



AARHUS UNIVERSITY



Coversheet

This is the accepted manuscript (post-print version) of the article.

Contentwise, the post-print version is identical to the final published version, but there may be differences in typography and layout.

How to cite this publication

Please cite the final published version:

Kousholt, K. (2016). Testing as social practice: Analysing testing in classes of young children from the children's perspective. *Theory & Psychology*, 26(3), 377-392. DOI: 10.1177/0959354316641911

Publication metadata

Title:	Testing as social practice: analysing testing in classes of young children from the children's perspective
Author(s):	Kristine Kousholt
Journal:	Theory & Psychology
DOI/Link:	https://doi.org/10.1177/0959354316641911
Document version:	Accepted manuscript (post-print)

General Rights

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

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

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Testing as social practice

Analysing testing in classes of young children from the children's perspective

Abstract

In this article, the implicit assumption that tests are a neutral tool for measuring the individual's learning achievement is challenged. Instead, testing is explored as a social practice which becomes part of children's conduct of everyday life. The theoretical foundation for the analysis is Danish-German critical psychology. This approach offers a dialectically developed set of concepts and hereby another basis for understanding school testing than the one implicit in the technology of testing. The analysis is primarily based on a case consisting of an observation conducted in a 2nd grade class at a low socio-economic school in Denmark. The analysis focuses on children looking for signs of assessment in the test situation. The discussion focuses on similarities between testing and computer gaming strategies. The article concludes by suggesting that we learn powerful lessons on assessment by taking the children's perspectives seriously.

Keywords: Educational standardised testing; conduct of everyday life; social practice; children's perspectives

Introduction

This article explores how we might understand standardised educational achievement testing as social practice by following school children's participation in test situations and investigating their perspectives on testing. The article is based on an ongoing qualitative study of the meanings children ascribe to their participation in the Danish national standardised tests, exploring children's perspectives on and participation in

test situations. The aim of this is to contribute with a nuanced analysis of the social meaning of tests from children's perspectives.

The article presents and analyses a rather unusual example of a test situation in a 2nd grade class (approx. 8-9 years of age) at a Danish school in a predominantly poor socio-economic area in Copenhagen which has scored low on the final exams compared to other Copenhagen schools. . It is the first time these children have taken this kind of test. Using the empirical analysis of this test situation, I seek to contribute to the research literature on testing by providing general theoretical considerations of the social meaning of testing.

The theoretical framework applied to the empirical material is Danish/German critical psychology because this framework offers a methodology that investigates testing from the perspectives of the participants, offering dialectically based concepts in order to expand dilemmas, contradictions and one aspect which is especially important for this analysis: the ability of participants to engage in redefining and transforming societal conditions.

Until recently, there were no standardised national tests in Danish schools. The Danish municipal primary and lower secondary school (*Folkeskole*) has traditionally focused on the development of a broad set of competences. In 1991, Denmark's participation in an international comparative reading test (IEA) placed considerable focus on basic academic skills (Pedersen, 2011). Denmark was placed lower than most political parties and other stakeholders expected, and these results triggered open criticism of the Danish school system (ibid.). The PISA tests, compiled by the OECD, once again found Danish pupils performing below expectations. Legislation to

introduce national standardised tests was passed in 2006 as part of a package of initiatives aimed at raising academic standards, and the tests were fully implemented in 2010. In Denmark, a lot of teachers are critical of the shift towards such standardisation, and the implementation of national testing has been accompanied by a good deal of critical debate.

Standardised testing has become an industry, and governments all over the world now put their faith in tests assessing the individual's academic achievements. Within education policy there is general acceptance that tests are a relevant assessment tool. This acceptance rests on the assumption that tests are a neutral tool which externally – and thereby objectively – can measure students' learning outcome. Proponents of educational testing emphasise the need for objective, so-called 'colour blind' assessment methods, which external testing is argued to be (Thorndike et al., 1991), and posit the goal of raising academic standards. Dorn (2007) presents a number of historically developed reasons for trusting in tests, one of which he defines as folk positivism, which is a modified form of positivism adopted by stakeholders.

The assumptions presented above have previously been challenged in various ways. In the following I will present research that in different ways point to the non-neutrality of testing in order to critically examine the powerful social consequences of testing. It is argued that tests (especially with high stakes) are powerful symbolic technologies of differentiation, sorting and control (Hanson, 1993; Shahomy, 2001; Madaus, Russell and Higgins 2009), and that tests produce what they are purported to measure, for instance fragmented knowledge and individuality (Hanson, 1993; Au, 2015). It has been posited that schools with many poor and/or ethnic minority pupils

are under more pressure to improve their test results (Study N.F.S., 1992; McNeil, 2000), thereby reducing the quality of education and enlarging the gap between the privileged and the non-privileged (McNeil, 2000). A well-known critique is called 'teaching to the test', which suggests that testing is not a neutral tool that leaves the tested practice unchanged. Instead, problematic washback effects become part of the tested practice because education is narrowed to meet the tested area and creativity is removed (Madaus, Russell and Higgins, 2009; Shohamy, 2001; Au, 2015).

Furthermore, it is argued that testing does not only measure students' abilities, but also produces children/enables children to produce themselves as objects: clever/slow/gifted/disadvantaged/with learning disabilities (e.g. Danziger, 1997; Lave & McDermott, 2002; Hanson, 1993), and that the pressure of high-stakes testing, especially in schools in poor socio-economic areas, can lead to a negative culture of blaming and labelling (Hempel-Jorgensen, 2009). Another criticism is that testing is part of educational structures of competition which make children view themselves as either a success or a failure (Varenne & McDermott, 1998; Reay & Wiliam, 1999), as well as believing that test results say something intrinsic about them (Reay & Wiliam, 1999; Danziger, 1997).

To challenge the assumed neutrality of testing, some researchers have previously pointed to the particularity of the test context itself as a key factor in determining test results. To exemplify, Nunes, Schliemann and Carraher (1993) demonstrate how Brazilian street children performed better at mathematical problem solving and calculation in a street setting than in a more formal test setting. This result shows that tests do not neutrally and directly measure individual academic abilities. As Ole Dreier

(2002 (1993)) states in relation to the experimental situation, participation in specially arranged contexts may also be a reaction to that special situation. Noble et al. (2012) propose that we term the gaps between the test scores of students from schools in low and high socio-economic areas 'test score gaps' instead of the usual 'achievement gaps'. It is suggested that pupils can interpret test items in various ways which do not always match the assumptions in the test design, and that social class is significant for children's test responses (ibid., p. 781; Cooper & Dune, 2000). It is stated that:

[...] some test items appear to function differently for different groups of students
[...] Thus, the students in our study who were from low-income households and ELLs [English Language Learners] students often answered the test items incorrectly despite knowing the targeted science content knowledge for the items, and the middle-class native speakers of English in our study often answered test items correctly even when they did not show evidence of knowledge of the targeted science content for the item. (Noble et al., 2012, p. 796).

This indicates that testing is not neutral but that it is a powerful tool. Dutro and Selland (2012, p. 342) investigate how children in a school in a high-poverty school make sense of testing. They also point to the non-neutrality of high-stakes testing, nevertheless finding that the young children in their study accepted the test as capable of accurately measuring their competences, and furthermore: "[...] its power to determine their trajectory in school" (ibid., 359). This shows how children adopt the implicit

assumptions of testing, underlining the need to discuss and challenge these assumptions.

In this article I will contribute to the body of test critique by analysing children's perspectives of and participation in test situations. This is done by researching testing from within and exploring it as a social practice. The literature reviewed above challenges the implicit assumption that tests are neutral, context-independent tools. Instead, in various ways the literature shows that testing becomes a significant part of the subjects' self-understandings and of the tested practice. On the basis of these findings, I wish to examine the following questions: How can we understand children's participation in test situations as more than a reproduction of test logic? How is this participation connected to test practice itself? And what can we learn about assessment if we consider the children's perspectives?

Theoretical and methodological foundations

The framework of Danish-German critical psychology has a historical dialectical materialistic foundation, meaning that the concept of critical psychology was developed in order to embrace the mutually constitutive processes of reproduction and transformation between subject and society, and to attempt to understand the role of conflict processes in this regard (for an overview see Mørck & Huniche, 2006; Kousholt & Thomsen, 2013). The framework of critical psychology builds on the work

of Vygotsky, Leontjev and Marx, with the more recently developed Danish tradition also being inspired by anthropology and ethnography (e.g. Lave & Wenger, 1991).

In this article, testing is understood as a social practice, carried out in an action context in which children participate as part of their school life (see also Dreier 2008). Understanding testing as a social practice makes it possible to analyse how subjects reproduce and transform conditions, based on how they are positioned and position themselves in their communities, thereby creating opportunities to act for each other (Højholt, 2011). Testing as a social practice carried out in concrete contexts (e.g. classroom, test situation) does not set the same conditions for all children, and the children have different 'reasons for actions' – and thus participate differently (ibid.). The term 'reasons for actions' is used as an analytical concept to help understand subjects' actions in terms of the conditions of everyday life, and not as merely rational or even expressed.

These considerations suggest that if we wish to challenge the implicit assumptions of testing, considering how conditions may look from the perspectives of children and following their participation in and across test situations might be a useful thing to do. This is relevant in this analysis because even though I want to explore testing as something which is done to children, thereby structuring their everyday lives and their understanding of being clever/not clever, I also wish to explore how children make sense of testing and transform testing. This points to the concept of generalised human agency as mediating between the individual and the social conditions and this: "refers to the human capacity to gain, in cooperation with others, control over each individual's own life conditions." (Holzkamp, 2013, p. 20).

In my analysis of children's participation in test situations, I am inspired by Holzkamp's (2013) critique that the standard experimental design of traditional psychology is detached from everyday life, and that this also applies to test situations. Holzkamp points out a kind of 'structure blindness' to the experiment itself which makes it difficult for subjects to act in opposition to what is expected of them (Holzkamp, 2013, p. 280-281). This perspective is discussed in the analysis because the children in the material presented here actually break with the structures of testing. As with the experimental setting, test situations can only be reproduced as such if the tested subjects do as the situation prescribes: "In reality, of course, each person involved knows that it is merely a matter of an agreement or – however seriously meant – an experimental game which only works as long as everyone 'plays along'" (Holzkamp, 2013, p. 271). However, the concept of structure blindness questions whether the involved parties are consciously aware that testing is a matter of agreement. Some of the literature referred to above indicates that this is not always the case for children in test situations. In the analysis I will explore how, in the case presented here, the children both play along and break with the expectations and assumptions of testing.

The empirical basis for this project consists of transcriptions and observational notes from semi-structured individual and group interviews, and of participant observations conducted in five school classes in four different Danish schools in the years 2011-2012. In this article I present the case of a particular test situation and relate it to this broader empirical material.

Presenting the Danish national standardised tests

The Danish national tests are mandatory and computer driven with an adaptive technical design. The test is delivered by the Ministry of Education, and children in the same school class are normally tested together. The adaptive design means that the level of difficulty is dependent on the pupils' answers; which means that when a child answers one task correctly/incorrectly, a more difficult/easier task will be presented to that child. The test is self-scoring and the goal of the test is to determine the individual pupil's academic aptitude in the tested content area. When the system has established the pupil's level within the tested area, the test stops, which means that the children both receive different test tasks and that the test finishes at very different times even though the overall time span for the test is 45 minutes with the option of extension if the pupil's academic level within the areas of the test has not been established. The test results are not dependent on time spent. The Danish reading test (which is represented in the empirical material) is divided into three different profile areas: decoding words, understanding of language and understanding of text. All the items are multiple choice. The political intentions are that the test results should be used for pedagogical purposes in order to raise the academic level. The tests are not used for giving grades or for granting access to further education, and neither tasks nor results are made public. The teachers receive the test results electronically and are obliged to inform the pupils' parents of the pupils' score, the minimum requirement in this connection being that the parents receive a standard letter from the ministry stating

the pupils' test score compared to the national average. The teachers are encouraged by the ministry to use the test results in their didactic planning, for instance via differentiated teaching. This means that the Danish national standardised tests are intended to be low-stakes testing. However, as Allen (2012) points out, even though there are important differences between low-stakes and high-stakes testing, they share a similar disposition towards power, and they both support learning in the form of individual progression towards rather narrow and prearranged outcomes. Recently (after the Danish school reform in 2014) the tests have been given more power since they are now supposed to measure whether two out of three national goals are being achieved.

Empirical case

In this section, I will present extracts from an observation of a test in Danish reading, focusing on the participation of two boys: Marius and Valdemar. The example is atypical because in the extracts the pupils do not participate as expected in test situations in a very observable, visible manner. However, as the broader empirical material show, it is not unusual for pupils to participate in test situations in unexpected ways. For instance, it becomes important for some pupils not to be one of the last to finish such tests, and some of the children in other test situations help each other and compare the number of tasks they have been set – even though this, in contrast to this example, is done in a rather concealed manner. In this case study I

have chosen to focus on the participation of those who do not just 'play along'.

However, there were also children in the school class performed very close to the expected norms of testing but the overall impression of the test situation is that it was a situation characterised by a lot of activity. I argue that this specific case analysis on the boundary of the norms of testing identifies some common aspects of testing. This case analysis illustrates: 1) that testing is characterised by lack of control of the participants; and 2) that participants can transform the meaning of testing in test situations.

Empirical case

In everyday school life the children and the teacher seem to have a very informal relationship with each other. The teacher (Bente) normally allows the children to talk quietly together and she seems to know the children well because she prioritises conversations with the children when there is enough time. Furthermore, the teacher encourages the children to help each other in everyday school life. Prior to the test, she explains the significance of this test to the children. Among other things, she tells them that they will have a nice and cosy little test today. She also tells the children that they are not allowed to work together or talk together. They should remain quiet, stay seated and raise their hand if they need help. Despite this framing of the situation, in the situation Bente allows some degree of talk and also talks to the children of other things than specific test questions alone. Prior to the test, she also says that the teacher is allowed to read the test task out loud for the children, and

that different colours on the teacher's screen will indicate if the child has just started, is on his/her way, or is finished. She tells them that they will receive the result of the test at an upcoming school-home meeting. The children and the teacher go to the computer room to carry out the test because this is the only place at the school where there are computers for all the children. The computer room is located on the second floor, and the classroom is located on the ground floor. The children are not used to being in the computer room, although they have been there before. Recently they have carried out a demo test delivered by the ministry in the same computer room. Bente has also done additional training of some of the profile areas in the test.

In the computer room there is a teacher's desk and a big screen on the wall behind the teacher's desk connected to the teacher's computer (the smartboard) at one end of the oblong room (the front of the room). The children's tables are placed in different ways in the room: either in small groups or beside each other in rows. Not all the children are facing the teacher's desk and screen. There are two teachers present during the test situation. One of them sits with four of the children the whole time because these children are considered to be in need of help. The other teacher (Bente), who informs and talks to the children during the test, both gives loud instructions to all the children and walks around from table to table, talking quietly to some of the children and helping some of them. She spends a lot of time at the back of the room helping children with their tests, while the observations are mostly carried out at the front of the room, where the teacher's screen is and where the children in the extracts go. This means that a lot of other interactions are going on in

the room which the observation notes do not describe.

The test starts. Some of the children are talking to each other, and the teacher, Bente, hushes them. Bente talks to all the children and says that if the children need help they must raise their hands. A lot of the children raise their hands.

One of the boys, Ruben, asks Bente: 'How many tasks are there?' Bente answers: 'endless'. Ruben asks 'When are we supposed to finish?' Bente answers: 'In 45 minutes'. Ruben asks in a suspicious tone: 'Are we supposed to do these tasks for 45 minutes?'

After some time, a boy says that he is on task number 12 now. Ruben says that he's on task number 15. Several of the other boys start telling each other how far they've got. Bente says loudly 'Stop' (talking).

Marius and Valdemar sit next to each other on the first row. They and another boy leave their computer and stand up in front of the teacher's screen as displayed on the smartboard (it is not common practice for the teacher's screen to be visible, but it is a possibility in these tests). Here they can see and compare the number of tasks that the children in the class have completed (but not the correctness). Valdemar says that he wants to see what percentages of the tasks are right.

They return to their seats. Marius helps Valdemar with a task (later on Valdemar receives help from other participants as well). Marius is given a task with a long text which he is supposed to read. He exclaims: 'Wow'. Valdemar says: 'I didn't bother to do that task. I skipped it.'

Valdemar looks briefly at a task. He says 'It's too difficult' and skips it.

Marius gets up and stands in front of the teacher's screen in order to see the different numbers better than he can from his desk. He cheers: 'Yes, I've got 21 rights!' Bente stands at this point near him and says that he cannot count on that.

Marius says, with an impressed voice, that Daniel has done 46. He afterwards asks another boy: 'How's it going for you, how many have you done?' Then he looks at the screen and cheers: 'I have 30 tasks, man.'

Valdemar now has his hand raised almost all the time and he says (not to anyone in specific): 'I can't do these tasks (he does not seem to read them) – I just skip them.'

Again, several of the children (mostly boys and also some girls) have left their seats and are standing in front of the teacher's screen. They look at the number of tasks they have done and compare their total with the other children's totals.

Valdemar says to Marius 'You just need to do four tasks then you are just as far as I am.'

The teacher's screen now shows in a green colour that some of the children have completed their test.

Valdemar bursts out (in a frustrated tone): 'Hey, I got 48 and my test isn't finished yet.' Valdemar then sees that Marius is finished and tells him. Marius cheers 'Yes!'

More of the children get up and look at the teacher's screen (mostly boys and also some girls).

Bente comes up to the screen (she has helped some children sitting at the back of the room), and says that the children cannot see how far they have got.

Valdemar and Bente stand next to each other. He tells her that he has a problem. She

asks him what it is. He says that he has now reached 45 and he is still not finished.

Bente says that he might need to have 90 before he is finished. Valdemar says 'I can't do any more.' He goes to his desk and sits at the computer. He quickly skips the next four tasks while saying 'too difficult, too difficult' to all of them (again not to anyone specific). Valdemar finishes shortly afterwards.

Bente says out loud that some of the children are done and can leave the room. The children who have finished run out of the door, rejoicing.

(Observation, 2nd grade, spring 2012)

What is apparent in the observation notes is that there is a gender division. In the extract above mostly boys are represented, even though a lot of girls were also present during testing. I also saw some girls compare their totals later on during testing albeit in a more quiet way. This gender division is undoubtedly relevant for further analyses, and I will elaborate a little on this in the discussion beneath.

However, it is not the central focus for this paper.

Empirically based analysis; children's participation in the social practice of testing

The analysis will centre on the social practice of testing and the children's reasons to act in certain ways.

The lack of control

The crucial point to be made about testing is that it is a social practice that implies lack of control for the participants, and that children participate in this special situation in different ways (cf. Dreier, 2002 (1993)). For instance, neither the children nor the teacher know when the children will finish or how many tasks they have to do, as apparent in the conversation between Ruben and Bente when Ruben asks: "How many tasks are there?" and Bente answers: "Endless". Ruben asks again: "When are we supposed to finish?" and Bente answers: "In 45 minutes". Here Bente refers to the overall time span. In practice, the children finish at different times with the last child taking about one hour and 15 minutes. The last child to finish is a girl who explains afterwards: "I couldn't understand anything, that's why". Given the adaptive nature of this test, Bente cannot provide Ruben with information on the number of tasks. This is also the case when Valdemar tries to engage Bente in his deliberations about when he will finish: Valdemar tells Bente that he has a problem. She asks him what it is. He says that he has now reached 45 and he is still not finished. For Valdemar this lack of control or overview with regard to when the test will finish is expressed as a problem that he has. This is also the case with the girl who finished last. Like Valdemar she does not define it as a problem of the system, but as her own problem (that she couldn't understand anything). The teacher's response to Valdemar can be interpreted as demonstrating her lack of overview. She says to Valdemar that he might need to do 90 tasks before he is finished. This utterance illustrates the teacher's lack of knowledge of

the number of tasks, and therefore she is not able to help Valdemar increase his control over the test situation in this manner. In a subsequent interview with the teacher she describes the test situation as stressful, and says that immediately after the test Valdemar had told her that he couldn't read. She says that Valdemar often benefits from reading with one of the more skilled readers in everyday school life, but that this was not an option during testing. However, in the test situation Valdemar receives help from Marius among others. In different ways the children and the teacher becomes part of each others' possibilities for control in the situation. In the following I will analyse how the children also look for signs of assessment as connected to the lack of control implicit in the test context.

Looking for signs of assessment

In the test situation, the children do not know the tasks beforehand and the test system does not provide them with feedback on how they are doing during the test. This means that some of the children in my material look for signs of assessment in different ways. In the extract above we see how some of the children perceive the number of tasks as feedback on the correctness of their responses. This is particularly apparent when Marius cheers: "Yes, I've got 21 right!" The teacher corrects his mistake, but this does not seem to lessen his preoccupation with the number of tasks. As a result, some of the boys, including Valdemar and Marius, concern themselves with the number of completed tasks. However, they are not only interested in getting a

high number – they are also interested in getting a high number compared to and together with other boys which points to testing as a social practice. Later, when I ask Valdemar about this during an interview, he says that he and the other children were trying to support each other through their interest in each other's totals. This is also apparent in the observation when Valdemar says to Marius: "You just need to do four tasks then you are just as far as I am"; and when Marius asks: "How's it going for you, how many have you done?" The latter sentence combines an interest for the other person (How's it going for you) with an interest in the number of completed tasks (how many have you done?). These utterances suggest that the boys' involvement in the number of tasks is a question of both competition and support. They are engaged in a joint activity trying to orient their participation through the other children's participation and through the number of tasks.

In accordance with Holzkamp's (2013) notion of structure blindness, tests in schools are a historically developed social practice with certain implicit power relations with respect to differentiation and individualisation, and where "certain possibilities to act are not even 'seen'" (ibid., p. 280). Even though, or perhaps because, some possibilities to act are not visible to the children, they engage in what is actually visible in the situation. This enables them to ascribe meaning to the test while drawing on other relevant contexts of their everyday lives together. The children seem to use the number of completed tasks to orient their participation towards that of other pupils, partly with the purpose of comparing progress. As Fischer (2011) states: "[...] *it is the child's interpretation of what matters that orients their actions.*" (ibid., p. 51). From Valdemar's and Marius' perspectives, it is apparent that completing a large number of

tasks matters to them. Completing a large number is a visible sign in a test situation that leaves little room for orientation. The number of tasks orients the children towards expectations of right/wrong.

The boys' engagement in the number of tasks changes during the test. At first it seems to be a way of navigating the test; a way to support and compete against each other. Later on it seems as if the boys use the number of tasks as a sign of when they will finish. For the teacher (and this is implicit in the test design), this is not an accurate indication: as the teacher tells Valdemar, he might need to reach 90 before he finishes. Valdemar sees Marius celebrate finishing before him, even though Valdemar has completed a larger number of tasks. And he sees that finishing fast is worthy of celebration. Valdemar's strategy of getting a high number of tasks seems to fail him and he expresses this as a problem that he has. The boys' preoccupation with finishing quickly, as well as the teacher's reply, might be among Valdemar's reasons for skipping four tasks in a row.

The collected material shows that the number of tasks is but one of several ways of looking for signs of assessment. Other children ask the teacher if a task is answered correctly before submitting. Some of the teachers provide them with some help and feedback, but not always and not in all classes (mostly in the younger classes). Some of the children try to guess the right answers from the teacher's reaction when they ask for help. Some of the children consult each other (in the older classes, this is done when the teacher is not watching because it is not allowed) or look at each other's screens, even though this kind of test makes it difficult to copy others. These activities have similarities to Powell, Danby and Farrell's (2006) investigation of children's

passing notes during teaching which (in their material for the girls' part) is done outside of teacher regulation. This at the same time points to the category of the teacher as an observer of students and one that expects students to do their work (ibid., p. 270) and in the test situations as both a maintainer of the order of testing and sometimes as someone who it is possible to negotiate that order with. The activities of looking for signs of assessment also point to testing as a context for the pupils' and teachers' social interaction even though testing is understood and arranged as a context for individual assessment. The children do not always know whether they are solving a particular task correctly, but they do know that they are participating in an exercise in which correctness and cleverness are at stake. When I subsequently interviewed the children from the test situation in the observation above, one of the girls (not the girl who finished last) stated that she is used to being one of the last ones in test situations because, as she says: "I'm not very clever, so...". This is an example of how testing and other school results become part of children's understandings of themselves as clever or non-clever, whereby testing is linked to the understanding of innate characteristics of the self (e.g. Grant, 2006, p. 111; Reay & Wiliam, 1999; Danziger, 1997).

As my interviews with children from different schools show, children feel connected to the other children in the test situation because of their communities across contexts and because of the shared activity in the test situation. For instance, they wonder how far the other children have got, if they have been given the same tasks, if they will finish first or last. The significance of the last point is apparent when some of the pupils say in interview situations that they became anxious when several classmates finished

and left the room because they saw this as reflecting the inadequacy of their own performance. The pupils also express in the interviews that they are aware that this is not actually the case in this type of test. Even though this is not important according to the test design, one might say that it is a logical ramification of testing practice because understandings of cleverness are at stake and speed is often perceived to matter in structures of competition.

One of the opportunities for assessment that the children look for in test situations is when they finish. Being among the first or last to finish is visible to the children, and they remember when they finished their test compared to the other children in the class. As well as numbers, time becomes an indicator of performance for the pupils even though neither of these aspects is intended to be part of this test type. Testing involves differentiating between test performances, which is generalised to differentiating between pupils' academic abilities. Aronsson and Hundeide (2002) present an understanding of children's (sometimes 'wrong') responses in test situations and elsewhere. They argue that children often view participation as their overriding goal in the situation. Whereas the goal of the examiner/the test is to generalise from the concrete situation to generalised phenomena or abilities, the children have other agendas (*ibid.*, p. 182). Aronsson and Hundeide apply the term 'relational rationality' to point to the local logic of children's participation. Like it is argued in relation to interview situations, (Danby, Ewing and Thorpe, 2015) children respond in meaningful ways in relation to the resources available in the context, and Danby, Ewing and Thorpe add that "Recognizing children as competent verbal and nonverbal communicators allows for new insights of how they construct their social

worlds [...]” (ibid., p. 82). Following these insights makes it possible to achieve understandings that go beyond viewing the children in the observed example as incompetent test-takers. Instead, the children in very competent ways apply situational signs such as numbers of tasks and time spent to indicate performance and cleverness. Even though this is not the intention of the test, it is a local logic of the children’s participation and can be termed ‘relational rationality’. What is more striking is that the test system itself delivers these opportunities for looking for signs of assessment.

What appears paradoxical is that we have developed certain kinds of assessment in which the assessment itself is invisible/intangible, and that this lack of assessment seems to encourage the children to find alternative ways to orient their participation, negotiate the meaning of the test and position themselves in comparison to their peers.

Children testing tests

The children in the observation reveal a paradox of testing itself: in the attempt to design an independent, neutral and objective test that can isolate children’s academic competences from what is going on in their lives, the test depends on the children who are being tested playing along (cf. Bernstein, 1971, p. 27; Holzkamp, 2013). This means that the children participate in ways that they find meaningful and, through this participation, they challenge the implicit assumptions of testing: they test the test.

What is intended as a precise measure of individual skills paradoxically becomes a part of the children's engagement in social practice, and they produce the test situation itself as something that the test designer did not intend. The test design becomes part of what constitutes the children's participation in the test itself. The test design is a constitutive part of Valdemar's and Marius' competition/shared engagement/number-counting, and it becomes a constitutive part of Valdemar's task-skipping. The children hereby participate in meaningful and competent ways, trying to learn how to orient their shared participation and to enhance their degree of control, namely by directing their participation towards what is actually visible to them.

The test situation offers a set of 'possibilities to act' (Holzkamp, 2013) which supports the idea of testing as a neutral and decontextualised measurement of individual academic skills. As Holzkamp (2013, p. 279) reminds us, these possibilities to act should not be misunderstood as deterministic. However, when subjects actualise the possibilities to act that testing offers, they contribute to the reproduction of social structures (following Holzkamp's [2013, p. 180] argument on the experimental setting). The boys in the observation reproduce some aspects of testing and children like Valdemar risk understanding themselves/being understood as pupils who do not perform well at tests, who are test-incompetent. However, they transform the local social structures as well as creating different meanings of the test situation. The intrinsic nature of test situations as social practice is apparent in the case extracts. In other words, the tested children interact in the historically produced test situation, creating possibilities and limitations for each other, and helping to reproduce and transform the test.

Discussion; testing and computer gaming strategies – acting ‘in between’

This section deals with similarities between testing and computer gaming. These similarities occurred in the analytical phase after the ending of the empirical production, which is why this section merely points to possibilities of interpretations and possible further research designs involving systematic empirical work on these relations.

The boys’ shared engagement in getting a high number of tasks in the test situation above can be linked to computer gaming strategies: partly because this particular kind of test is computer based, and partly due to the boys’ familiarity with computer games. I have previously observed some of the same boys playing computer games in school, with them engaging in the same forms of activities and communities as in the test situation. They did not sit still for long. Instead they consulted each other and took an interest in each other’s gaming. Aarsand and Aaronsson (2007) point to children’s (boys’) different kinds of communicative actions (response cries among others) during gaming in order to secure and display “joint involvement and collaboration” in gaming (ibid., p. 1567). This has similarities to the boys’ joint involvement in test taking. For instance Valdemar’s response to the test saying ‘too difficult’ while skipping tasks and when the boys start telling each other how far they’ve got in the beginning of the example. The boys share their number of tasks as shared insights in signs of assessment and as joint involvement as some of the boys in Aarsand and Aarronssons

empirical material on computer gaming. In some of the computer games the children play, the point is to move fast and quickly skip tasks that are not immediately clear. Aarsand and Aaronsson (2007) identify various gaming strategies and participation. Some of these are combined in ways that “create a sense of speedy action”. Speed is central to computer gaming, and for the way in which some of the children understand the test situation. This is also how Valdemar participates in the test situation. He would rather quickly skip some of the tasks than spend a lot of time on them.

In Aarsand and Aaronsson’s (2007) empirical examples on computer gaming there are only boys represented. Boys’ engagements in communities involving certain games might also be part of the constitution of the gender division in the observation presented above. Powell, Danby and Farrell (2006) point to a gender division in children’s passing notes during teaching which could be included in a further investigation of gender division in different activities. The authors argue on basis of interview material that: “These elaborate descriptions reveal how the girls competently participate in the covert activity of passing notes, whereas the boys competently participate in the overt activity of passing notes [...] (ibid., p. 261) and that the boys’ overt activity is “[...] a way for these boys to ‘maintain, organize and articulate their activities to display their gender to others’ [...]” (ibid., p. 271). This could point to the boys’ participation in the test situation as both reminiscent of their interaction in their gaming communities and as gendered practice, which need further investigation. At the same time the children are also aware that this is a situation where individual performances are measured.

Søndergaard (2013) points to one key difference between mastering school tasks and mastering computer games. Whilst peer learning is accepted by the children as an element of computer gaming, in part because it is more fun to play with and against other skilled players, school tasks are largely seen as an individual activity establishing hierarchical differences (ibid., p. 121). Some of the boys in the empirical example above seem to act 'in between' individual school activities and gaming communities in this particular test situation. They are using computer gaming skills, interacting with each other as in gaming communities, and they apply situational cues to gain control in the situation, but they are also aware that this is a situation supposed to measure individual performances. The teacher is an important participant in the boys' opportunity to do this 'acting in between'. She allows the children a certain degree of interaction (talking and wandering around), even though this goes against her pre-test instructions. As mentioned previously, a lot of teachers in Denmark have been critical of national standardised tests because they fear they restrict their teaching and damage the self-esteem of some pupils. In a subsequent interview, the teacher explained to me that she told the children that the test would be "nice and cosy" because she wanted to downplay the importance of the test and create a relaxed atmosphere. As mentioned in the introduction, Denmark has a tradition for more formative assessment methods and less focus on performance than other countries, even though this has changed recently. The teacher could be criticised for being unable to maintain the proper test conditions, or respected for her willingness to support learning in communities even in test situations.

There are boundaries that distinguish this situation from other situations, and the teacher explains these to the children prior to testing. The purpose of testing is to categorise hierarchical levels of cleverness/aptitude, the level of control is normally quite high, the skills to be tested are not new but curriculum-driven and already known, and it is possible to categorise the children's answers as either right or wrong. The boys act 'in between' because they are simultaneously reproducing structures of testing, for instance hierarchical structures of cleverness as a competition against each other, and redefining what the test is about, for instance taking care of each other and computer gaming. Following Holzkamp's (2013) understanding of reproducing and/or transforming experimental settings as a matter of 'playing along', I posit that the children do play along, even though they – in cooperation with the teacher – change what the test is about.

Conclusion

The empirical material and the analysis demonstrate how children draw on their competences and communities as they attempt to make sense of the test situation. The form of the test itself co-constitutes the boys' participation in the test context. The situation is highly ambiguous and the form of the test itself co-constitutes this particular intermingling of computer gaming and test taking; of support and competition; of test reproduction and test transformation. The children strive to make sense of the test situation as an action context connected to other action contexts in their shared conduct of everyday life. In doing so, they challenge the implicit

assumptions of testing as a neutral measurement of isolated skills which is able to provide context-free test results.

However, what happens in the context, i.e. the boys' orientations, engagement, support, competition, computer gaming strategies and frustration, is not included in the standard presentation of the test results. Instead, the test results will appear as a context-free measurement of the individual's academic aptitude in the tested area – no more and no less. What seems to be a neutral tool for measuring the individual learner's achievement level in isolation becomes part of children's attempts to make sense of the test with respect to looking for signs of assessment, engagement in numbers and computer gaming. Some of the children seem to strengthen their communities during test taking, and some seem to strengthen an understanding of themselves as not clever/incompetent.

On the basis of the analysis, I suggest that instead of viewing the children's and teacher's participation in the test situation as wrong, and instead of trying to create test practices that look more like what is intended from the test developers' point of view, for instance by regulating test situations more, we could learn powerful lessons on testing from the participants. We could learn that the children's participation is competent and reasonably connected to the logic of testing itself and to other everyday life situations of the children. We could also learn about the meaning of testing from the children's perspectives and take this into account in more fundamental ways. For instance, the children are engaged in redefining and transforming the test as a relevant assessment method that could be done in communities, that involves possibilities of joint engagement and possibilities for

orienting one's participation. Furthermore, the children's participation points to the potentiality of developing assessment methods that increase the participant's agencies, and perhaps assessment methods that give children the chance to transform the present condition. This is how the children try to redefine the test during testing, and perhaps we should take the children's redefinition seriously when developing and re-configuring tests and other assessment methods in education. This means that instead of negating and sorting out this test situation as 'wrong', we could learn how to develop relevant assessment methods from the perspectives (and the engagement) of the children involved.

References

- Aarsand, P. A. & Aronsson, K. (2009). Response cries and other gaming moves – Building intersubjectivity in gaming. *Journal of Pragmatics* 41: 1557-1575.
- Allen, A. (2012). Cultivating the myopic learner: the shared project of high-stakes and low-stakes assessment. *British Journal of Sociology of Education* 33, no. 5: 641-659.
- Aronsson, K. & Hundeide, K. (2002). Relational rationality and children's interview responses. *Human Development* 45: 174-186.

Au, W. W. (2008). Devising inequality: a Bernsteinian analysis of high-stakes testing and social reproduction in education. *British Journal of Sociology of Education* 29, no. 6: 639-651.

Bernstein, R.J. (1971). *Praxis and Action*. Philadelphia: University of Pennsylvania Press.

Cooper, B. & Dunne, M. (2000). Constructing the 'legitimate' goal of a 'realistic' math item. A comparison of 10-11 and 13-14 year-olds. In *Assessment. Social practice and social products*, ed. A. Filer, 87-109. London & New York: Routledge/Falmer.

Danby, S., Ewing, L. and Thorpe, K. (2011). The novice researcher: Interviewing young children. *Qualitative Inquiry* 17, no. 1: 74-84.

Danziger, K. (1997). *Naming the mind. How psychology found its language*. London: Sage Publications.

Dorn, S. (2007). *Accountability Frankenstein. Understanding and taming the monster*. Charlotte: Information Age Publishing.

Dreier, O. (2002(1993)). *Psykosocial behandling. En teori om et praksisområde*. [Psychosocial treatment. A theory of a realm of practice.] (2. ed.). Copenhagen: Dansk Psykologisk Forlag.

Dreier, O. (2008). *Psychotherapy in Everyday Life*. New York: Cambridge University Press.

Dutro, E. & Selland, M. (2012). "I like to read, but I know I'm not good at it": Children's perspectives on High-Stakes Testing in a High-Poverty School. *Curriculum Inquiry* 42, no. 3: 340-367.

Fischer, R. (2011). Failing to learn or learning to fail? The case of young writers. In *Vygotsky and special needs education. Rethinking support for children and schools*, ed. H. Daniels and M. Hedegaard, 48-64., London & New York: Continuum International Publishing Group.

Grant, B. (2006). Disciplining students: the construction of student subjectivities. *British Journal of Sociology of Education*, 18, no. 1: 101-114.

Hanson, F. A. (1993). *Testing testing. Social consequences of the examined life*. Berkely & Los Angeles: University of California Press.

Hempel-Jorgensen, A. (2009). The construction of the 'ideal pupil' and pupils' perceptions of 'misbehaviour' and discipline: contrasting experiences from a low-socio-economic and a high-socio-economic primary school. *British Journal of Sociology of Education*, 30, no. 4: 435-448.

Holzkamp, K. (2013). Psychology: Social self-understanding on the reasons for action in the conduct of everyday life. In *Psychology from the standpoint of the subject. Selected writings of Klaus Holzkamp*, ed E. Scraube & U. Osterkamp, 233-341. Basingstoke: Palgrave Macmillan.

Højholt, C. (2011). Cooperation between professionals in educational psychology – Children's specific problems are connected to general dilemmas in relation to taking part. In *Vygotsky and special needs education. Rethinking support for children and schools*, eds. H. Daniels & M. Hedegaard, 67-85. London & New York: Continuum International Publishing Group.

Kousholt, K. & Thomsen, R. (2013). Dialectical approaches in recent Danish critical psychology. *Annual Review of Critical Psychology* 10, no. 1.

Lave, J. & McDermott, R. (2002). Estranged labor Learning. *Outlines* 4, no. 1: 19-48.

Lave, J. & Wenger, E. (1991). *Situated Learning. Legitimate peripheral participation*. New York: Cambridge University Press.

Madaus, G., Russell, M. and Higgins, J. (2009). *The paradoxes of high-stakes testing. How they affect students, their parents, teachers, principals, schools and society*. Charlotte: Information Age Publishing.

McNeil, L. M. (2000). *Contradictions of school reform. Educational costs of standardized testing*. New York: Routledge.

Mørck, L. L. & Huniche, L. (2006). Critical Psychology in a Danish context. *Annual Review of Critical Psychology*, 5.

Noble, T., Suarez, C., Rosebery, A., O'Connor, M.C., Warren, B. and Hudicourt-Barnes, J. (2012). "I never thought of it as freezing": How students answer questions on large-scale science tests and what they know about science. *Journal of Research in Science Teaching* 49, no. 6: 778-803.

Nunes, T., Schliemann, A. D. and Carraher, D. W. (1993). *Street mathematics and school mathematics*. New York: Cambridge University Press.

Pedersen, O.K. (2011). *Konkurrencestaten*. [*The competitive state*.] Copenhagen: Hans Reitzels Forlag.

Powell, K., Danby, S. and Farrell, A. (2006). Investigating an account of children 'passing notes' in the classroom. How boys and girls operate differently in relation to an everyday, classroom regulatory practice. *Journal of early childhood research* 4, no. 3: 259-275.

Reay, D. & William, D. (1999). 'I'll be a nothing': structure, agency and the construction of identity through assessment. *British Educational Research Journal* 25, no. 3: 343-354.

Shohamy, E. (2001). *The power of tests. A critical perspective on the uses of language testing*. Harlow: Pearson Education Limited.

Study, N.F.S. (1992). *The Influence of testing on Teaching Math and Science in Grades 4-12*. Chestnutt Hill, MA: NFS Study.

Søndergard, D. M. (2013). Den distribuerede vold: om computerspil, mobning og relationel aggression. [The distributed violence: on computer games, bullying and relational aggression.] In *Mobning gentænkt. [Rethinking bullying]*, ed. J. Kofoed & D. M. Søndergaard, 116-160. Copenhagen: Hans Reitzels Forlag.

Thorndike, R. M, Cunningham, G. K., Thorndike, R. L. and Hagen, E. P. (1991) (5th ed.). *Measurement and evaluation in psychology and education*. New York: Macmillan Publishing Company.

Varenne, H. & McDermott, R. (1998). *Successful failure. The school America builds*. Colorado: Westview Press.

