

Chapter 12

Educational Assessment in Norway—A Time of Change

Sverre Tveit

Introduction

This chapter discusses the context, problems, historical background and new approaches to student assessment in Norway. The beginning of the 21st century can be characterised as a time of change in Norwegian education. Based on the disappointing results obtained by Norwegian students on international comparative tests, the educational reforms of the 1990s were determined to have failed. Following implementation of a new reform, ‘the Knowledge Promotion’ of 2006, the regulations and practices for student assessment are among the areas that are being questioned.

Norway is distinguished from many countries by not grading students through formal assessments until they are 13–14 years old. Few countries have seen the ideological fights over student assessment that have occurred in Norway; conflicts which reached a peak in the 1970s when a national committee suggested the further abolition of formal grading in lower secondary school. Although the fight against formal grading did not result in any significant immediate changes to practices of the time, the controversy has remained a latent conflict that typically ignites when new educational reforms are introduced. In 2004, assessment again became a matter of considerable national policy debate when a controversial new framework for national testing was introduced. As authorities and schools are developing new regulations and practices for student assessment for the new curricula of 2006, controversy continues.

The first section of the chapter briefly presents the context of educational assessment in Norway. The next section discusses problematic issues related to student assessment, such as comparability of teachers’ judgments, external examinations and formative feedback. The third section is a short, retrospective analysis of the evolution of assessment regulations in Norway in the 20th century. I argue that the lack of theoretical foundation of the approaches to student assessment is one of the main causes for the problems now faced. The fourth section presents and comments on the recent initiatives by the Norwegian government to address identified

S. Tveit (✉)

Postgraduate student, University of Oslo, Norway and Griffith University, Queensland, Australia
e-mail: sverrtv@student.uv.uio.no

problems and to develop a thorough approach to student assessment. The chapter concludes by identifying three key strategies considered crucial for Norway to develop a successful approach to student assessment for the 21st century.

Investigation of regulations, government documents and research studies in Norway form the framework for the chapter. The text reflects my background as a student representative, calling for reforms in the systems for educational assessment in Norway and continued work in learning and understanding assessment theories and practices in other nations.

The Context of Educational Assessment in Norway at Glance

Being previously a part of Denmark for more than 400 years (1380–1814), and for almost a century in a union with Sweden (1814–1905), Norway shares important parts of its history with its Scandinavian neighbours. ‘The countries are tied together through history, common cultural traditions, the same basic values and the same democratic ideals’ (Lysne, 2006, p. 329). Continuous assessment, conducted by the teacher, is the most significant type of summative assessment in Norway. This allows emphasis to be placed on students’ development over the course of the school year, and the development of global cognitive areas such as creative and collaborative skills. This clearly reflects social constructivist theories of learning. The constructivist theoretical rationale for teaching and learning in Norway has its basis in the work of Piaget and Vygotsky. The ‘Core Curriculum’, which was introduced in 1993 and still forms the basis for all subject-specific curricula, has a number of references to these theoretical perspectives (NOU, 2003, p. 15).

Organisation of Education in Norway

For most Norwegians, education is free throughout the compulsory and upper secondary years, and even tertiary years. There are 10 years of compulsory education, starting when children are 6 years old. In 2007, 97 per cent of Norwegian students undertook the compulsory years in schools governed by local municipalities; 93 per cent attended upper secondary schools governed by the regional municipalities¹ (Statistics Norway, Undated-a). In 2007, 96 per cent of students continued to upper secondary education after completing year 10 (Statistics Norway, Undated-b). Fifty-six per cent of these students attended 3-year upper secondary programs for general studies, which qualify for university or university college enrolment, while the rest

¹ The term compulsory education refers to the rights and obligations children have for education. Parents may teach their children themselves, but in practice almost all children undergo the compulsory years in schools. Private schools in Norway can get 85 per cent public funding; however this is only granted to schools that represent religious or pedagogical alternatives to the public schools. As the extent of private schooling is limited, the arrangements for assessment in these schools are not discussed.

attended vocational educational programs, which are normally 2 years of schooling followed by 2-year apprenticeship programs (Statistics Norway, Undated-a).

Democratic Education

In order to understand the reactions of students to changes in national assessment policy, and the government response to the student actions, it is necessary to understand the power that students in Norway have in affecting policy. An important aspect of Norwegian education is the emphasis on democratic education. Not only are there powerful teacher unions in Norway, but school students also have their own union protecting their civil rights and allowing the expression of students' opinions in public debates and towards the government and regional authorities.

Norwegian students have the opportunity to have their views taken into consideration through student councils comprised of class representatives. These councils comprise students starting as early as year 5 (when the students are 10–11 years old), with members gaining more responsibility as they get older. Interestingly, Norwegian students have excelled in international comparative studies of students' basic knowledge and skills about the democracy and democratic institutions (Mikkelsen, Buk-Berge, Ellingsen, Fjeldstad, & Sund, 2001, p. 240).

The school-student union of Norway (hereafter called 'the student union'), for which policy is decided by student council representatives from all 350 member schools and groups (mainly public upper secondary schools), is an important participant in the public debates over education and has showed strength in two particular cases related to student assessment: the fights over national testing and the calls for a thorough review of the regulations and practices for student assessment. Both are discussed further in the last sections of the chapter.

General Concept of Student Assessment

New national curricula were introduced with the educational reform known as 'the Knowledge Promotion' in 2006. The new curricula state competence aims (Norwegian terminology) the students are to achieve, but provide few regulations or requirements on how to organise teaching and how to assess student learning. The previous curricula had more comprehensive statements about what the students were meant to do. Standards for assessing the students' learning outcomes, however, have never been stated on a national level. This may change following the Norwegian government's initiative in 2007 to conduct a thorough investigation of the entire assessment system, including trialling national criteria for student assessment, as discussed later.

A strong tradition in Norway is the absence of grading of students until they attend lower secondary school (13–14 years old). Since 1972, the policy in Norway has been not to grade students until it is necessary for selection purposes in lower

secondary schools (NOU, 1974, p. 16). There have been strong advocates for the abolition of student grades at this schooling level, as well. The fight against formal grading in the 1970s did not result in any significant immediate changes of existing practices, but the controversy has been a latent conflict that has had a major impact on the evolution of the assessment systems present today (Lysne, 2006, p. 327). The absence of formal grading in primary school is different from many other countries, and calls for an explanation. Rationales for grading students typically follow three lines—selection, motivation and information (Wikström, 2006)—all of which appear to have been generally rejected in Norway.

In Norway, most students go to the local school they ‘belong to’ throughout the compulsory years of education, hence no selection procedures are necessary until towards the end of lower secondary school. The notion that students should be graded for the sake of their motivation has been repeatedly rejected over concerns about the negative impacts on low achievers (Tønnessen & Telhaug, 1996, p. 25). In discussions about assessment principles, the information purpose is typically being addressed as a key argument for preferring criterion-referenced assessment over norm-referenced assessment. The argument is that grades should express what the students can achieve, and how they progress, rather than comparing their performances relative to that of others (Wikström, 2006, p. 118). As discussed below, Norway has never had a distinct criterion-referenced approach; hence grades are less likely to provide information about progress and achievement levels.

High-Stakes Assessment

Referring to their importance in qualification procedures for education and jobs, grades reported on the students’ transcripts are often called ‘high-stakes’ assessment. The Norwegian approach to high-stakes assessment is to a large extent dependent on teachers’ assessment. Students receive one grade for each subject they study, apart from language subjects, which have separate grades for written and oral achievements. There are no national regulations on how these grades should be determined; typically they are based on a number of tests, assignments and other student work that have been graded throughout the year.

In Norway, the national authorities’ responsibility for high-stakes assessment has traditionally been limited to the examination system. On the one hand, the national authorities have produced and organised the grading of *external* examinations; on the other hand, they have been responsible for regulation of the *local* examinations. While the external examinations are in written form only, local examinations can take written, oral or practical forms. The local and regional municipalities are responsible for producing, implementing and grading local examinations, responsibilities that typically have been delegated to schools or responsible teachers within schools.

In principle, students can be sampled to undertake examinations in every subject; however, in practice, each student only undertakes a few examinations. Students are

normally not notified about what examination they have been sampled to sit until two days before the examinations occur, providing an incentive to prepare well for all subjects throughout the whole school year.

In *lower* secondary school, a student undertakes one external exam and one oral local examination at the end of year 10. In *upper* secondary school, the numbers of examinations vary according to programme of study. In the programme for general studies, 20 per cent of the students are sampled 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 to four examinations. The type of examination varies according to the programmes of specialisation the students are enrolled in. However, all students undertake the external examination in the Norwegian Subject Curriculum (Utdanningsdirektoratet, 2007b). The examinations are reported separately on the students' transcripts and typically make up less than 20 per cent of the grades on their transcripts, while the continuous assessment conducted by teachers in each subject forms the rest of the transcripts (Tveit, 2007b, pp. 194-195, 212).

Global Influence on Education and Student Assessment Policy

International comparative tests of student achievement, which have been conducted over the past 5–10 years (PISA—Programme for International Student Assessment; PIRLS—Progress in International Reading Literacy Study; TIMSS—Trends in International Mathematics and Science Study), have had a major impact on the public debate about schooling in Norway in general. As Norway is among the countries in the world that spend the most money on education (Utdanningsdirektoratet, 2007a, p. 34), one would expect that students in Norway would be among the highest achievers in these tests. However, this is not the case. Despite a few comparative studies conducted in the 1980s and 1990s, which could have triggered an alert, it was not until the results of the PISA2000 tests were published in 2001 (Lie, Kjærnsli, Roe, & Turmo, 2001) that the quality of the education system became a great public concern in Norway.

PISA2000 placed Norway as number 13 of the 31 OECD countries in reading and science, and number 17 in mathematics. These rankings made the front pages of all the major newspapers in Norway, with headlines such as 'Norway is a school loser' and 'Typical Norwegian is average'. The fact that the 'winner' in this 'contest', Finland, is one of Norway's neighbouring countries made the results even harder to accept. The Norwegian Ministry of Education was famously quoted as saying 'This is disappointing, almost like coming home from a winter Olympics without *one* Norwegian medal. And this time we cannot accuse the Finns of being drugged' (my translation) (Bergesen, 2006, p. 41). The PISA2003 and PISA2006 studies confirmed the dismal picture (Kjærnsli, Lie, Olsen, Roe, & Turmo, 2004). In PISA2006 Norway generally scored significantly lower than the mean of the OECD countries in science, reading and mathematics (Kjærnsli, Lie, Olsen, & Roe, 2007, pp. 18, 24).

For Norway, these results are identified as not acceptable by authorities and the general public and have become a driving force towards reforms in the education

systems in general. So, for assessment systems, in 2004, the Norwegian Minister of Education introduced new systems for monitoring the quality of the education system, of which national testing was a key component. As the outcomes of the evaluations of the educational reforms of the 1990s were published, possible causes for the disappointing results were addressed, providing the foundation for a comprehensive reform of primary and secondary schooling. The name of the reform, 'the Knowledge Promotion', reflects the main goal of the reform: to improve the learning outcomes for all students (Kunnskapsdepartementet, 2007, p. 2). A thorough investigation of the regulations and practices for student assessment were to accompany the reform after its implementation in 2006.

Problematic Issues in Educational Assessment in Norway

In this section, some problematic issues in educational assessment in Norway are discussed. The discussion is limited to comparability of assessments and teachers judgments, confusion about what to assess and lack of formative feedback.

Comparability of Assessments and Teacher Judgments

In 2003, the Norwegian authority responsible for the development of primary and secondary education (now called the Norwegian Directorate for Education and Training) concluded that 'there appear to have evolved different cultures for grading students across schools, subjects and teachers' (my translation) (Læringscenteret, 2003, p. 35). In a white paper presented to the Norwegian Parliament in 2006, the government stated that the regulations for student assessment are not clearly understood, there is not sufficient competence in student assessment, in teacher education and in schools, and little research on student assessment has been conducted (St.meld. nr. 16, 2006–2007, p. 79).

There seem to be few studies of comparability of teachers' judgments across schools in Norway (Tveit, 2007a). Two statistical studies that compare the grades given in continuous assessment with grades achieved on external examinations indicate significant differences in the interpretations of standards across schools (Hægeland, Kirkebøen, & Raaum, 2005; Kristensen, 1999, Unpublished.). Neither of these studies, however, investigates teachers' judgments of student work; there appear to be no such studies of how continuous assessment is conducted in Norway. A study of teachers' judgments of achievement levels in the test of writing in English, which formed part of the evaluation of the national testing framework in 2005, revealed that the teachers lacked references to an overall standard and used the group as a reference (Lie, 2007, p. 89).

As previously mentioned, examinations play a significant role in high-stakes assessment in Norway, supplementing the continuous assessment conducted by the teachers. A problem, however, is that the purposes and theoretical rationales for

these examinations, and how they are linked to the continuous assessment, are not defined (Tveit, 2007a, p. 215). Rather, the combination of the two can be seen as reflecting a compromise between two ideological positions towards student assessment. Those opposing external examinations typically argue that these tests do not measure relevant subject content. From this perspective, capacity to assess student work produced throughout a whole school year is a key argument for emphasising continuous assessment. Those calling for more external examinations typically argue that such tests provide a more reliable measure and prevent biased assessments undertaken by the teacher.

Those in favour of keeping or expanding the external examination system appear to base their point of view on an assumption that the external examinations represent an important incentive for the teachers to make valid and reliable judgments in their continuous assessment of students' achievement levels. This concept, however, has not been supported in research and can be questioned logically. Students' learning outcomes measured on a one-day examination, in a restricted assessment mode, can simply not be compared to their achievements measured throughout a number of assignments, tests and collaborative projects through the course of a year. Teachers are likely to argue that they know their students' achievements and abilities better than an examination can measure. This notion is understandable, considering the lack of theoretical linkage between the grades achieved in continuous assessment and the external examinations. This also appears to be the reason why there are no procedures for sanctioning teachers who may be making wrong judgments of the students' achievement levels (Tveit, 2007a, p. 215). Students have the right to lodge complaints on continuous assessment grades and examinations. For complaints on the continuous assessment, however, the court of appeal only considers whether the formal regulations have been followed—it does not judge the teacher's judgments on achievement levels. One reason for this lack of legal protection of students is the absence of requirements for collecting and storing evidence (Tveit, 2007a, p. 212; 2007b, p. 196).

While there is weak empirical evidence on the comparability of teacher judgments in the continuous assessment, a study of the external comparability of judgments in the external examination at the end of year 10 in the Norwegian subject curriculum indicated relatively high levels of comparability of judgments (Vagle, Berge, Evensen, & Hertzberg, 2007, p. 76). It can be assumed, however, that the level of comparability across schools in continuous assessment is considerably weaker than in the external examinations, as the national authorities' programmes for training teachers in making judgments on achievement levels historically have been reserved for the examination assessors only. As the vast majority of the students' school leaving certificates, which form the basis for tertiary education qualification in Norway, is based on continuous assessment conducted by the teacher, this is a major concern. The available evidence indicates that consistency in assessment has not yet been achieved. The absence of national criteria or standards, and the fact that the regulations do not state any requirements for moderation (neither social nor statistical), imply that significant injustice to individual students may occur.

Confusion About What to Assess

The lack of comparability across schools in Norway is due to a large extent to the lack of formal regulations for student assessment. This has resulted in confusion about what should be assessed: Should the students' efforts be assessed? How should their 'order' and 'behaviour' be assessed? Should verbal activity and participation in class be included in the judgments?

Grades for effort were abandoned in Norway in the 1920s. Since then there have been several attempts to reintroduce the concept, but those attempts have never succeeded, as such assessment clearly is difficult to do without bias (Lysne, 2004, p. 200; Tveit, 2007a, pp. 204–207). However, although such assessment has been politically discussed and rejected several times, it has not been clearly stated in the assessment regulations that students' effort should not be taken into consideration when assessing their work.

A study of upper and lower secondary school teachers' and students' views on assessment conducted by Dale and Wærness (2006) showed that teachers to a large extent take students' effort into consideration when grading (pp. 192–193). Verbal activity in class is often regarded as an indication of good effort. When placing too much emphasis on verbal activity without stating explicit criteria, one risks valuing activity regardless of the content and level of the students' contribution. Some students are better at this game than others, and receive better grades from the teachers by simply being verbally active. This disadvantages students who are well prepared for the class, but are less extroverted (Tveit, 2007a, p. 210).

According to Wynne Harlen (2004), strong evidence shows that 'using grades as rewards and punishment is harmful to students' learning by encouraging extrinsic motivation' (p. 5). Hence, it is important that subject grades are not used as a disciplinary tool. It should be clear to students that subject grades are judgments and feedback on the quality of their work only. In the assessment regulations in Norway, there are two specific grades for assessing students' 'order' and 'behaviour'. One should therefore expect that such factors were not an issue when grading students' work within each subject.

Dale and Wærness (2007), however, showed that many teachers tend to mix these two grading processes. One of the teachers in their qualitative study was quoted '[...] Positive attitude in the classroom is important. [This] *'Trynefaktor'* [Norwegian term for a sense of 'face factor'] should not be underestimated. It counts in the working life [...]' (my translation) (p. 194). Many students report that the so-called *trynefaktor*—the concept of judgments being biased based on whether one likes the appearance and attitude of the person or not—is a big problem. Although there are no known studies that have investigated this particular problem, it is a topic of general concern among students, parents and teachers.

Lack of Formative Feedback

Rumours about 'face factor' gain ground if students are not provided with good explanations for the grades they receive. In the Dale and Wærness (2006) study, a

student said that ‘I don’t really know what their assessment is based on. I never really get the criteria stated. We’ve asked for it before and then we’ve been told that it is an overall assessment’ (my translation) (p. 197). A reference to ‘an overall assessment’ does not help students understand what they should have been doing different to achieve better results. The image drawn by this student was confirmed in a survey undertaken in 2005. A majority of the students said that the teachers to a limited extent told them what they were good at and what they needed to improve (Furre, Danielsen, & Stiberg-Jamt, 2006, p. 62). An evaluation of the previous reform of education for the compulsory years also identified that teachers are not good at expressing their expectations to students. Considerable positive feedback was given; however, this feedback was not sufficiently based on the quality of the students’ achievements (Klette, 2003, p. 53).

The lack of explicitness in assessment criteria must be understood in the context of national curricula, which historically do not state how to assess levels of achievement. In guidelines developed for the grading of state-wide external examinations, however, there are nationally stated criteria. Dale and Wærness (2007) have also examined the criteria used in some of these examinations, which are commonly used by teachers in the continuous assessment in the following years. They found that the criteria used for describing achievement levels on all levels but the two highest are written and characterised with reference to the highest level and the lack of obtaining those standards. If these criteria were to be used in the continuous assessment, low achievers would be described in terms such as ‘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’ (my translation) (Dale & Wærness, 2007, p. 106). Such an approach to stating expectations to students has limited formative assessment effect and is not useful for assessment practices in the classroom.

Underlying Factors for the Problems of Educational Assessment in Norway

In this section I argue that one of the underlying causes of the confusion about student assessment in Norway is an indistinct adoption of fundamental principles for student assessment. Historically, two distinct approaches to student assessment have been applied in the Western world: *norm referencing* and *criterion referencing*. Generally, one can say that in terms of theoretical development of the concepts in international literature, the former had its peak in the first half of the 20th century, while the latter concept gained ground in the second half and was dominating towards the beginning of the 21st century, whether approached through assessment or measurement paradigms.

The Evolution of the Assessment Regulations

In Norway, grade inflation became a major concern in upper secondary education in the 1920s, when a statistical review showed that the distributions were very

much skewed, a concern shared by most of the countries in the Western world. The first national regulations for grading came with the national model plan of 1939 (*Normalplanen*), in which a principle of norm referencing based on a normal distribution was implemented for a five-step scale: *Outstandingly good*: 4%, *Very good*: 24%, *Good*: 44%, *Fairly good*: 24%, *Good*: 4%. It was made clear that the distribution guide should apply only to large groups, and not small groups such as single classes. ‘The so-called “relative grading system” had never been critically discussed, not even the fact that it doomed a certain percentage of the students to fail’ (Lysne, 2006, p. 343). An obvious problem was that most teachers did not have any references to the levels of other students than their own and, therefore, to a certain extent, applied normal distributions within their own groups. This concept typically results in its being easier to get a high grade in a low-achieving class and vice versa (Wikström, 2006, p. 118). In the 1960s and 1970s, new concepts of grading, inspired by United States theories, collided with calls from the political left wing in Scandinavia for the total abolition of formal grading (Egelund, 2005, p. 208). According to Lysne (2006), Bloom and Tyler had a great influence on theory and practice of education in Scandinavia in the 1960s. In 1963, Robert Glaser defined the distinction between criterion-referenced and norm-referenced testing and assessment, and within a few years a significant body of literature had been written about the concept of criterion-referenced assessment (Lysne, 2006, p. 348).

At this time, the frustration with regulated distribution of grades reached its peak in Norway and formed the ground for what can be argued was the most passionate debate over school politics in Norway ever, perhaps unparalleled in the world (Lysne, 2004, p. 113). The ultimate consequence of the system—that it made classmates compete against each other—was used as one of the arguments for the total abolition of formal grading. A way out of the problems with norm-referenced grading had to be found. For those opposing formal grading, most commonly among the influential politicians and researchers on the left wing in the 1970s, being compared to an objective standard, as criterion-referenced assessment was represented, was no better than being compared with other students. They feared too much emphasis on knowledge as a means to monitor and control (Lysne, 1999, pp. 37, 39).

The 1980s in Norway was the time for recovering from the debate over formal grading of the 1970s. The upper secondary school students’ union had in 1978 collected 60 000 signatures for retention of a grading schema in both lower and upper secondary schools. The Labour party lost the majority in the Parliament after the election in 1981 and attempts to abolish formal grading at these levels were not continued by the incoming government (Lysne, 2004, p. 113).

Instead, the new conservative government used the opportunity to introduce what Lysne (2004, p. 120) calls ‘adapted goal-referenced assessment’ (my translation of the Norwegian term ‘*tillempet målrelatering*’) (p. 120), which had already been applied in upper secondary school since 1968. The concept of ‘adapted goal-referenced assessment’ was to abandon the norm-referenced principle but not to entirely apply a criterion-referenced system: absolute learning outcomes or evidence of specific skills should not be required in the curricula. Emphasis was placed on the

choice of learning content, teaching aids and learning methods that were focusing on the students' personal development towards a wide range of goals (Lysne, 2004, pp. 120, 197). The teachers, however, had doubts about how to put this into practice. In an attempt to clarify the regulations 'the norm-referenced model was involved again, and thus the new system became sort of a bastard of those two principles' (Lysne, 2006, p. 352).

Assessment Regulations in the Reforms of the 1990s

When new curricula for the entire school system were introduced in Norway in the 1990s, it was attempted to formalise the principle of 'adapted goal-referenced assessment' by the introduction of a core curriculum that stated the general objectives of schooling in 1993. In 1994, a form of holistic assessment was prescribed in an operational directive for the new reform of upper secondary school. A concept of 'overall competence' (my translation of the Norwegian term '*helhetlig kompetanse*'), which referred to a broader understanding of competence than subject-specific competence, was to be emphasised when grades were given in each subject. This concept included students' ability to cooperate with each other and develop their personality and character (Lysne, 2006, p. 353). It is easy to agree that these are important aspects of schools' and teachers' mandates. A problem, however, is that if there is no clear-cut distinction between these general goals and the assessment of specific learning objectives, there is again a great risk of the assessment being biased (Tveit, 2007a, p. 206).

Sweden had been experimenting with regulations for holistic assessment in the 1970s; however, the concept was abandoned there a few years later. According to Lysne (2006), a special feature of Norwegian school politics is that '[m]ost of the ideas Sweden had tried and found not to work, Norway was inclined to try over again, and as a rule to come to the same conclusion as in Sweden' (p. 354). In Norway, the principle of 'overall competence' made the confusion about what the learning objectives were and how to make judgements on the students' achievements even worse. As Lysne (2006) noted, (p. 353) teachers and parents protested spontaneously because they could not understand what was meant—the directive was withdrawn and replaced by a new one that was more in accordance with traditional prescriptions for grading. However, the regulations remained obscure. In the preparations for the educational reform of 2006, the Norwegian government acknowledged that the concept of emphasising holistic assessment had contributed to the obscurity of the regulations for student assessment (St.meld. nr. 30, 2003–2004, p. 39).

Despite the alleged rejection of the concept of norm-referencing, the prevailing regulation during the educational reform for upper secondary school of 1994 (Reform 94) had distinct references to norm-referenced assessment: the six grade levels were grouped in three levels 'Above average' (6 and 5), 'Average' (4 and 3) and 'Below average' (2 and 1). This way of expressing the achievement levels can be seen as legitimating the concept of applying a norm-referenced strategy to grade

small groups of students (Dale & Wærness, 2007, p. 105). While other parts of the assessment regulations were changed when implementing the new reform in 2006, this remained unchanged. I argue that the government's reluctance to change this essential part of the regulations was owing to apprehension of reigniting the debate over formal assessment—and thereby jeopardising the generally broad consensus about the new reform.

New Approaches to Students' Assessment in the 21st Century

The new educational reform of 2006 implied a wide range of changes in response to the disappointing results on the international comparative tests in the first years of the 21st century. The introduction of *basic skills* and *competence aims* (learning outcomes) are among the most significant changes in the curricula. The basic skills are defined as being able to *express oneself in writing and orally, being able to read, do mathematics and use digital tools* (Utdanningsdirektoratet, 2007a, p. 73). All basic skills were attempted to be included in all subject curricula; the concept being that all teachers share a responsibility for supporting the students' development of the basic skills.

The curricula state the competence aims the students are expected to achieve at the end of years 2, 4 and 7 in primary school, the end of lower secondary school (year 10) and for each year in most subjects in upper secondary school. The introduction of competence aims is a response to the lack of explicitness in the learning objectives addressed earlier; a move accompanied by a range of changes in the regulations of student assessment. It was made clearer that subject-specific achievements, on the one hand, and behaviour, order and effort, on the other hand, should be kept apart, clarifying for students the basis for judgment of students' achievement levels. Only evidence of achievement of the competence aims should form the judgment when grading students. The curricula, however, include no shared criteria or standards for judging whether the competence aims are being achieved. Instead, the local authorities and schools were encouraged to develop criteria themselves.

New Approach to National Testing

As mentioned in the introduction, an initiative for national testing of students' basic skills had been introduced already in 2004, 2 years before the implementation of new curricula. Back in 1988, a report from the OECD (1988) had questioned whether Norway had sufficient tools for monitoring the quality of its education system. Throughout the 1990s, a system for national evaluation of schooling was discussed in several documents by the government and the Parliament, but no centrally coordinated system was implemented until an official committee suggested so in 2002. Then, the first results of the PISA tests had already ignited a heated debate about the quality of schooling in Norway. National tests, a key element of

the national quality monitoring system, were implemented hastily, and the government highly misjudged the controversies that evolved from their initiative. Not only were there substantial ideological battles over the principle of publishing the schools results and thereby promoting competition, but the quality of the tests was also found by researchers to be poor (Hølleland, 2007).

The government, however, continued its approach despite critique from researchers, teacher unions and the student union. When the second circle was introduced in 2005, it became clear that the student union had achieved significant support for the boycott actions it had started the year before. Between 36 and 45 per cent of students boycotted the tests in mathematics, reading and English writing in upper secondary school (Hølleland, 2007, p. 30). The boycott established the controversies over national testing as one of the key issues addressed by politicians in debates over education policy before the parliamentary elections later the same year. Following the election, a new government was formed, which soon announced that it would follow the advice of the researchers who had conducted an even more critical evaluation, and institute a one-year moratorium in order to develop a more solid framework for national testing. In this process, more influence has been given to the stakeholders, and the teacher and student unions have been making a significant contribution to the debate about how the testing framework can be improved (Hølleland, 2007, p. 37; Lie, 2007, p. 88).

While merely the question as to whether publishing schools' results on the tests along with their mean grades was the main concern addressed in the media, the Norwegian 'experiment' was interesting from a research perspective. It revealed a number of problems related to student assessment in general, which provided strong arguments for a broader examination of the systems and regulations of student assessment that were to come (Tveit, 2007a, p. 202).

The problems of the first tests, acknowledged later, was indistinctiveness on at least two levels: the purposes of the tests and the design of the different types of tests (Utdanningsdirektoratet, 2007a, p. 82). Furthermore, there were mixed messages on whether the results should be taken into account when giving final grades for the individual student, or whether they were measures of group performance only (Hølleland, 2007, p. 37). The new framework introduced in 2007 was limited to year 5 and year 8 in primary and lower secondary school, in mathematics and in reading in Norwegian and English (Utdanningsdirektoratet, 2007a, p. 82). The implementation of these tests was more in accordance with the regulations than the implementation of 2005 (Kavli, 2008, p. 4).

By having the test at the outset of the mid-years and lower secondary, it is made clear that the tests for the students are low stakes. The purpose of the test framework is now primarily defined as giving 'information about the group and the year set to teachers, school owners, local authorities and the regional and national level as the basis for improvement and development activities' (Utdanningsdirektoratet, 2007a, p. 82). The results, based on a common scale, will be published and made available to the public; however, the ministry will not introduce a ranking of schools.

New Approach to Criterion-Referenced Assessment?

While the approach to student assessment for quality monitoring purposes was hastily implemented *before* the new curricula, the policy on how to assess the individual students' achievements was not discussed until *after* schools had started to use the new curricula.

In the white paper to the Parliament of 2003, which prepared the new educational reform, it was stated that the continuous assessment should be 'standard based' (St.meld. nr. 30, 2003–2004, p. 40). The theoretical rationale for this assessment principle, however, was never explained in any government documents (a task that was still unaccomplished at the time of publication). Perhaps the government had neither the courage nor the time for the ideological battles over student assessment that surely awaited it; a somewhat latent conflict it indeed was a part of itself. As a result, these fights were postponed until after the new curricula had been implemented. Although it was predictable for those involved in preparing the reform, it was not until the examinations were to be prepared and the teachers were to conduct their final assessments based on the new curricula that controversies over student assessment reached the surface.

In a white paper to the Norwegian Parliament in December 2006, the new government mandated a thorough review of the regulations and systems for student assessment in order to achieve 'more equal and fair student assessment' (St.meld. nr. 16, 2006–2007, p. 79). In August 2007, additional changes were made to the regulations, acknowledging the critique that had been raised about legitimatising the concept of norm-referencing by the way achievement levels were described in the regulations. The achievement levels were to be described on a continuum scale from 'very low competence in the subject' (grade 1) to 'outstanding competence in the subject' (grade 6). At the same time, the Directorate for Education and Training introduced a project called 'Improved Assessment Practices' (*Bedre Vurderingspraksis*), which aims to investigate four models for developing a shared understanding of achievement levels across schools within the subjects Norwegian, mathematics, social science and food and health. The three models being trialled in primary schools range from the schools themselves developing and trialling criteria for one (high) or two (high and low) achievement levels, to trialling pre-stated achievement levels for high and low achievement. In the fourth model, lower and upper secondary schools are developing and trialling achievement levels for each of the six levels of competence stated in the general regulations for grading (Utdanningsdirektoratet, 2007c, p. 12).

This strategy is a step in the right direction, in order to facilitate shared understanding across schools and thereby more consistent and fair grading. However, the approach still suffers heavily because of the lack of theoretical rationale for assessment in the fundamental basis for making judgments—the curricula. A study of experiences with the early launch of the new curricula showed that student assessment was the field where most teachers (65 per cent) expressed that they needed more training in relation to the new reform (Bergem, Båtevik, Bachmann, & Kvangarsnes, 2006, p. 28). Student assessment was therefore stated as one of the national

priority areas for professional development in 2007 (Utdanningsdirektoratet, 2007a, p. 36). Projects for trialling assessment criteria and professional development programs are crucial; however, for this to be successful the approach has to be grounded in theory. Whether Norwegian politicians and stakeholders are able to overcome the ideological controversies that have characterised the debate on student assessment since the 1970s and develop a distinct approach to criterion-referenced assessment remains an unanswered question.

Conclusion

What can we learn from the history of educational assessment in Norway? The past decades' ideological controversy over assessment policy can be seen as one of the reasons Norwegian politicians have failed to address the practical problems in the present assessment systems. The theory and practice of educational assessment in Norway can be seen as weak, owing to ideological rather than informed debates over fundamental principles regarding student assessment. While different theories may be seen as underpinning the different styles of assessment that have operated since the early 20th century, these links and bases have not been made overt. The lack of theoretical foundation of the assessment systems may reflect a fundamental fear of acknowledging what formal grading essentially expresses—that some students perform better than others.

The absence of formal grading in primary schools implies that student assessment, particularly in terms of grading, is not likely to form an important part of the teachers' competence. Teachers in Norway appear not to have been provided with sufficient training in student assessment. While the wide concerns for comparability of teacher judgments across schools appear to be owing to a fundamental theoretical and conceptual problem of the approach to student assessment, concerns about the teachers' practical competence is of equal importance.

For Norway to develop a successful approach to student assessment for the 21st century, three strategies can be seen as crucial. First, assessment regulations and practices need to be grounded in theory. This includes both the continuous assessment and the examinations, and not least the linkage between the two. By applying a distinct approach to criterion-referenced assessment, Norway can benefit from the experiences and theoretical concepts of the international community.

Second, considerable changes in the method of administration and reform of education should be applied. In Norway, problems in the educational system tend to have accumulated over a number of years, and often after approximately 10 years an extensive reform has been initiated. In a constantly more global and competitive world, politicians are inclined to introduce major reforms to address the problems their country is facing. Educational systems, particularly student assessment, are vulnerable in such reform strategies. It would be preferable to plan the education system on predictable cycles of revision of the curricula and to establish permanent arenas for professional development in relation to this.

Third, but not least, new strategies for giving the teaching profession the instruments and confidence necessary to make judgments on students' achievement need to be introduced. Resources should be provided to ensure that all teachers can develop quality assessment instruments and arenas where teachers can share experiences and moderate each other's judgments on the quality of students' achievements.

The concepts and processes of grading students are among the most powerful institutions in society—future generations' dreams and ambitions rely to a large extent on the grades they receive in school. Acknowledging that student grading can never be perfectly just, policy makers should provide extensive resources to improve concepts and procedures for student assessment.

References

- Bergem, R., Båtevik, F. O., Bachmann, K., & Kvangarsnes, M. (2006). Tidleg oppstart med nye læreplanar: Kartlegging av erfaringar med førebuing og iverksetjing. *Arbeidsrapport no. 196*. Volda: Høgskulen i Volda.
- Bergesen, H. O. (2006). *Kampen om Kunnskapsskolen*. Oslo: Universitetsforlaget.
- Dale, E. L., & Wærness, J. I. (2006). Vurdering og læring i en elevaktiv skole. Oslo: Universitetsforlaget.
- Dale, E. L., & Wærness, J. I. (2007). Tilpasset opplæring og elevvurdering. In S. Tveit (Ed.), *Elevvurdering i skolen—grunnlag for kulturendring*. Oslo: Universitetsforlaget.
- Egelund, N. (2005). Educational assessment in Danish schools. *Assessment in Education: Principles, Policy and Practice*, 12(2), 203–212.
- Furre, H., Danielsen, I. J., & Stiberg-Jamt, R. (2006). *Analyse av den nasjonale undersøkelsen 'Elevundersøkelsen' 2006*. Kristiansand: Oxford Research.
- Harlen, W. (2004). *A systematic review of the evidence of the impact on students, teachers and the curriculum of the process of using assessment by teachers for summative purposes*. London: EPPI-Centre, Social Science Research Unit, Institute of Education.
- Hægeland, T., Kirkebøen, L. J., & Raaum, O. (2005). *Skoleresultater 2004. En kartlegging av karakterer fra grunn- og videregående skoler i Norge*. Oslo: Statistics Norway.
- Hølleland, H. (2007). Nasjonale prøver og kvalitetsutvikling i skolen. In S. Tveit (Ed.), *Elevvurdering i skolen—grunnlag for kulturendring*. Oslo: Universitetsforlaget.
- Kavli, H. (2008). *Nasjonale prøver 2007—Brukernes evaluering av gjennomføringen*. Oslo: Synovate.
- Kjærnsli, M., Lie, S., Olsen, R. V., & Roe, A. (2007). *Tid for tunge løft. Norske elevers kompetanse i naturfag, lesing og matematikk i PISA 2006*. Kortversjon. Oslo: Universitetsforlaget.
- Kjærnsli, M., Lie, S., Olsen, R. V., Roe, A., & Turmo, A. (2004). *Retten eller ville veier? Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2003*. Oslo: Universitetsforlaget.
- Klette, K. (2003). Lærernes klasseromsarbeid; Interkasjons- og arbeidsformer i norske klasserom etter Reform 97. In K. Klette (Ed.), *Klasserommets praksisformer etter Reform 97*. Oslo: Pedagogisk Forskningsinstitutt, Universitetet i Oslo.
- Kunnskapsdepartementet (2007). *Oppdragsbrev nr. 06 om tiltak knyttet til individvurdering i skole og fag- og yrkesopplæring*. Oslo: Kunnskapsdepartementet.
- Lie, S. (2007). Evalueringen av de nasjonale prøvene og hva vi kan lære av dem. In S. Tveit (Ed.), *Elevvurdering i skolen—grunnlag for kulturendring*. Oslo: Universitetsforlaget.
- 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. Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

- Lysne, A. (1999). *Karakterer og kompetanse. Stridstema i norsk skolehistorie*. Bind 1. Haslum: AVA Forlag.
- Lysne, A. (2004). *Kampen om skolen. 1970–2000. Karakterer og kompetanse. Stridstema i norsk skolehistorie*. Bind 2. Haslum: AVA Forlag.
- Lysne, A. (2006). Assessment theory and practice of students' outcomes in the Nordic Countries. *Scandinavian Journal of Educational Research*, 50(3), 327–359.
- Læringssenteret (2003). *Karakter—mer enn karakterer*. Oslo: Læringssenteret.
- Mikkelsen, R., Buk-Berge, E., Ellingsen, H., Fjeldstad, D., & Sund, A. 2001. Civic Education Study Norge 2001. Demokratiske beredskap og engasjement hos 9.-klassinger i Norge og 27 andre land. Oslo: Institutt for Lærutdanning og Skoleutvikling, Universitetet i Oslo.
- NOU (1974: 41). *Karakterer, eksamener, kompetanse m.v. i skoleverket. Norges offentlige utredninger*. Oslo: Statens forvaltningstjeneste.
- NOU (2003: 16). *I første rekke. Norges offentlige utredninger*. Oslo: Statens forvaltningstjeneste.
- OECD (1988). *OECD-vurdering av norsk utdanningspolitikk*. Oslo: Kirke- og undervisningsdepartementet.
- Statistics Norway (2008). Færre grunnskolar. Retrieved May 23, 2008, from <<http://www.ssb.no/emner/04/02/20/utgrs/>>.
- Statistics Norway (Undated-a). Emne:04 Utdanning. Tabell: 05326: Elever i videregående opplæring, etter studieretningstype og skolen sitt eierforhold (F). Reviderte tall for 2007. Retrieved May 29, 2008, from <<http://statbank.ssb.no/statistikkbanken/Default.FR.asp?PXSid=0&nvl=true&PLanguage=0&tilside=selecttable/hovedtabellHjem.asp&KortnavnWeb=vgu>>.
- Statistics Norway (Undated-b). Uvalgte nøkkeltall, fylkeskommuner – nivå 1. Gjennomsnitt alle fylkeskommuner 2007. Retrieved May 23, 2008, from <<http://www.ssb.no/kostra/stt/index.cgi?spraak=norsk&nivaa=1®ionstype=fylkeskommune&faktaark=101040642883067®ioner=default@default&kolonne=0&event=ny&mal=region&cookie=0>>.
- St.meld. nr. 30 (2003–2004). *Kultur for læring*. Oslo: Utdannings- og forskningsdepartementet.
- St.meld. nr. 16 (2006–2007). *... og ingen stod igjen*. Oslo: Kunnskapsdepartementet.
- Tveit, S. (2007a). Analyse av retningslinjer og praksis for elevvurdering. In S. Tveit (Ed.), *Elevvurdering i skolen—grunnlag for kulturendring*. Oslo: Universitetsforlaget.
- Tveit, S. (2007b). Formål og retningslinjer for elevvurdering i Kunnskapsløftet. In S. Tveit (Ed.), *Elevvurdering i skolen—grunnlag for kulturendring*. Oslo: Universitetsforlaget.
- Tønnessen, L. K. B., & Telhaug, A. O. (1996). Elevvurderingen i norsk skole i etterkrigstida. *Karakterboka. Om karakterer og vurdering i ny skole*. Oslo: Universitetsforlaget.
- Utdanningsdirektoratet (2007a). *The Education Mirror 2006. Analysis of primary and lower and upper secondary education in Norway*. Oslo: Utdanningsdirektoratet.
- Utdanningsdirektoratet (2007b). Høringsbrev om forslag til endringer i forskrift til opplæringsloven 23.06.2006 nr. 724 og endring i Læreplanverket for Kunnskapsløftet. Oslo: Utdanningsdirektoratet.
- Utdanningsdirektoratet (2007c). *Vurdering. Et felles løft for bedre vurderingspraksis—en veiledning. Revidert utgave*. Oslo: Utdanningsdirektoratet.
- Vagle, W., Berge, K. L., Evensen, L. S., & Hertzberg, F. (2007). Det umulige er nesten mulig. Om norskeksamnen på ungdomsskolen og utvikling av vurderingsnormer. In S. Tveit (Ed.), *Elevvurdering i skolen—grunnlag for kulturendring*. Oslo: Universitetsforlaget.
- Wikström, C. (2006). Education and assessment in Sweden. *Assessment in Education: Principles, Policy and Practice*, 13(1), 113–28.