

INHALT

BILDUNGSSYSTEME UND NEUE STEUERUNG: GOVERNANCEANALYTISCHE PERSPEKTIVEN

Herausgeber: Hans-Georg Kotthoff/Anatoli Rakhkochkine

Hans-Georg Kotthoff/Anatoli Rakhkochkine

Einleitung zu diesem Heft 283

BEITRÄGE

Herbert Altrichter/David Kemethofer

Neue Ansätze der Steuerung des Schulsystems und die Einstellung
von Schulleitungen 291

Claudia Rupp

Subjektive Deutungen neuer Steuerungsformen: Eine Fallstudie zur
,inneren Logik‘ der Handlungskoordination in einer englischen Schule 311

*Karen Andreassen/Peter Kelly/Kristine Kousholt/Elizabeth McNess/
Christian Ydesen*

Standardised Testing in Compulsory Schooling in England and Denmark:
a Comparative Study and Analysis 329

Arnd-Michael Nohl

Erziehungswissenschaftliche
Governance-Forschung und qualitative Mehrebenenanalyse –
methodische Überlegungen am Beispiel einer Curriculumstudie 349

Marcelo Parreira do Amaral

Der Beitrag der Educational Governance zur Analyse der
internationalen Bildungspolitik 367

ELEKTRONISCHER SONDERDRUCK

BuE 68 (2015) 3

279

VARIA

Andrea Richter

Die Ausweitung des Bildungsauftrags in der Ganztagslernkultur
am Beispiel Denkmalpädagogik: ein Plädoyer 385

REZENSIONEN

*Kwiatkowski-Celofiga, Tina: Verfolgte Schüler. Ursachen und Folgen
von Diskriminierung im Schulwesen der DDR. Göttingen:
Vandenhoeck & Ruprecht, 2014 (Gert Geißler) 399*

TABLE OF CONTENTS

NEW STRATEGIES OF GOVERNANCE IN EDUCATION:
ANALYSES FROM EDUCATIONAL GOVERNANCE
PERSPECTIVES

Editors: Hans-Georg Kotthoff/Anatoli Rakhkochkine

Hans-Georg Kotthoff/Anatoli Rakhkochkine

Introduction 283

ARTICLES

Herbert Altrichter/David Kemethofer

School Leaders' Attitudes towards New Governance of Schooling 291

Claudia Rupp

Subjective Interpretations of Evidence Based Reforms: a Case Study
on the 'Inner' Logic of Action Coordination in an English School 311

*Karen Andreasen/Peter Kelly/Kristine Kousholt/Elizabeth McNess/
Christian Ydesen*

Standardised Testing in Compulsory Schooling in England and Denmark:
a Comparative Study and Analysis 329

Arnd-Michael Nohl

Educational Governance Research and Qualitative Multi-level Analysis –
Methodological Reflections with Reference to a Study on Curriculum
Reform 349

Marcelo Parreira do Amaral

The Contribution of Educational Governance to the Analysis
of International Education Policy 367

VARIA

Andrea Richter

Historical Monument Pedagogy and All-day Schools:

Conditions and Opportunities 385

BOOK REVIEWS

*Kwiatkowski-Celofiga, Tina: Verfolgte Schüler. Ursachen und Folgen
von Diskriminierung im Schulwesen der DDR. Göttingen:*

Vandenhoeck & Ruprecht, 2014 (Gert Geißler) 399

*Karen Andreasen/Peter Kelly/Kristine Kousholt/
Elizabeth McNess/Christian Ydesen*

STANDARDISED TESTING IN COMPULSORY SCHOOLING IN ENGLAND AND DENMARK: A COMPARATIVE STUDY AND ANALYSIS

Summary: Within education, national testing is flourishing and, considering the important role which assessment plays in the production and reproduction of culture, it is important to examine further the possible impact of such practice. While England has a long tradition of national educational testing, Denmark, in contrast, has only recently introduced such tests. Thus the two countries present excellent cases for comparison and analysis in order to gain knowledge about the possible consequences of such testing schemes on pedagogy and the content of the school curriculum, as well as the impact they may have on pupils' perceptions of their potential academic skills. This article draws on research into the national testing of reading conducted in England and Denmark in spring 2013 and draws on the work of Basil Bernstein to compare and contrast both sets of national assessment practices.

Zusammenfassung: In vielen Bildungssystemen erleben standardisierte nationale Lernstandserhebungen und Tests einen starken Aufschwung. In Anbetracht der bedeutenden Rolle, die schulische Leistungsbewertung für die Produktion und Reproduktion von Kultur spielt, scheint es ratsam, die Wirkungen dieser Praxis näher zu untersuchen. Während England über eine lange Tradition von nationalen Lernstandserhebungen verfügt, wurden solche Erhebungen in Dänemark erst vor kurzem eingeführt. Die beiden Bildungssysteme repräsentieren insofern zwei ‚ideale‘ Fälle für einen internationalen Vergleich, dessen Ziel darin besteht, mögliche Konsequenzen dieser Lernstandserhebungen auf den Unterricht und die Inhalte des Schulcurriculums zu erfassen und die Wirkungen zu analysieren, die nationale Tests auf die Wahrnehmung der SchülerInnen in Bezug auf ihre eigene potentielle schulische Leistungsfähigkeit entfalten. Der vorliegende Beitrag stützt sich auf Studien zu nationalen Lernstandserhebungen in England und Dänemark, die im Frühjahr 2013 durchgeführt wurden. Er basiert in theoretischer Hinsicht auf den Arbeiten von Basil Bernstein und vergleicht und kontrastiert die Praxis der nationalen Lernstandserhebungen in den beiden Untersuchungsländern.

1. Background and context of standardised testing in schools

Since the 1960s the effects of globalization and international cooperation, as well as the rise in international comparisons and competition, have greatly influenced educational governance in Northern Europe (LAWN/GREK 2012). It is common for policymakers across Europe to view education almost solely with regard to its economic potential, as a means of maintaining economic growth and stability (BALL 2013). This vision of education, closely related to the competitive state (PEDERSEN 2010), is clearly reflected in recent educational policies and practices across economically developed countries. Such policy and practice is firmly rooted in the drive for the production of attainment data, thus making national and international comparisons possible while keeping strong control over educational resources. At the same time, this is also concerned with increasing the benefits of schooling for individual pupils. Against this background many countries have implemented standardised testing in order to increase both national educational output and the quality of schooling for individuals (AU 2008; LAWN/GREK 2012).

2. Objectives of basic schooling and the lifelong learning imperative

In an evolving world with its constant calls for adaptability, contemporary basic schooling is challenged by the demands of the labour market and relevant life skills. Education policies across Europe address these challenges in different ways, although there are important and defining common denominators. The European Union (EU) points to eight interrelated so-called key competences for lifelong learning (EUROPEAN PARLIAMENT 2006). One key competence is that of 'learning to learn'. This, it argues is, 'the ability to pursue and organise one's own learning, either individually or in groups, in accordance with one's own needs, and awareness of methods and opportunities' (EUR-LEX 2011, 1). Pupils should also develop, 'an awareness of the impact of language on others and a need to understand and use language in a positive and socially responsible manner', and they should have a 'sense of initiative and entrepreneurship' and 'the ability to turn ideas into action' which 'involves creativity, innovation and risk-taking, as well as the ability to plan and manage projects in order to achieve objectives' (Ibid.). Basic schooling, the EU maintains, should support the development of these key competencies in pupils, 'to a level that equips them for adult life, and which forms a basis for further learning and working life' (Ibid.). In the field of educational evaluation, the EU asks for, 'measures aimed at ensuring equal access to both

lifelong learning and the labour market, and support for learners that recognises the differing needs and competences of adults' (EUROPEAN PARLIAMENT 2006, 8).

Currently, schools and societies across Northern Europe are charged with the political requirements of keeping expenses low, providing high quality education, and being able to demonstrate good results in comparison to other schools and countries. Such demands can be experienced as contradictory and, consequently, schools find themselves under a great deal of pressure to 'perform' (MCNEIL 2000). At the same time, standardized testing is being used to demonstrate educational quality and to assess teaching activity in order to improve individual outcomes for pupils. Sociological and historical analyses have argued that modes of assessment can have a strong influence on both educational practice and the way in which various stakeholders perceive and understand teaching, learning and pupil potential (MADAUS/RUSSEL/HIGGINS 2009; MCNESS/BROADFOOT/OSBORN 2003; REESE 2013). Pupil experience of such activities can also strongly influence the formation of learner identity, and pupils' and teachers' understanding of their academic potential (ANDREASEN/HJORNE 2014; KOUSHOLT 2013). However, there is also some concern that such influences may work against the stated objectives of basic schooling and lifelong learning as described above.

3. Standardised testing in England and Denmark

Although the proliferation and use of national standardized testing has become a common feature of most North European school systems, some countries have stronger and more widespread traditions for such practices than others. England has a long tradition of national educational testing from the introduction of the 11-plus examination (1944 Education Act) to the testing of curriculum subjects at the end of compulsory schooling. While the introduction of comprehensive schooling from the late 1960s saw the abolition of the 11-plus requirements for most children, a new national testing regime was re-introduced as part of the 1988 Education Reform Act at the ages of 7, 11 and 14 years.

In contrast, Denmark, a relative newcomer to the field, only introduced national tests in 2010. It is, therefore, interesting to compare the two national testing schemes and their influence on pupils, teachers, parents and school leaders. Though both testing regimes display some similarities, they are also characterized by distinct differences in terms of their design, the circumstances under which they are taken, how they are marked, and the manner in which the results are used to inform individual and general classroom practice. Bernstein's theory of pedagogic practice (BERNSTEIN 2000) will be used to analyse these differences and their

possible consequences for pedagogy and the content of the school curriculum, as well as pupils' perceptions of their potential academic skills.

In both countries, mandatory assessments are linked to the national curricula. In England, this link is explicit though assessment is limited to English, maths, and science. In Denmark the relationship is not stated explicitly although there is a connection. In both countries, the link between the national test and the national curriculum is particularly visible in subject areas where the test makes it possible to establish such a link and less visible in subjects where the content of the national curriculum cannot be assessed directly using the specific technology. In order to explore the workings and impact of the two national testing schemes the article begins with the contextual background of the schemes, followed by an introduction to the relevant parts of Bernstein's framework which have been used for comparative analysis. This is then followed by a presentation of the findings, a conclusion and a discussion in the light of questions related to the EU aims for lifelong learning mentioned earlier.

4. Background, history and origins of the national testing practice in England and Denmark

While the implementation of national standardised testing in both England and Denmark can be seen as influenced by globalisation and general developments in the Western world, it is important to understand how the process and practice has developed differently within each country.

4.1 England

During the 1970s the United Kingdom, in common with other industrialised countries, experienced major changes in its economic circumstances. Increasing economic globalisation and a sharp increase in oil prices led to the decline of traditional industries, resulting in large scale unemployment and dwindling international competitiveness. Education began to be seen as the key to future prosperity (BALL 1990). The election of a Conservative government in 1979 extended this thinking by challenging the educational establishment and calling for the introduction of a neoliberal agenda to open up what was referred to as the stranglehold of 'producer capture' (especially with regard to the local education authorities). It established the beginnings of market mechanisms to ensure accountability and improve educational outcomes. Such discussion culminated in

the 1988 Education Reform Act, which established a national curriculum and national testing in the form of Standard Assessment Tasks (SATs) (POLLARD et al. 1994).

It was these new national testing arrangements that were the greatest source of dissatisfaction for teachers and educationalists. Children were to be tested at the end of four Key Stages: KS1 (7 year olds) in English and mathematics, KS2 (11 year olds), KS3 (14 year olds) and KS4 (16 year olds) in English, mathematics and science. KS4 tests were never introduced as pupils continued to take their national exams, the General Certificate of Secondary Education, in each of their individual school subjects. KS3 tests have since been dropped as a result of teacher protests, but KS1 and KS2 tests remain at the primary level of education in England. Importantly, prior to the introduction of SATs, a Task Group on Assessment and Testing devised a system which combined teachers' own formative assessment with SATs (DES 1987). The latter would, it was hoped, be perceived by pupils not as tests but as normal classroom tasks, perhaps carried out in groups, and they would be closely supervised by teachers for gathering assessment evidence. Though this approach was applied to KS1 tests, from the outset, KS2 tests, which are the focus of this research, relied on paper and pencil tests, carried out in formal examination conditions and marked externally, with no teacher input.

In summary, the national tests in England have been operating in their present form since 1989. Though there have been some changes, they have been generally championed by successive Labour and Conservative governments for their ability to improve education quality and make schools and teachers accountable for what they do. However, school leaders, teachers and educationalists continue to be concerned about their impact on learning and pupil motivation.

4.2 Denmark

Introduced in 2010 as a mandatory test, the Danish national standardized tests marked a departure from the type of standardized testing that had originally been part of the Danish education system from the 1960s up till then (YDESEN 2013). The Danish national standardized tests were carried out for the first time in 2010 in the Danish municipal primary and lower secondary school [*Folkeskole*]. Pupils in grades 2–8 are tested in between one and four different subjects each year, with the exception of the 5th Grade. The subjects tested through these years are: literacy, maths, English, geography, biology, physics/chemistry and Danish for second language speakers.

The implementation of Danish national standardized tests is the result of several political initiatives and national, as well as international, processes. As part of these processes, Danish school children's participation in international tests plays an important role in the introduction of national testing. The participation in an international reading test is significant in the development of the Danish educational system. Danish school children were placed lower than expected, triggering open criticism of the Danish school system. The former minister of education, Bertel Haarder, has since said, 'I remember it like it was yesterday. It was simply the most powerful blow to the Danish school's self image ever'. (quoted in GUSTAFSSON 2012). The right-wing parties, to some extent, used the poor results to criticise the progressive foundations of Danish schooling and emphasized, among other things, a need for tests (ibid.). Subsequently, the PISA tests carried out in 2000, 2003, 2006, 2009, 2012 all found Danish pupils performing below expectations. Denmark has been described in the media as a 'PISA shock' country because, it is argued, the tests have had a big impact on school policy (FOLKE-SKOLEN.DK 2012).

Danish right-wing governments from 2001–2011 increased the focus on subject knowledge and promoted an active break with the ideas of progressive education (GUSTAFSSON 2012). The acquisition of skills was now considered not primarily to serve the pupils' own individual development but to serve the labour market, in order to ensure the nation's prosperity and ability to compete internationally (PEDERSEN 2010). The national tests were signed into law in 2006 during the right-wing government period, but the implementation of national tests was also supported by the social democrats, even though standardized testing is not a traditional item on their agenda (GUSTAFSSON 2012). As such, the introduction of the Danish national tests was an indication of a combination of different political agenda. This is, for instance, the case in the adaptive design of the test, which means that they adapt to individual academic differences and thereby make every child testable (KOUSHOLT/HAMRE forthcoming). There is also the emphasis on the test as a pedagogical tool to help the weakest pupils and to break with negative social inheritance, rather than as a means of control.

5. Setting up a Bernsteinian analytical framework

Bernstein's pedagogical model was chosen because, unlike other models, it provides a coherent analytical framework for understanding the selective regulation and distribution of different forms of knowledge and achievement through testing (AU 2008). Whilst it was developed in England, his work has been widely applied

to other European contexts (OSBORN/McNESS/BROADFOOT 2000). It enabled us to identify similarities and differences between the two national testing schemes in order to gain knowledge about their workings and impact (MORAIS/NEVES 2010; SINGH 2002).

BERNSTEIN (2000) introduced a basic distinction between competence models of teaching, which emphasize ‘the realisation of competences’, and performance models which place an emphasis ‘upon a specific output of the acquirer.’ He further argued that these different ways of understanding competences and knowledge become visible in assessment procedures (BERNSTEIN 1997). Thus this distinction is of significant relevance to our analysis, and Bernstein’s description of what characterises the two models provides a useful framework. In operationalizing his analytic gaze, Bernstein developed the pedagogic models in Table 1.

Applying this model to the central concern of the article, it becomes apparent that the six analytical focus points clearly relate to pedagogy and teacher activity and the content of the school curriculum, with clear implications for pupil perception of their potential academic skills. Pedagogies aligned to the competence model of teaching and learning are characterised by weak classification of knowledge and thus less strong boundaries between subject content and teaching activities, while pedagogies of the performance model correspond to strong classifications of knowledge and frames of teaching activities.

BERNSTEIN (1997, 68) further describes the visible pedagogy as characterized by explicit hierarchy in which, ‘space and time are regulated by explicit principles, there are strong boundaries between spaces, times, acts, communications’ and the pedagogic act is highly framed. The opposite applies to invisible pedagogy where space/time are weakly classified and hierarchy implicit. These two pedagogies also correspond to different kinds of assessment. In visible pedagogy BERNSTEIN (1997, 70) argues that:

... an ‘objective’ grid exists for the evaluation of the pupils in the form of (a) clear criteria and (b) a delicate measurement procedure. The child receives a grade or its equivalent for any valued performance.

No such ‘grid’ exists in invisible pedagogy and the evaluation procedures are multiple, diffuse and not easily subject to apparently precise measurement, which makes the comparison between pupils and schools complex. Employing Bernstein’s analytical distinctions between competence and performance models on the horizontal level, and categories, evaluation orientation, control, pedagogic text, autonomy and economy on the vertical level, this article will discuss these distinctions in relation to the empirical findings from the two national contexts.

Table 1: Bernsteinian Pedagogic Models (2000, 45)

	Competence model	Performance Model
1. Categories – space – time – discourse	weakly classified	strongly classified
2. Evaluation orientation	presences	absences
3. Control	implicit	explicit
4. Pedagogic text	acquirer	performance
5. Autonomy	high	low/high
6. Economy	high cost	low cost

Using the themes and focus points of Bernstein's analytical framework in the analysis will inevitably shape the treatment of the questions, but the explicit manner in which this shaping is actualized will make its epistemological gains and losses clear in the analysis.

6. The empirical study

The study was carried out in a small sample of schools in England and Denmark during spring 2013. It employed a qualitative approach which included an analysis of documentary evidence, specifically national legislation, together with interviews with teachers and pupils in Year 6 (10/11 year olds). Three teachers from three different schools in each country were interviewed together with thirty six of their pupils; six from each of the sample schools. The teachers were the main or sole teachers of their pupils' classes. The selection of pupils and schools was designed to ensure diversity in the study (DENZIN 2000). The three schools in each country represented, respectively, one city school, one situated in a suburban setting, and one rural school. The six pupils from each school represented both genders (three girls and three boys), and different academic potential and skills as assessed by their main teacher.

Table 2: Descriptions of Schools

	England	Denmark
Teacher A	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.	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.
Teacher B	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.	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.
Teacher C	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; he is also deputy head of the school.	This is a medium sized village school with some 250 pupils on roll. The teacher is female with 15 years teaching experience.

Interview questions focused on an exploration of the life-worlds of the teachers and pupils and their experiences of testing, such as the related procedures, the importance of the test, and their understanding of what was assessed (DENZIN 2000; KVALE/BRINKMAN 2008). The supporting documentary evidence was accessed on the official homepages of the Ministries of Education in the two countries.

7. Comparative analysis of the two tests and test regimes

At first sight, mandatory national testing in lower secondary schools in England and Denmark differ in almost every possible way with regard to the design of test, the way it is marked and how it is used. Making a deeper analysis, however, based on Bernstein's theory, we find some similarities. In the following section we provide a comparative analysis of the two mandatory standardized national tests and test practices as expressed in pieces of legislation and other relevant policy documents.¹

7.1 England

The overall purpose of teaching English at this level of basic schooling is described in the National Curriculum (English Years 5 & 6) as:

... to ensure that all pupils: – read easily, fluently and with good understanding [...] – write clearly, accurately and coherently, adapting their language and style in and for a range of contexts, purposes and audiences – use discussion in order to learn; they should be able to elaborate and explain clearly their understanding and ideas – are competent in the arts of speaking and listening, making formal presentations, demonstrating to others and participating in debate. (DFE, 2013a, 3)

The English SAT test assesses pupils' skills in reading and writing, while competencies related to spoken language are not assessed. The test is integrated into a system which also includes regular mandatory teacher assessments of pupils' work and reports to parents. This is in addition to the obligations on local authorities to ensure appropriate standards with regard to their schools' performance. The whole arrangement is administered by the Standards & Testing Agency (STA), an executive agency of the Department for Education (DFE).

Practice with regard to SATs has a clear and explicit orientation towards school performance, since school results are made public on the homepage of the DFE (2014). The average score of the school is reported in three categories: Low attainers, middle attainers, high attainers, and there is an average KS2 score for the whole school. In parallel with these national tests, class teachers have an obligation to 'prepare annual reports for the parents of every child and make arrangements for parents to discuss the report with the child's teacher, if they so wish' (STA 2013, 33). The report should cover, 'the child's achievements; general progress; and attendance record. It must also include the following results, where appropriate: results of any national curriculum tests, by level; and results of any public examinations, by subject and grade' (Ibid.).

The interviews give some evidence that the national testing process makes teachers feel under pressure for their pupils and classes to perform well. As one teacher said:

... your reputation as a teacher relies on that quite a lot even if it is a key stage combined effort it doesn't feel like that to the teacher. (Teacher A)

In addition to preparing students for national testing, teachers also have an obligation to undertake regular teacher assessments, which are more qualitative in nature and based on the teacher's experience of an individual learner's progress. However, the national standardised tests are perceived as more important due to the potential consequences of a school's poor performance. A teacher further reflects on the problems related to the standardised testing as experienced:

The tests aren't an accurate reflection on how well the children have achieved in the school as a whole. It is particularly unfair to the not so academic children who are still very intelligent but the tests are only testing a particular type of intelligence (...) it's a very narrow one. (Teacher C)

The test, its design and content, is also perceived to have a profound impact on teaching, encouraging a need to ‘teach to the test’. For instance, as one teacher explains:

Towards the middle of the spring term we started building SATs type questions into our guided reading, so we would look at a comprehension, we’d look at a question, and we would talk about possible answers and the kind of answers which the children would need to give and they got quite good at saying, well, this question is three points so I need to say three things to answer. It’s kind of teaching them to do the tests, really” (Teacher B)

The teachers also concur that SATs can be stressful for some children and argue that the testing regime leads to a narrowing of the curriculum to those areas to be tested, so that the children do not receive a balanced curriculum:

It can be unfair for the less able children, the pressure they feel. We don’t do heavy coaching which other schools do. That’s the bit I don’t like. It reduces creativity in the curriculum, and affects the way the curriculum is run. Yes, we give them a broad balanced curriculum, but it is skewed towards the test subjects. (Teacher C)

In summary, the teachers expressed worries about the influence of national testing, especially with regard to how the pressure to perform creates dilemmas between teaching to the test and maintaining creativity. They also have reservations about the kind of intelligence which is being measured and appreciated.

7.2 Denmark

The target levels [*trinmålene*] describe the overall aims and goals of teaching Danish in grades 5 and 6 of the public school. Apart from describing reading goals, the target levels also contain goals for written language, language, literature and communication. As a consequence of its character and limitations due to the technique – as will be described later – the national test only assesses, ‘a few target levels in the subject of Danish (UNDERVISNINGSMINISTERIET 2011). These are (Ibid, 5):

- Language comprehension. Pupils’ knowledge of words and idioms
- Decoding. Pupils’ proficiency at identifying written words and letters. In contrast to language understanding, decoding is something specific to reading
- Text comprehension. Pupils’ understanding of written texts

Other aims described in the target levels, which are not directly tested, include (UNDERVISNINGSMINISTERIET 2009, 6):

- Read linguistically developed texts and use different reading comprehension strategies
- Seek explanation of words for understanding words and expressions
- Adapt reading speed, precision and reading technique to purpose, genre, and level of difficulty
- Express understanding of text orally and in writing
- Read with knowledge of one's own benefit from the read
- Seek and chose fiction and professional literature at the library and in digital media

The test is presented as one intended for teachers' continuous assessment in school classes (UNDERVISNINGSMINISTERIET 2011, 3). However, the system in which they are embedded supports an orientation towards performance for both schools and pupils. The teachers describe such processes:

It is certainly embarrassing [for pupils] to just sit there and finish last (Teacher C)

And it is clear that when you compare results, then you say, I have achieved 62% with class 6A, so if I compare with class 6C with only a 47% score [...] then it is natural for head teachers to say that your class has done darn well; they have achieved 62% compared to the other class (Teacher A)

Such quotes point to the pressure of performance which is found both among the pupils in the actual test situation, and in comparisons between classes. Tests are made for differentiation purposes and these purposes create competition and comparisons in different ways at different levels.

The system in which the national test is integrated is similar to the SATs in England in some respects, but differs on important points. Danish teachers need to prepare a report (the pupil plan) for parents, describing pupil achievements, results of national tests, and results from other assessments, and English teachers have to prepare a similar report also used for discussions with parents. But scores for each school from the national standardised test in Denmark are not made public and accessible for anyone to see, for instance on the homepage of the Ministry of Education. But results are made public and can be seen at school level, since it is integrated into a mandatory *Quality Report* made by schools (UNDERVISNINGSMINISTERIET 2014b). The education act states that the municipal council should make such a report once a year (UNDERVISNINGSMINISTERIET 2014a). The quality report should account for, 'results from each individual school within the municipal jurisdiction and for education in total'. As well as the council's response to the latest quality report, it should also describe:

1) grading in the school-leaving examination, 2) results of the tests mentioned in § 13, para 3 of the act according to § 55b, and 3) results from any other type of evaluation broadly used in the municipal education system, e.g. to examine pupils' knowledge and skills in areas not covered by the tests mentioned in section 2.

In the event of the quality of the school being considered unsatisfactory by the council, the staff of the school ‘must produce a plan of action aimed at improving the level of the school’ and the report should also consider the effects of such described plans of action.

The test results are also used in relation to a ‘mandatory pupil plan’ made by the class teacher. The pupil plan ‘must contain information about the on-going evaluation of teaching in all subjects’, and ‘information about the decided response to the results of the on-going evaluation, including response to the tests in selected subjects at different class levels mentioned in § 13 para 3 of the education act, and about possible agreements with the parents and how the pupil contributes to the pupil reaching the formulated learning goals’. The teachers express it in this way:

One might say that the test is a tool in the school-home collaboration. [...] then you can get a talk about how things have been and..., but in our everyday work things are normally not aimed at the test, we have our common goals, our year plan which structure our teaching and we evaluate continuously (Teacher B) Parents use the results with the advice we give. When we sit down and talk with them and they get to take the results home. (Teacher C)

8. Testing models and the implications for assessment

The overall *technology* of the test has profound implications both for what can be tested, its content and forms of questions, for how it is or can be marked, for reporting and for the costs of the test. The two tests differ significantly in relation to these issues.

The SATs in England are ‘paper-and-pencil’ tests, where teachers hand out booklets for pupils to complete. These are received in schools some time in advance, but only opened at the start of the test (STA 2013, 22). Pupils answer questions by completing multiple-choice questions or by writing extended answers. Aspects of reading comprehension are tested (which of these statements [about the text] are true?) and areas such as inference (Was [a character from the story] likeable? Give three pieces of evidence from your text to support this view.) Hence tests are complex and often require pupils to give their opinions. This characteristic means that practising for the tests is both possible and beneficial, and such practice is an integrated part of the preparation for SATs. The SAT test technology makes external manual scoring necessary, which also means that it is cost-intensive and requires a great deal of administration. It also means that there can be discrepancies, mistakes and a time-consuming appeal process.

In contrast, the Danish National Test is an *IT-based multiple-choice test*. All questions are answered by ticking an answer on the computer screen. The ques-

tions in the test are of a specific design and drawn from a limited ‘test item bank’. In the reading test, pupils respond to questions concerning proverbs, idioms, hyphenation and understanding text. No one sees the test in advance, and individual pupils do not answer the same questions since questions are given randomly. The tests are also adaptive, meaning that each pupil’s academic level is, ostensibly, assessed and questions adapted automatically to this level. This makes the possibilities for practicing in advance more limited than in the English SATs. Teachers are however able to practice how to handle various elements, for instance, proverbs, how to understand sayings and other sorts of problem solving tasks. Our research shows that Danish teachers have integrated such practising into their teaching, and did so very quickly in the year when the tests were originally implemented. The technology of the test makes it low-cost, with low workload and very little administration required. However, the development of the test was very demanding and expensive, taking several years.

In both tests *time* is a strong factor and plays a role in the final scores, but this is done in two different ways. In Key Stage 2, pupils have 15 minutes reading time and 45 minutes to answer the questions, though schools can apply for additional time for pupils with special needs (STA 2013, 5). In the Danish National Test pupils have 45 minutes for the test, but might finish sooner or later than that, since time can be extended by 15 minutes at a time, or up to a maximum of 3 hours if the test cannot determine the pupil’s proficiency level (UNI-C 2014). Less than 45 minutes can be used if the computer determines the pupil’s assessment very quickly (Ibid.).

The physical *spaces* in which the tests take place are also organised and controlled in two different ways as a result of the differing test technologies. In England, teachers are given, ‘detailed information including how to prepare rooms and what equipment is needed. Schools that do not comply with the published procedures could be subject to investigations of maladministration.’ (STA 2013, 24). All pupils take the test together on a set day under examination conditions. Aid is generally not allowed, except for some children with, for instance, special educational needs. For instance, a spellchecker was not allowed in the English reading test researched, despite spelling not being part of the assessment (DFE 2012). In Denmark, pupils’ test-space is organised and controlled by the computers. Pupils sit in front of a computer, and the teacher has his/her own computer on, so that she/he can follow the number of items the pupils are solving. There are no precise descriptions of the required physical space, and computers can be placed in very different ways in the room. In fact, research has shown that it is sometimes possible for pupils to follow other pupils’ problem-solving progress on their screens (ANDREASEN 2011). However, since they are not answering the same ques-

tions at the same time, cheating is difficult, even though they might be able to ‘consult’ each other on the different questions. Pupils are allowed to use aids. As a rule all pupils should have access to the same kinds of aid as used during ordinary teaching. The teacher can decide which kinds of aid are made available.

It can be argued that the *pedagogic text*, or what pupils produce in the two sets of tests, has a clear orientation towards performance (of reading). Any question or problem is asked in a form which makes it possible to score answers. Scoring is complex, which indicates a high degree of categorization and precise indications of what counts as the right and legitimate form of this knowledge. Making answers ‘scorable’ also adds very strict limitations on which kinds of questions *can* be asked and on the criteria for *evaluation* which is focused on deficits. Scoring also means that pupil performance can be ranked, and so tied to *discourses* and the use of specific *categories*. These are then established by, and used in, legislation and documentation. The Danish national tests are self-scoring, and teachers get an automatically generated report showing pupils’ results in five categories: average, above average, significantly above average, below average and significantly below average (UNDERVISNINGSMINISTERIET 2010, 7). This discourse about pupils’ skills, potential and performance is also used actively by the Ministry in the examples given in the material for teachers to read ahead of the test. For instance, below is an example from the Ministry of Education (UNDERVISNINGSMINISTERIET 2011, 7):

In the example above Anne has scored very good results – perhaps even surprisingly good results compared to what the teacher had expected. Perhaps Anne is not very active in class and the teacher has assessed her as an average pupil.

The interviews showed that such categories were integrated into the discourses of pupils and teachers in different ways. For instance a Danish pupil explained that:

You get some butterflies in your stomach, will you do well or extremely poorly... will you just be average. And also you feel bad if you do not want to take the test. (Pupil, boy, school C)

A Danish teacher points to the importance of finding the pupil’s correct attainment level. By doing so, she also adopts the discourse expressed in the public documents about the national tests and the practice of ranking pupils:

Yes it does, I mean it says something about ehh, normally you know who is placed where without the test, right. Of course it says something about that too. And then you might sometimes be surprised if someone is in an unexpected place and then you check “but, what happened here?” and perhaps you have a talk with that pupil and find out “how was your day really?” and “How was this test?” and learn if there were any other conditions. (Teacher B)

The quote indicates that the test becomes part of a practice of ranking, and that teachers often categorize pupils according to an invisible scale of cleverness

before they take the test, which impacts on how they interpret the results. With reference to Bernstein's analytical framework this pedagogy resembles the performance model.

In terms of the use to which the outcomes of the SATs are put, English pupils discussed them in the following way:

They're for the teachers so they know what level you are working at and so they can challenge you; they are important to us as well, it helps tell us how good we are and whether we should push ourselves a bit more. (Pupil, girl, School A)

The next school (secondary school) are interested in the results and they look at them and see which form to put you in. Our school will find it useful because they might give us some different sheets to help us. (Pupil, boy, School C)

For the teacher, the SAT results represent a measure of her/his teaching ability. As one teacher explained:

The test results (...) are important to me in that I am judged on that. Everything is geared towards that data – how many level 4s are we going to get, and so on. (Teacher A)

Scoring is done manually by official Key Stage 2 markers engaged by the DfE and using an official mark scheme (UK GOVERNMENT 2014). Often the markers are teachers, but they do not mark their own pupils' work. It is possible to apply for a review of marking 'if schools believe that there is evidence that a mark scheme has not been applied correctly or a clerical error has occurred' (STA 2013, 26).

9. Concluding discussion

At first glance the two standardised tests appear to be very different, showing only few similarities as to their general design. However, a detailed analysis has made similarities visible at a deeper level; similarities which will be of interest to many including policymakers and teachers.

The differences displayed are related to, among other things, the test technique (respectively, paper-and-pencil (England) and IT (DK), the administration of time, control of space, use of aids (no aids allowed in SAT testing, aids allowed in the Danish national tests), types of questions and problems to be solved, scoring and reporting, and consequences. They also differ in their focus. The English reading tests focus on comprehension and inference, whereas the Danish tests are more specific, looking at particular literacy devices.

Nonetheless, there are also important similarities. Of special interest are the similarities related to the kind of knowledge being assessed, and the way in which

the test scores are used and the practices of which they are a part. These aspects of the tests are important because of the profound impact assessment has on pedagogy and discourses about pupils' skills and potential.

Both national testing schemes legitimize and add power to specific discourses and practices which thus become dominant. It can be argued that both tests correspond to the "performance model" of pedagogy (BERNSTEIN 2000) and, as such, do not assess competencies – many would argue that the English tests do test the competencies of reading comprehension and inference, though pupils have to perform these in order to gain marks. Further, both national testing schemes assess only a limited part of the curriculum. Since assessment practice clearly influences teaching, this is an important concern as there is potential for the curriculum to be narrowed and creativity restricted. Both tests are integrated in a test practice, defined by a legislative framework, which introduces a performance-oriented discourse about pupils' skills and potential. Teachers and pupils clearly adapt to this discourse, which is a significant result. The results are reported using specific categories to schools, teachers, pupils and parents. These categories dominate discourses about pupils' potential, and might influence, for instance, the formation of pupil learner identity, the way they understand themselves, their potential to succeed, and opportunities for future education. In consequence, this focus on 'performance' hinders the development of national versions of The European Union's key competences for lifelong learning as defined above (EUROPEAN PARLIAMENT 2006). Finally, compared to the Danish teachers, UK teachers describe a much more specific and planned teaching-to-the-test practice, largely as a result of the high stakes nature of the results, which is also an expected part of the teaching and supported by different kinds of teaching resources.

Notes

- 1 In both England and Denmark, pupils take the reading test in year 6, at approximately 11 years of age. However, due to differences between the ages at which children start school, the Danish pupils' average age might be a few months higher than that of the English pupils. There are also some differences in the way the tests are administered. While all English pupils, in a particular year cohort, sit the Standard Assessment Tasks (SATs) in the same week in May, Danish pupils might sit the Danish National Test several months earlier, due to teachers' freedom to book a time of their own choice for testing during a three-month-interval in spring from late January to late April.

References

- ANDREASEN, K. E./HJORNE, E. (2014): Assessing Children in the Nordic Countries - Framing, Diversity and Matters of Inclusion and Exclusion in a School for All. In: BLOSSING, U./MOOS, L. (eds.): *The Nordic Education Model: 'A School for All' Encounters Neo-Liberal Policy, Policy Implications of Research in Education 1*. Dordrecht, 155–172.
- ANDREASEN, K. E. (2011): Summative test i formativ praksis [*Summative testing in formative practice*]. In: ANDREASEN, K. E./FRICHE, N./RASMUSSEN, A. (eds.): *Målt & Vejet: Uddannelsesforskning om evaluering*. Aalborg, 301–322.
- AU, W. W. (2008): Devising inequality: a Bernsteinian analysis of high stakes testing and social reproduction in education. In: *British Journal of Sociology of Education*, 29 (6), 639–651.
- BALL, S. J. (1990): *Politics and Policy Making in Education: Explorations in Policy Sociology*. London.
- BALL, S. J. (2013): *The Education Debate*. Bristol.
- BERNSTEIN, B. (1997 [1975]): *Class and Pedagogies: Visible and Invisible*. In: HALSEY, A. H./LAUDER, H./BROWN, P./WELLS, A.S. (eds.): *Education, Culture, Economy, and Society*. Oxford, 59–79.
- BERNSTEIN, B. (2000 [1996]): *Pedagogy, Symbolic Control and Identity*. Oxford.
- DENZIN, N. K./LINCOLN, Y. S. (2000): *The Discipline and Practice of Qualitative Research*. In: DENZIN, N. K./LINCOLN, Y. S. (eds.): *Handbook of qualitative research*. London, 1–45.
- DES (1987): *Report of the Task Group on Assessment and Testing: Three Supplementary Reports*. London.
- DFE (2012): *Key Stage 2 Modified test administrators' guide*. Standards and Testing Agency. Crown copyright 2012.
- DFE (2013a): *English programmes of study: key stages 1 and 2*. National curriculum in England. Crown copyright 2013.
- DFE (2013b): *Assessment innovation fund*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/268361/assessment_innovation_fund_launch_doc.pdf (01.05.2015).
- DFE (2014): *School performance tables*. [http://www.education.gov.uk/schools/performance/\(01.05.2015\)](http://www.education.gov.uk/schools/performance/(01.05.2015)).
- EUROPEAN PARLIAMENT (2006): *Recommendation of the European Parliament and of the Council of 18 December 2006 on key competences for lifelong learning*. 2006/962/EC. <http://www.europarl.europa.eu/sides/getdoc.do?pubref=-//ep//nonsgml+tc+p6-tc1-cod-2005-0221+0+doc+word+v0//en> (01.05.2015).
- EUR-LEX (2011): *Key competences for lifelong learning. Summary*. <http://eur-lex.europa.eu/legal-content/en/txt/html/?uri=uriserv:c11090&rid=1> (01.05.2015).
- FOLKESKOLEN.DK (2012): *Danmark er et Pisa-chok-land*. [Denmark is a Pisa-shock country] <https://www.folkeskolen.dk> (01.05.2015).
- GUSTAFSSON, L.R. (2012): *What did you learn in school today? How ideas mattered for policy changes in Danish and Swedish schools 1990–2011*. PhD dissertation. Aarhus.

- KOUSHOLT, K./HAMRE, B. (forthcoming): The pupils that had to be as clever as they could. To appear in: *The global testing culture: Oxford studies in comparative education series*. Oxford.
- KOUSHOLT, K. (2013): Deltagelse på tværs af sociale lærings- og testsammenhænge [Participation across social learning- and test contexts]. In: *Psyke & Logos*, 34 (1), 180–202.
- KVALE, S./BRINKMAN, S. (2008): *Interviews*. London.
- LAWN, M./GREK, S. (2012): *Europeanizing Education – Governing a new policy space*. Oxford.
- MADAUS, G.F./RUSSELL, M./HIGGINS, J. (2009): *The Paradoxes of High-Stakes Testing*. Charlotte, NC.
- MCNEIL, L.M. (2000): *Contradictions of school reform. Educational costs of standardized testing*. New York.
- MCNESS, E./BROADFOOT, P./OSBORN, M. (2003): Is the Effective Compromising the Affective? In: *British Educational Research Journal*, 29 (2), 243–257.
- MORAIS, A./NEVES, I. P. (2010): Basil Bernstein as an Inspiration for Educational Research. *Specific Methodologies and Approaches* (2010). In: SINGH, P./SADOVNIK, A./SEMEL, S. (eds.): *Toolkits, translation devices and conceptual accounts: Essays on Basil Bernstein's Sociology of Knowledge*. New York.
- OSBORN, M./MCNESS, E./BROADFOOT, P. with POLLARD, A./TRIGGS, P. (2000): *What Teachers Do: changing policy and practice in primary education*. London.
- PEDERSEN, O. K. (2010): Institutional Competitiveness - When Nations came to Compete. MORGAN, G. et Al. (eds.) *Oxford Handbook in Institutional Analysis*. Oxford, 625–658.
- POLLARD et Al. (1994): *Changing English Primary Schools? The Impact of the Education Reform Act at Key Stage One*. London.
- REESE, W.J. (2013): *Testing Wars in the Public Schools – a forgotten history*. Harvard.
- SINGH, P. (2002): Pedagogising Knowledge: Bernstein's theory of the pedagogic device. In: *British Journal of Sociology of Education*, 23 (4), 571–582
- STA (2013): *National Curriculum Assessment. Assessment and Reporting Arrangements. Key Stage 2*. 2014. Crown Copyright 2013.
- UK GOVERNMENT (2014): *Key stage 2 tests: how to become a marker*. <https://www.gov.uk/key-stage-2-tests-how-to-become-a-marker> (01.05.2015).
- UNDERVISNINGSMINISTERIET (2009): *Fælles Mål 2009*. Dansk. Undervisningsministeriets håndbogsserie nr. 3 – 2009.
- UNDERVISNINGSMINISTERIET (2010): *National test*. Dansk. Læsning. Skolestyrelsen.
- UNDERVISNINGSMINISTERIET (2011): *Brug testresultaterne*. Skolestyrelsen.
- UNDERVISNINGSMINISTERIET (2014a): *Bekendtgørelse af lov om folkeskolen*. Act nr. 665 of 20/06/2014.
- UNDERVISNINGSMINISTERIET (2014b): *Bekendtgørelse om kvalitetsrapporter i folkeskolen*. Act nr. 698 of 23/06/2014.
- UNI-C (2014): *Test- og prøvesystemet – De nationale test. Brugervejledning for skoler*. Undervisningsministeriet. UNI-C. Styrelsen for IT og Læring.

YDESEN, C. (2013): Educational Testing as an Accountability Measure: Drawing on Twentieth-century Danish History of Education Experiences. In: *Paedagogica Historica*, 49 (5), 716–733.

Biography

Dr. *Christian Ydesen*, MA in History and Philosophy (2007), PhD (2011), assistant professor and post.doc. at Aalborg University (since 2012). Main focus: history of educational testing, accountability, ethnicity and education and international organisations.

Address: Christian Ydesen, PhD, Department of Learning and Philosophy, Aalborg University, Nyhavnsgade 14, 9000 Aalborg, Denmark. E-mail: cy@learning.aau.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.

Address: Elizabeth McNess, PhD, Graduate School of Education, University of Bristol, 35 Berkeley Square, Clifton Bristol BS8 1JA, UK. E-mail: elizabeth.mcness@bristol.ac.uk.

Dr. *Kristine Kousholt*, MA in Danish literature and Psychology (2004), PhD (2009), associate professor (since 2015). Her dissertation concerns assessment practice in primary and lower secondary school. She has afterwards completed a post doc project with focus on teachers' and pupils' perspectives on and participation in the Danish national standardized tests.

Address: Kristine Kousholt, PhD, Department of Education, Aarhus University, Tuborgvej 164, 2400 København NV, Denmark. E-mail: krko@edu.au.dk.

Dr. *Peter Kelly* is Reader in Comparative Education at Plymouth University, UK, where he has worked since 2003. Main focus: comparing pedagogy and its effect on educational inequality across Northern European countries.

Address: Peter Kelly, PhD, Institute of Education, Plymouth University, Drake Circus, Plymouth PL4 8AA, UK. E-mail: peter.kelly@plymouth.ac.uk.

Dr. *Karen E. Andreasen*, MA in Educational Studies. PhD (2008), associate professor at Aalborg University (since 2012). Main focus: Education, educational assessment, history of education, educational sociology, in- and exclusion.

Address: Karen Andreasen, PhD, Department of Learning and Philosophy, Aalborg University, Nyhavnsgade 14, 9000 Aalborg, Denmark. E-mail: karena@learning.aau.dk.