Beyond washback effect: A multi-disciplinary approach exploring how testing becomes part of everyday school life focused on the construction of pupils' cleverness

Lars Holm and Kristine Bagge Kousholt

Abstract

The growth in international assessments within education, and the frequently intense attention the results garner, make it highly relevant to examine how these assessments are affecting the ways in which education is understood, defined, and practiced. In order to analyze some of the complex consequences of educational assessment, we propose a multi-disciplinary approach that combines perspectives from Latour´s actor-network theory and critical psychology – a combination chosen in order to contribute to the critique of testing in a way that goes beyond common understandings of washback effects. This multi-disciplinary approach is applied to data from longitudinal fieldwork in five schools in Denmark. We analyze how the Danish national test in literacy becomes part of the social construction of cleverness and, in this way, interrogate the traditional understanding of positive or negative washback effects. The findings reported in this article reveal some of the significance of assessment with regard to how pupils´ cleverness is negotiated and constructed in and around literacy testing, and at concomitantly demonstrate how cross-disciplinary concepts based on theories of everyday life can contribute to the field of testing research with a critical perspective.

Keywords

testing, washback, test as social practice, multi-disciplinary approach, critical psychology, ANT,

Introduction

The political desire to increase learning in primary and lower secondary education forms the basis for the development and implementation of national strategies in which more comprehensive testing plays a prominent role. In recent time, in the USA, it was the national education strategy, "No Child Left Behind" (NCLB) that triggered intensified testing (Nichols & Berliner 2007); in Norway, increased testing was embedded in a 2006 government resolution regarding a so-called "Kunnskapsløft" [Knowledge Lift] (Kulbrandstad 2010); and in Denmark, national tests are part of a comprehensive program, adopted in 2005, with the title "Promoting assessment culture" (Pøhler & Sørensen 2010). In Norway and Denmark, the development of such strategies has a common background: poor placements in the Programme for International Student Assessment (PISA) ranking list, in which literacy levels are compared among students in a number of countries (Elstad & Sivesind 2010; Holm 2014). Thus, the increased use of tests is simultaneously anchored in local, situated practice that plays a role in individuals' everyday lives and as a contemporary social phenomenon in international comparisons and increased international competition (McNamara 2005; Nichols & Berliner 2007; Extra, Spotti & Avermaet 2009; Elstad & Sivesind 2010; Holm & Laursen 2011).

However, according to McNamara and Roever (2006), conducting research in a way that reflects the social, societal dimensions of tests and testing practices presents significant theoretical challenges:

> …to investigate the social meaning of tests and their social impact takes us into relatively uncharted waters theoretically. The problem here is that we need an adequate social theory to frame the issues that we wish to investigate. (McNamara & Roever 2006, p. 253)

We accept this theoretical challenge and, on the basis of a critical review of existing research on testing, argue for a multidisciplinary, critical approach, in which testing is analyzed as a social practice. The research problem in this article is how to grasp how testing becomes part of an everyday construction of cleverness. Furthermore, we hope to move beyond the common understanding of positive/negative washback. We acknowledge the need for exploring the effects of testing on teaching, as addressed in research on washback, but believe it is equally important to more broadly interrogate the ways in which testing affects the understanding and construction of pupils' cleverness - something which occurs both as part of test practice (formal construction of cleverness) and as part of a more complex everyday life (informal construction of cleverness). By addressing these issues theoretically and analytically, we wish to contribute to a

dynamic and context-sensitive understanding of testing as part of everyday school life.

The theoretical inspiration for this approach is, on the one hand, critical psychology's focus on how structures of cleverness are created in a dialectical interaction between the participants and the conditions of local communities, and on the other hand, actor-network theory (ANT), with its focus on the interaction between actors and objects. This combination of different theoretical perspectives allows a research approach in which analytical attention is paid to the interaction between humans and materiality in constructions of cleverness. Both theoretical perspectives share the view that materiality is created by human beings while at the same time becoming part of practice and/or designing practice in unpredictable ways. As Sørensen (2006, p. 149) states: "Design is never just about making a technology but always also about designing practices, whether intentionally or not." Sørensen (2006) furthermore argues that the actors or participants in a practice will always also do something with the design or materiality in unpredictable ways, so one could say that materiality is never fulfilled with the design of the technology. Analyzing testing on this basis directs research attention, not to constructions of cleverness as a general phenomenon, but to structures of cleverness in relation to testing, and to the local constructions of cleverness in a community. This points to a need for a twin focus; on both structures of cleverness that have developed historically and are embedded in materiality, for instance as testing, and on the ongoing constructions and negotiations of cleverness in specific contexts, such as a classroom.

A multidisciplinary approach to testing, first requires a clarification of the scientific traditions´ complementarity and differences, of which the research questions will be central. We will then demonstrate the potential of this research approach through an analysis that focuses on the structures of cleverness that emerge in the interaction between the test, teachers, and pupils in different classrooms in Denmark. The purpose of this analysis is to challenge a dualistic understanding of washback effect as either positive or negative and point to the ambiguity and complexity of the context in which testing is situated: everyday school life.

The concept of cleverness is our attempt to translate the Danish concept 'dygtighed', but it is not entirely adequate. 'Dygtighed' is a rather imprecise concept and the meaning depends on the contextual use. It can mean skill, ability, aptitude, cleverness, or competence. However, the concepts of 'dygtighed' and 'cleverness' share a focus on the individual's innate categories and are both often understood as something quantifiable. 'Dygtighed' is a frequently used empirical concept in Danish schools, and the practice of cleverness in accordance with test practice will be analyzed and discussed throughout this article. By examining the

interaction of structures of cleverness that are constructed through and around national tests, we analyze the complex meanings that assessments have for teachers and students in Danish classrooms. We begin by unfolding the theory-based arguments supporting a multidisciplinary analysis of testing before exemplifying such an approach through two concrete analyses.

An approach to research in testing that goes beyond washback effect

In a general ontological perspective, tests may be viewed from different paradigmatic standpoints[1]. The first such paradigm that we will present views tests as primarily a measuring instrument which is societally helpful and can be used to generate knowledge about, and compare, academic levels among individuals, schools, and nations. The theoretical point of departure for this paradigm is theories of validity and psychometric theory, including "item-response theory" (McNamara & Roever 2006). "Validity" refers to the question of whether a test measures what it is intended to measure. Over the years, the concept of test validity has been refined and differentiated to cover different types of validity; for example, content validity, construct validity, and consequential validity (Bachman & Palmer 1996; Weir 2005; McNamara and Roever 2006). The concept of reliability is often associated with assessing validity, and indicates the precision with which tests measure. The reliability of a test is often measured by repetition; an approach that may be recognized from studies within the natural sciences where repeatability ensures that the result does not depend on subjective disturbances (bias), but that the "domain is closed" (Christensen 2002) and the measurement precise. Within this area of test research, it is stated that reliability may be difficult to attain when people's academic skills are the object of the measurement (see Thorndike, Cunningham, Thorndike & Hagen 1991). In the context of this approach to testing, the research interest is directed towards the development and discussion of the test as an objective, precise, and fair measuring instrument (Bachman & Palmer 1996), and towards development of meta-analyses of quantitative test results, in order to provide valid comparative statements (Allerup 2015; Kreiner 2009), or to establish possible causal connections between test results and social conditions and backgrounds (Egelund & Nielsen 2012).

The second paradigm we identify directs research attention to "the influence of testing on teaching and learning" (Alderson & Wall 1993). This research interest, often referred to as "washback," has typically focused on measurable

---

[1] This research overview takes its point of departure in research on testing of language and literacy

educational effects of using a particular test. According to Cheng, Watanabe & Curtis (2004, p. 4), the basic assumption within the framework of this paradigm is that, "tests or examinations can and should drive teaching, and hence learning." The test is seen as an educational technology to enhance "efficiency" and/or generate "improvement" of teaching and education and as potentially beneficial – both for individuals and for society as a whole (Pettersvold & Østrem 2012). Based on this assumption, washback is typically analyzed as either positive or negative washback (Alderson & Hamp-Lyons 1996; Alderson & Wall 1993; Cheng 1999: Watanabe 1996; Ryumon 2007). Negative washback is described as the phenomenon seen when, for example, a test is based on a narrow definition of "language skills," and therefore has, or is expected to have, a limiting or negative effect in relation to teaching and learning outcomes (also referred to as "teaching to the test"). In contrast, positive washback describes a situation in which a test results in "good" teaching practices that have, or are expected to have, a positive effect, such as when an oral test is introduced in language subjects with the expectation that it will encourage a greater focus on spoken language in teaching and hence improve measurements of students' oral language proficiency (Taylor 2005). It is central to the research on washback that, at its core, it is a paradigm where the test is seen as an instrument that has the potential to improve a teaching situation by creating "measurement-driven instruction" (Popham 1987). As Alderson & Wall (1993) highlight and critically discuss, it has previously been argued that washback could be linked to validity in the sense that a test could be considered valid if it promotes positive washback in teaching.

Thirdly, we identify a paradigm in which the test is considered as a power and management instrument, from a political and societal perspective. The test is seen as part of political decision-making processes and as the result of the implementation of specific global and national strategies (Spolsky 1995; Shohamy 2001, 2006; Menken 2008). This field of research typically applies sociological theory, including the work of Bourdieu, Bernstein, and Foucault (Shohamy 2001; McNamara & Roever 2006). Within this paradigm, inscribing and analyzing the test in relation to discursive struggles over social values and norms is central (Shohamy 2006). Research is first and foremost directed towards the logic and values that are embedded in the political processes surrounding testing and towards exploring the more general (intended as well as unintended) societal and educational effects of the test, including their function in relation to categorizing and constructing social identity (McNamara 2005). Thus, it is typically the effects of the test as part of broader political and societal contexts which are the focus of the research.

The three paradigms briefly outlined above represent very different theoretical approaches to test research, but, at the same time, have one common characteristic: in all three paradigms, the test is viewed as an instrument which functions in a given social structure - as a measuring instrument, as a pedagogical instrument, or as a political instrument. The three approaches each contribute important knowledge about testing, but at the same time, the common understanding of tests as instruments results in a tendency to overlook the interaction between tests and actors and therefore to overlook the space for individual interpretation and the role of norms and values in local test practices (Lissovoy 2013; Holm 2007). As with the research on washback effect, we wish to focus on how testing influences teaching. Alderson & Wall (1993) critically discuss the term washback effect, arguing that its dichotomistic and deterministic nature needs to be explored. The concept of washback presupposes that testing will more or less automatically influence teaching in either a positive or negative direction, depending on either the quality of the test or the focus of the research. In this article we wish to open up and nuance the concept of washback. To this end, we have chosen firstly to analyze a lesson intended to prepare pupils for a test in Danish literacy and secondly to examine teachers' constructions of cleverness after their pupils have participated in a test in Danish literacy. The first analysis is conducted within a critical psychology framework and the second in an ANT framework. Both focus on the interplay between the test as materiality and the network/community of actors/participants and on the construction of cleverness in this interplay.

Inspired by Sørensen (2006), Forchhammer (2006), and Olesen & Markussen (2006), testing is understood as a technology that materializes in different ways (the computer screen, the things used for practicing test etc.) and is attributed with some kind of intention and with meaning-making constructions that are processual parts of social practice.

We will contribute to the body of research on testing and to the related theoretical discussions by examining the local interaction between tests and actors. We will now turn our attention to a further exploration of this multidisciplinary approach to testing.

Critical psychology as a theoretical perspective for analyzing testing

In our first analysis, we draw on the Danish/German school of critical psychology, developed with strong references to the work of Leontiev, Vygotsky and Marx, among others. Critical psychology was developed at Freie Universität in Berlin and was since the publication of Klaus Holzkamp's "Grundlegung der

Psychologie" (Foundation of Psychology) in 1983 also known as the 'Science of the subject' (Nissen 2000). Critical psychology was developed as a critique of what is referred to as the "control sciences". Among these control sciences are approaches to psychology that naturalize certain properties of human beings – for example, intelligence - and thereby reproduce social inequalities (Dreier 1979). The aim of the subject science is to develop adequate knowledge of the meaning of societal conditions explored from subjects' first-person perspectives.

Holzkamp (2013) distinguish between conditioning discourse and reason discourse. Conditioning discourse takes its point of departure in the understanding that human beings are only subject to certain conditions and that it can be presumed that human beings act according to these conditions. On the other hand, reason discourse opens up for an understanding of human beings as acting subjects whose actions are to be understood connected to the societal and socio-material conditions of which they are part and connected to the meaning that human beings ascribe to these conditions in everyday life; hence first-person perspectives (Holzkamp 2013b; Schraube 2013).

Critical psychology is developed with the ambition of creating a psychology linked to the historical, materialist dialectics of Marxism and identifying basic categories with an analytical and methodological grounding in the subjective as part of societal practice. As part of this approach, everything is considered in its development, and the material (rather than the ideal) is understood as primary (Dreier 1979, p. 10). In continuation of this, Marx' 1845 theses on Feuerbach (Marx 1973) as a new materialism has been an inspiration in the development of Critical psychology not least its methodology (practice research) (Nissen 2000). Marx agreed with Feuerbach on being critical towards religion. However, he considered Feuerbach's critique to be far too abstract, not considering the religious mind as part of a specific societal formation (Marx 1973, the seventh thesis). The theses on Feuerbach directs further criticism of any abstraction, mystification, individualization and/or isolation of human beings and of their products; For instance, knowledge production and categories are to be understood as part of acting in practice (Marx 1973; Juul Jensen 1999; see for instance Højholt 2011 taking on this aspiration regarding children's perspectives). That is to say, psychological processes are not viewed in isolation, but as psychological aspects of social practice, and as aspects of the necessity of development, of which people are a part. Critical psychology's unique contribution (of which the theses on Feuerbach has provided a stepping stone) is to transcend the dichotomy between the individual and the social by analyzing reasons for actions as grounded in social structures and conditions (Osterkamp 2009, p. 168).

Within critical psychology, the concept of communities is a form of specification of the concept of social practice (Kousholt 2006). The concept of community contains a dialectical duality that refers to participation in something common while at the same time implying a dynamic action - it is something we constitute communally. Dorte Kousholt (2006) also clarifies that it is crucial to note that the concept of community does not connote agreement or harmony, but is, rather, a perspective that implies that human beings create conditions for each other, and that these conditions may be conflictual as human beings have different standpoints and, therefore, different perspectives on what is common (see also Axel 2011 on conflictual cooperation as anchored in praxis). Critical psychology is a subject science, which implies that considering human beings as only subject to social conditions is highly reductive, disregarding human agency (Holzkamp 2005; see also Marx 1973, the 9th, 10th and 11th theses on the need for considering human agency, the interconnection between subject and society and the aspiration for changing society through human agency). In the analysis below, the meaning-making of testing will be analyzed as part of a complex practice and from different participants' perspectives; it will be shown how subjects also act with the technology of testing in creative ways (e.g. Kousholt 2016b). Furthermore, the concept of learning as connected to the distinction between conditioning discourse and reason discourse presented above will be discussed in the analysis.

Based on the theoretical framework of critical psychology, subjects should not be understood as isolated individuals, but as "social subjectivities," whereby the dichotomy between the subjective and the objective is transcended (Nissen 2002, p. 70). Related to the ambition of transcending the dualism of the individual and society starting from the first-person perspective, Ute Osterkamp (2009, p. 170, with reference to Holzkamp 1983) reminds us that the contradictions of social conditions are not external to human beings, but run straight through us. As such, critical psychology is not a psychology that looks into human beings, but a standpoint that, together with the relevant participants, explores the meaning of the generalized capacities of things as possibilities to act in concrete practice. Hence, meaning is part of acting with things in the world (Holzkamp 2013b, p. 279; Forchhammer 2013).

Critical psychology as a subject science is based on first-person perspectives on historical and socio-material conditions, and on reasons to act with them, in them and on them (e.g. Schraube 2013). Socio-material conditions, such as the reproduction of the capacity of a thing or a technology[2], is considered

---

[2] In the analysis, we use the concept 'technology' instead of 'thing' as we like Forchhammer (2006) find that the former can imply materiality, strategies, and situated and transformative meaning-making processes.

as generalized possibilities to act and become a meaning constellation of the generalized capacity of the thing/technology (Holzkamp 2013b, p. 279). Holzkamp (2013b p. 279) point to Leontiev's classic example of both the practical value and the generalized and reproduced meaning of the axe's capacity of axing as an embodied idea objectified in the axe. This is crucial, because things are more than their properties (Olesen & Markussen 2006). However, we must not forget that possibilities to act are not freely given possibilities; power structures, obstacles, and restrictions are part of the relation between subject and material things in the world (Holzkamp 2013b, p. 280). Schraube (2013) focuses on the power of technologies to avoid the risk of perceiving them as merely a means to an end. Sometimes, technologies unintentionally become an end in themselves and it is important to be aware of the transformative power of technologies for human beings' sociality. Material things must be understood as part of practice and as part of subjects' creative, transformative, and/or reproductive actions with the things (e.g. Costall & Dreier 2006; Forchhammer 2006; Holzkamp 2013b). As Costall & Dreier (2006 p. 10) state, this underlines:

> "[…] the need to understand things not in isolation but in relation to the human activities in which they are involved, and the different, even conflicting, meanings things can have for the people involved with them."

It is this need that our analysis of testing from a critical psychological perspective will address. Namely, how subjects act with and around the materiality of testing and how this materiality becomes part of teaching in unpredictable ways and part of subjects' constructions and negotiations of cleverness.

Actor-network theory (ANT) as a theoretical perspective in analyzing testing

Law defines ANT as "a disparate family of material-semiotic tools, sensibilities, and methods of analysis that treat everything in the social and natural worlds as a continuously generated effect of the webs of relations within which they are located" (Law 2009, p.141). Thus, ANT is not an abstract theory that aims to explain why something happens, but instead a wide array of research approaches that, on the basis of case studies, make it possible to describe how "relations assemble or don´t" (Law 2009, p. 141). As a result, there is no clear distinction between theory and method in ANT, because the theory in itself represents a general method that must be operationalized in relation to the specific research questions that are formulated (Arnoldi 2003). Besides the central unit of analysis

– the actor-network – there is a broad variation in concepts and categories used in ANT-inspired research (Schraube 2013).

Actor-networks are dynamic entities, characterized by the fact that actors and networks constantly constitute and redefine each other, because everything may be considered both an actor and a network – it is simply a matter of perspective. As Callon puts it, an actor-network is:

> "reducible to neither an actor alone nor to a network… An actor-network is simultaneously an actor whose activity is networking heterogeneous elements, and a network that is able to redefine and transform what it is made of." (Callon 1987, p. 93)

Latour terms the interaction between human and non-human actors in the network "translation" and he defines this concept as "the work through which actors modify, displace, and translate their various and contradictory interests" (Latour 1999, p. 331). Latour's argument for this basic assumption is that objects may play a central role as actors in a network, through the meaning people delegate them and that, additionally, objects may mediate interaction and create meaning that is connected to other places and other times, thereby exceeding "the limits of the local" (Brandt & Clinton 2002). Actors are defined by Latour as "entities that do things" (Latour 1992, p. 241).

It should be emphasized that Latour's theorizing does not reject conceptual differences between human and non-human actors. His juxtaposition of human and non-human actors is rather an expression of a research strategy based on "the principle of general symmetry", which, according to Latour, "simply means to not impose a priori some spurious asymmetry among human intentional action and a material world of causal relations" (2005, p. 76). Thus, the principle of general symmetry is an expression of openness in relation to analyzing the actors involved in a network's processes of translation, without giving priority to certain actors beforehand. Reckwitz argues clearly and precisely for this theoretical standpoint:

> "…objects are necessary components of many practices- just as indispensable as bodily and mental activities. Carrying out a practice very often means using particular things in a certain way…When particular things are necessary elements of certain practices, then, contrary to classical sociological arguments, subject-subject relations cannot claim any priority over subject-object relations, as far as the production and reproduction of social order(lines) is concerned." (Reckwitz 2002, p. 252-253)

Actor-network theory seems to be particularly appropriate as an approach to research on testing because it makes it possible to direct research attention to how, through translation processes, human and non-human actors construct what might be called "calculative spaces" (Fenwick & Edwards 2010, p. 117) – spaces in which the construction of cleverness is central for the translation process - or carry out what Hamilton refers to as "ordering projects" (Hamilton 2001), in which actors seek to give something a certain hierarchical order.

Theoretical similarities and differences

The two theoretical perspectives presented above are, to a certain extent, complementary. Firstly, both are scientific traditions characterized by an open and investigative ontology and presented as critical counter-theories to prominent scientific traditions within psychology and sociology, respectively. Secondly, there is a common understanding of "connectedness" as something that is created through interaction and is locally and socially situated, although different concepts are used to describe this interrelatedness and connectedness: respectively, "network" and "community."

ANT and critical psychology share a methodological orientation towards ethnographic research designs in which long-term fieldwork plays a key role. They likewise both view materiality as something that must not be taken for granted as a "thing", but viewed as something created by people, and as something that "returns" and co-creates people through an interactional meaning-making practice (Nissen 2002) where people can delegate responsibility and roles to objects. This focus on that the returning of materiality is central in our analysis of the technology of testing as leading to unforeseen aspects of practice. We find that both theoretical perspectives transcend the dichotomy between subject and materiality, albeit in different ways. Whereas critical psychology insists on first-person perspectives in order to understand reasons for acting in and with certain socio-material conditions, ANT rejects such a privileging of the human, considering humans and non-humans as equally interesting in an analytical perspective. The role of testing as materiality that co-constructs cleverness will be central in our analysis.

As suggested above, there are also differences between the two theories. Indeed, they may be said to represent different standpoints in relation to anthropocentrism, and in relation to epistemology. Critical psychology has the intention of presenting conscious and explicit societal critique, as well as an emancipatory ambition through the analysis of different ways in which social inequality is (re-)produced. Critical psychology may also be characterized as a

science of the subject, with a focus on inter-subjectivity, and on generalized societal agency, studied from a first-person perspective (Holzkamp 2005), where materiality is also a part of subjects' dialectical conditions. Therefore, critical psychology can be understood as a theory of the mutual constitution of subject and society (Kousholt & Thomsen 2013). The epistemology of ANT, meanwhile, is first and foremost directed toward the processes "in which networks combine in particular ways to produce objects, knowledge, or facts through the displacement or suppression of dissenting voices, or of those facts unfit to fit" (Clarke 2008, p. 152). The analytical and critical dimensions of ANT focus first and foremost on how processes of knowledge production are part of the creation of inequality and social distinctions, and on how knowledge production must therefore be analyzed in relation to power. Latour describes this relation between power and knowledge as follows: "What is called knowledge cannot be defined without understanding what gaining knowledge means" (Latour 1987, p. 22). ANT directs attention to the analysis of the content of differentiations. In an analysis of testing, it is important to pay attention to this dimension because an analysis of content dimension shows us which forms of knowledge are socially valued and demanded of students at a given time within different school subjects. A new test of literacy, for instance, appends certain priorities and values to an existing assessment culture and must be seen in this light.

Despite the differences between the scientific traditions of critical psychology and ANT, we do not believe it impossible to combine the analytical perspectives they offer. On the contrary, we believe that such a combination can enrich the science of testing and be an important supplement to existing research on testing.

Combining the theoretical perspectives offered by ANT and critical psychology with the intention to analyze testing directs analytical attention to how testing becomes part of teaching and how testing co-constitutes different constructions of cleverness in certain communities/networks. With these theoretical perspectives, we shift our analytical lens from either positive or negative washback effects to how testing becomes part of everyday school practices. A common criticism of tests with a long history (as highlighted by Hanson 1993) points out that the introduction of tests affects teachers' organization of lessons, narrowing the curriculum, impeding creativity, and letting the tests determine content ("teaching to the test"). We find this criticism important as a generalized understanding, but we also see the need to challenge a deterministic understanding of tests as instruments that create specific practices. Exploring the constructions of pupils' cleverness in the social community of the classroom and in the testing network will demonstrate the complex localization of the test in the everyday life of pupils and teachers.

We would like to emphasize that we do not see our multi-disciplinary approach to testing as a new paradigm that can or should replace other research paradigms. We acknowledge the need to ground our research and practice in theory, but at the same time find it important not to let theory turn into grand narrative considered beyond critique or to understand theories as isolated entities with no need for explorative conversations with other theoretical orientations. We believe critical research on testing must understand theory as an ever-changing framework, rather than a fixed point of reference. A multi-disciplinary research approach is aligned with this understanding of theory. Hence, this article should be understood as a step on the path to integrating theoretical perspectives analytically.

Methods and methodological considerations

The material that forms the basis for this article comes from two different research projects on the implementation of national tests in Danish schools. The first of these projects is a post-doc project that examined national testing as part of teachers' and pupils' school lives in the years immediately following the implementation of national tests. This project focused on how national tests become part of pupils' communities and learning processes. Based on a critical psychological perspective, implying a focus on connections between conditions such as tests and the meanings participants ascribe to them, attention was primarily paid to children's participation in, and perspectives on, national tests. This was investigated by following children using anthropologically inspired participant observations during lessons, recess, and test situations, supplemented by interviews conducted both before and after participation in national tests. The pupils' teachers were also interviewed, but the primary focus was on the meanings the children ascribed to the tests. Four socio-economically diverse schools were selected and different age groups were represented in the project (two 2nd grade classes, one 6th grade class and two 8th grade classes distributed at the four different schools).

The second project is a longitudinal study (2009-2015) focused on how the national literacy tests are embedded in existing testing practices and on how the national tests change existing testing practices regarding literacy. This research project draws on Latour's ANT and focuses on translation processes within the network: how "poor" test results in literacy are constituted as a problem to be solved; how actors enter into the resolution of this problem; how different roles are defined in relation to each other; and the extent to which the different actors are transformed "into manageable entities" (Clarke 2008). Attention was first and

foremost focused on the teachers as actors. The study was based on anthropologically inspired participant observations of test practices, test preparations, and the use of test results in school-home conferences. Additionally, interviews were conducted with teachers throughout the longitudinal study regarding literacy and literacy testing. The material in this study was generated by three classes at three schools from grade two to four in three major Danish cities. The three schools are all located in so-called "socially deprived areas", and the percentage of pupils with ethnic minority background in the classes is between 50% and 90%.

The two research projects share an approach to research on testing inspired by ethnography and both, to different degrees, carried out longitudinal studies based on a belief that long-term classroom presence is necessary to investigate, in a nuanced way, how testing becomes part of everyday life. In order to investigate children's perspectives, a certain amount of knowledge is required of the children's participation in the classroom, of their positions, and of how the conditions look from their perspectives. A focus on how new tests change existing testing practices, which is central to the ANT-inspired study, likewise requires a longitudinal research design. The research focus of both studies is not the test itself, but how it becomes a part of everyday school life, and this requires a research-based understanding of the participants' or actors' everyday lives because we cannot know from without, or beforehand, how "new" tests become part of everyday school life.

In the following analysis, we have firstly tried to analyze the material on the basis of our theoretical standpoints in order to be able to contribute with knowledge of how testing becomes part of children's and teachers' everyday lives. Secondly, the analysis is intended to be exemplary, highlighting how the chosen theoretical perspectives may contribute to an alternative approach to research on tests. To this end, the analysis presented here is not an exhaustive analysis of the totality of material, but instead presents a number of examples illustrative of the theoretical approaches and arguments.

Analysis with a focus on communities and constructions of cleverness

From the approach of critical psychology, there is a focus on the social practice of which the tests become part and how they become meaningful/non-meaningful in this specific practice. Testing is considered to be part of conditioning structures, which at the same time have been constituted at other times and in other places. This means that the tests may also be regarded as part of a social

practice that is carried out, reproduced, and transformed in concrete action contexts (e.g. Nissen 1996).

Such a theoretical perspective adds to test research an alternative understanding of tests, test performance, test situations, and constructions of cleverness. Instead of using tests themselves as the jumping-off point for analysis, a critical psychological perspective analyzes the social practice of which the test is a part. Furthermore, practice is conceived as ambiguous, conflictual, and composed. In the following analysis, we explore how parts of the Danish national tests in literacy (concerning proverbs and sayings) become significant for classroom practice. Testing is understood as a technology that materializes in different ways and that can be reproduced and transformed in practice (see above). The question at the center of the following analysis is: How does the technology of testing become part of the community's construction – and negotiation - of cleverness?

Before testing: How the technology of testing materializes and becomes part of school practice with respect to the construction of cleverness

The Danish national tests are computer-based, adaptive, and have a multiple-choice design. The children perform the test individually in a room with their classmates. In this way, the test is a special materiality that imposes certain possibilities for action (Holzkamp 2013b). The children are supposed to read text on the computer screen, engage in the text, and tick what they believe to be the one correct answer to the question. There are intentions with the materiality (e.g. Forchhammer 2013; Schraube 2013) and these intentions are directed towards finding the children's level of cleverness in order to differentiate the teaching in this regard. So how does this materiality become part of school practice outside the concrete test situation? The first section below will focus on the imitation of test practice in lessons where tests are rehearsed while the second section will focus on the transformation of test practice in these same lessons.

Imitating test practice

In one of the 2nd grade classes from the empirical material, the teacher has introduced a new topic on proverbs and sayings, as this is part of the test in literacy which the children are supposed to take some weeks later. The school is located in a socioeconomically privileged area. It is a school that focuses on maintaining a high academic level. In an interview, the teacher of the 2nd grade class explained that, because of the perception of the school as high performing,

he experiences the test as a source of pressure. Prior to the national test in literacy, the teacher explained to the researcher that he was going to practice the topic of sayings before testing. He said that he doubted whether it would help the children and that it is mostly these types of question that the children would find difficult. The teacher also said that he thought that sayings were something that the children might learn at home; if the parents often used sayings at home, it would probably be a little easier. He expressed doubt as to whether it was the school's task to teach children sayings, but he would nevertheless spend time on the topic. The teacher and the researcher talked about how the meaning of some sayings seemed to change over time. The teaching about sayings was directed towards the national test in literacy as test rehearsal. Despite his doubts, the teacher chooses to work with the sayings in his teaching.

The classroom is located on 2nd floor, there are huge windows pointing at the school yard. The room is square-formed. The teachers' desk is located in the same end of the room as the door (the front of the room, one might say). Beside the teacher's desk there is a whiteboard that can both be connected to the internet and can be written on (while rehearsing sayings, the whiteboard is not connected to the internet but functions as an old-fashion board). The walls are richly decorated mostly with drawings and posters that have an academic or a social message (e.g. social rules). There are four seats at each of the children's tables. They are placed in such a way that most of the children are not facing the teacher directly, but sideways. In the classroom, the teacher and the children talk about the Danish saying "to get cold feet". The teacher says that now, the children must sit for a moment and think carefully, and try to see whether they could contribute something regarding the meaning of this phrase. Several of the children (especially the boys, which was also observed to be the case in other lessons in this class) raise their hands immediately. The teacher says that everyone must take time to think about this, as they all have to do the test soon. In this way the teacher introduces a certain relation between the test, the children and the test practice in the class room. He introduces the test as a goal for the teaching in sayings and this seems to have implications for the lesson further on.

The teacher starts by asking Mikkel, who did not raise his hand. Bo says that he has raised his hand and asks why he was not chosen. The teacher says that it is all well and good, but that he will also ask those who have not raised their hands for their thoughts. Mikkel says that he did not really have time to think about anything. The teacher asks some of the other children, who offer a number of different answers. The teacher writes these answers on the board. One of the answers is "a guilty conscience." The teacher once again asks one of the children who do not has raised their hands. The child does not have an answer. The teacher asks what they are going to do during the test when they are confronted

with a saying. He says that it is important to try. One of the girls, Maria, begins to talk about other sayings, but does not provide an answer to the question of "cold feet". She contributes with the saying "hand on heart", among others. The teacher acknowledges Maria's contribution and says that she and another girl have studied sayings, and Maria agrees (Maria receives extra language training where sayings, among other things, are taught, which she also talks about in an interview). Some of the children raise their hands, making suggestions that are already on the board. There still are some boys with their hands raised, and the teacher has now written several answers on the board. A boy says "to be nervous," which immediately prompts two of the other boys to cry out "uh, that's what I was going to say". A third boy says "becoming afraid", adding that it is almost the same. The teacher writes "afraid" on the board, just under "nervous". The two boys who gave the impression of also wanting to say "nervous" keep their hands raised. The teacher asks them, and they say "nervous". The teacher says that he does not want to ask any more children. He starts to cross out the answers on the board that are not correct. Some say "boo" (when their answers got crossed out), others say "YES" (when their answers were still present). Finally, only "nervous" and "afraid" are left on the board. The teacher says that if you have to choose between them, it is probably "nervous", crossing out "afraid". The three boys who answered "nervous" cheer and say aloud, "that was what I said", and "me too".

Before the lesson on sayings, the teacher expresses some doubts regarding teaching something that he thinks should be learnt at home. At the same time, the materiality of testing, the test items on sayings and proverbs, introduces the topic to school practice, thereby becoming something the teacher has to relate to because all the children face the same test regardless of how their parents might use sayings in their everyday communication at home. In this way, the materiality of the test items on sayings becomes part of the teaching practice even though the teacher is not sure that this is a relevant task for the school to teach. Following Holzkamp (2013b p. 279-281), one might say that the generalized and reproduced meaning of testing; the idea that individual cleverness should be identified and differentiated through the use of test items such as sayings and proverbs, is part of creating conditions for possibilities to act in the classroom. These possibilities are not freely given but part of power structures and conditions (the national test are compulsory and national; all children shall take them) that the teacher does not see alternative possibilities to act than to rehearse testing. However, these power structures seem to disappear and in this way the teacher seem to have freely chosen this practice of rehearsal despite his doubts regarding this (ibid., p. 281). Furthermore, as he introduces the teaching (test rehearsal) as means to and end (testing), he is also part of

creating/reproducing certain meaning constellations about individual cleverness that is to be tested (ibid.).

In the teaching practice, the materiality of the test on the computer screen is not directly present in the classroom. From the teacher's first-person perspective, the test offers different possibilities for action. He could, for instance, choose not to teach his students about sayings, refusing to "teach to the test". However, these possible lines of action are not innocent or without consequences. As the teacher expresses, not practicing these test items might mean that the children would be unprepared for the upcoming literacy test. As stated, the teacher stated that he felt a degree of pressure that his pupils perform well on the test and the materiality and content of the test thus become powerfully constitutive in the classroom. In the classroom, the technology of the test does not materialize as a computer screen presenting different test items; instead, the materiality consists of a whiteboard and a pen, among a lot of other materiality in the classroom (e.g. children sitting on chairs in groups at four at one table, not all directly facing the teacher). This materiality constitutes other possibilities to act and other possibilities in the construction and negotiation of cleverness.

The next day, the teacher is once again working on sayings with the class. Again, there are a lot of boys with their hands raised and only a few girls. The teacher asks a question of a girl who has not raised her hand. She does not know the meaning of the saying "to have hair on your chest". The teacher asks whether there is anything she can rule out and, in line with what the teacher has told the children about these items, she says that it does not mean to actually have hair on your chest; that is, it does not have anything to do with either hair or the chest. In this way, the children learn testing strategies that are useful for this test item, which involves knowing that sayings do not have a literal meaning, but are metaphors for something else. In this way, the possible correct answers to a multiple-choice test are reduced, improving the likelihood of answering correctly. The children seem to understand this strategy, whereby the number of distractors is reduced (and also, they might learn something about sayings in general). In this way, it is not certain that they have a better understanding of the concrete saying, but they will most likely be more competent to answer this type of test item correctly. After the children had taken the test, the teacher stated that the class has done well in this category.

In the classroom observations above, the teacher writes a saying on the board, and then the children try to guess or answer what it means. The lesson in sayings could be said to simultaneously both imitate and differ from a test context. The imitation of the test context may be identified through the teacher's questioning of pupils who did not raise their hands, because he wishes to ensure that those who do not have their hands raised also reflect on the sayings. His

reason for doing this is that they will all be tested. The test makes all the pupils testable and asks all the students at the same time, regardless of whether they want to be tested or not (see for instance Kousholt & Hamre 2015 on the pupils as testable). In this way, the teacher imitates the logic of the test and, at the same time, clarifies the demands of a specific item. He uses the test to motivate the children to think about sayings, asking what they are going to do in the test situation when they are confronted with a saying, and he argues that it is important to try. However, it is not only important for the children to try in the actual test situation, but also during the lesson imitating the test context.

Another imitation of the test context is reflected in the fact that the teacher brings together pupils' answers and then performs a kind of multiple-choice process, crossing out the wrong answers (distractors). In this way, he shows a test strategy that is reminiscent of the process of elimination. The boys' participation in the teaching situation is also an imitation of the test context. It seems to be more important for them to demonstrate their own cleverness and the correctness of their answers than it is to contribute to the overall knowledge regarding the possible meanings of sayings (this is also evident in other observations in this class, but in particular in the work with sayings). This is seen when some of the children persist in answering "nervous", irrespective of the fact that this answer has already been given and written on the board. In a test situation, it makes sense that all choose the same test answer if it is the correct answer because testing often operates with one correct answer (multiple choice format). However, this does not contribute to the collected learning processes of the classroom.

To understand learning we have to understand the subjects' reasons for learning as Holzkamp (2013a) emphasizes. This follows the distinction between conditioning structures and reason structures as presented in the theoretical section above. In crucial areas this teaching cannot be seen as a learning context – at least not expansive learning. Holzkamp (2013a) makes a distinction between defensive learning and expansive learning. Expansive learning is a process that depart from one-sidedness, fixations and narrow goal-directedness (ibid., p. 125; see also Schraube & Marvakis 2016 for interesting discussion). It should be noted that the conceptual pair restrictive/expansive learning should not be used to evaluate other's learning but as analytical categories that can help gaining knowledge on reasons for learning and not for learning which are connected to the conditions for learning. Following this, the test rehearsal could instead be viewed as an action problem that can be overcome without learning (Holzkamp 2013a). For both the teacher and the children the test rehearsal does not seem to be a meaningful practice of learning but instead as a mean to an end – a

possibility for the children to handle the items on sayings and proverbs in the test. This could be illustrated with the following quotation:

> "I am not interested in learning something (I cannot see the use of it), but in coping with particular demands by learning in such a way and just as much as is needed, I am largely other-directed; by demonstrating learning and learning success I smoothly accommodate myself to expectations imposed upon me by third parties." (Holzkamp, 2013a, p. 124-125).

This quotation illustrates very well some of the aspects in the observation above. Both the teacher and the children seem to deal with "extraneous learning demands" (ibid. p. 123) as an action problem; cleverness is practiced as individualized as the individual's correct/in-correct answers to single items.

On the other hand, this is not the whole story to be told; the children do seem to gain new knowledge on sayings and how to understand sayings and many of them participate in the teaching in an engaged manner. This knowledge and this engagement could be part of new action contexts where the children could follow new learning problems. In such ways, the conceptual pair restrictive/expansive learning can also be seen as fluid, open and moving (see also Schraube & Marvakis 2016, p. 212-213 on this).

In this way, the technology of the test materializes in different ways in the classroom than in the test situation; however, some of the intentions implied by the technology seem to be reproduced by the participants and it becomes a powerful constituent in this classroom.

Following the theoretical discussion above, it can be stated that a test with a multiple-choice design does not impose multiple choices for action. It instead imposes more restricted possibilities of action as it has to do with the children's relation to a screen and their relation to a set of questions to which they have to tick the right answer. This multiple-choice format is imitated in the classroom, where the sayings are used as examples to teach the children how to correctly answer items on sayings in the test situation. In this way, the technology of testing could seem to impose an either positive or negative washback, depending on one's perspective, and it could seem as though testing becomes a goal in itself; learning to do the test right. At the same time, the technology of testing as part of a complex practice also offers other possibilities to act, as the next section will show.

Beyond washback in practice

This section focuses on how the teaching situation is not only an imitation of a test situation. This is partly because, despite the initiatives to mimic testing, a

sense of community regarding the content of the teaching is constituted in the class during teaching. In the observation above, the community and the meanings of the sayings seem to engage Maria in the lesson in ways the researcher did not observe in lessons with other content (in the following days, where the class continued to work with sayings, Maria was similarly engaged). This engagement is due to the fact that Maria is an expert in this subject, as she has worked with sayings in the school's language center, where she receives extra instruction. While it can be regarded as a weakness to be in need of extra support in other contexts, in this case it becomes a strength for her.

In the teacher's explanation of the meanings of sayings, he draws on his knowledge of the children's lives outside school. He uses one of the boys, who has been winter bathing with his father, as an example. The teacher says that this requires "hair on your chest". As already mentioned, the empirical material shows that there is a big difference between the apparent levels of engagement in the topic among the girls and boys. The substantial gendered difference in hand raising reflects a special way of being a boy pupil or a girl pupil in this class. The teacher attempted to even out the difference by asking the girls when they raised their hands, and the boys sometimes were not asked even though they raised their hands, thus also constituting a gender differentiation performed by the teacher. In this analysis, gender is not a systematic analytical category, but we can see gender differences in the material. The high level of activity in this class also becomes gendered, active participation, with the boys most visibly active. Powell, Danby, and Farrell (2006) find that girls are competent in covert activities (passing notes discretely) in the class without attracting the teacher's attention, whereas the boys pass notes and other activities in overt ways. This also gives the boys an opportunity to form relationships with their male teacher (ibid.).

Something similar might be at stake in this material where the boys also are those who seem to have the opportunity to be active in a visible way, forming a connection with their male teacher. In the example the teacher simultaneously draws on his knowledge of one of the boys´ winter bathing, introduces a saying that responds to this knowledge in a respectful and engaged manner and ascribes a certain understanding of masculinity to this form of father-son activity. On the other hand, and due to the materiality of testing where sayings are among the test items, Maria seems to transgress this gendered participation and become more engaged than otherwise observed. The materiality of testing and the teacher's test rehearsal offers Maria new possibilities of action in this lesson where it is possible for her to become expert on sayings. In relation to this, Schraube & Marvakis (2016) point to that some fluidity between teacher- and pupil positions can be fruitful so that these positions do not become unambiguous and frozen.

The fluidity or movement can be found in small everyday displacements for instance Maria's new action possibilities as part of the conditions reproduced and created in the test rehearsal context (see also Kousholt 2016a for other examples as part of "expansive assessment").

In another observation, where the class was also working with sayings, some of the students gradually seem to become so certain of what the topic is about that they experiment with the form. For example, one of the boys says that "to get cold feet" means to cut off one's feet and put them in the freezer. And then he says that it of course does not mean this. A boy comments on another boy who does something the teacher does not appreciate, saying that he "swims against the tide". In the latter example, the boy uses the test training to reformulate an action that the teacher did not appreciate as a positive action. To 'swim against the tide' has a positive meaning of taking an active and conscious stance against what is commonly accepted as the norm. One could say that he negotiates and transforms the meaning of accepted school participation using the technology of schooling in the sense of differentiation of cleverness. In the first example, the boy does exactly what the teacher tells the pupils not to – to take the sayings at face value. Instead, sayings mean something else that one needs cultural access to understand. The boy knows the two different meanings of the saying and plays with them in a humoristic manner and in this way, it seems that they (momentarily) expand their possibilities to act, developing agency, and that the test rehearsal also becomes something else. Following Willis (2002), one might say that he is doing "split epistemology" as he plays with the sayings double meaning, creating a new and creative meaning of the rehearsal of test items. As Forchhammer (2006, p. 133) suggests: "The subject is seen as a mediating authority who transforms the objective sociohistorical action conditions into subjective meaning and concrete action." These examples illustrate that there are different ways to participate in test practice and that the children are also in possibility relations to the test, and at the same time that testing is a powerful technology to relate to (e.g. Holzkamp 2013b; Schraube 2013; Kousholt 2016a; 2016b).

In addition to this type of work with sayings, the children also work on a task where they are asked to find sayings, which is actually a task for 4th grade classes. The children work on this task in groups, helping each other both within and among the groups. However, there are also groups that refuse to help other groups because, as one boy states, his group should complete the task faster than the other groups. As they finish the task, a group of girls says that "now the exam papers only need to be submitted." As the examples show, the work on the tasks is characterized by a sense of community and helpfulness, but, at the same time, there are also test-like connotations of differentiation, speed, and individual

cleverness for some of the pupils. There seems to be a duality in the conditions of school- and test practice that the children are participating in, negotiating – trying to make sense of, reproducing and transforming test practice as part of everyday life.

During a recess, the researcher sits with the teacher in the staffroom. The teacher and the researcher talk about the lesson on sayings with a female teacher. The female teacher laughs, saying "totally teaching to the test." The male teacher blushes and looks embarrassed. He says that he does not think that is true. This exchange shows that some of the teachers distance themselves from "teaching to the test," and that test rehearsal is not something that they want to be seen as characteristic of their practice. The teacher gives the impression of not intending to "teach to the test," even though the observations show that there are test-like connotations and imitations in the teaching, partly due to the fact that the content reflects the test's content, and partly because the form occasionally imitates the test situation. As such, even though this test rehearsal seems to be a free choice of the teacher, this might not be a free choice after all. As stated above the teacher feels a form of test pressure and he wants to prepare his pupils properly for the upcoming test as he feels responsible for them having a good test experience. In this way certain possibilities to act can be difficult to even see and as Holzkamp (2013b, p. 281) states: "[…] the individuals are virtually swimming in a fictitious free space between real walls." At the same time, there is much more going on than mere test imitation. This diversity indicates that, if the teaching could be categorized as 'washback', then this is not an unambiguous washback practice. For instance, it is impossible to say if this washback is negative or positive as this depends on the first-person perspectives of the participants and their different positions in the community. When we follow the tests into everyday school life and see their meanings from participants' perspectives, we see that the test sets certain conditions for the participants; however, these are not to be understood as unambiguous 'teaching to the test', but also as potentially constitutive of new communities and new ways of becoming a child being able to teach about sayings and developing agency. In this way they can be part of new and other learning processes. At the same time, we see how teaching becomes test-like, not only preparing students for the content of the test, but imitating test practice in several ways.

The materiality of testing consists of the computer it is carried out on, of the content of the items, where sayings are but one content area, and of the multiple-choice format - to mention some. These contents are not directly present in the lessons above, but the generalized meanings of this kind of testing - which could be said to imply that there is only one correct answer as to what the sayings mean, as in the multiple-choice format, and the differentiation of pupils'

individual cleverness, - become constitutive powers in the lessons. In the lesson, the generalized meaning of testing becomes part of other materialities: the whiteboard, the teacher's pen, the children sitting on chairs at tables etc. The materiality of testing becomes "*dynamic elements* in a continuous flow of activities" (Olesen & Markussen 2006, p. 188) and, even though one should not neglect the power of testing in the rehearsal to testing, it is worth noting that the children respond to the rehearsal on the topic of sayings in different ways. Some play with the topic in humoristic and creative ways, and it becomes possible for a girl to enter a new position as a knowing and contributing pupil.

After testing: Analysis with a focus on practices of testing and teachers' constructions of cleverness

To build an empirical understanding of the meaning of the test that draws on ANT, our analytical attention in the following analysis is focused on the role and responsibility the teachers delegate to the national tests in literacy and on how "calculative spaces" are constructed and "ordering projects" carried out. From the analysis, it appears that the national tests result in an increase in calculative spaces and introduce a number of dilemmas for teachers, which shed light on conditions and opportunities for teachers' construction of pupil cleverness. These dilemmas relate to the use and validity of different tests and test results and to the validity of formal and informal constructions of cleverness. This analysis highlights that materiality is more than just a thing in the world. As Olesen and Markussen (2006) state, there is more to a thing than its parts and properties; it transcends its parts and properties and its meanings should be found in its contexts in the intermingling between materiality and people acting with it and around it (e.g. Olesen & Markussen 2006, p. 171-173): "Material things are never 'just material' […]. Materiality also comprises fictional constructions" (Olsesen & Markussen 2006, p. 189). In this respect, the fictitiousness should not be understood as 'pure invention', but as part of a construction of a social practice. The analytical question is: How do teachers delegate roles and responsibility to the national test as an actor in regard to the construction of calculative spaces and pupils' cleverness?

Construction of calculative spaces

The network thinking in ANT makes it possible to follow the effect of a test - a non-human actor – in relation to the processes and dynamics in the testing network, in which the construction of cleverness is central. One effect of the

introduction of the national tests was a considerable increase in calculative spaces. When the national tests were introduced, so-called demo-tests, intended to allow students and their teachers to practice prior to the 'real thing', were also made accessible to schools and used frequently in the weeks before the national test, hereby constructing a number of new calculative spaces in the everyday life of the school.

The technology of a test, and its materiality, relates not only to the test and the demo-test, but also to a test result. In relation to the national Danish test in literacy, it should be noted that "the teacher must communicate the test result verbally to the pupil," and that the test result "must be printed out and handed out to the parents" (Skolestyrelsen 2010, p. 10). In this way, the test and the test result are delegated power to construct a calculative space in the pupil´s home and to inscribe the parents in the evaluative network. The teachers were also required to comment on and discuss the test result with parents at the compulsory annual school-home conference, hereby constructing other calculative spaces. In a Danish context, an increased number of persons are thus inscribed as actors in an increased number of calculative spaces in relation to the test and its results, and, therefore, test and test results become parts of people's everyday lives to a greater degree than previously, thereby co-producing reality/conditions (Hanson 1993). The national test did not only affect the construction of calculative spaces, but also became an important – but also ambiguous – actor in the construction of pupils' cleverness.

Competing constructions of cleverness

In interviews with teachers, ambivalence emerges in relation to the results of the national literacy tests. In the following excerpts 1 and 2, two examples of this ambivalence are presented. In excerpt 1, the teacher responds to a question about why three different tests have been conducted in the 2nd grade class. In example 2, the teacher answers a question on the differences between informal constructions of cleverness and formal constructions of cleverness (Holm 2015).

Excerpt 1:

> *Anne: And the result from the national test is not an accurate picture of what she is capable of.*
> *Interviewer: No?*
> *Anne: And I believe that this is not possible in the test. But you can say that, when we have carried out two or three different tests - perhaps you can obtain at least a slightly broader perspective.*
> *Interviewer: Yes.*

> *Anne: That is how it is with the tests. They will be snapshots of the student*
> *at that exact point in time, right?*

It appears in this example that two ordering projects are at work at the same time: Partly a formal construction of cleverness, which is based on the national tests and other formal test instruments, and partly an informal construction of cleverness, which is based on the teacher's perception of the individual pupil's abilities. The opposition between the formal and informal constructions of cleverness is, however, quite complex. In example 1, it appears that different test tools, in each their own way, may contribute to the construction of cleverness. The basic assumption seems to be that test results may be multiplied, and the more test tools, the broader the perspective on the individual child. At the same time, however, the teacher questions this assumption by using the word "perhaps", and goes on to talk about the fragmented character of test knowledge by emphasizing that the test tools provide a "snapshot" of the student, and therefore appear as a construction of cleverness with only momentary validity, whereas the teacher's informal construction of cleverness is given more permanent validity and connotes a kind of "sure knowledge". In the example above, it also seems as though the teachers trust some test results, using them in their informal construction of the pupil's cleverness. In that respect, teachers' informal construction of cleverness may be an assemblage of both formal constructions of cleverness and informal constructions of cleverness. The more permanent validity given to the teacher's informal construction of pupil cleverness is clarified in the excerpt below, where two teachers answer a question about whether they were given cause to question the correctness of the national test results.

Excerpt 2:

> *Birgitte: So, I have to say that the results of the national test didn't*
> *surprise me.*
> *Anne: And, um, there is, of course, nothing that comes - there is, of course,*
> *no great surprise for you, because you know the test, and you know what*
> *the pupils can do.*

The interviews with the teachers clearly show that two different "ordering projects," in the form of formal and informal constructions of cleverness, exist simultaneously. One is described as a "snapshot" anchored in the test material, and the other as broader and representing the teachers' knowledge about the student (which may well be *partly* based on tests). The formulation in excerpt 2 is interesting because it appears to be a strong and indisputable indication of consensus, and a naturalization of the point of view that of course teachers know

their students' levels of academic cleverness. The use of the word "of course" indicates that all interlocutors in the interview are expected to accept the validity of the teachers' informal constructions of the pupils' cleverness. This can be interpreted both as an implicit and a more general critique of tests as representing valid knowledge – "the whole truth" - and as emphasizing the value of their professional judgement as teachers.

The basis for the informal construction of cleverness appears in the teacher interviews to be an assemblage of several factors, including assumptions about which languages and language varieties are central to the individual pupil's everyday life outside school, assessments of the individual pupil's writing, and general observations from everyday life at the school (Holm 2015). Additionally, formal test results are often part of the informal construction of cleverness, if they do not contradict the informal construction of cleverness. What clearly emerges from our material is that there are (at least) two different constructions of cleverness. These differ in terms of what is considered to be relevant knowledge about the individual child for assessing academic cleverness in literacy. It is clearly important to the teachers that they base their constructions of cleverness on their professional knowledge of the individual child. The partial absorption of the national test meanwhile demonstrates not only that the test has changed the network, but also an acceptance of the construction of cleverness in the national test, provided that the test results are in accordance with the teacher's informal construction of cleverness. In this way, the boundary between the informal and formal constructions of cleverness is permeable, and the understanding of cleverness put forward in the national test has thereby become an important co-actor in the network.

Literacy testing in Danish primary and lower secondary schools has been a system-wide practice for decades. In the first year after the introduction of the compulsory national test, the majority of schools continued to also use the traditional test for literacy. Thus, national tests became a supplement to existing testing practices. The use of several different test formats resulted in a number of cases where a given pupil received divergent results. In the interviews, several teachers told of pupils whose scores are "at" or "over" the age-related norm in the traditional literacy tests, but in the national tests, the same pupils placed "below average". These categories offer certain possibilities for the teachers' interpretation of the pupils' cleverness as something that could be understood related to an average. The inconsistency between the different test results raises the teachers' criticisms regarding the validity and reliability of the national tests. The national tests are divided into three categories: decoding, comprehension, and language comprehension. Each of these categories is further divided into a number of item types (see above). The teachers' criticisms are directed towards

specific item types; that is, items that test sayings, and items that test students' knowledge of individual words using visual representations. From the interviews with the teachers, it appears that the teachers' criticisms relate to ideas about the linguistic practices the pupil is part of in their everyday lives. In excerpt 3, the teachers discuss the items that test an understanding of sayings (Holm 2015).

Excerpt 3:

> *Birgitte: I don't know what the idea is with these items.*
> *Anne: I think of it as very cultural. In fact, mono-cultural ... And one might say that it might be fair to require the pupils to understand this at some point. But I think that it is very demanding to expect this of the second grade....*

This excerpt illustrates the reasons for the teachers' reluctance to ascribe validity to the national tests' formal construction of cleverness. They are critical towards certain parts of the age-related literacy levels in the test, but their critique is related to the children's age and not to the test itself. The general construction of the test, and its three categories and various item types are not criticized in this case. This demonstrates that the teachers' criticism of the test is not an expression of a systematic and ontologically critical stance in relation to testing. They seem to accept basic assumptions about tests, with their criticism focused on concrete challenges they encounter in their everyday practice. The teachers point to a need for tests and items that are more sensitive to the everyday lives of migrant pupils with Danish as a second language during their early years of schooling, but apparently not later. The teachers' criticisms may therefore be addressed by introducing new or different test types, and test practices can thus be reproduced.

The introduction of the national literacy test in Danish schools in 2009 may be seen as introducing a new ordering project in the testing network, and at the same time as an extension of the formal construction of pupils' level of cleverness in literacy, because the schools continued to use the various literacy testing instruments they used prior to the introduction of the national tests. The new national tests have accentuated teachers' reflections on validity and have created dilemmas for teachers - partly in relation to varying test results, and partly owing to the interaction between formal and informal constructions of cleverness. Since the national tests in literacy have been made central in relation to the formal construction of pupils' literacy levels, both institutionally and discursively, one effect of the tests is that teachers' informal constructions of cleverness (namely, the part that is also based on everyday knowledge) have been put under strong pressure as a valid and legitimate construction of cleverness. In spite of this pressure, teachers assign most validity to the informal construction

of cleverness in cases where the two constructions of cleverness are not consistent. However, looking at the increase in calculative spaces established to practice the demo-test or to present its results, the national test appears not only to have influenced everyday school life, but also to have inscribed more actors in the testing network by making it compulsory to inform parents and pupils about test results. Thus, the national tests in literacy generally appear as a powerful and influential actor in the network that, through its materiality, creates new practices whose authority is partly accepted and partly questioned.

Conclusion

We can conclude that the national tests have created both new and complex calculative spaces and ordering projects. The national tests have become part of the teachers' constructions of children's cleverness and of the children´s constructions of cleverness: who is assigned how much cleverness by whom and why? The national tests generate teaching practices imitating test practice aiming at individualized cleverness constructions. In this way all the pupils are made testable. At the same time, pupils and teachers can act in test-like practices in creative and inclusive ways, creating new and attractive positions for some pupils.

The national tests have also created an interesting complexity in the way teachers construct and navigate between formal and informal constructions of cleverness. In addition, the analysis shows that a particular element of the national test in literacy, namely proverbs and sayings, is delegated a special role, and thus becomes a framework for the practice of teaching. The analysis has also shown that the test has become part of how teachers experience their own practice and how they distance themselves from certain definitions of this practice (a "teaching to the test" practice). The analysis shows and discusses how cleverness is at stake in different ways in the material. Even though teachers wish to transcend the narrow measure of testing and the narrowing of their teaching practice as 'teaching to the test', they seem, in some respects, to be caught in an individualistic and fixed understanding of cleverness, as well as reproducing test-like elements in their teaching. At the same time, this should not be interpreted unambiguously. The teachers in the material do much more than that; for instance, the teacher in the first analysis also supports the pupils and strives to create inclusive communities through his teaching. However, as Osterkamp (2009) argues, the individualistic approach is convincing. One way to understand this is through the concept of cleverness, which aligns with testing and other educational assessments. One of the tasks of the school is to ascribe students a

certain degree of cleverness, and national testing is but one possible part of this process. This might be one reason why teachers' criticisms of the test are not an expression of an ontological approach to testing, but instead represent a more reduced critique of the possibility that a specific individual's cleverness might have been wrongly measured. As such, cleverness is assumed to exist and be measurable, regardless of the suitability of any concrete test to accurately do so. The schools' more general connection to measures of cleverness – including the teachers' tasks – might be among the reasons for not challenging the individualistic view of cleverness. Testing could be said to both constitute and become part of ongoing differentiation processes and ongoing understandings of cleverness. The technology of testing becomes part of conditions for understanding cleverness in school.

In writing this article, we have experienced that an interdisciplinary research approach creates a fruitful complexity, generating discussions of the different academic concepts from various scientific traditions, such as communities/networks, actors/participants. We find that critical psychology and ANT have different scientific ontologies, but also many similarities. The similarities include (as previously outlined) the fundamental understanding that human beings create materiality, which, in turn, returns and creates conditions for human beings' participation (Nissen 2002). In this way, materiality does something and human beings do something with materiality in different and indeterminate ways. A critical psychological view is occupied with how materiality offers different possibilities for action and with understanding this from first-person perspectives. ANT offers the opportunity to view materiality as a non-human actor interacting with human actors. The different ontologies are articulated through different concepts, which we consider it important to use. As a result, this article may sometimes give the impression of different concepts with nearly the same meaning being used in parallel. However, it makes a difference whether people´s situated actions and interactions are termed as "communities" or as "networks". Where "community" draws on the philosophy-derived concept of praxis, and thereby relates to the understanding of humans' constant reproduction and transformation of historically developed social conditions, the concept of "networks" focuses on multiple networks that constantly form between human and non-human actors, and on how this contributes to constructions of meaning.

On the basis of our work on this article, we believe that cross-disciplinary research projects, rather than being a scientific bricolage, have the potential to open a window for the development of theories, methods, and concepts. This article does not identify one particular social/social-psychological theory as the primary theoretical framework for understanding meanings of testing, as asked

for by McNamara & Roever (2006). We instead propose a theoretical approach based on theories of everyday life and, through the application and discussion of – in this case - critical psychology and ANT, we have suggested concepts and presented a number of analytical findings related to the meaning of testing; something which is rarely the focus of research on testing. We see this article as an example of a multi-disciplinary research approach to testing, and we believe that other multi-disciplinary research projects that draw on other theories might reveal other meanings of testing in people's everyday lives.

References

Alderson, J. C. & Hamp-Lyons, l. (1996). TOEFL Preparation Courses: A Study of Washback. *Language Testing 13*(3), 280-297.

Alderson, J. C. & Wall, D. (1993). Does Washback Exist? *Applied Linguistics 14*(2), 115-129.

Allerup, P. (2015). PISA og TIMSS – internationale undersøgelser af samme slags? [PISA and TIMSS – international surveys of the same kind?] In Andreasen, K. E., Buchardt, M., Rasmusen, A & Ydesen, C. (eds) *Test og prøvelser.*[Test and trials] (pp. 51-71) Aalborg: Aalborg Universitetsforlag,

Arnoldi, J. (2003). Aktør-netværksteori: a-moderne (sociologisk?) teori. [Actor-network theory: a-modern (sociological?) theory] *Dansk sociologi 14*(3), 9-23.

Axel, E. (2011). Conflictual cooperation. *Nordic Psychology 63*(4), 56-78.

Bachman, L. F. & Palmer, A. S. (1996). *Language Testing in Practice: Designing and Developing Useful Language Tests.* Oxford: Oxford University Press.

Brandt, D. & Clinton, K. (2002). Limits of the Local: Expanding Perspectives on Literacy as a Social Practice. *Journal of Literacy Research, 34*(3), 337-356.

Christensen, G. (2002). *Psykologiens Videnskabsteori.* [The epistemology of Psychology] Frederiksberg: Roskilde Universitetsforlag.

Clarke, J. (2008). Assembling "Skills for life": Actor-network theory and the New Literacy Studies. InPrinsloo, M. &Baynham, M. (eds), *Literacies, Global and Local.* Amsterdam: John Benjamins.

Callon, M. (1987). Society in the Making: The Study of Technology as a Tool for Sociological Analysis. In Bijker, W. E., Hughes T. P. & Pinch, T. J. (eds), The *Social Construction of Technological Systems. New Directions in the Sociology and History of Technology* (pp. 83-106). Cambridge, MA: MIT Press.

Cheng, L. (1999). Changing assessment: washback on teacher perceptions and actions. *Teaching and Teacher Education 15*, 253-271.

Cheng, L., Watanabe, Y. & Curtis, A. (2004). *Washback in language testing: Research contexts and Methods.* Mahwah, NJ: Lawrence Erlbaum Associates.

Costall, A. & Dreier, O. (2006). Introduction. In: Costall, A. & Dreier, O. (eds.), *Doing things with things. The design and use of everyday objects* (pp. 1-12). Hampshire: Ashgate.

Dreier, O. (1979). Den kritiske psykologis overvindelse af psykologiske teoriers videnskabelige vilkårlighed [The critical psychologies overcoming of the scientific arbitrariness of psychological theories]. In: Dreier, O. (ed.), *Den kritiske psykologi*. [Critical Psychology] (pp. 7-26). København: Rhodos.

Egelund, N. & Nielsen, C. M. P. (2012). PISA Ethnic 2009: Immigrant and native Danish students results in PISA 2009. InEgelund, N. (ed.), *Northern Lights on PISA 2009: focus on reading* (pp. 91-105). Copenhagen: Nordic Council of Ministers.

Elstad, E. & Sivesind, K. (eds.). (2010), *PISA: sannheten om skolen?* [PISA: the truth about the school?] Oslo: Universitetsforlaget.

Extra, G., Spotti, M. & Van Avermaet, P. (2009). *Language testing migration and citizenship. National perspectives on integration regimes*. London: Continuum.

Fenwick, T. & Edwards, R. (2010). Actor-*network Theory in Education*. New York: Routledge.

Forchhammer, H. B. (2006). The woman who used her walking stick as a telephone: The use of utilities in practice. In:Costall, A & Dreier, O. (eds.), Doing things with things. The design and use of everyday objects (pp. 131-146). Hampshire: Ashgate.

Hamilton, M. (2001). Privileged Literacies: Policy, Institutional Process and the Life of the IALS. *Language and Education* 15(2-3),178-196.

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

Holm, L. (2015). De nationale test og læreres konstruktion af dygtighed i læsning. [The national tests and teachers' constructions of literacy cleverness] *Pædagogisk Psykologisk Tidsskrift* 3, 86-99.

Holm, L. (2014). De nationale test som redskab i folkeskolen. [The national test as instrument in elementary schooling] In: Krejsler, J. & Moos, L. (eds.) *Klasseledelsens dilemmaer. Fortsatte magtkampe i praksis, pædagogik og politik* [The dilemmas of classroom mangement. Ongoing power games in practice, pedagogy and politics] (pp. 340-358) Frederikshavn: Dafolo.

Holm, L. (2009). Who evaluates the evaluation? An analysis of computer-adaptive testing using literacy as an example. *Nordisk Pedagogik 29*(1), 137-148.

Holzkamp, K. (1985). *Grundlegung der Psychologie*. Frankfurt: Campus.

Holzkamp, K. (2005). Mennesket som subjekt for videnskabelig metodik. [The human being as subject for scientific methodology]. *Nordiske Udkast* 33(2), 5-33.

Holzkamp, K. (2013a). The fiction of learning as administratively plannable. In: Schraube, E. & Osterkamp, U. (eds.) Psychology from the standpoint of the subject. Selected writings of Klaus Holzkamp (pp. 115-132). Basingstoke: Palgrave Macmillan.

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

Højholt, C. (2005). Præsentation af praksisforskning. [Presentation of research in practice] In Højholt, C. (ed) *Forældresamarbejde. Forskning i fællesskab.* Gylling: Dansk Psykologisk Forlag, 23-46.

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

Højholt, C. (2016). Situated inequality and the conflictuality of children's conduct of life. In: Schraube, E. & Højholt, C. (eds.), *Psychology and the conduct of everyday life* (pp.145-163) London & New York: Routledge.

Juul Jensen, U. (1999). Categories in activity theory: Marx's philosophy just-in-time. In Chaiklin, S. & Hedegaard, M. & Juul Jensen, U. (eds.), *Activity theory and social practice* (pp. 79-99). Aarhus: Aarhus University Press.

Kousholt, D. B. (2006). *Familieliv fra et børneperspektiv. Fællesskaber i børns liv.* [Family Life from a Child Perspective. Communities in childrens lives.] Roskilde: Roskilde Universitetscenter.

Kousholt, K. (2016a). Evalueringer som mulige engagementer i fælles læreprocesser [assessments as children's potential engagements in shared learning processes]. *Nordiske Udkast 43*(1), 49-71.

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

Kousholt, K. & Thomsen, R. (2013). Dialectical approaches in recent Danish critical psychology. *Annual Review of Critical Psychology 30*(1) (Online)

Kousholt, K. & Hamre, B. (2016). Testing and school reform in Danish education: an analysis informed by the use of 'the dispositive'. In: Smith, W C. (ed.), *The global testing culture, shaping education policy, perceptions, and practice* (pp. 231-247). Oxford studies in comparative education. Southhampton: Symposium Books.

Kreiner, S. (2009). Om udvikling og afprøvning af pædagogiske test. [About development and testing of pedagogical tests] In: C. Bendixen & S. Kreiner (eds), *Test i Folkeskolen* (pp. 43-81). Viborg: Hans Reitzels Forlag.

Kulbrandstad, L. I. (2010). Leseopplæring på ungdomstrinnet før og etter PISA 2000. [Teaching literacy before and after PISA 2000] In Elstad, E. & Sivesind, K. (eds.), PISA – sannheten om skolen? (pp. 176-198). Oslo: Universitetsforlaget.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society.* Cambridge: Harvard University Press.

Latour, B. (1992). "Where Are the Masses?" The Sociology of a Few Mundane Artifacts. In Bijker, W. & Law, J. (eds.), *Shaping Technology / Building Society Studies in Sociotechnological Change*(pp.225-258). Cambridge, MA: MIT Press.

Latour, B. (1999). *Pandora's hope: essays on the reality of science studies*, Harvard University Press, Cambridge Mass.

Latour, B. (2005). *Reassembling the social: an introduction to Actor–network theory,* Oxford, Oxford University Press.

Laursen, H. P. & Holm, L. (2011). Migrants and literacy crises. Journal of Applied Language Studies – *APPLES 5*(2): 3-16.

Law, J. (2009). Actor-Network Theory and Material Semiotics. In Turner, B. S. (ed.) *The New Blackwell Companion to Social Theory* 3rd Edition (pp.141-158). London: Blackwell Publishers.

Lissovoy, N. D. (2013). Pedagogy of the Impossible: neoliberalism and the ideology of accountability. *Policy Futures in Education*, *11*(4), 423-435.

Marx, K. (1973). Teser om Feuerbach. In: Marx, K. & Engels, F. *Udvalgte skrifter i to bind.* Tidens forlag. Online: http://www.marxisme.dk/arkiv/marx-eng/biblio/meus.htm#bind2

McNamara, T. (2005). 21st Century Shibboleth: Language Test, Identity and Intergroup Conflict. *Language Policy 4*(4), 351-370.

McNamara, T. & Roever, C. (2006). *Language Testing: The Social Dimension*. London: Blackwell Publishers.

Menken, K. (2008). *English Learners Left Behind. Standardized Testing as Language Policy*. Clevedon: Multilingual Matters.

Messick, S. J. (1996). Validity and washback in language testing. *Language Testing 13*(3), 241-256.

Nichols, S. L. & Berliner, D. C. (2007). *Collateral Damage: How High-Stakes Testing Corrupts America´s Schools*. Cambridge, MA: Harvard Education Press.

Nissen, M. (1996). Undervisning som handlesammenhæng. [Teaching as context of action] In: Højholt, C. & Witt, G. (eds.), Skolelivets socialpsykologi. Nyere socialpsykologiske teorier og perspektiver [The social psychology of school life. Contemporary theories and perspectives of social psychology] (pp. 209-252). København: Unge Pædagoger.

Nissen, M. (2000). Practice research: critical psychology in and through practices. *Annual Review of Critical Psychology* 2, 145-179.

Nissen, M. (2002). Det kritiske subjekt. [The critical subject] *Psyke & Logos 23*(1), 65-85.

Olesen, F. & Markussen, R. (2006). Working with material things: From essentialism to material-semiotic analysis of sociotechnical practice. In: Costall, A & Dreier, O. (eds.), Doing things with things. The design and use of everyday objects (pp. 167-191). Hampshire: Ashgate.

Osterkamp, U. (2009). Knowledge and Practice in Critical Psychology. *Theory & Psychology 19*(2), 167-191.

Pettersvold, M. & Østrem, S. (2012). *Mestrer – mestrer ikke. Jakten på det normale barnet*. [Can –cannot. Chasing the Normal Child.] Oslo: Res Publica.

Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan* 68, 679-682.

Powell, K. Danby, S. & 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*(3), 259-275.

Pøhler, L. & Sørensen, S. A. (2010). *Nationale test og anden evaluering af elevernes læsning.*[The national tests and other forms of evaluation og students literacy] Frederikshavn: Dafolo.

Reckwitz, A. (2002). Toward a Theory of Social Practices. A Development in Culturalist Theorizing. *Euorpean Journal of Social Theory 5*(2), 243-263.

Ryumon, A. (2007). Washback of high-stakes assessment: Year 12 Japanese. In: Marriott, H., Moore, T. & Spence-Brown, R. (eds.). *Learning Discourses and the Discourses of Learning*. Melbourne: Monash University Press.

Schraube, E. (2013). First-person perspective and sociomaterial decentering: Studying technology from the standpoint of the subject. *Subjectivity 6*(1),12-32.

Schraube, E. & Marvakis, A. (2016). Frozen fluidity. Digital technologies and the transformation of students' learning and conduct of everyday life. In: Schraube, E. & Højholt, C. (eds.), *Psychology and the conduct of everyday life.* London & New York: Routledge.

Spolsky, B. (1995). *Measured Words: The Development of Objective Language Testing*. Oxford: Oxford University Press.

Shohamy, E. (2001). *The Power of Tests: A Critical Perspective on the Uses of Language Tests*. London: Longman.

Shohamy, E. (2006). *Language Policy: Hidden Agendas and New Approaches*. London: Routledge.

Skolestyrelsen, (2010). *Nationale test. Dansk, læsning 2.,4., 6. og 8. klasse.* [The national test. Literacy. Grade 2, 4, 6 and 8]. København: Skolestyrelsen.

Sørensen, E. (2006). Politics of things: The interplay of design and practice in a design workshop with children. In: Costall, A. & Dreier, O. (eds.), *Doing things with things. The design and use of everyday objects* (pp. 147-163)*.* Hampshire: Ashgate.

Taylor, L. (2005). Key concepts in ELT. Washback and Impact. *ELT Journal 59*(2), 154-155.

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

Watanabe, Y. (1996). Investigating washback in Japanese EFL classrooms: problems of methodology. *Applied Linguistics Association of Australia*. Series S. *13*, 208-239.

Weir, C. J. (2005). *Language Testing and Validation: an evidence-based approach*. Basingstoke: Palgrave.

Willis, P. (2001). "Tekin' the Piss". In D. Holland & J. Lave (eds.), *History in Person. Enduring Struggles, Contentious Practice, Intimate Identities* (pp. 171-216). Santa Fe: School of American Research Press.

Lars Holm is an Associate Professor and holds a PhD in literacy and globalization. His general research interests are in language, literacy and testing practices in multilingual and postcolonial educational settings.

Kristine Bagge Kousholt, PhD, is an Associate Professor in educational psychology. Her main research interests are children's perspectives on – and participation in standardized testing, children's communities and bullying.