsibility although teachers were increasingly directive. Yet regardless of this, it is clear that whatever the status of the measure, the nature of the measure had some shaping effect on curriculum and pedagogy.

Given their importance in England, the shaping effect of testing was much greater as the whole of Year 6 was planned towards the tests. Indeed, to some extent, the whole of the school was geared towards tests with children regularly taking 'optional' tests (that is, optional to the schools, not the students), often termly, and always at the end of each school year from the age of seven. The three English teachers expressed confidence in the national tests, regarding them as robust and reliable indicators of student reading comprehension focussed particularly on making inferences from the texts and finding and using evidence from the texts to support their answers. They agreed that the additional tests aimed at the highest achievers, which used more mature texts and vocabulary,

required more sophisticated student responses. Their instructional discourse was directly focussed on the tests, and students confirmed that schools followed similar approaches, beginning in the term before the tests. Approaches were strongly classified as test preparation and included the teaching of techniques and strategies that aimed to ensure children were familiar with and did as well as they could on the tests. Pedagogy was highly framed by teachers, and in terms of the regulatory discourse, teachers positioned themselves as being largely responsible for student learning. Teachers ensured there was something relating to the tests in every literacy lesson. These occurred once a day for about an hour at a time. Students sat regular, often weekly, practice tests, given under test conditions so that they could experience what it was like to do the tests in the time given. Tests were followed-up by teachers working with a different group each day, helping the children talk about the questions and their responses, and marking these together. The children learned about the best ways of approaching the tests as a whole, how to make best use of their time, the kind of questions they would be asked including the style of individual questions and what markers were looking for. All were given regular homework involving the completion of booklets of practice tests, targeted at different groups depending on their assessed level. In addition, individual students identified as in danger of not achieving the required minimum level had further sessions, sometimes at lunch time, working closely with the school's special needs coordinator and followed up in class by a teaching assistant. Weekly classes before school, called booster groups, targeted groups of children identified at the end of the autumn term as on the borderline between levels. Students described these as fun and relaxed. Clearly, by the time they came to the tests, the children in England had done so many tests before that they knew what to expect, although many were still anxious beforehand.

So, curriculum and pedagogy were responsive, with adaptations made by school managers and teachers in England to maximise test performance. Here it appeared that the more a single measure was focussed on, the greater the shaping effect; and the greater the stake attached to a single measure, the more it was focussed on. However, even in an emergent, low stake context like Denmark where a number of assessment approaches were used, testing had a noticeable shaping effect, and it seemed that whether the expressed purpose of the tests was summative or formative was less important, something we will now consider further.

7. Distributive rules for the national tests: what counted as success and who could best achieve it

7.1 National tests and the regulation of success

The emerging performativity and associated management practices in Denmark allowed student test results to be partially constructed as marketable or exchangeable, their exchange value - which was fixed in England - being negotiable, depending on the individuals involved. Being formative in intent, Teacher C made it clear that Danish teachers could decide the conditions under which the test was conducted; electronic aids such as devices that could read aloud were used to assist the weakest students, and it was up to individual teachers to assess whether their use was taken into account when assessing the results. Teacher B confirmed that sometimes teachers allowed students to sit in isolation to aid their concentration. However the tests were conducted, though, the teachers only gave general feedback to the children and parents and did not discuss specific test items, whilst parents tended to ask why children's scores deviated from their expectations and what they could do to help improve them. Hence, on such occasions reading was reduced to test results alone, and these were then exchanged by teachers for parental recognition or concern. The outcomes of such discussions were implemented in a 'pupil plan' for working towards the academic goals of individual students in the subsequent school years. These plans were a central tool for sharing information

with parents about where their children needed help. But students described that results were reported not as scores, but using the categories of median performance and below or above. As one student said, '[We were told] about our proficiency levels ... there were three things, it was spelling, language understanding, and ... I do not remember the third thing, but here you were rated above average, or those grades or evaluations', and the first said, 'Understanding, spelling and language understanding'. This was remarkably summative in character, albeit with some flexibility, and involved the early commodification of test performances as labels partially used to demonstrate student worth largely for parents. Objectifying a measure thus and thereby investing in it greater singular significance than is warranted by the many mitigating factors detracting from its veracity is the first step towards commodification. In contrast the English teachers largely accepted the accuracy and fairness, objectivity even, of the tests - although not always the reliability of external marking - and their use in making summative judgements and comparisons without contextualisation. For the most part it was important for these teachers that children took the tests within the rules and without flexibility so as to maintain the credibility of their scores and allow legitimate comparisons to be made. This helped sustain the belief that test results faithfully captured the worth of students, teachers and schools. Otherwise, teachers suggested, the English national tests were of limited formative value to children as they were taken too late to inform their primary teachers, and secondary schools used primary school teacher assessments in order to group students on admission, although they conceded that taking the tests in some ways acted as a preparation for future tests. Hence in England, the increased stake or exchange value of particular scores over others and the opportunity for scores to be exchanged for school, teacher and student success and all that went with each of these clearly constructed test results as valued entities in their own right or commodities.

The apparent objectivity of results in England increased their legitimacy, allowing the value of and opportunities for exchange to increase. But clearly their objectivity and, by implication, their validity and reliability could be challenged. In terms of what they were thought to measure, the Danish teachers agreed that the national reading test gauged students' capacities to decode language and to understand a text. Test items like the word chains described earlier focussed on decoding whilst others required students to read and respond to a text. However, the English tests focused solely on reading understanding. It is perhaps not surprising that teachers identified Danish students from more privileged social circumstances and those who spoke with their parents in Danish, which was often not spoken in the homes of bilingual children of minority ethnic heritage, as doing better on the reading test. This fits with OECD reports (Nusche et al., 2010) and may explain why teachers in Denmark were particularly sensitive about cultural bias in the test. Teachers considered bilingual children to be normally very good at decoding but severely challenged when it came to understanding, and it was Teacher B and C's view that children's knowledge of proverbs and idioms in particular was dependent on their home environment; indeed, Teacher C stressed how important the understandings children brought from home was for how well they performed in the test. Teacher B argued that bilingual children of minority ethnic heritage did not have the same knowledge of the Danish culture, history, language and environment as those from families established in Denmark, identifying this as a cultural bias inherent in the test and suggesting this was another reason why teachers' knowledge of individual children was so important when considering the test results. It is important to recognise that, whilst such views might be benign, they could also reflect more deeply set cultural preconceptions. One response had been to place bilingual children of minority ethnic heritage in special reception classes which introduced them to areas considered culturally significant such as proverbs and idioms; an act of separation rather than inclusion. This

targeted strategy was surprising given that mixed teaching groups were the norm; more so, given the formative intent of the tests, because the wish to counter cultural bias in the tests would appear to have been motivated by a summative need for judgements to fairly reflect students' reading attainments. In addition, with the time limit for the test being 45 minutes Teacher A emphasized the importance of students' reading speed for all. Further, Teacher B suggested children's like of different genres could affect their test performance, with those preferring fiction doing better than those who like non-fiction. However, teachers did not comment in depth on the role either of these played in giving students advantage in the tests or in relation to those groups of students who needed particular support.

The English teachers interviewed focussed on the impact of socio-economic status and did not mention ethnicity. Again, this was perhaps unsurprising because, in recent years, national debate has been concerned largely with the achievements of white students from socio-economically poorer homes which are broadly lower than those of minority ethnic students (OFSTED, 2015). Teachers felt that fluency was the most important thing students need to be successful on the tests. They shared the same analysis: children needed to be able to read quickly, write quickly and remember key points from the texts. To achieve the higher levels they required a wide vocabulary, which teachers believed would probably be best gained through reading a lot of fiction including some aimed at young adults along with some non-fiction. Reading fiction, especially more adult fiction, teachers suggested, allowed children to better understand and use descriptive forms of expression, including those relating to characterisation and emotion.

There is a lot of reading within the time limit so the weaker readers who can't read at speed don't get a fair reflection of what they can do. We tend to use scribes for our very slow writers because we know they can read but can't always get their ideas down on paper. Overall they don't reflect the abilities of the children because it is all about speed.
(Teacher D)

Children who read and understood slowly were disadvantaged; Teacher F speculated, 'it would be interesting to see if there would be a different outcome if there was no time restraint; I think there definitely would be'. So, whilst their Danish counterparts focussed on ethnic heritage, English teachers identified how the characteristics of social and gender groups could provide an advantage or not in the tests; many boys tended to read mainly non-fiction whilst girls often had a more balanced reading diet, and low achievers who only read that with which they were comfortable were not challenged with more sophisticated texts and vocabulary; as Teacher D suggested, 'if you read fiction you've got a far better vocabulary and better understanding of characters' emotions'.

So, in both cases, stratification in test results aligned with social trends originating outside school and concerning the cultural knowledge or reading preferences, gendered or otherwise, which children brought to school. These complemented reading and writing speed in aiding the achievement of higher test scores. It is interesting here that, in both countries, allowing variations in students' speed and efficiency to lead to stratified performance against the measure was relatively straightforward and much less complicated than trying to grade their level of understanding. Certainly it would be difficult to address such social and fluency differences within the constraints of short term test preparation. But whereas in Denmark, alongside some teaching of how to do the tests, teachers could also emphasise their own judgement and downplay the tests because they carried little weight and were not considered objective, in England teachers' only option was to teach children to be strategic to maximise their scores. Indeed, the English teachers suggested that teaching test technique could partly compensate for lack of reading fluency, taking those below the national norm up a level by helping them use their time well; as Teacher D said, 'the weaker readers

might skip all the big questions and do all the one mark questions, and sometimes the stronger readers might do that if they're not good at finding the evidence'.

Nevertheless, it is also clear that the relative autonomy of teachers in Denmark and their reliance on professional judgement allowed the preconceptions of teachers to influence their decision making; potentially, this could lead them to contribute further to the marginalisation of some of their students. The actions of teachers in England, however, were more likely to be in response to external measures.

8. Testing and neoliberal governance

From the foregoing analysis it is clear that testing regimes played an important role in the governance of education by mobilising policy technologies (Ball, 2013). Although practitioners often regarded them as neutral, helpful even, tests changed the meaning of practice and thereby shaped associated social relationships. We now identify features of the processes by which policy technologies shaped curriculum, pedagogy and thereby students through standardised testing.

Our analysis supports Newton's (2007) assertion that the distinction between formative and summative assessment is fluid with both having similar effects; in Denmark assessment was used to identify and summarise for learners and their parents, whether qualitatively or quantitatively, what learners knew, and in England, summative judgements were used formatively to inform future teaching. Nevertheless, the design of the tests varied depending on whether policy makers regarded them primarily as formative or summative and for whom they assumed the data was produced; education administrators, teachers, parents, students or combinations of these. Attempting to design tests with these intents brought unintended consequences in policy enactment (Ball & Bowe, 1992; Ball et al., 2012). Now, what mattered was not so much who testing was for as for whom the outcomes were important. Here, stakes for individuals were relative, defined locally and relationally, and could be real or imagined; but the higher the perceived stake, the more an interested party would seek to shape test outcomes in their favour. The extent to which testing was important for individuals depended on whether results could hinder or forward their interests, whilst how they could influence these outcomes depended on the freedoms, constraints and opportunities which were available to them. With little real benefit yet to be gained, Danish school managers or teachers focussed on using tests largely to serve the interests of their students, something they found professionally satisfying. But in England, external evaluation outweighed professional fulfilment. As senior managers, teachers and students were all objects of the externally controlled and standardised tests, all had an interest and all worked for high student test success together as part of a team, albeit an unequal one. But to reduce risk, senior managers were often reluctant to depend too much on the judgement and actions of individual teachers and so made use of external guidance, particularly that of Ofsted, and spread responsibility across a group of staff wherever they could. Similarly, teachers were reluctant to rely on children performing well on the test day and sought to increase the likelihood of this through thorough and comprehensive training. Thus senior managers strongly framed teachers and teachers strongly framed students.

Yet despite this apparent collegiality, depending on their roles, clearly some individuals or groups were better positioned than others to ensure that national testing processes and outcomes worked in their interests. In Denmark it was still largely teachers who held sway in decision making over parents and students, reflecting the broad toleration of professional autonomy and confidence in teacher expertise there and leaving senior managers with only limited influence; the testing regime fitted these relations. Here professional expertise used in the service of students still held sway over market choice. As a result, for the teachers and students interviewed in this study at least, it

appeared that, in the three years since their introduction, the changes to both curriculum and pedagogy resulting from the national standardised tests in Denmark had not become as significant in the lives of children as they were in England. Teachers regarded the tests as partially flawed and so used them alongside other approaches to provide formative and summative appraisals of students. As such, neither teachers nor students regarded the test results as the principal measure of student achievement and the basis for their performative use was shaky. Indeed, it was teachers' insistence on the primacy of formative assessment which brought about the development of adaptive tests, albeit with some technological shortfalls. Whilst these made comparisons of students and schools difficult, they were, nonetheless, used to provide summative reports of student progress for parents indicating some appetite for these and similar measures. Meanwhile, within the less individuated education culture, the traditional focus was on nurturing students' subject and pastoral development together within a view of *Bildung*, the idea that education brings about the personal formation of pupils. Teaching to improve test performance was not targeted at particular individuals, although ethnic heritage students whose home language was not Danish received additional support, an intervention designed to improve their test performance. It is possible here that teachers' relative autonomy allowed personal predilections and biases to influence decision making, raising the concern that little protection is offered against these when they are marginalising. Nevertheless, the knowledge required for successful test performance was becoming identified and classified as such. And whilst teachers seemed to have resisted an overt emphasis on teaching to the test and maintained a more holistic approach, there had been some increased teacher framing with the adaption of curriculum and pedagogy to cover tested elements in the form they were tested and, following external government guidance, offer advice to students and the opportunity to sit a mock test.

In contrast, suspicions about the secret garden of curriculum and pedagogy, raised by Prime Minister James Callaghan in a speech at Ruskin College, Oxford in October 1976, had challenged teacher autonomy and mobilised neoliberal reform in England many years before; reforms which had asserted the hand of the market over the expertise of professionals. Ironically, though, this demanded confidence in another group of experts, albeit respected for their neutrality, the schools inspectorate, who, using apparently objective measures of student attainment in the form of standardised tests, could inform school evaluation and parent choice. Now it was the views and expectations of Ofsted that were commonly accepted and dominated school agendas; as such, senior managers and teachers co-opted and developed Ofsted's individualised and data-led account of schooling to ensure they benefitted from the national tests whilst maintaining the compliance of parents and students. In so doing, they accepted the legitimacy of the tests as accurate and precise measures of learning and their comprehensiveness in capturing the worth of schooling. Together, this led to the dominance of management; using test data to identify underperforming groups and shared difficulties, responding to these with appropriately targeted pedagogic adaptations and interventions, and later evaluating these responses using further data. As a result, the schools we studied focussed their teaching almost entirely on test knowledge in the form it would be tested, particularly during the final year of primary school, using an approach which was entirely technological and reductionist. Their pedagogic response, set within this highly individualised and outcome oriented culture, largely involved instrumental interventions underpinned by universalist, linear and causal assumptions (Adams, 2008) about what works. The normal school routine became centred on test preparation and regular testing, with students and their work subject to strong classification and framing. Detailed and regular pupil tracking and the targeted teaching of test strategies dominated as teachers focussed on short-term gains in test performance during this final

year. This concern for student management prevailed over one for learning as the approaches used had little to do with improving children's reading, and a skills-coaching view of teaching pervaded; the strategies taught encouraged students to make the most of their current resources rather than seeking to address those factors which would improve reading fluency and comprehension and increase familiarity with a range of genres, all of which would take time. Technology allowed children's progress to be monitored closely, so this training could be targeted very specifically; without technology such instrumental micromanagement would have been impossible. This ensued largely because the tests were less about the appraisal of students and more important as the primary accountability measure for schools and teachers. As a result, teachers' classification of curriculum, pedagogy and students using notions like 'borderlines' and 'booster groups' legitimated practice in the form of highly framed interventions, allowing teachers the opportunity to demonstrate their worth and complex expertise. Some were better at this than others, some students responded to this better than others, and through such processes particular classroom practices were seen to be more effective than others. But whilst tracking software allowed the categorisation and monitoring of students which underpinned the instrumental response in England, testing software in Denmark controlled formative adaptations to student responses; those whose patterns of answers accurately matched their understanding, as predicted by the writers of the computer algorithm, were advantaged to such an extent that teachers encouraged all students to attend to this. Hence, in both countries, technology privileged normative behaviours, allowing teachers to correct the behaviour of students lying outside norms in ways which are worthy of further analysis.

Since 2016, tests for 11 year olds in England no longer reference national curriculum levels (STA, 2015). This is in line with other developments in school assessment. However the tests remain externally set and marked, with results used to measure school and teacher performance by comparing students' raw test scores to national averages. Hence it is our view that the analysis provided in this study still applies. This study is novel in comparing extended policy cycles in contexts contrasting in the extent to which they have been shaped by neoliberal reform. It is clear from our analysis that, in contexts where neoliberal policy technologies dominate, high-stakes testing is mediated by those individuals for whom outcomes are important and who can exploit the opportunities they have to improve results. In emerging neoliberal contexts, where professional judgement informs teachers' decision making, teachers deliberately deploy pedagogical approaches to service what they believe to be their students' best interests, even if this might not always be so. But without such authority in established neoliberal contexts, teachers try to accommodate a concern for both their students' and their own interests, aligning both with test success. Focused on these pragmatic concerns and mindful of the high-stakes attached to failure, teachers tend to use pedagogical approaches which will achieve success with minimum risk. The narrowing of curriculum and pedagogy resulting in neoliberal contexts was identified long ago. Ball (2003), for example, discusses how policy technologies have a direct impact on teachers, whilst Lingard (2009) amongst others discusses the reconstruction of teacher professionalism in neoliberal contexts as it relates to pedagogy. Here we have gone further, showing how, as their influence increases, policy technologies undermine a deliberate pedagogy rooted in ideas legitimated through scholarship and experience.

Significantly we have charted the changing landscape of neoliberal governance, showing how, as neoliberal education agendas become more established and testing more influential, so opportunities for teachers to draw on their professional knowledge of teaching, learning and students are constrained by their need to prioritise immediate demands. Teachers in neoliberal contexts are no less caring or conscientious than those elsewhere and some have sought to

negotiate and resist (for example, see Gewirtz et al., 2009). But for many, the approaches they have adopted stem from the circumstances in which they work and demands placed upon them; their wish to be proactive is compromised by their need to be responsive, as they become less reliant on professional judgement and more on external measures. It is the tension and confusion between two distinct forms of pedagogy, crudely put, the deliberate and the responsive, and a wish for a return to the former which causes educators considerable angst. Indeed, the move from professional judgement to market regulation came at great cost for teachers in England (for example, see Jeffrey & Woods, 1998) and the long term consequences for teachers and students are not yet fully understood. It remains to be seen whether Denmark will follow a similar path.

References

- Adams, P. (2008) Considering 'best practice': The social construction of teacher activity and pupil learning as performance, *Cambridge Journal of Education*, 38(3), 375-392.
- Alexander, R. (2001) Border Crossings: Towards a Comparative Pedagogy, *Comparative Education*, 37(4), 507-523.
- Andreasen, K. & Hjørne, E. (2013) Assessing children in the Nordic countries: Framing, diversity and matters of inclusion and exclusion in a school for all. in: U. Blossing, G. Imsen & L. Moos (Eds.) *The Nordic education model: 'A school for all' encounters neo-liberal policy*. Dordrecht: Springer, 155-172.
- Andreasen, K., Kelly, P., Kousholt, K., McNess, E. & Ydesen C. (2015) Standardised testing in compulsory schooling in England and Denmark: A comparative study and analysis, *Bildung und Erziehung*, 68(3), 329-348.
- Au, W. (2008) Devising inequality: A Bernsteinian analysis of high-stakes testing and social reproduction in education, *British Journal of Sociology of Education*, 29(6), 639-51.
- Ball, S. (2003) The teacher's soul and the terrors of performativity, *Journal of Education Policy*, 18(2), 215-28.
- Ball, S. (2012) *Global education inc*. London: Routledge.
- Ball, S. (2013) *The education debate* (second edition). Bristol: Policy Press.
- Ball, S. (2015) Education, governance and the tyranny of numbers, *Journal of Education Policy*, 30(3), 299-301.
- Ball, S. & Bowe, R. (1992) Subject departments and the 'implementation' of National Curriculum policy, *Journal of Curriculum Studies*, 24 (2), 97-115.
- Ball, S., Maguire, M. & Braun, A. (2012) *How schools do policy: Policy enactment in secondary schools*. London: Routledge.
- Bernstein, B. (1990) *The structure of pedagogic discourse*. London: RoutledgeFalmer.
- Bernstein, B. (1996) *Pedagogy, symbolic control and identity: theory, research, critique*. London: Taylor Francis.
- Breakspear, S. (2012) *The policy impact of PISA: an exploration of the normative effects of international benchmarking in school system performance*. OECD Education Working Papers 71. Paris: OECD Publishing.
- Busemeyer, M.R. & Trampusch, C. (2011) Review article: Comparative political science and the study of education, *British Journal of Political Science*, 41(2), 413-443.
- Egelund, N. (2008) The value of international comparative studies of achievement, a Danish perspective, *Assessment in Education: Principles, Policy & Practice*, 15(3), 245-251.

- Ekholm, M, Mortimore, P, David-Evans, M, Laukkanen, R & Valijarvi, J (2004) OECD-rapport om grundskolen i Danmark 2004 [OECD report on basic schooling in Denmark 2004]. Copenhagen: Undervisningsministeriet.
- Esping-Andersen, G. (1990) *The three worlds of welfare capitalism*. Princeton NJ: Princeton UP.
- Gewirtz, S., Mahony, P., Hexall, I. & Cribb A. (Eds.) (2009) *Changing teacher professionalism; international trends, challenges and ways forward*. London: Routledge.
- Goodson, I. & Lindblad, S. (Eds.) (2011) *Professional knowledge and educational restructuring in Europe*. Rotterdam: Sense Publishers.
- Hall, P. & Soskice, D. (2001) *Varieties of capitalism: the industrial foundations of comparative advantage*. Oxford: Oxford UP.
- Harvey, D. (2007) Neoliberalism in creative destruction, *Annals of the American Academy of Political and Social Science*, 610, 22-44.
- Hermann, S. (2007) *Magt og oplysning. Folkeskolen 1950-2006 [Power and enlightenment. Folk Schools 1950-2006]*. Copenhagen: Unge Pædagogers Forlag.
- House of Commons (2008) *Testing and assessment: Third report of session 2007–08*. Available at: www.educationengland.org.uk/documents/pdfs/2008-testingand-assessment.pdf [Accessed 01-05-2016].
- Jeffrey, B. & Woods, P. (1998) *Testing teachers: The effect of school inspections on primary teachers*. London: Falmer.
- Kelly, Peter (2014) *Intercultural comparative research: rethinking insider and outsider perspectives*, *Oxford Review of Education*, 40(2), 1-20.
- Knodel, P., & Walkenhorst, H. (2010) *What's England got to do with it? British underestimation of international initiatives in education policy*. in: K. Martens, N. Alexander Kenneth, M. Windzio & A. Weymann (Eds.) *Transformation of education policy*. Basingstoke: Palgrave.
- Lingard, B. (2011) *Policy as numbers: Ac/counting for educational research*, *Australian Educational Researcher*, 38(4), 355–382.
- Lingard, B. (2009) *Pedagogizing teacher identities*, in: S. Gewirtz, P. Mahony, I. Hextall and A. Cribb (Eds.) *Changing teacher professionalism; international trends, challenges and ways forward*. London: Routledge, 81-93.
- Lingard, B., Martino, W. & Rezai-Rashti, G. (2013) *Testing regimes, accountabilities and education policy: commensurate global and national developments*, *Journal of Education Policy*, 28(5), 539-556.
- Martens, K. (2007) *How to become an influential actor – The ‘comparative turn’ in OECD education policy*. in: K. Martens, A. Rusconi & K. Lutz (Eds.) *Transformations of the State and Global Governance*. London: Routledge. 40–56.
- McNess, E., Broadfoot, P. & Osborn, M. (2003) *Is the effective compromising the affective? British Educational Research Journal*, 29(2), 243-257.
- Mundy, K., Green, A., Lingard, B. & Verger, A. (Eds.) (2016) *The Handbook of Global Education Policy*. Oxford: John Wiley & Sons.
- Newton, P. (2007) *Clarifying the purposes of educational assessment*, *Assessment in Education: Principles, Policy & Practice*, 14(2), 149-170.
- Nusche, D., Wurzburg, G. & Naughton, B. (2010) *OECD reviews of migrant education: Denmark*. Paris: OECD Publishing.
- OECD (Organisation for Economic Cooperation and Development) (2001) *Knowledge and skills for life: First results from PISA 2000*. Paris: OECD Publishing.

- OFSTED (Office for Standards in Education) (2015) Key Stage 3: The wasted years? Available at: www.gov.uk/government/organisations/ofsted, [Accessed 01-01-2016].
- Osborn, M. (2004) New Methodologies for Comparative Research? Establishing 'Constants' and 'Contexts' in Educational Experience, *Oxford Review of Education*, 30(2), 265–285.
- Ozga, J. (2008) Governing knowledge: Research steering and research quality, *European Educational Research Journal*, 7 (3), 261–272.
- Sellar, S. and Lingard, B. (2014) The OECD and the expansion of PISA: new global modes of governance in education, *British Educational Research Journal*, 40(6), 917–936.
- Singh, P. (2002) Pedagogising knowledge: Bernstein's theory of the pedagogic device, *British Journal of Sociology of Education*, 23(4), 571-582.
- Standards & Testing Agency (2015) National curriculum assessments: 2016 sample materials. Available at: www.gov.uk/publications, [Accessed 01-01-2016].
- Standing, G. (2011) *The Precariat: The new dangerous class*. London: Bloomsbury.
- Stevenson, H. & Wood, P. (2013) Markets, managerialism and teachers' work: the invisible hand of high-stakes testing in England, *The International Education Journal Comparative Perspectives*, 12(2), 42-61.
- Tomlinson, S. (2005) *Education in a post-welfare society*. Buckingham: OUP.
- West, A. (2010) High-stakes testing, accountability, incentives and consequences in English schools, *Policy and Politics*, 38(1), 23–39.
- Wiseman, A. (2013) Policy responses to PISA in comparative perspective, in: H.-D. Meyer & A. Benavot (Eds.) *PISA, Power and Policy: The emergence of global educational governance*. Oxford: Symposium Books. 303-322.