

## PROFILES OF EDUCATION ASSESSMENT SYSTEMS WORLDWIDE

### Educational assessment in Norway

Sverre Tveit\*

*Department of Education, University of Oslo, Oslo, Norway*

*(Received 21 November 2012; accepted 24 July 2013)*

Norway has seen major changes in the field of educational assessment over the past decade, following the 2001 ‘PISA shock’ that stimulated reform of the entire primary and secondary education systems: new outcome-based curricula with cross-disciplinary basic skills were accompanied by major revision of assessment regulations, comprehensive government projects promoting formative assessment, national tests as a main component in a new national quality assessment system and new regulations for examinations and teacher reporting of overall achievement marks. The paper provides a historical context to the country’s prohibition of formal marking in primary education and the recent tensions determining how assessment criteria should be stated and used for formative and summative purposes. It is argued that Norwegian primary and secondary education is riddled with unresolved tensions as to the role of assessment criteria and national tests, sparked by incremental implementation of assessment policies and principles accompanying the new outcomes-based curricula.

**Keywords:** Norway; teacher assessment; assessment policy; criteria; examinations

### Introduction

The first decade of the twenty-first century can be characterised as a time of change in educational assessment in primary and secondary education in Norway. In light of disappointing test results in international comparative surveys such as PISA and TIMSS at the beginning of the century, the educational reforms of the 1990s were considered to have failed. PISA 2000 placed Norway as 13th among the 31 OECD countries in Reading and Science, and 17th in Mathematics (Lie, Kjærnsli, Roe, & Turmo, 2001). In a nation expecting to be among the best in the world, these rankings made for unflattering newspaper front pages.

Although achievement studies had revealed disappointing results in previous years, it was the ‘PISA shock’ that became a driving force for reforming the education system. ‘Almost like coming home from a winter Olympics without *one* Norwegian medal’, the Minister of Education and Research said. ‘And this time we cannot accuse the Finns of being drugged’, she added laconically as an expression of the nation’s shame, referring to neighbour country Finland’s top ranking<sup>1</sup> (Bergesen, 2006, p. 41, author’s translation). The name of the reform implemented in

---

\*Email: [sverre.tveit@ped.uio.no](mailto:sverre.tveit@ped.uio.no)

2006, *The Knowledge Promotion*, reflects the main goal of the reform: raising achievement for all students (Ministry of Education and Research [MER], 2007).

This paper overviews the context of educational assessment in Norway, with emphasis on the changes undertaken since the preparations of the 2006 reform. First, a general introduction to the Norwegian education context is provided with emphasis on how the education system is governed, how the current curriculum is organised and a broad outline of recent innovations in assessment policy and practice. This forms the backdrop for a more detailed presentation of assessment policy and procedures with emphasis on formative assessment, national testing, overall achievement marks and examinations. The article concludes by providing a historical context for Norway's prohibition of formal marking in primary education, current tensions as to how to state assessment criteria in the new outcomes-based education system and ambiguity of purposes of the national tests. Finally, it is argued that an implication of the past decade of innovation in assessment policy and practice may revoke old ideological disputes over formal marking in primary schools.

### **The context of educational assessment in Norway**

Norway has a strong economy and welfare system and is among the highest-spending countries in the world where education is concerned (Directorate of Education and Training [DET], 2011a). Education is free throughout primary and secondary education. The year they turn six years old, children begin 10 years of compulsory education,<sup>2</sup> which starts with seven years in primary school supplemented by three years in lower secondary school. While not compulsory, students have the right to be admitted to one of three chosen upper secondary programmes. Nearly all students enrol either in three-year upper secondary general studies programmes, which certify them for enrolment in higher education, or in one of the nine upper secondary vocational programmes consisting of two years' schooling supplemented by two years' apprenticeship or alternatively a third year of schooling.

### ***Governance and curriculum reform***

Primary and lower secondary education in Norway is governed by the local municipalities, while the upper secondary schools are governed by regional municipalities. The MER is responsible for national education policy. The DET is MER's executive agency and has overall responsibility for supervising education, governing the education sector and implementing Acts of Parliament, regulations and other national policies.

Norway is a geographically long, narrow and sparsely populated country with about 5 million inhabitants. There are 428 local municipalities who are responsible for primary and lower secondary education, and 19 counties that are responsible for upper secondary education. One-third of the schools have fewer than 100 students and less than one-third (27%) more than 300 students (DET, 2012). In order to reach out to all schools, the state is represented by county Governors that oversee, support and inspect the counties' and local municipalities' implementation of and adherence to policy and regulations.

The 2006 reform, with new curricula for primary and secondary education, represented a move from an input- to output-orientated policy with more emphasis on measurable outcomes (Skedsmo, 2011a). Cross-disciplinary basic skills (*express oneself orally; express oneself in writing; read; do mathematics*<sup>3</sup>; and *use digital*

*tools*) were among the most significant innovations in the curriculum. Rather than outlining the subject content and classroom activities that teachers were previously responsible for each year, the curriculum now focuses on *competence aims*<sup>4</sup> that students are expected to achieve by the end of Year 2, 4 and 7 in primary school, by the end of lower secondary school (Year 10), and by the conclusion of subjects in upper secondary school. How to organise the teaching is considered part of teachers' professional judgement (Karseth & Sivesind, 2010).

### ***Innovations in assessment policy and practice***

Educational assessment has for decades been a political battleground in Norway. In the 1970s, a national committee suggested the total abolition of formal marking in the lower secondary school (Lysne, 2004). While not having much effect on policy at the time, the debates in the 1970s expressed how formal marking, examinations and national testing are controversial ideological themes in school politics (Lysne, 2006) – disputes that became highly visible when the 2006 reform was implemented. A coalition government of the Conservative, Christian Democratic and Liberal parties from 2001 to 2005 prepared the education reform and achieved wide consensus across Parliament for most parts of the reform. When a new coalition government of the Labour, Socialist Left and Centre parties was formed after the 2005 election, it proceeded to implement the reform as planned in most respects. Assessment of student achievement and aspects of the national testing framework were, however, among controversial or unresolved parts of policy and regulations during reform preparations. Therefore, substantial deliberations with regards to educational assessment were undertaken by the new government, MER and DET during and after implementation of the 2006 curriculum. It has been a major challenge that the new outcomes-based curriculum's implications for assessing student achievement were not sufficiently substantiated when implementing the reform.

Schools that participated in an early launch of the new curricula discovered that student assessment was the area that most teachers (65%) needed more training in if they were to put the ideas of the reform into practice successfully (Bergem, Båtevik, Bachmann, & Kvangarsnes, 2006). Student assessment was in 2006 identified as one of the national priority areas for professional development. The new government admitted a range of problems regarding educational assessment: the regulations were not clearly understood; the assessment literacy of teachers and teacher educators was poor; and research in the field was limited (Stortingsmelding (White Paper) nr. 16, 2006–2007). These challenges were further confirmed in several studies evaluating the reform implementation. This formed the basis for several innovations with regard to regulations and professional development in subsequent years. The assessment regulations had been changed when the reform was implemented in 2006; however, further changes to the regulations were necessary when detailing assessment policy in subsequent years. In 2007, the government initiated a national project called 'Improved Assessment Practices' in part as a response to the identified problems of interpreting the assessment regulations. These were then briefly modified in 2007 before being subject to a further review and hearing in 2008 that led to a thorough revision of the regulations, implemented in 2009.

The most important changes were a greater emphasis on formative assessment, more explicit requirements for documenting the basis for assessment and strengthening of students' legal protection by stating clearer requirements for evidence

underpinning teacher judgements (DET, 2008). Further, in 2010, DET published a 95-page circular letter interpreting the assessment regulations, in response to the difficulties schools and teachers – particularly of small municipalities – experienced with interpreting the practical application of the assessment regulations (DET, 2010). The most important changes in the assessment regulations are detailed in the following sections.

### Assessment procedures and instruments

Table 1 outlines procedures and instruments that form part of the assessment context in Norwegian primary and secondary education: formative assessment, national tests, teachers' determination of overall achievement marks, and examinations. Formative assessment in Norway should not be viewed as a formative-summative dichotomy,

Table 1. Assessment instruments and procedures.

Procedure	Procedure years and description
Formative assessment	Teachers provide feedback to students on their achievements throughout all levels of education, based on the national curriculum's competence aims stated for Year 2, 4, 7 and 10 and the conclusion of subjects in upper secondary
End-of-term assessment	For all levels of primary and secondary education students receive an end-of-term assessment which reports on their achievements and challenges with regards to subjects and skills as part of the basis for student and parent conferences each term. A formalised part of formative assessment
National tests	Compulsory tests undertaken to monitor students' basic skills in English and cross-disciplinary skills in reading and mathematics at the beginning of Year 5 and 8. Reading and mathematics in Year 9. Teachers are encouraged to use individual students' results to adapt their teaching; while aggregate results for schools, municipalities and counties are used for governing purposes
Mapping tests	Tests used to enable early intervention for students with learning difficulties by identifying the 20% with lowest skills (intervention benchmark). Compulsory reading tests towards the conclusion of Year 1, 2 and 3. Compulsory tests of understanding numbers and arithmetic skills in Year 2; voluntarily in Year 3. Compulsory tests in reading and mathematics in the beginning of the first year of upper secondary; voluntarily for English
Order and conduct	Students' order and conduct, which express the degree to which students comply with the schools' regulations, are assessed each term. Starting in Year 8 separate marks for order and conduct are reported on a three-point scale
Overall achievement marks	Starting in Year 8, students' achievements are marked based on local interpretation and year alignment of the national curriculum competence aims stated for Year 10, and for the conclusion of subjects in upper secondary. The school-leaving certificate includes 16 overall achievements marks for lower secondary education and minimum 20 for upper secondary education (programme dependent). A six-point marking scale, from 1 to 6, is used, of which the first is failure
Examinations	Students are drawn to sit one external written and one local oral examination in Year 10. In upper secondary education, students undertake 5–6 examinations of which 3–4 are external written examinations and 2–3 local oral or practical examinations (programme dependent). Same six-point marking scale

which is sometimes misinterpreted both in theory and in policy contexts (e.g. Taras, 2009). In Norway, the DET distinguishes between formative (undervisvurdering) and final assessments (sluttvurdering). Summative assessments do occur in primary education (mapping tests and end-of-term assessments), but without explicit marking. For all levels, summative assessments undertaken during the terms should also be used for formative purposes. In secondary education, student achievement is reported with marks both continuously for assignments and tests and as overall achievement marks for end-of-term assessments and the school-leaving certificates.

### *Formative assessment*

The purpose of formative assessment is to support students' learning on an on-going basis by determining the quality of their work and providing feedback. This has been strengthened in the assessment regulations, in response to research that reported a substantial potential for improving these practices, which coincided with increased international attention to formative assessment (e.g. Black & Wiliam, 1998a, 1998b). While the term featured in earlier curricula in Norway (e.g. 'Curriculum guidelines', 1987), a white paper to Parliament in 2006 was a milestone in terms of the Norwegian government's recognition of unsatisfactory formative assessment practices and new policies for supporting students' learning (Stortingsmelding (White Paper) nr. 16, 2006–2007).

In the evaluation of the previous reform of primary and lower secondary education (L97), Klette (2003) observed that teachers were not good at making their expectations clear to students. Considerable positive feedback was given; however, this feedback was not sufficiently based on the quality of the students' achievements and there were few specific advices for how to improve. According to Dale and Wærness (2006), students in both lower and upper secondary education reported that teachers did not articulate their criteria for assessing achievement. In a national survey, a majority of students reported that teachers to a very limited extent told them what they were good at and what they needed to do to improve (Furre, Danielsen, & Stiberg-Jamt, 2006). Compared to many other countries, Norwegian teachers tend not to articulate learning intentions for the benefit of their students or to follow up students systematically (OECD, 2008; Vibe, Aamodt, & Carlsten, 2009). In a recent OECD review, the need to further improve formative assessment practices, especially in terms of providing feedback to students, was emphasised (OECD, 2011).

The failure to pay attention to the provision of explicit statements of learning intention can be understood in the context of national curricula that do not state levels of achievement. The guidelines developed for the marking of state-wide external examinations, however, have a long-standing tradition of stating criteria and standards. Secondary teachers in the classroom have to various extents used these examinations and associated rubrics in subsequent years. Dale and Wærness (2007) argued that the criteria in these rubrics were shaped by the norm-referenced tradition that was still present in the assessment regulations, by expressing levels of achievement in relative terms, i.e. 'Above average' (6 and 5), 'Average' (4 and 3) and 'Below average' (2 and 1). In some of these rubrics, all but the two highest marks were characterised by criteria indicating lack of attainment of these highest standards. Thus, it was argued that low achievement could be characterised in ways that do not support learning; e.g. 'Your text has no clear structure, incoherent logic, the message has little relevance, the content is poor and your language is imprecise

and characterised by many mistakes' (Dale & Wærness, 2007, p. 106, author's translation). Thus, while secondary examinations are developed for final assessments, the norm-referenced terminology of the marking scale (inherited from previous marking schemes) had the potential to distort formative assessment practices when used in the classroom. In 2007, these assessment regulations were changed and achievement levels are now described on a continuum from 'very low competence in the subject' (1) to 'outstanding competence in the subject' (6). Schools are encouraged to use criteria that describe what students can do rather than what they cannot do.

In 2009, further changes to the assessment regulations were made, as part of the national project 'Improved Assessment Practices', established in 2007. Among these were requiring teachers to specify the learning goals and a strengthened emphasis on students' skills and subject achievement for the end-of-term assessments. The projects further involved development of various models for facilitating a shared understanding of achievement levels (Hopfenbeck, Throndsen, Lie, & Dale, 2012). In response to further demands from the teaching profession and in order to ensure proper application of new policies, the project was prolonged with the national government effort *Assessment for Learning*, involving almost 900 schools across 187 local municipalities and all 19 counties, from 2010 to 2014. The DET were particularly inspired by the Scottish government's programme *Assessment Is For Learning*, which drew considerably on the work of Black and Wiliam (1998a), and also the Canadian state of Alberta's approach to assessment for learning. Stobart (2012) and other Assessment Reform Group members have been important contributors to national and regional Assessment for Learning conferences in Norway. Drawing on theory and research on formative assessment (in particular Black & Wiliam, 1998a, 1998b), current assessment policy states that students learn best when they:

- understand what they are to learn and what is expected of them;
- receive feedback that tells them about the quality of their work or performance;
- receive advice on how they can improve;
- are involved in their own learning by, for example, assessing their own work and development (DET, 2011b, p. 70).

The evidence of a weak culture for formative assessment was an important motivation for establishing the assessment for learning effort. DET points to Black and Wiliam (1998a), Hattie (2009) and other meta studies that suggest 'assessment for learning (formative assessment) is one of the most effective ways to raise students' achievements and their opportunities to learn' (DET, 2011c, p. 2, author's translation).

### ***National tests***

Over the past 10 years considerable effort has gone into establishing a comprehensive national quality assessment system for the school sector, holding schools and municipalities accountable for the quality of their schools' and students' achievements. In 1988, an OECD report questioned whether Norway had sufficient tools for monitoring the quality of its education system and proposed several strategies for ensuring this (OECD, 1988). Throughout the 1990s, a system for national evaluation



of schooling was discussed by government committees, MER and Parliament (Granheim, Lundgren, & Tiller, 1990; Moe, 1997; NOU (Green Paper), 2002; Stortingsmelding (White Paper) nr. 47, 1995–1996; Stortingsmelding (White Paper) nr. 28, 1998–1999). A national system, however, was not established until 2004 when national tests were first introduced as one of several components of the national quality assessment system.<sup>5</sup> While undergoing major revisions and complementation following the first implementation (see Discussion), the system now includes national tests of the basic skills in English and cross-disciplinary skills in reading and mathematics in Year 5 and 8; and reading and mathematics in Year 9. All students are required to sit these tests.<sup>6</sup> Further, some students are sampled to sit for tests in writing as a basic skill for Year 5 and 8 and in the Social Studies and Science subjects for Year 10.

While the national tests are the best-known components of the national quality assessment system, it also includes participation in international comparative achievement surveys (e.g. PIRLS, PISA and TIMSS) which helps monitor the quality of the education system in relation to new reforms and policy, and in comparison with other countries. Further, *user surveys*, where students, teachers and parents express their opinions, provide important information about the learning environment. DET monitor municipalities' compliance with the statutory requirements for selected themes for *supervision*. Finally, DET uses the web service *The School Portal* (*Skoleporten*) to provide schools, municipalities and the public with relevant statistical information about the education system (e.g. examination marks, overall achievement marks, national tests and mapping tests) on national, regional and school levels.

Currently, the official purpose of the national tests is on the one hand to provide information to students, parents and other stakeholders, school leaders, municipalities and national authorities as to the quality of students' basic skills, in order to monitor the quality of the education system (accountability purpose). On the other hand, the students' attainment on individual and group level should help teachers and schools to identify students' strengths and weaknesses which teachers can take into account in their teaching (formative purpose) (DET, 2011b). Evaluations show that while school owners and school leaders find the tests useful for decision-making, many teachers report that the national tests are not helpful for adapting their teaching and giving feedback to individual students (Aasen et al., 2012; Allerup, Kovac, Kvåle, Langfeldt, & Skov, 2009; Seland, Vibe, & Hovdhaugen, 2013). These shortcomings may be partly caused by the conflicting purposes of the national tests (see Discussion).

### **Overall achievement marks**

Teachers' determination of overall achievement marks is the most extensive form of assessment used for certification purposes in Norway. Starting in Year 8, students' achievements are marked based on local interpretation and year alignment of the national curricula's competence aims stated for Year 10, and for the conclusion of subjects in upper secondary. The school-leaving certificate for lower secondary education includes 16 overall achievements marks while that for upper secondary (programme dependent) includes a minimum of 20.

There are few national regulations for how these marks should be determined; typically they are based on a number of tests, assignments and other student work

that have been marked throughout the year. These assessments put the emphasis on students' growth over the course of the school year, including the development of global cognitive areas such as creative and collaborative skills that traditional tests and examinations cannot measure. Students are awarded one overall achievement mark for each subject, apart from language subjects, which have separate marks for written and oral achievements. While there are no national mandatory tests that form part of the basis for determining overall achievement marks, DET has recently developed optional standardised tests in Science and Social Studies – two of the lower secondary subjects that do not have any external examination. Apart from these, standardised tests used for marking student achievement are almost unprecedented in Norwegian classrooms.

A deep-rooted problematic practice among Norwegian teachers has been to credit students' effort when awarding overall achievement marks (Dale & Wærness, 2006). Since the 2006 reform several changes have been made to the assessment regulation, in particular clarifying that such practice is inappropriate and that only students' achievements should be emphasised when marking students (DET, 2008). Despite this, in a later study Prøitz and Spord Borgen (2010) observed that while teachers generally credited achievement when marking high-achieving students, effort and attitude were still rewarded for low achievers.

Few studies have investigated the comparability of teachers' determination of overall achievement marks across schools in Norway. A study by Statistics Norway comparing overall achievement marks and examination marks in Year 10 from 2001–2002 to 2007–2008 did however identify systematic disparities. Lenient marking (compared to the external assessments) was identified as a cross-subject factor, which may imply that not only individual teachers but schools' assessment cultures give rise to deviation between overall achievement marks and examination marks (Galloway, Kirkebøen, & Rønning, 2011). This implies that further attention to the issues of consistency and comparability of teachers' marking practices and the basis for final assessment is necessary.

### *National and local examinations*

There are two types of examinations in Norway; national external examinations that are in written form only and local examinations that can take written, oral or practical forms. DET organises the marking of external examinations while only regulating the procedures for local examinations. Students do not sit external examinations for all subjects; however, they are not notified about what examination they are drawn for until two days before the examinations occur, providing an extra incentive to prepare well for all subjects throughout the school year. The local municipalities and the counties are responsible for producing, implementing and marking local examinations, responsibilities that typically are delegated to schools or teachers.

Students are drawn for one external examination and one local oral examination at the end of lower secondary education (Year 10) that form part of the school-leaving certificate. In the upper secondary education programme for general studies, 20% of the students are drawn for one examination after the first year and all students for one examination in the second year. In the final year, all students undertake three or four examinations. The type of examination varies according to the programmes of specialisation; however, all upper secondary students undertake the external



examination in the Norwegian language subject and in total five or six examinations' marks form part of the school-leaving certificate.

DET applies several strategies for ensuring high-quality centrally produced examinations and marking procedures: annual comprehensive examiner seminars; annual examination board meetings; and assessment guides with descriptions and characteristics of achievement levels (standards). The external examiners, who work in pairs, are practising teachers that have been recommended by their own school principals and undertaken additional formal training in marking examinations (DET, 2011b). Local examinations are assessed by the student's own teacher and by an external examiner; the quality of these examinations is the local responsibility of the involved municipalities, schools and teachers.

For many years there have been considerable variations in schools' application of the local examinations with respect to students' opportunities to prepare a presentation based on examination tasks drawn several days in advance. As of the school year 2013–2014 stricter national guidelines for schools' application of the local examinations are being implemented in order to ensure more comparable assessments (DET, 2013).

### **Challenges prompted by outcomes-based curricula and cross-disciplinary national testing**

The previous sections have described Norway's tradition for educational assessment and many innovations that have been incrementally implemented following the outcome-based 2006 curricular emphasis on competence aims and cross-disciplinary basic skills. These fundamental changes in the curriculum coincided with the international trend towards formative assessment, which demands increased attention to the provision of clear learning expectations and the identification of student attainment as the basis for supporting student learning. The implementation of these policies prompted tensions as to how criteria for student achievement should be articulated, and purposes and uses of national tests of basic skills.

### ***Resistance to formal marking and national involvement***

Current assessment policy tensions draw on similar ideological arguments that once shaped the Norwegian tradition of prohibiting mark-based assessments in primary education. Rationales for formal marking typically follow three lines of argument: motivation, selection and information (Wikström, 2006). In Norway, formal marking has been prohibited in primary school since 1973, when a Parliamentary decision brought an end to a long-standing tradition of marking students for the sake of motivation (NOU [Green paper], 1974). Concerns about the negative impact on low achievers have since been a main reason for continuing the prohibition of formal marking in the primary sector (Tønnessen & Telhaug, 1996).

A second, and in Norway probably the most prevailing rationale for marking students, is the purpose of selection. However, as students go to designated schools in their districts throughout the compulsory years of education, no selection procedures are necessary until transfer from lower to upper secondary education. In the 1970s, there were strong advocates for abolishing formal marking in secondary education as well; however, these attempts fell short of addressing plausible alternatives for selecting students to further education (Lysne, 2004). The disputes over

formal marking in the 1970s did however have far-reaching implications for the *basis* for determining overall achievement marks in secondary education. It substantiated scepticism towards explicitly stated learning objectives that could be measured and controlled; a notion both left- and right-wing governments complied with when revising the curricula in the 1980s and 1990s. The curricula and regulations emphasised *holistic competence* (helhetlig kompetanse), which can be defined as a wider interpretation of student achievement, emphasising collaborative skills and cross-disciplinary perspectives (Dale & Wærness, 2006; Lysne, 2004). The 2006 reform brought an end to the concept of holistic competence, which MER concluded had brought uncertainty as to the basis for assessing student achievement and inappropriate emphasis on students' effort, order and conduct when determining overall achievement marks for subjects (Stortingsmelding (White paper) nr. 30, 2003–2004).

The third argument, that achievement marks provide helpful information to students and parents as to where they are in their learning, may also have less relevance in the Norwegian education system, as these marks have not been determined based on clearly stated learning objectives. Whilst changing to outcome-based curricula, these still do not state level-specific achievement expectations and as we have seen schools and teachers struggle to develop sound assessment criteria themselves. Several years after the 2006 reform was implemented, many secondary teachers were still observed to use marks to motivate their students, despite the fact that this practice is now explicitly prohibited (Prøitz & Spord Borgen, 2010). If student work is marked lower or higher than the actual attainment, e.g. by rewarding effort for low achievers or sustaining high-achieving students' motivation by making them strive more for a higher mark, it becomes less plausible to refer the judgements to information detailing what the mark represents. Thus, along with the notion of holistic competence, the widespread practices of emphasising students' effort and motivation may have distorted the potential information value of the marks in Norwegian secondary education, which in turn makes the information argument for marking students in primary years less relevant. In other words, the potential of using summative assessments for formative purposes is distorted when the numerical marks do not reflect valid statements for what characterises such achievements.

### *The criteria dilemma*

The move to an outcomes-based education system with subject curricula stating competence aims by default implies more emphasis on defining the basis for marking students' achievements. The vast proportion of teacher judgements (80%) underpinning upper secondary students' school-leaving certificates, which form the basis for a national system for higher education enrolment, is based on teachers' determination of overall achievement marks of their own students. Thus, upper secondary teachers' judgements, in particular, are extremely high-stakes in Norway compared to many countries. Given the evidence indicating substantial differences in teachers' and schools' expectations of students when determining overall achievement marks, and the high-stakes nature of these judgements, it is imperative to strengthen comparability. On the other hand, as discussed above, there is a deep-rooted scepticism in Norway towards detailing national assessment criteria and standards. Both within and between the MER and the DET, there have been considerable disagreements as to how to approach this 'criteria dilemma' and the most

influential teacher union, *Union of Education Norway*, has consistently opposed national assessment criteria and standards (Utdanning, 2006, May 5).

In the white paper that prepared the 2006 reform, MER first stated that it would provide the necessary tools for ensuring that teachers' judgements are 'standards based' (Stortingsmelding [White paper] nr. 30, 2003–2004, p. 40). Emphasis on assessment criteria and standards was primarily aimed at improving comparability and consistency in teachers' determinations of students' overall achievement marks. In 2005, a working group comprising members from MER and DET recommended development of mandatory national assessment criteria for all subjects; however, this was not pursued by the MER. Later, DET recommended a less controversial approach with development of national criteria in four subjects (Norwegian, Mathematics, Social Science, and Food and Health) detailing two or three levels of achievement instead of six mark-specific level statements (DET, 2006). MER did not adopt these recommendations either, and the new curricula were therefore introduced without accompanying assessment criteria and standards. Local authorities and schools were encouraged to develop these themselves.

As mentioned earlier, the project 'Improved Assessment Practices' formed the start of several measures for improving formative assessment practices. The project also included a modified approach to developing shared understanding of achievement levels in the four subjects for which DET initially proposed to develop standards accompanying the curricula. These were called *characteristics of levels of achievement*,<sup>7</sup> which essentially is what in both Norwegian and English terminology can be described as a *standard* (DET, 2009).<sup>8</sup> Terminology that may sound inflexible for local interpretation and having connotations associated with standardising the curricula was avoided – an extension of the Norwegian reluctance towards national statements of learning expectations.

The project evaluations observed that many schools found it hard to see the connection between characteristics of levels of achievement and the curriculum's competence aims. The researchers proposed to develop national assessment criteria and characteristics of levels of achievement, with the opportunity for schools to detail these criteria themselves (Throndsen et al., 2009).<sup>9</sup> Accordingly, DET proposed to MER that national recommendations for characteristics of levels of achievements should be established for the subjects that had formed part of the project (DET, 2009b). This was as controversial as the similar proposals made three years earlier, and again rejected by MER. Instead, optional 'assessment guides' were developed for curricula with accompanying examples which to various extents include level-specific criteria. Studies show that the use of these guides is limited (Aasen et al., 2012; Hodgson, Rønning, Skogvold, & Tomlinson, 2010) and further studies have recommended mandatory criteria and standards (Prøitz & Spord Borgen, 2010).

DET continues to support teachers and schools in developing a shared understanding of characteristics of achievements through the national effort *Assessment for Learning*. While the main purpose of developing assessment criteria and standards initially was to ensure comparability and consistency in teachers' marking practices, the label of the national effort reflects a shift in focus with more attention being paid to formative assessment. Comparable interpretations of criteria and standards are essential for ensuring valid formative assessments, otherwise feedback may guide students in the wrong direction in their learning (Stobart, 2012). Nevertheless, the extremely high stakes of students' overall achievement marks implies

that the summative use makes stronger demands for consistency and comparability of teachers' judgements in order to ensure fairness and social justice in selection to further education. This leads to the contestable yet compelling argument that emphasis on summative use of assessments (i.e. certification and selection) warrants national or mandated criteria and standards; whilst schools' own development of criteria and standards is emphasised for formative assessment.

An important dilemma, however, is the potential of pre-stated assessment criteria and standards to lead to practices where students and teachers may check off criteria rubrics instead of engaging in deep learning. Such instrumental approaches may ultimately lead to instrumental learning where criteria compliance replaces learning (Torrance, 2007). The above discussed reluctance of stating clear and measurable learning outcomes that has characterised the Norwegian assessment tradition for decades relies partly on such a fear of instrumental learning. This scepticism towards national mandatory criteria has a solid basis in the education sector in Norway and particularly in the teacher union, despite studies observing more teachers calling for national involvement in stating criteria and standards.

Another argument for principally opposing emphasis on criteria and standards is that it takes time that otherwise could be used for teaching (DET, 2005). On the other hand, many teachers and schools now call for national criteria and standards precisely because they would then need to spend less time developing these themselves (Hopfenbeck et al., 2012; Prøitz & Spord Borgen, 2010).

### *Ambiguous purposes of national tests*

As for the criteria dilemma, a historical account is necessary to understand current policy tensions with regards to the purposes and uses of national tests. The national tests and the School Portal saw a rough start when introduced in 2004. The government was accused of facilitating publication of league tables and inexpedient control of schools (DET, 2011b). In 2005, many students and teachers in upper secondary education boycotted the test, which established national testing as one of the key issues in political debates on education before the 2005 parliamentary elections (Hølleland, 2007). Evaluation of the tests revealed several problems to do with test purpose, design, development and administration (Lie, Hopfenbeck, Ibsen, & Turmo, 2005). Following the election, the new centre-left government instituted a one-year moratorium in order to develop a more solid framework for national testing and avoid publication of league tables.

When the new testing framework was introduced in 2006, the purpose was primarily defined as, 'providing information to students, teachers, school leaders, parents, school owners and the regional and national level as basis for improvement and development' (MER, 2006, p. 1). It was emphasised that the tests primarily should provide information about groups of students and 'not diagnostic information about individual students' (DET, 2011b, p. 17). Accordingly, the strongest practical implications of the national tests have been associated with the purpose of holding municipalities and schools accountable for their students' results and the education they offer, and to provide information necessary for the national governance of the education sector. As described earlier, research suggests that it achieves this purpose relatively well.

The first unsuccessful implementation of the national tests – and all the commotion it involved – gave rise to unrealistic expectations as to what purposes

such tests can serve. Much time and effort has since been put into communicating and informing the public about the main objectives of the tests (DET, 2011b). Studies show that the legitimacy of the testing has increased as schools have gained experience with using the tests and interpreting the results; however, teachers continue to want the tests to provide more information about the students than they do today (Allerup et al., 2009).

In more recent frameworks for the national tests, DET expands the official purpose of the tests to additionally ‘contribute to strengthen the schools’ and teachers’ formative assessment’ (DET, 2012, p. 4, my translation). In a recent evaluation of the testing framework, DET leaders reports that initial controversies over publishing and ranking school results may have come to overshadow the prospects of using the tests for formative purposes. Analyses of how the purposes of the tests are presented in a hierarchy of documents show that emphasis is moved from emphasising accountability purposes when addressing the governing structures, to emphasising formative use when addressing schools and teachers. While it may be plausible to address different purposes to different audiences, the researchers warn that such a simplification may obscure important aspects and prompt unrealistic expectations as to what purposes the tests can serve (Seland et al., 2013; Skedsmo, 2011b). It can be questioned whether it is feasible to continue attempting to combine two conflicting purposes in single instruments.

## Conclusion

The implementation of new curricula and accompanying assessment regulations in Norway can be viewed as an extreme example of a universal criteria dilemma – whether, and if so to what extent, criteria or standards should be (nationally) mandated. The significance of teachers’ judgements as the basis for the school-leaving certificates, along with a national admissions policy that mainly relies on students’ GPA, implies that teachers’ judgements are extremely high stakes. This strongly accounts for *national* criteria and standards in Norway, in order to improve consistency and comparability and contribute to more fairness and social justice of certification and selection. Such a push for national prescriptions of assessment clashes with Norwegian policy and practice traditions opposing clear and measurable learning goals detailed by the state that have characterised the education system since the 1970s (Prøitz, 2013).

Whether national or not, the increased emphasis on shared assessment criteria and standards underpinning final assessments and guiding formative assessment may prompt surprising and unexpected policy changes – a controversial reintroduction of formal marking in primary schools. Pursuing both the information and motivation arguments, right-wing parties and some of the largest municipalities (such as the capital city Oslo) have requested permission for formal marking of students in Year 5–7; however, such requests have repeatedly been rejected by the national government (DET, 2009). Changing this policy would require a parliamentary law decision, which may occur with a change of government following the 2013 Parliament election.

Ironically, the centre-left government’s push for formative assessment may have facilitated a change in the public’s perceptions of formal marking. As the basis for numerical marks becomes more tangible, the information arguments for formal

marking become more prominent. The remaining argument will then be concerns about the negative impact on low achievers, which initially prompted the prohibition of formal marking in the 1970s. Thus, the past decade of innovations in policy and practice of educational assessment has established a new practice context that has revoked ideological disputes over educational assessment.

## Notes

1. Finland had been exposed to doping scandal when Lahti hosted the 2011 World Championship in Nordic skiing.
2. Compulsory education refers to the rights and obligations children have for education. While parents can choose to organise education themselves almost all children undertake the compulsory years in schools.
3. In Norwegian, the term 'regning' is used for the basic skill *mathematics*, however an official translation is not established. This skill should not be confused with the Mathematics *subject* (distinguished by capitalisation).
4. The official translation of the Norwegian term *kompetansemål*. Learning objectives, attainment goals and learning outcomes are versions of the same concept, which have various connotations in different contexts.
5. The presentation partly draws on the DET's (2011b, pp. 14–15) official description.
6. Exemptions are made only for rigidly justified reasons, e.g. special needs diagnoses.
7. The term 'characteristics of achievement levels' is a direct translation of the term *kjenne-tegn på måloppnåelse*.
8. In the evaluation (Throndsen, Hopfenbeck, Lie, & Dale, 2009), it was suggested that characteristics of levels of achievement should be called *criteria*, complying with internationally established terminology instead of coining a new term. In their report to MER, DET (2009b) pointed out that the appropriate international term rather would be a standard; however, it was decided to keep the new terminology.
9. For an English summary, see Hopfenbeck et al. (2012).

## Notes on contributor

Sverre Tveit is a PhD research fellow candidate at Department of Education at the University of Oslo. His main interest lies in the validity, comparability and legitimacy of assessment systems that underpin certification and selection in secondary education.

## References

- Aasen, P., Møller, J., Rye, E., Ottesen, E., Prøitz, T. S., & Hertzberg, F. (2012). *Kunnskapsløftet som styringsform – et løft eller et løfte?* [The Knowledge Promotion as governance – a lift or a promise?]. Oslo: NIFU.
- Allerup, P., Kovac, V., Kvåle, G., Langfeldt, G., & Skov, P. (2009). *Evaluering av det Nasjonale kvalitetsvurderingssystemet for grunnsopplæringen* [Evaluation of the national system for quality assessment in primary and secondary education]. Kristiansand: Agderforskning.
- Bergem, R., Båtevik, F. O., Bachmann, K., & Kvangarsnes, M. (2006). *Tidleg oppstart med nye læreplanar: Kartlegging av erfaringar med førebuing og iverksetjing* [Early start with new curricula: Mapping experiences on preparation and implementation]. Volda: Høgskulen i Volda.
- Bergesen, H. O. (2006). *Kampen om Kunnskapsskolen* [The fight over the "knowledge school"]. Oslo: Universitetsforlaget.
- Black, P. J., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5, 7–74.
- Black, P. J., & Wiliam, D. (1998b). *Inside the black box. Raising standards through classroom assessment*. London: King's College.



- Curriculum guidelines for compulsory education. (1987). Oslo: Ministry of Education and Research.
- Dale, E. L., & Wærness, J. I. (2006). *Vurdering og læring i en elevaktiv skole* [Assessment and learning in a student active school]. Oslo: Universitetsforlaget.
- Dale, E. L., & Wærness, J. I. (2007). Tilpasset opplæring og elevvurdering [Adapted teaching and student assessment]. In S. Tveit (Ed.), *Elevvurdering i skolen – grunnlag for kulturrendring* [Student assessment in school – a basis for change of culture] (pp. 103–110). Oslo: Universitetsforlaget.
- DET. (2005). *Elevvurdering i Kunnskapsløftet – anbefaling* [Student assessment in the Knowledge Promotion – recommendations]. Oslo: Utdanningsdirektoratet.
- DET. (2006). *Elevvurdering i Kunnskapsløftet* [Student assessment in the Knowledge Promotion]. Oslo: Utdanningsdirektoratet.
- DET. (2008). *Høringsbrev endring i forskrift til opplæringsloven kapittel 3 og 4 og endring i forskrift til privatskoleloven kapittel 3 og 4* [Consultation paper for proposed changes in the regulations for the Education Act chapters 3 and 4]. Oslo: Utdanningsdirektoratet.
- DET. (2009a). *Ad Søknad om forsøk – vurdering med karakter på barnetrinnet* [Application for trialling marking in primary education]. Oslo: Utdanningsdirektoratet.
- DET. (2009b). *Sluttrapport. Oppdragsbrev nr 6, om tiltak knyttet til individvurdering i norsk skole og fag- og yrkesopplæring* [Final report: Commissioning letter 6–2007 about actions for individual assessment in primary and secondary education]. Oslo: Utdanningsdirektoratet.
- DET. (2010). *Rundskriv Udir-1-2010 – Individuell vurdering i grunnskolen og videregående opplæring etter forskrift til opplæringslova* [Circular letter: Individual assessment in primary and secondary education]. Oslo: Utdanningsdirektoratet.
- DET. (2011a). *The education mirror 2011*. Oslo: Utdanningsdirektoratet.
- DET. (2011b). *OECD country background report*. Oslo: Utdanningsdirektoratet.
- DET. (2011c). *Grunnlagsdokument. Satsingen Vurdering for læring 2010–2014* [Basic document for the Assessment for Learning national effort 2010–2014]. Oslo: Utdanningsdirektoratet.
- DET. (2012). *Nasjonale prøver – Veiledning til skoleeier og skoleledere med retningslinjer for gjennomføring* [National tests – guidelines to school owners and school leaders with guidelines for implementation]. Oslo: Directorate of Education and Training.
- DET. (2013). *Utdanningsdirektoratets anbefaling – Forskriftsendring muntlig eksamen* [The recommendation from the Directorate of Education and Training – Regulation changes oral examinations]. Oslo: Utdanningsdirektoratet.
- Furre, H., Danielsen, I. J., & Stiberg-Jamt, R. (2006). *Analyse av den nasjonale undersøkelsen 'Elevundersøkelsen' 2006* [Analysis of the national student survey 2006]. Kristiansand: Oxford Research.
- Galloway, T. A., Kirkebøen, L. J., & Rønning, M. (2011). *Karakterpraksis i grunnskoler. Sammenheng mellom standpunkt- og eksamenskarakterer* [Marking practices in primary and lower secondary schools. Relations between overall achievement marks and examination marks]. Oslo-Kongsvinger: Statistics Norway.
- Granheim, M. K., Lundgren, U. P., & Tiller, T. (1990). *Målstyring og evaluering i norsk skole. Sluttrapport EMIL-prosjektet* [Goal steering and evaluation in Norwegian school. Final report from the EMIL project]. Oslo: Norsk råd for anvendt samfunnsforskning.
- Hattie, J. (2009). *Visible learning*. London: Sage.
- Hodgson, J., Rønning, W., Skogvold, A. S., & Tomlinson, P. (2010). *På vei fra læreplan til klasserom* [On the way from curricula to classroom]. Bodø: Nordlandsforskning.
- Hølleland, H. (2007). Nasjonale prøver og kvalitetsutvikling i skolen [National tests and quality development in school]. In S. Tveit (Ed.), *Elevvurdering i skolen – grunnlag for kulturrendring* [Student assessment in school – a basis for change of culture] (pp. 29–43). Oslo: Universitetsforlaget.
- Hopfenbeck, T. N., Throndsen, I., Lie, S., & Dale, E. L. (2012). Assessment with Distinctly Defined Criteria: A research study of a national project. *Policy Futures in Education*, 10, 421–433.
- Karseth, B., & Sivesind, K. (2010). Conceptualising curriculum knowledge within and beyond the national context. *European Journal of Education*, 45, 103–120.

- Klette, K. (2003). Lærernes klasseromsarbeid; Interkasjons- og arbeidsformer i norske klasserom etter Reform 97 [Teachers' classroom work: Interactions and ways of working in the Norwegian classroom after Reform 97]. In K. Klette (Ed.), *Klasserommets praksisformer etter Reform 97* [Classroom practices after Reform 97] (pp. 39–76). Oslo: Pedagogisk Forskningsinstitutt, Universitetet i Oslo.
- Lie, S., Hopfenbeck, T. N., Ibsen, E., & Turmo, A. (2005). *Nasjonale prøver på ny prøve* [National tests retested]. Oslo: Department of Teacher Education and School Research, University of Oslo.
- Lie, S., Kjærnsli, M., Roe, A., & Turmo, A. (2001). *Godt rustet for framtida? Norske 15-åringers kompetanse i lesing og realfag i et internasjonalt perspektiv* [Well prepared for the future? Norwegian 15-year-olds' competence in reading and scientific subjects in an international perspective]. Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.
- Lysne, A. (2004). *Kampen om skolen. 1970–2000. Karakterer og kompetanse. Stridstema i norsk skolehistorie* [The fight over school 1970–2000. Marks and competence. Controversies in Norwegian school history]. Haslum: AVA Forlag.
- Lysne, A. (2006). Assessment theory and practice of students' outcomes in the Nordic countries. *Scandinavian Journal of Educational Research*, 50, 327–359.
- MER (Ministry of Education and Research). (2006). *Oppdragsbrev til Utdanningsdirektoratet* [Commissioning letter to the Directorate of Education and Training]. Oslo: Ministry of Education and Research.
- MER. (2007). *Oppdragsbrev nr. 06 om tiltak knyttet til individvurdering i skole og fag- og yrkesopplæring* [Commissioning letter No. 06 on actions related to individual assessment in school and vocational education and training]. Oslo: Ministry of Education and Research.
- Moe, O. (1997). *Nasjonalt vurderingssystem for grunnskolen. Forslag fra utvalget* [National system for quality assurance in compulsory education: Committee proposals]. Oslo: Kirke-, utdannings- og forskningsdepartementet.
- NOU (Green Paper). (1974). *Karakterer, eksamener, kompetanse m.v. i skoleverket* [Marks, examinations, competence, etc. in the education system]. Oslo: Statens forvaltningstjeneste.
- NOU (Green Paper). (2002). *Første klasser fra første klasse* [First class from Year 1]. Oslo: Statens forvaltningstjeneste.
- OECD (Organisation for Economic Co-operation and Development). (1988). *OECD-vurdering av norsk utdanningspolitikk* [OECD-assessment of Norwegian education policy]. Oslo: Kirke- og undervisningsdepartementet.
- OECD. (2008). Teaching and Learning International Survey (TALIS). *Resultater fra OECDs internasjonale studie av undervisning og læring* [Results from OECD's international study of teaching and learning]. Paris: OECD.
- OECD. (2011). *OECD reviews on evaluation and assessment in education – Norway*. Paris: OECD.
- Prøitz, T. S. (2013). Variations in grading practices – subjects matter. *Education Inquiry*, forthcoming.
- Prøitz, T. S., & Spord Borgen, J. (2010). *Rettferdig standpunktavurdering – det (u)muliges kunst?* [Fair overall achievement marking – the (im)possible art?]. Oslo: NIFU STEP.
- Seland, I., Vibe, N., & Hovdhaugen, E. (2013). *Evaluerer av nasjonale prøver som system* [Evaluation of national tests as a system]. Oslo: NIFU Nordic Institute for Studies in Innovation, Research and Education.
- Skedsmo, G. (2011a). Formulation and realisation of evaluation policy: Inconsistencies and problematic issues. *Educational Assessment, Evaluation and Accountability*, 23, 5–20.
- Skedsmo, G. (2011b). Vurdering som styring av utvikling og overvåking av resultater [Assessment as steering of development and monitoring of results]. In J. Møller & E. Ottesen (Eds.), *Rektor som leder og sjef. Om styring, ledelse og kunnskapsutvikling i skolen* [The principal as leader and boss. About steering, leadership and knowledge development in school] (pp. 74–94). Oslo: Universitetsforlaget.
- Stobart, G. (2012). Validity in formative assessment. In J. Gardner (Ed.), *Assessment and learning* (2nd ed., pp. 133–146). London: Sage.

- Stortingsmelding (White Paper) nr. 47. (1995–1996). *Om elevvurdering, skolebasert vurdering og nasjonalt vurderingssystem* [Student assessment, school-based evaluation and national quality assurance system]. Oslo: Kirke-, Utdannings- og forskningsdepartementet.
- Stortingsmelding (White Paper) nr. 28. (1998–1999). *Mot rikare mål. Om einskapsskolen, det likeverdige opplæringstilbodet og ein nasjonal strategi for vurdering og kvalitetsutvikling i grunnskolen og den vidaregåande opplæringa* [Towards richer goals. About the unity school, equivalent provision of education and a national strategy for assessment and quality development in primary and secondary education]. Oslo: Kirke-, Utdannings- og forskningsdepartementet.
- Stortingsmelding (White Paper) nr. 30. (2003–2004). *Kultur for læring* [Culture for learning]. Oslo: Utdannings- og forskningsdepartementet.
- Stortingsmelding (White Paper) nr. 16. (2006–2007). *... og ingen stod igjen* [... and nobody was left behind]. Oslo: Kunnskapsdepartementet.
- Taras, M. (2009). Summative assessment. The missing link for formative assessment. *Journal of Further and Higher Education*, 33, 57–69.
- Thronsdén, I., Hopfenbeck, T. N., Lie, S., & Dale, E. L. (2009). *Bedre vurdering for læring* [Improved assessment for learning]. Oslo: Institutt for lærerutdanning og skoleforskning.
- Tønnessen, L. K. B., & Telhaug, A. O. (1996). Elevvurderingen i norsk skole i etterkrigstida [Student assessment in Norwegian school post World War II]. In K. Skagen (Ed.), *Karakterboka. Om karakterer og vurdering i ny skole* [The marking book. About marks and assessment in new school] (pp. 15–34). Oslo: Universitetsforlaget.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, 14, 281–294.
- Utdanning. (2006, May 5). *Skepsis til nasjonale vurderingskriterier fra 2. trinn*. [Sceptical to national assessment criteria from Year 2]. Oslo: Utdanning. Retrieved from <http://utdanningsnytt.no/4/Meny-B/Grunnskole/Pedagogikk/Skepsis-til-nasjonale-vurderingskriterier-fra-2-trinn/>
- Vibe, N., Aamodt, P. O., & Carlsten, T. C. (2009). *Å være ungdomsskolelærer i Norge. Resultater fra OECDs internasjonale studie av undervisning og læring* [To be a lower secondary teacher in Norway. Results from OECD's international study of teaching and learning]. Oslo: NIFU STEP.
- Wikström, C. (2006). Education and assessment in Sweden. *Assessment in Education: Principles, Policy & Practice*, 13, 113–128.