The Quality Turn. Political and Methodological Challenges in Contemporary Educational Evaluation and Assessment

The Turn and the Paths. School External Evaluation in England, France and Switzerland: A Sociological Approach

Standards-Based Accountability in the United States: Lessons Learned and Future Directions

Constitutive effects as a social accomplishment: A qualitative study of the political in testing

Classroom assessment and education: Challenging the assumptions of socialisation and instrumentality

Assessment, evaluation and quality assurance: Implications for integrity in reporting academic achievement in higher education

Admission policy of Swedish teacher education favouring men: Discussion in Parliament in 1962

Being a skilled reader: Reception patterns in vehicle engineering students’ literature discussion

Predicting the use of praise among pre-service teachers: The influence of implicit theories of intelligence, social comparison and stereotype acceptance

“Own work” in primary school – A teaching and learning practice in the context of administration and control
EDUCATION INQUIRY

Education Inquiry is an international on-line, peer-reviewed journal with free access in the field of Educational Sciences and Teacher Education. It publishes original empirical and theoretical studies from a wide variety of academic disciplines. As the name of the journal suggests, one of its aims is to challenge established conventions and taken-for-granted perceptions within these fields.

Education Inquiry is looking for lucid and significant contributions to the understanding of contextual, social, organizational and individual factors affecting teaching and learning, the links between these aspects, the nature and processes of education and training as well as research in and on Teacher Education and Teacher Education policy. This includes research ranging from pre-school education to higher education, and research on formal and informal settings. Education Inquiry welcomes cross-disciplinary contributions and innovative perspectives. Of particularly interest are studies that take as their starting point, education practice and subject teaching or didactics.

Education Inquiry welcomes research from a variety of methodological and theoretical approaches, and invites studies that make the nature and use of educational research the subject of inquiry. Comparative and country-specific studies are also welcome.

Education Inquiry readers include educators, researchers, teachers and policy makers in various cultural contexts.

Every issue of Education Inquiry publishes peer-reviewed articles in one, two or three different sections. Open section: Articles sent in by authors as part of regular journal submissions and published after a blind review process. Thematic section: Articles reflecting the theme of a conference or workshop and published after a blind review process. Invited section: Articles by researchers invited by Education Inquiry to shed light on a specific theme or for a specific purpose and published after a review process.

Education Inquiry is a continuation of the Journal of Research in Teacher Education, which is available in printed copies as well as electronic versions and free access at http://www.use.umu.se/forskning/publikationer/lof/

Editors
Associate Professor Nafsika Alexiadou, Umeå University, Sweden
Assistant Professor Linda Rönnberg, Umeå University, Sweden

The editorial board
Professor Marie Brennan, Victoria University, Australia
Professor Bernard Corru, Directeur de la Formation - CNED, Directeur de CNED-EFAD, France
Professor Per-Olof Enson, Umeå University
Professor David Hamilton, Umeå University, Sweden
Professor Brian Hudson
Professor Gloria Ladson-Billings, University of Wisconsin-Madison, USA
Professor Martin Law, University of Edinburgh, UK
Assistant Professor Eva Lindgren, Umeå University, Sweden
Professor Lisbeth Lundahl, Umeå University, Sweden
Assistant Professor Linda Rönnberg, Umeå University, Sweden
Professor Kirk Sullivan, Umeå University, Sweden
Associate Professor Manya Sundström, Umeå University, Sweden
Professor Gaby Weiner, University of Edinburgh, UK
Professor Pavel Zgaga, University of Ljubliana, Slovenia

Language Editor
Murray Bales, Ljubljana, Slovenia

Guidelines for Submitting Articles
See Education Inquiry’s homepage: http://www.use.umu.se/english/research/educationinquiry
Send Manuscripts to: EducationInquiry.Editor@adm.umu.se

©2012 The Authors. ISSN online 2000-4508
Constitutive effects as a social accomplishment: A qualitative study of the political in testing

Peter Dahler-Larsen*

Abstract

Although it is often recognised that testing has unintended effects, not enough intellectual curiosity has been invested in a deeper questioning of what “unintended” means. While this concept often remains dependent on an image of good, underlying, rational intentions, it underestimates the contested, controversial and socially productive nature of testing. A conceptual alternative is offered, termed “constitutive effects”, as well as a theoretical framework in which this concept makes sense. Four domains in which constitutive effects are likely to unfold are carved out. A qualitative case study of how a national testing system of Danish for immigrants influences practices in one school in Denmark is offered as an illustration of such constitutive effects. Finally, a discussion unfolds concerning in what sense constitutive effects of testing can be seen as political (in a way that is not captured by the more conventional conceptualisation of “unintended effects”).

Keywords: unintended effects, testing, constitutive effects, the political in testing

Introduction

The literature on indicators, tests and evaluation systems recognises that such systems often have unintended effects. It is a rich theme around which many contributions circle (Brennan, 2006; Doran, 2001; Courtney, Needell and Wulczyn, 2004; Espeland and Sauder, 2007; Hood, 2007; McBriarty, 1988; McNeill, 2000; Osterloh, 2010; Osterloh and Frey, 2010; Robinson, 2003; Smith, 1995; Suen, 2003; Talbot, 2005; van Thiel and Leeuw, 2002; Vulliamy and Webb, 2001: 366; Weingart, 2005; Wilson, Wilson et. al., 2006). A well-known example in tests and assessment is “teaching-to-the-test” where teachers reorganise their teaching practices to maximise their students’ test scores, perhaps at the expense of a deeper understanding of the material taught, and perhaps at the expense of parts of the curriculum that are ignored or taken out because they do not fit into the format of the tests. The operational side of measuring or testing is pointed out as problematic because it leads to a deviation from an original or overarching idea of what ought to be measured. More often than not, these unintended effects are assumed to be negative, while the underlying or original intentions are assumed to be positive. The term “unintended effects” is thus aligned with

*Department of Political Science and Public Management, University of Southern Denmark. E-mail: pdl@sam.sdu.dk
©Authors. ISSN 2000-4508, pp.171–186
particular normative standpoints which may or may not be stated explicitly. Sometimes, these normative judgments are expressed in strong derogatory terms, such as “misuse” (Perrin, 1998), “perverse effects” (de Bruijn, 2001), “dysfunctional effects” (Feller, 2002) or marred by “diseases” (Bouckaert and Balk, 1991). It is questionable whether analysis is actually promoted through the use of such terms.

Since the term by definition assumes a set of original and authentic intentions against which the unintended effects can be identified, it is relevant to ask how it is possible to identify these intentions in rigorous theoretical and empirical analysis. Whose intentions count? At what point in time? What if various stakeholders configure changing intentions in a process of testing or measurement? How consistent do intentions need to be, and how clearly must they be conceived or expressed?

These questions are not always answered carefully and completely by those who subscribe to the term “unintended”. Sometimes the term “unintended effects” merely assumes a set of original good intentions. Perhaps it is assumed that these intentions are held by a benevolent architect of a testing system, or it is assumed that these good intentions are shared by all, based on a consensual and unproblematic view of, say, “what constitutes school quality” or “what is a good achievement”.

Based on careful philosophical attempts to understand quality (Stake and Schwandt, 2006) and other central concepts that circulate in culture and politics (Gallie, 1955; Koselleck, 2004), it seems unfair to assume an unproblematic consensus of what it was intended to measure. To call an effect “unintended” is perhaps, although it points to some problematic sides of testing or measuring, to not attend sufficiently to the contestable, performative, socially productive and constitutive capacities of tests, measurements, assessments and indicators, and thereby perhaps also to ignore their genuinely political ramifications. The term unintended effects represents in fact only one of several ways, and perhaps not the best way, in which these phenomena can be conceptualised and thereby understood.

This article sets out to do three things. Firstly, it offers a conceptual alternative to “unintended effects”, termed “constitutive effects”, and explains the theoretical framework in which this alternative makes sense. Secondly, it offers a case study in which a testing system had a serious impact on teaching practices if not the wider context in which teaching takes place. Through qualitative data, an attempt is made to find out which configuration of actions, rules, incentives, interpretations and interactions it takes to produce such effects and “keep them in place” as social constructions in the context studied. Special attention is paid to how teachers see themselves as professionals in combination with how they understand testing and its effects. Third, and finally, it is discussed in what sense constitutive effects of testing can be seen as political (in a way that is not captured by the more conventional conceptualisation of “unintended effects”).
Constitutive effects

Constitutive effects is a term that seeks to capture the way tests, measurements and indicators help define the social realities of which they are a part. Consistent with a social constructivist view, these realities are seen as both factual and as created through processes of social construction (Berger and Luckmann, 2009). Broadly speaking, and without going into more specific schools of thought within social constructivism, there are two major aspects of social life that are also the major building blocks in processes of the social construction of reality: language and social interaction.

Tests, of course, enter both of these domains. Tests provide a language of criteria, standards, items, test formats, skills and scores. Before something is measured quantitatively, a fundamental set of categories, distinctions and criteria are established in language (Porter, 1994). The paraphernalia of testing also becomes a way of describing test persons, their skills and achievements. Based on the social significance of the categories, “abilities”, “skills” and “results” that tests measure, people use tests in a number of ways in their interaction with each other. People imagine that what the tests measure is an objective part of reality, which in fact it is, after tests measure it. The social constructions of test-based reality flow out of the way people interact with reference to this constructed reality (Hanson, 2002). Tests sort people, and the sorting has social consequences. People strive to achieve good test results. The ambition to score well in tests influences students, teachers, school managers and others in particular ways. How people talk about tests, how they react to them and how they react to each other’s reactions is the social stuff that “effects of tests” are made of when viewed in a social constructivist perspective.

It is in such a perspective that the term “constitutive effects” of tests has been coined in order to capture how social reality emerges out of tests. More specifically, I suggest that constitutive effects concern:

a) a particular content, such as “traits”, “skills”, “competencies”, “qualities” or “results” that tests claim to measure objectively;

b) a particular definition of time frames in which certain results are expected;

c) a configuration of social relations among actors around tests and testing based on an attribution of roles; and

d) a broader world view (or set of world views) in which the meanings related to a)-c) are knit together.

Still the “unfinished” or processual nature of constitutive effects should be emphasised. They are not things but social accomplishments, to borrow the words of Thomas Schwandt. They are not effects of physical causes and should not be analysed as such. We may be looking more at rhetorical and discursive effects and at “attempts to define” a) through d) rather than “objects” which are already there. The added value of the term “constitutive effects” lies perhaps not so much in identifying a whole new
category of consequences of testing, but in showing that the language and interactions related to tests have broad human, social and political ramifications on a number of domains, such as those suggested above.

If a constructivist analysis lives up to our expectations, it is able to specify those aspects of a given social context that plays a role in the attempts to define reality by means of tests, and it should specify the social processes that help us understand how tests lead to more or less stable and fixed realities in the circumstances.

For example, “purification” (Skærbæk and Christensen, 2010) is a term constructivists use to describe how a particular technology for measuring, sorting or testing becomes “cleaner” and more legitimate while what is perceived to be random and subjective elements are gradually removed from it. For example, medical education, expertise and vocabulary hopefully “purify” medical diagnosis, and the prohibition of bribes to judges hopefully helps “purify” legal verdicts. “Lock-in” is the effect of the institutionalisation of certain practices of tests and other measurements on the actions of individuals. When legal, financial or normative sanctions are institutionalised, people are motivated to score well on measurements that release the sanctions. People who do not score well in a test or indicator system run the risk of being ridiculed or seen as incompetent, even if they deliberately choose not to participate. All this supports the “lock-in” effects that stabilise tests as part of reality. “Overflow” (Callon, 1998) is also possible. Overflow means side effects that are dislocated in time and space (like externalities in economics). For example, tests may create stress in families or indirectly affect how textbooks are written. Due to the complexity of social interaction in modern life, the effects of tests are not narrowly confined in time and space; the same is thus true about potentially negative reactions to overflow effects.

In this way, constitutive effects is a “working” concept with a broad focus, leaving room for empirical analysis that is specific and contextualised. Next, we shall explore how constitutive effects unfold in a particular case. We shall seek to attend to the “effects” of testing as both a process and result, and as constituted by both “objective” and “subjective” qualities.

A case study of testing of Danish for immigrants

Immigration policy is a highly politically charged policy area in Denmark. Every immigrant into Denmark has the right to receive funding for three years’ participation in a programme on Danish as a second language offered by special schools called “Sprogskoler” (“language schools”). The Danish education for immigrants is divided into levels and modules (app. 20), every one of which is followed by a particular test. Municipalities, which finance the programmes according to law, delegate to schools the right to carry out the programmes depending on the outcome of a tender among bidding schools at regular intervals. According to the contract between the municipality and the school, the financial framework for a school may depend totally on the number of tests passed by its students. Further, certain civil rights granted to
Constitutive effects as a social accomplishment: A qualitative study of the political in testing

immigrants, such as citizenship or a residence permit, depend on particular scores in particular tests.

Since testing is institutionally regulated and is followed by sanctions and rewards, we will not be surprised if testing has a strong impact on life in the school. But how? And how do teachers see themselves in that process?

I studied one school and interviewed one principal, six teachers, two testers and two administrators. My purpose – reflected in semi-structured interviews – was to understand the impact of the testing system on the working life of employees or, if you will, its constitutive effects as far as these can be reported and exemplified through interviews that tap into the employees’ perspective. (The data collection is originally part of a larger project which explores the impact of measurement, testing and indicators on the psycho-social well-being of employees in the public sector in Denmark).

Both managers and teachers greeted my study with enthusiasm. The school had been through administrative and organisational turmoil over the years and much concern had been expressed about the testing system, as well as the role of language for immigrants in our society. Both teachers and the manager welcomed a discussion of the testing system and of the working environment of teachers.

Most of the teachers were not young anymore. Their faces and sentences were characterised by that kind of experience which only some decades in the job can deliver. Their personalities and moods varied quite a lot and, while some of them had envisioned themselves in other careers such as academic careers, the majority said they had really come to love the job as a teacher of language for immigrants. There were both men and women among the interviewees, and both positive and negative views of the testing systems were offered. The interviews lasted about one hour each and took place between lessons or after lessons in a quiet room in the language school. In that sense, the teachers were interviewed in their own environment. The study did not include ethnographic observations in class rooms.

After reading and coding my interview transcripts, I organised my qualitative data in the following categories:

- how teaching is organised at the school;
- how the school organises information to teachers about test results;
- teachers’ views on tests;
- effects of testing as reported by teachers;
- how teachers view themselves as professionals in a world of tests; and
- a summary of how reality is kept in place around testing.

These are also the categories in which I will report my findings in the following sections.
How teaching is organised

Since some of the teaching takes place at uncomfortable hours, it is a delicate matter of how to allocate teachers to the various classes. Management attends to the preferences of each teacher in that respect. Teachers can also express their individual wishes to teach at a particular proficiency level, and their wishes are respected to the extent possible, they report. Perhaps even more importantly, they have a wide degree of discretion when it comes to teaching methods. They appreciate this discretion very much. Several of them are happy to tell about how they have developed their own approach in ways they think are successful. Over time, an increasing number of teaching materials has also been made available to them, but they are still to be used at their discretion.

The testing itself is not in the hands of teachers. Two special testers are appointed at the school. They take pride in securing the integrity of the testing system. The teachers, not testers, inform the students about test results. Testers report that test results are largely respected by teachers. A tester says she “does not know how teachers take care of the weakest students”. The testers are not faced directly with the human consequences of testing. The teachers are.

How the school organises information to teachers about test results

We can imagine this issue is important and controversial and that a difficult balance must be struck. Management used to send out a monthly “barometer” describing how the whole school performed compared to the budgeted results. Perhaps because this was perceived as a rough and uncomfortable instrument, today there is instead some “talk” about financial results, partly differentiated into the different sections of the school, each responsible for different teaching levels in the whole programme. Management claims it is easy to measure the performance of each individual teacher. However, the teachers say that such individual results would be both “unpleasant” and “difficult”. Students with different skills and backgrounds are unevenly distributed into different classes with different teachers. For that reason, it is “difficult” to measure the results of each teacher’s work fairly and precisely. (Note here that “difficult” may refer both to a methodological problem and to an interpersonal problem. These are intertwined). It is now common practice that the administration regularly provides teachers with a list of students which describes when the student is normally expected to take the test. It is up to the teacher’s informal assessment of each student (as well as the student herself or himself) to determine when is the best time to sign up for the test. But, as one teacher says, “it would be weird if your students never take the test”. The teachers report that because of the way teachers are informed about the need for students to be tested, “it is mentioned and it is not mentioned”. Although information about each teacher’s ‘performance’ is not directly available, the teachers are acutely aware of the importance of sending students to tests.
What teachers think of tests

The teachers hold both positive and negative views of tests. They report that in the “old days” students could “study forever.” It was a “cowboyland” with little formal structure, loosely defined goals and no indications of what students have learned along the way. Now, the test system helps define goals and thus provides a structure to the programme in combination, of course, with the elaborated differentiation of all teaching into levels and modules. Further, the difficult decision of how to allocate a student to the appropriate level and module is made easier and more legitimate when test results objectively determine the proficiency level of the student. Although validity can be discussed, the test is “not completely crazy”, says a teacher. “There is something meaningful about it”, “Students have a clear sense of what they should do”.

Still, teachers express strong negative viewpoints about the tests, too. One believes that is “insanely unjust” to couple tests to citizen’s rights. Others claim that the tests are too narrow because they do not attend to important aspects such as vocabulary and pronunciation. The tests favour “parrot talk”, as one says. They do not allow for attention to Danish culture and society, which would otherwise be useful for the students. Finally, the test format is too “school-oriented”. A test item such as giving a summary of a novel is only remotely understood by immigrants who are not already well-trained in a school system. It takes a long time merely to explain the test format. Thus, the test is a test of the school’s adeptness rather than of proficiency in Danish.

Constitutive effects of tests as reported by teachers

The teachers report that the testing system has a broad impact on the content of what is taught. “Tests guide more than 50% of our teaching”, says one. The teaching-to-the-test includes the preparation of students for item formats.

Another important aspect of testing is how it regulates the time available for various purposes during teaching. As a certain focus must be kept on tests, “there is no time to teach more ‘action-oriented’ and useful knowledge”, says one, “And no time to work with students’ integration into Danish society”.

The testing system also influences how teachers relate to students, or at least how they talk about them. Two typifications are used to describe students that are somehow problematic and/or culturally remote from Denmark. One such type is “weak”. After probing, teachers explain that “weak” can refer to a lack of schooling, a lack of alphabetisation or a lack of memory. The latter is sometimes a consequence of torture or other traumatic experiences that immigrants had before coming to Denmark. Another typification is “slow”. “Slow” is the direct term for those who have difficulties in passing tests. “Slow” does not refer to the wider set of aspects which explain the “weak” student’s situation. “Slow” is what a student appears to be when looked at through the lenses of testing. In other words, one of the constitutive effects of testing
is to provide a way of looking at the students that defines them as “slow”, and perhaps not drawing attention to the wider set of cultural and personal meanings that would otherwise be invoked by the term “weak”.

**Professionalism among teachers**

Teachers say that their professionalism has increased over the years. Objectives, topics and teaching materials have become more structured, and “we have become more clever”, as a teacher says. One fact that helps in understanding how a sense of professionalism is developing among language teachers is that there was no old well-established group among them with a self-identity, tradition, privileges and other norms, values and frameworks that would otherwise constitute pillars of professionalism. Historically, there was little official teaching of Danish for foreigners. As such programmes were installed (in recent decades), school teachers and university graduates entered the field. They were united by function rather than by values, norms or traditions. It was only after some formalisation of the programmes into levels and modules, after an official 1-year additional education, and after the establishment of tests that a sense of common structure and professionalism gradually developed. In other words, there was no strong pre-existing professional basis for any organised opposition against the testing system. Instead, the testing system was, at least partly, integrated into what it means to be a professional teacher of Danish as a second language.

It is perhaps the case that, as professionalism increases, “sensitivity towards other humans becomes less outspoken”, as one says or, put in another way, professionalism means finding a way to relate to students without getting overly involved emotionally. “If you are a good teacher”, says one, “if you are able, then you can make it work”. The sense of professionalism includes living with the test system. “If it goes well in the tests, you are a good teacher”; “If it happens too many times that students flunk the test, it probably has something to do with the teacher”. These statements indicate that teachers incorporate living with the testing system into their sense of a good teacher. In this way, they also become able to live with the system’s potential stressful nature. “I am sure, I have done everything I can, so I am not sleepless”, says one. “I think I am a good teacher”.

A professional attitude comes “over time” in relation to the students as well as in relation to the system, explains one. A professional attitude means “not saying from the beginning that management is stupid”. “It is like you have to be several ears and eyes”. The statement implies that the professional teacher has at least some understanding of the demands of management, too.

“It is a vulnerable thing to be a teacher”, says another. “It is up to the individual how to handle it. It is a matter of being on top of things, and having experience, and personality”.
Summary of the case: How reality is kept in place

The testing system is kept in place through classical mechanisms of institutionalisation such as legal rules and financial systems. On that basis, tests regulate the flow of the school’s daily life through two great organising principles in modernity: time and money. Tests have become social facticity. “It is just the way it is”, as a tester says. Institutional “lock-in” into this reality at the school level has happened in a number of ways.

The school deliberately secures the integrity of testing as much as possible. Testers take pride in living up to the rules of objective testing, and they are personally relieved of the pressure to inform students about the results. Despite discussions among teachers of the content and the politics of tests, there is no critique of the way the school operates the test system (in the interview data). There is no indication that the local legitimacy of testing is undermined. Further, the school takes care not to over-emphasise the responsibility of each teacher for a quantitatively defined target, which would be perceived as unfair. Instead, teachers generally perceive in a number of formal and informal ways that the number of tests that is passed is important for the whole school.

There was quite a critical debate among teachers at the beginning, but “the debate died out”, as one said. People realised that the tests were “not completely crazy”. As the tests were connected to the financial basis of the school, and since they did not go away, it was considered less meaningful to continue criticising them. A handful of years ago, there was a financial crisis in the school where staff numbers were diminished dramatically. Those who left were “messy”, “always sick”, “did not have their paperwork in order” or were simply “troublemakers”, says one teacher. Then people were “careful with what they said”, explained another. Whether or not these statements accurately describe the teachers who left or stayed, respectively, they are indicative of what teachers think is required to be a member of today’s workplace culture at the school. They also illustrate “lock-in”. After a while, it makes little sense to object to institutionalised facticity.

Still, not everyone takes part in all aspects of the school’s life around the testing system. One teacher reports to have declined to work as an external tester at other schools. It is too demanding and stressful to declare human beings “second rate citizens” if they do not pass the test. A tester has no more patience with all the administrative work and will step back in a few months. So, life with tests is not for all. The most critical voices or those who cannot live with its functions step out of their role entirely or partly.

For those who stay, the teacher’s sense of professionalism now incorporates a pragmatic way of living with tests. Teachers seem to accept the idea that a good, professional teacher also knows how to guide the way through testing for a sufficient number of students, although both teachers and the school are careful not to specify the “sufficient” number in an overly accurate way. Teachers observe that it is demanding
for the individual teacher to cope with the pressure, but they also believe it is possible to become a professional, i.e. to meet the demands. One’s professionalism develops over time, they say, but only people with a certain personality can do it. The fact that not everyone can live with the demands of the testing system underscores the idea that it requires special personal traits as well as experience to handle the pressures upon the teacher. Teachers’ tales about professionalism thus incorporate stories about the personal traits that are required to become professional. Sociologically speaking, pressure on individuals helps create a belief that it takes individuals of a particular kind to handle that pressure. The belief is, objectively speaking, not untrue, but it creates a particular form of attention to individual characteristics as factors which explain why some make it and others do not.

The teachers who make it take concrete steps that in fact help solidify the role of tests in daily life. For example, teachers ask for exemplary tests so their students can practice being tested. However, as official tests cannot be taken out of the testing room, it is necessary for teachers to produce “home-made” tests that imitate the formats used in official tests. The teachers even argue that students have a right to know more or less how they are being tested. Several claim that the opposite would be unethical testing practice. It would also be stressful for new teachers, it is argued, if there were no predictability in the tests. In other words, to teach to the test is, at least partly, combined with a positive ethical argument. In the circumstances, it is seen as ethical to comply with the demands of the testing system.

On top of that, a great deal of pragmatism seems to help keep the daily operations of the testing system in place. The institutionalisation of testing means that alternatives to the testing become difficult to imagine. The people in the ministry who make the tests can do nothing else than what they actually do, says one, because no test can cover everything. And “once this system is made, it must also be maintained”, as one says.

Some students also help sustain a focus on the tests. Some need the test results for objective reasons such as a residence permit. Others want tests because they like an objective sign of the progress they are making with the Danish language.

Thus, in multiple ways the testing system has helped install a particular strategic landscape in which teachers must act, and in which they see themselves. Institutional mechanisms help with “lock-in” effects that solidify this particular strategic landscape as reality.

But the reality shaped by testing is not without overflows. Not all teachers can live with the pressure. Teachers develop strategies to handle the tests that influence how they manage teaching, which includes finding less time for activities they feel would help better integrate students into society. Testing influences how teachers define themselves as professionals, and how they look upon students (seeing some as “slow”). Teachers develop a sense of professionalism according to which they do not become too psychologically close to the students’ problems. Still, to be a language teacher continues to be a job full of tensions. It is especially difficult to inform failing
students about their results, and shameful to disappoint those who have worked really hard with a language that is difficult for them.

**Limitations and implications**

Constitutive effects may not be equally intense everywhere. When testing is institutionalised and supported by legal and financial arrangements, such as in the case above, stronger constitutive effects may occur than would otherwise have been the case. So it is not testing itself which makes all the difference, but how the testing fits into a larger testing system which is, in turn, embedded in a larger social and political structure. (However, in principle it might be possible, as one referee pointed out, to provide a more fine-grained analysis of how various aspects of the tests themselves are reacted upon by students and teachers before as well as after the moment of testing, although that goes beyond the scope of the present study).

It is also possible to formulate further hypotheses about how various aspects of the social structures undergirding testing systems (publication, power, professionalisation etc) prepare the ground for variations in the constitutive effects of testing, but that exceeds the scope of a single case study. What the remaining space allows is a reflection on the utility of the very concept “constitutive effect”.

The term constitutive effects is an open, broad, sensitising concept without a full operational definition (Schwandt, 2001: 237). It only provides a sense of direction for our attention when confronted with the social effects of testing. It is, as one referee correctly noted, difficult to define negatively. What is *not* a constitutive effect? Probably, if the language involved in test items, results and indicators makes little difference for students and teachers, and if testing induces no significant interactions among human beings, the constitutive effects of testing would be negligible. But that may not be the issue. The concept of constitutive effects does not point to a whole new category of effects that is either existing or not. Instead, it points to the processual, contested and socially constructed/constructive nature of the effects of testing. And by departing from a framework of rational, good-hearted intentions (that may be tragically betrayed, thus resulting in so-called unintended effects), it opens a liberating perspective on the truly political in testing.

**The political in testing**

It was argued in the opening section of this article that the term “unintended effects” of testing is not sufficiently attentive to the contestable, performative, socially productive and constitutive capacities of tests. By calling something “unintended”, its genuinely political ramifications are ignored. It is now time to consider in what respect the constitutive effects of testing found in the case can be said to be political.

They are, of course, not directly related to politics in any classical sense, such as “the games of interests and log-rolling leading up to a decision in an official authoritative body determining who gets what, when and how”. But policies today are not only made
in relation to official decisions. Policies are implemented at various levels, in various networks and by various actors. We know from implementation theory and empirical findings that the implementation of policies is inherently political and cannot be separated from politics. How tests are carried out at the school level and how teachers manage teaching must be understood as an integrated part of the political in testing.

This political aspect is, of course, manifestly clear from the embeddedness of the testing system in the larger policy framework of Danish immigration policies. The testing system which is objective and at the same time creates serious obstacles for “slow” students is largely consistent with an immigration policy which favours immigrants who are well-educated and come from cultures and language areas close to Denmark. It would be political blindness not to see that parallel. Even though several teachers expressed some resentment to the present Danish immigration policy, the present strategic landscapes with tests determining school finances de facto makes it strategically beneficial for schools and teachers to look favourably at “fast” students and unfavourably at “slow” students. Perhaps the Danish immigration policy in a way needs schools and teachers to adapt to the strategic landscape defined by the same policy. The point is not that this constitutive effect can be solely attributed to tests. The point is that tests within a larger social context contribute to a politically loaded categorisation of human beings.

If the test system is a policy instrument, it carries, transports and incorporates a particular view of the world, of the problem to be solved, and of the people to be tested (Lascoumes and Le Gales, 2007).

The constitutive effects of testing are also clearly political in the sense that they are controversial and contested. Although teachers largely live up to the demands of the testing system and realise that it is part of the reality that does not go away, they describe some of the effects of the testing system as both unfair and counterproductive. Unfair, because weak students are treated shamefully, despite their efforts and trauma and other obstacles that make it difficult for them to remember the strange Danish words. Counterproductive, because teachers find there is no time to teach students useful things about how to understand Danish culture and how to get along in Danish society on a daily basis. With a strict focus on test results, “action-oriented knowledge” as it is called, is squeezed out of the curriculum, say the teachers. If this is correct, the testing system may have practical consequences that are inconsistent with the official policies of integration, even though the ideology of the testing system is not. Overflows are political realities, but they are not simple functions of ideology.

Perhaps the most far-reaching perspective on the political dimension in testing (as well as in other forms of assessment, evaluation and quality measurement) is found with reference to Rosanvallon’s (2009) view of the political. Rosanvallon sees the political as the terrain in which society works upon itself (p. 91). Society uses representations of its norms, values, goals and categorisations, and these representations are both descriptive and constitutive of what society is up to. In our modern type
of society, an evaluation society if you will (Dahler-Larsen, 2012), goals, objectives, evaluation criteria, quality criteria and test results are important and significant such representations. They describe what counts as both existing and desirable as society seeks to form itself. They are tools with which society works on itself. In that deep sense, they are political.

According to Rosanvallon, there is a close link between the concept of the political and democracy. Rosanvallon’s contribution lies in emphasising that the representations upon which members of a democratic society work are always imaginary, historical and therefore somewhat open-ended. The way that representations such as quality indicators, measurements and test scores are worked with in practice, and how they are turned into reality through constitutive effects, therefore have an inescapable political and democratic relevance. This is perhaps one of the richest perspectives in which we can view the constitutive effects of tests.

Professor Peter Dahler-Larsen, PhD, is director of the Master Program in Evaluation at University of Southern Denmark. He is particularly interested in cultural, organizational, and democratic aspects of evaluation. He has contributed to several international handbooks on evaluation. Together with Jenny Ozga, Hannu Simola, Christina Segerholm, he has published “Fabricating Quality” (Routledge 2011), a book on an international research project on quality assurance and evaluation mechanisms in schools. Other publications include “Kvalitetens Beskaffenhed” (University of Southern Denmark Press 2007) and “The Evaluation Society” (Stanford University Press 2012). He is presently doing research on the impact of performance indicators on public employees in Denmark.
References

Berger B., and Luckmann T. (2009) *Den Samfundsskabte Virkelighed*. København: Lindhart og Ringhof.

Bouckaert G., and Balk W. (1991) Public productivity measurement: Diseases and cures. *Public Productivity & Management Review* 15(2), 229–235.

Brennan R. (2006) *Perspectives on the Evolution and Future of Educational Measurement*. Educational Measurement. Westport, CT: Praeger Publishers.

Callon M. (1998) An essay on framing and overflowing: Economic externalities revisited by sociology. In Callon M. (ed.) *The Laws of the Markets*. Oxford: Blackwell.

Courtney M., Needell B., and Wulczyn F. (2004) Unintended consequences of the push for accountability: The case of national child welfare performance standards. *Children and Youth Services Review* 26, 1141–1154.

Dahler-Larsen P. (2012) *The Evaluation Society*. Palo Alto, CA: Stanford University Press.

De Bruijn H. (2001) *Managing Performance in the Public Sector*. New York: Routledge.

Doran H. (2001) *Dissertation: Evaluating the Consequential Aspect of Validity on the Arizona Instrument to Measure Standards*. Arizona: University of Arizona.

Espeland W., and Sauder M. (2007) Rankings and reactivity: How public measures recreate social worlds. *American Journal of Sociology* 113(1), 1–40.

Feller I. (2002) Performance measurement redux. *American Journal of Evaluation* 23(4): 435–452.

Gallie W. (1955) Essentially Contested Concepts. *Proceedings of the Aristotelian Society* 56.

Hansen A. (2000) How tests create what they are intended to measure. In (ed.) *Assessment: social practice and social product*. London: RoutledgeFalmer.

Hood C. (2007) Intellectual obsolescence and intellectual makeovers: Reflections on the tools of government after two decades. *Governance* 20(1), 127–144.

Koselleck R. (2004) Modernity and the planes of historicity. In Koselleck R. (ed.) *Futures Past. On the Semantics of Historical Time*. New York: Columbia University Press.

Lascoumes P. and Le Gales P. (2007) Introduction: Understanding public policy through its instruments – from the nature of instruments to the sociology of public policy instrumentation. *Governance: An International Journal of Policy, Administration, and Institutions* 20(1), 1–21.

McBriarty M. (1988) Performance appraisal: Some unintended consequences. *Public Personnel Management* 17(4), 421–434.

McNeil L. (2000) *Contradictions of School Reform. Educational Costs of Standardized Testing*. New York: Routledge.

Osterloh M. (2010) Governance by numbers: Does it really work in research? *Zeitschrift für Sozialtheorie* 2, 267–283.

Osterloh M., and Frey B.S. (2010) *Academic Rankings between the “Republic of Science” and “New Public Management”*. Working Paper. Zurich: CREMA – Center for Research in Management, Economics and the Arts.

Perrin B. (1998) Effective use and misuse of performance measurement. *American Journal of Evaluation* 19, 367–379.

Porter T.M. (1994) Making things quantitative. *Science in Context* 7(3), 380–407.

Robinson P. (2003) Government accountability and performance measurement. *Critical Perspectives on Accounting*, 171–186.
Rosanvallon P. (2009) *Demokratin som Problem*. Hägersten: Tankekraft Förlag

Skærbæk P. and Christensen M. (2010) Consultancy outputs and the purification of accounting technologies. *Accounting, Organizations and Society* 35(5), 524–545.

Smith P. (1995) On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration* 18(2&3), 277–310.

Stake R.E., and Schwandt T.A. (2006) On discerning quality in evaluation. In (eds.) *The Sage Handbook of Evaluation*. London: Sage Publications: 404–418.

Suen H. (2003) *Some very long-term and some chronic effects: The invisible dimension of the consequential basis of validity*. Unpublished paper.

Talbot C. (2005) Performance management. In (eds.) *The Oxford Handbook of Public Management*. Oxford: Oxford University Press.

van Thiel S., and Leeuw F. (2002) The performance paradox in the public sector. *Public Performance and Management Review* 25(3), 267–281.

Vulliamy G., and Webb R. (2001) The social construction of school exclusion rates: Implications for evaluation methodology. *Educational Studies* 27(3), 357–370.

Weingart P. (2005) Impact of bibliometrics upon the science system: Inadvertent consequences? *Scientometrics* 62(1), 117–131.

Wilson D., Crox B. and Atkinson A. (2006) “What gets measured gets done”: Headteachers’ responses to the English secondary school performance management system. *Policy Studies* 27(2), 153–171.
I wish to thank Christina Segerholm for organising this stimulating symposium and the participants for their positive contributions to it. I also wish to thank Barbara Crossouard for leading me to interesting references.
EDITORIAL

THEMATICAL SECTION

Christina Segerholm The Quality Turn. Political and Methodological Challenges in Contemporary Educational Evaluation and Assessment

Xavier Pons The Turn and the Paths. School External Evaluation in England, France and Switzerland: A Sociological Approach

Laura S. Hamilton, Brian M. Stecher & Kun Yuan Standards-Based Accountability in the United States: Lessons Learned and Future Directions

Peter Dahler-Larsen Constitutive effects as a social accomplishment: A qualitative study of the political in testing

Barbara Crossouard Classroom assessment and education: Challenging the assumptions of socialisation and instrumentality

D. Royce Sadler Assessment, evaluation and quality assurance: Implications for integrity in reporting academic achievement in higher education

Thomas A. Schwandt Quality, Standards and Accountability: An Uneasy Alliance

OPEN SECTION

Maria Hedlin Admission policy of Swedish teacher education favouring men: Discussion in Parliament in 1962

Stig-Börje Asplund Being a skilled reader: Reception patterns in vehicle engineering students’ literature discussion

Anna-Carin Jonsson & Dennis Beach Predicting the use of praise among pre-service teachers: The influence of implicit theories of intelligence, social comparison and stereotype acceptance

Kerstin Bergqvist “Own work” in primary school - A teaching and learning practice in the context of administration and control