

## *Developing national standards for the assessment of writing*

### *A tool for teaching and learning*

The goal of this project is to carry out in-depth research on specified assessment norms for writing and their effects on the quality of pupils' writing and teachers' rating. It builds on a 2010 pilot carried out with the support of the NFR. The pilot was a national investigation of teacher perception of norms<sup>(1)</sup> for the assessment of writing as a key competency, and was recently reported in Matre et al (2011).

#### **1 Relevance**

The ability to write is a key competency in a knowledge society (DeSeCo 2000) – in social, public and working life as well as in education. It is an important prerequisite for active participation in a modern democracy, but simultaneously an important source of current educational inequality. There is also a positive correlation between poor writing competency and marginalization in the labour market. It is thus important that every child is given the opportunity to develop his/her writing competency in order to participate fully.

Still, there is a mismatch between this strategic role of writing as well as its position in the Norwegian 2006 curriculum, on the one hand, and our current knowledge about how this key competency develops and can be assessed, on the other. Admittedly, Norwegian writing research has been gradually developing since the late 1980s (see for instance *Written Communication*, Vol. 19, 3/4, 2002, *Special issue on Norwegian Writing Research*). However, research on writing as a key competency is still relatively fragmented.

When the National Centre for Writing Education and Research opened at Sør-Trøndelag University College, Trondheim, in August, 2009, this situation improved. In addition to the 2010 pilot for the present project, a long-term project is now underway at the centre to develop a national test of writing as a key competency, at the start of grades 5 and 8. The long-term project is funded by the Norwegian Directorate for Education and Training as a new governmental tool for implementing the national curriculum. It has been decided that the aim of this test shall be double – to yield a) a national summative picture of learning outcomes and b) a formative basis for more targeted local teaching efforts. In this political context, the national centre is developing the conceptual and organizational basis for a new form of assessment in Norway. Here, the findings from our NFR 2010 pilot are now being integrated, in order to define a valid reference criterion for the assessment scale as well as to achieve higher inter-rater reliability. During 2011, the effect of this empirical criterion will be measured, as new assessors are being trained for membership in a new national panel of raters. The present proposal thus shares several research premises with the planned national test and its raters, and four members of the present proposal's research group are directly involved in developing the new test, one of the present group as that project's leader.

There is still, however, an urgent need for empirically based insight into how educational efforts may translate more systematically into learning outcomes. The current lack of empirically standardized, specified national norms of expected writing proficiency is an important case in point, and a core example of educational relevance, where several challenges converge.

#### **2 Scientific outline**

##### **2.1 Background and status of knowledge**

The national *Quality Assurance of Learning Outcomes* project (Berge et al 2005) was part of the NFR national evaluation of the Norwegian 1997 curriculum. This project collected statistically representative data from four years (1998-2001) surrounding the transition between exams under an old and a then new national curriculum, and offered elaborate empirical analyses of learning outcomes in one school subject (Norwegian as a first language) at one grade level (10<sup>th</sup> grade) in

one writing situation (final exam). Equally systematic studies are still lacking, however, for other subjects, grade levels and writing situations. Within current Norwegian writing research there is a particularly evident lack of studies relating to key competency developments at grade levels 5–7 in primary school, even if partial studies are underway (eg Hundal 2010). This lack is all the more crucial in a situation where Norwegian educational authorities have decided that a new set of writing tests are to be developed.

Subsequent to the QAL project, a full-scale population test of writing competency was developed in the period 2003–2005 and run in 2005. The analysis of results from this national test revealed both the need for and the importance of shared national norms of expectation (standards). In their final report to The Norwegian Directorate for Education and Training the test developers (Thygesen et al 2007; Fasting et al 2009) pointed out that the absence of a nationally shared set of specific standards for each of the grade levels involved might be a major reason why inter-rater reliability scores were generally low, even for aspects that are technically easy to assess (like orthography and punctuation).

This situation contrasts with what has been achieved in international studies. A decade of work on norm based teaching in US schools resulted in improved results on the National Assessment of Educational Progress test (Quality Counts 2006). Doherty & Hilberg (2007) report from a quasi-experimental study where 23 teachers and 394 pupils in grades 3 and 5 at two US primary schools took part. Their results show that a) teacher use of carefully planned, objective-oriented teaching predicted learning outcomes in comprehension, reading, orthography and vocabulary in a standardized test of reading and writing, and b) progress was higher for those pupils where teachers systematically used specified standards. The pupils who profited most from this approach were those with fairly low English language skills.

Today, the development and constitution of national standards for writing are stimulated by an international trend towards accountability and documented learning outcomes (Hamp-Lyons 2011), where Norwegian research should take part. For a long time international assessment research has been divided into two opposing strands; one of these (assessment *of* learning) emphasizes traditional validity and reliability issues, whereas the second strand (assessment *for* learning) views assessment as part of learning and teaching processes, with correspondingly wider validity concerns, including consequential validity (Messick 1989). An increasing proportion of research, however, combines teaching and learning processes with the product dimension of testing, emphasizing standards development based on professional practice for the benefit for pupils, teachers, leaders as well as assessment research (James et al 2006; Timperley & Parr 2010). This newer, third strand (assessment *as* learning (Dobson, Eggen, & Smith 2009; Eggen 2011), is based on Deweyan inquiry learning circles and combines a Tyler testing and standards setting tradition with a progressive, bottom up capacity-for-assessment tradition. Educational ideological contexts play a major role in this strand (Biesta 2008). Hence, multiple relations between ideological/semiotic resources like curricula and their development (Broadfoot & Black 2004) and professional capacity building programs are important (cf. Timperley & Parr 2010).

Different theoretical frameworks for assessing writing are discussed by George Hillocks jr. in his overview of writing assessment models and solutions in different states in the USA (Hillocks 2002). In his study, different norms and standards in local writing assessment models are presented, and the different frameworks are critiqued. More recently, the New Zealand literacy project with its well developed “National Standards” has created a stir internationally. The necessity of transparent norms is emphasized also by the *Conference on College Composition and Communication (CCCC)* in their “Position Statement” of writing assessment (CCCC 2009).

The development of standards is considered a fundamental improvement, but similarly a complex challenge (Brown, Glasswell & Harland 2004). The challenges are equivalent to those involved when developing valid instruments for essay testing (Coffman 1971, Cooper & Odell 1977, Breland et al 1987). These traditional challenges were recognized already in the early

literature of writing assessment (Diedrich, French & Carlton 1961), and were considered systematically in the IEA study of writing (Gorman, Purves & Degenhart 1988, Purves 1992). In a Nordic research context, challenges of assessment norms have been presented in different publications and research projects (Berge 1996, 2002; Fasting et al 2009).

## **2.2 Objectives and research questions**

On the basis of national needs and international research developments, we see it as a strategically important task for education in Norway to formulate three objectives that may be empirically investigated:

- to define which specified levels of writing competency should reasonably be expected at specified grades as a basis for setting explicit standards
- to investigate what effects a shared set of explicit standards for educational writing, across nationally distributed local learning ecologies, may have on the quality of teacher assessment, in particular for rater agreement
- to investigate what effects an integration of shared explicit standards in teaching, across nationally distributed local learning ecologies, may have on the quality of pupils' writing.

These objectives will be met by answering the following three empirical research questions:

- which specified norms of assessment can sustainably serve as explicit standards for writing in an educational assessment-as-learning environment?
- which specific effects do the educational integration of shared explicit standards have on the quality of teacher assessment?
- which specific effects do the educational integration of shared explicit standards have on the quality of pupils' writing?

Within the *Utdanning 2020* programme, this set of research questions places the current project largely under topic area B, but with clear relevance also to the interface between topic areas A and B. In particular, the project will contribute to investigating how assessment may promote learning at individual and aggregate levels, as well as forging new insight about connections between teacher practices of assessment and pupils' learning outcomes.

Valid standards depend on a theoretical construct of writing that may assure both construct and content validity. One of the major long-term contributions of the project that developed the full-scale 2005 national test of writing was to develop such a construct, based on a social semiotic framework (Halliday 1978). The core of the construct was built on the historical developments behind current writing practices outside school. On the basis of this conceptual platform a didactically transparent 'wheel of writing' was developed, where six underlying purposes in the unmarked case coincide with six immediately situated 'acts of writing' (Thygesen et al 2007).

This theoretical construct makes it possible to develop standards that control for one major pitfall in early writing assessment research – that of assuming that writing competency is a uniform, or narrow, phenomenon. To the contrary, international research has shown that writing competency is differentiated according to purposes, situations, educational cultures and types of writing (Purves 1992).

## **2.3 Research approach and choice of methods**

The present project has two stages – a pilot study carried out in 2010 and a main study which is the focus of the current application. In the main stage, a national sample of schools will be investigated, with a combination of quantitative and qualitative methods.

The first stage was a focal case study designed to explicate teacher standards of expected competency as an empirical basis for a full-scale second stage. The focal aim of the first stage was to empirically define national criterion-referenced indicators of competency for two grade levels, in

close collaboration with teachers from a small set of nationally distributed schools. The methodological approach for stage 1 was qualitative. Here, schools were selected to represent differences with respect to their geographical and demographic embedding; their teaching agendas, teaching materials, and assessment practices. Methodologically, data was collected through semi-structured small-group, dyad and individual teacher interviews, supplemented with pupil scripts and (where available) local assessment documents.

The pilot study yielded tentative standards for assessing writing as a key competency at the beginning of grades 5 and 8 (Matre et al 2011). Still, a significant finding was that the explicated norms divide into several subsets with different characteristics. One subset consists of fairly explicit shared norms that seem stable (for instance for spelling); a second subset consists of norms that are only partially shared (for instance for narrative organization) and a third subset consists of largely implicit norms that may only partially surface when confronted with researcher questions (for instance on communication and multimodality). Norms in the third subset may probably not be constituted at an inter-individual level. Neither the second nor the third subset seems stable.

This finding has significant implications for the methodological design of the second, main stage. In our 2009 application, the main stage was planned to be a largely quantitative, nationally representative study. Based on the pilot study, however, we realize that further development of the standards needs to be an integrated, initial part also of the main stage, as a Part 1. It is also important to study (as a Part 2) how writing is taught, what kind of texts are written and how these are assessed by teachers, as specified norms are being integrated in different local learning ecologies at a large set of intervention schools.

## **2.4 Design and sub-projects needed to answer three research questions**

### **Sample**

24 integrated primary schools (with grades 1–10), with two classes and their teachers at the focal grades will be selected according to the principles used in the pilot study (cf. 2.3). At these intervention schools, specifically designed teaching of writing (see below) will form the background for post-testing.

As a preparation for the main stage intervention, a first subset of 8 of the 24 schools will be invited to start work on refining standards during the autumn of 2011. The costs involved in 2011 will be covered by HiST/the National Centre for Writing Education and Research, and are not included in the budget for the present application.

### **Project plan**

In order to address the project's three research questions, several sub-studies are needed. These studies, presented as A through G below, are presented in relation to the three questions, but as two consecutive parts. Two doctoral studies are included.

### **Part 1: Refining standards – research question 1**

In order to address the first research question – *which specified norms of assessment can serve as explicit standards for writing in an assessment-as-learning educational environment?* – we will define specified standards for both carrying out different 'acts of writing' and using relevant semiotic resources for carrying out those acts (cf. the 'wheel of writing'). Building on the tentative standards of Matre et al (2011), the process of refining those will be performed in collaboration with teachers at 8 intervention schools, as well as in detailed matching against the national curriculum. Empirical data will at this stage be supplemented with a clearly normative aspect: Which level of proficiency do we find it reasonable for average pupils to reach at two specified grade points, granted good teaching? The national researchers in the consortium will all be involved in Part 1.

Empirical data for refining standards will be collected from the assessment of pupil portfolios that consist of texts from intervention-based classroom work, representing various acts of writing according to the writing wheel, in different situations. The result of this analysis will feed a second step – to confront the resulting specified norms for assessment with the national 2006 curriculum as a normative source.

School year 2011/12 (subset of eight schools):

A. Refine standards for communicative and semiotic aspects of writing in terms of a) specified goals and b) criteria for goal attainment in assessment of pupils' scripts.

The development of shared norms at individual schools will form the qualitative empirical basis for one of the subprojects within the study. In addition, the norms will be refined according to an analysis of the national curriculum. This will be coordinated with an ongoing external, but similar doctoral study (financed by HiST) investigating how shared norms develop and are negotiated within the national panel of raters.

School years 2012–15 (all 24 intervention schools)

B. Study the developing didactic integration of standards at the intervention schools.

Pupils' texts will be collected on the basis of writing tasks that have been/are being piloted for the forthcoming sample-based national assessment, which is to take place at the beginning of grades 5 and 8. To obtain content validity, all participant pupils will write different tasks to cover the wheel of writing. Their texts will be collected and stored as a data base to be used in participant network workshops and in between-schools analysis of outcomes. These texts will be assessed locally, and yield data about teachers' formative and summative assessment practices using the same multi-dimensional scoring instrument as the one used by the national assessment panel.

C. Writing competency development at graphological and textual levels of linguistic analysis will be followed. The specific challenge of developing multimodal competency will be emphasized as well as the development of encoding, vocabulary, grammar and punctuation, with a joint focus on whether integration of standards leads to improved learning outcomes in these selected areas.

D. Competency development will be studied across levels of analysis in a separate longitudinal case study, where focal pupils are followed over a one-year period of time (Jan–Dec in grades 3/4 and 6/7). The result will be individual, competency profiles, summed for each dimension across several writing tasks/forms of writing.

### **Part 2: Integrating standards – research questions 2 and 3**

In order to answer research questions 2 and 3, Part 2 (2012–2015) focuses on whether the use of a specified writing construct and defined standards in a number of intervention schools' differing learning ecologies will

*(2) lead to improved quality of teacher assessment, in particular for rater agreement*

*(3) lead to improved quality of pupils' writing.*

In order to investigate local integration (vs. 'implementation'; see Timperley & Parr 2010) of the newly defined standards across a second subset of 16 schools, a repeated-measurement field study will be carried out, where empirical data will be collected from a corpus consisting of a large set of texts from classroom work throughout the school year for grades 3/4 and 6/7.

At each intervention school, teaching and assessment practices will be planned in concert between principals, teachers, the project's intervention coordinator and one member of the research consortium, so that each plan may fit its local learning ecology. In the intervention, three sets of intellectual instruments will be central – the social semiotic construct of writing (appendix 1), the standards for the grades to be tested (5 and 8) and the multi-dimensional scoring instrument (appendix 2). At each intervention school, the core instruments of our study will thus be used as a practical pedagogical nave. One general aim will be to jointly create consecutive writing situations so that all acts of writing are eventually included and sampled.

Learning outcomes will be measured (at the beginning of grades 5 and 8; by the national test of writing, see below) after one year of intervention teaching for the grade 4/7 classes and two years of teaching for the grade 3/6 classes. Here, the forthcoming national assessment will make it possible to combine intervention school portfolios with results from the annual national sample (from 2012 onwards) as a nationally representative control group. This will allow for quantitative comparison with the results from the intervention school pupils and their teacher raters.

The field studies approach is controversial in that it opens for intervening variables, although it yields robust external validity and more in-depth within- and across-case insights into learning processes, workability and sustainability (Snow 1974, Cicourel 1997). At the same time, however, it should be noted that *post hoc* comparative testing will be available, granted the forthcoming national test.

Research questions 2 and 3 will both be answered on the basis of comparisons between the selected intervention school classes and a paired sample of non-intervention school classes, both sets of school classes taking the national test. Two sets of pupils will thus be included in the study – an intervention group will be compared with a “naturally given” control group. A design like this is well suited to post-test the project's intervention strategy in terms of teacher assessment as well as to throw light on the question of what impact a norm-focused teaching and assessment regime will have on pupils' learning outcomes in terms of text quality.

School years 2013–15 (a subset of 16 intervention schools)

- E. To be able to answer the research question of whether defined standards will lead to improved inter-rater agreement as a result of shared norms, intervention class teachers will take part in the assessment of national writing test scripts. These class teachers will constitute a separate subsample of raters. Their ratings will be compared with those of the national panel of specially trained raters as well as a control group of non-trained teacher raters.
- F. To be able to answer the research question of whether defined standards will lead to improved quality of writing, intervention classes will use the national writing test of the year. Their results will be assessed by the national panel rating the national test scripts. These results will be compared as an aggregate learning outcome with the outcome from the same year's nationally representative sample. As in the national test, assessment will be carried out using a multi-dimensional scoring instrument that yields (individual and aggregate) text quality profiles.
- G. As a way of linking research questions 2 and 3, changing formative assessment practices as a result of standard integration will be studied at the intervention schools.

Across sub-studies, analytical strategies will be based on theoretically derived primary trait analyses and an abductive grounded theory approach. All participant teachers will meet each term during the project period to share their experience and address common challenges. Some of these meetings will be piggy-backed to the Writing Centre's annual network conference for practitioners.

## 2.5 Project management, organization and cooperation

In terms of management, organization and coordination, the National Centre for Writing Education and Research will play the pivotal role that has this far been lacking in Norwegian writing research. Under the centre, earlier informal researcher networks will be further developed and formalized through bilateral agreements between the centre and the involved research institutions:

- Department of Linguistics and Scandinavian Studies at the University of Oslo (UiO),
- Department of Language and Communication Studies at NTNU,
- Program for Teacher Training, NTNU,
- Department of Education at the University of Agder (UiA),
- Faculty of Education and International Studies at Oslo University College (HiO) and
- Faculty of Education and Interpreter Education at Sør-Trøndelag University College (HiST).

These agreements will by implication guarantee a continued interdisciplinary collaboration between researchers. In terms of interdisciplinary research efforts, the proposed project will be carried out by a consortium of scholars representing first language didactics, education, special education, text linguists/discourse analysis and applied linguistics. CVs for the most central involved researchers are attached.

The project will be administered at HiST, the host institution of the writing centre, professor Synnøve Matre (also Prof II at the centre) will head our proposed project, with associate professor Randi Solheim as co-director. HiST and the national writing centre can draw on a solid local network of researchers, teacher trainers and teachers that has been active since the early 1990s. Furthermore, the NFR-supported *SKRIV* project (Smidt et al 2010) was located at HiST and carries out research on how the teaching and learning of writing develops across the curriculum. This project will be developed further within HiST's new, strategic research commitment on writing as knowledge and professional development (2011–2015).

In terms of international cooperation, agreements have been made with Hong Kong Institute of Education, the AsTTle group at the University of Auckland and the Educational Assessment Research Unit at University of Otago, New Zealand (appendix 3). These agreements include theoretical and methodological collaboration, researcher exchange, comparative research and joint publications.

The international research group will be supported by an international reference group having all agreed to participate. This group will represent the IEA writing assessment (prof. Sauli Takala, Jyväskyläns Yliopisto); literacy teaching and testing (prof. Caroline Liberg, Uppsala universitet; prof. John Hattie, University of Melbourne) and second language assessment (prof. Lyle Bachman, UCLA).

In terms of target groups/users the three most important ones are The Norwegian Directorate for Education and Training, teacher education and school leaders as well as practitioners. The Directorate is the official body in charge of carrying out national assessments, but like the national centre has no direct research agenda of its own. It thus depends on institution-external research data. Also, teacher education has so far not focused on assessment. The programme plan for *Utdanning 2020* recognizes that there is a lack of assessment competency in elementary education (p.15f), but this also holds for teacher education. Results from the project have implications for training according to the new national curriculum for teacher training. Results will also feed directly into the teacher training programme at HiST, as the national centre's host institution, as well as NTNU's, UiO's, UiA's and UCO's programmes for teacher training. Simultaneously it will enhance assessment among participating teachers and their schools.

## 2.6 Budget

Please consult the Application form.

### **3 Perspectives and compliance with strategic documents**

#### **3.1 Strategic anchoring**

Within a national Norwegian frame the development of assessment capacity has been increasingly emphasized as crucial for researchers, teacher educators and teachers. According to a number of national policy documents the connection between pupils' learning outcomes on the one hand and teachers assessment practices on the other, is in need of strengthening (*Knowledge Promotion*, St.meld. nr 31, *Kvalitet i skolen*; St.meld. 30, *Kultur for læring*).

The National Centre of Writing Education and Research has been given the responsibility for developing and running the new national sample-based writing assessment system according to these political initiatives. For this reason, assessment is a core part of the strategic aims of the centre. The centre will act as an important partner for the present project in several ways: It will finance preparatory work at eight intervention schools during the 2011 autumn term; it will top finance the planned project coordinator (<50 %) so that this position can be advertised as a full position, and it offers an annual venue for piggy-backed workshops in national networks. As mentioned, HIST has made writing as a main research priority, and the institution will sponsor one of the doctoral projects within the study. It will also carry out the necessary administrative work for the project (budgeting, hiring and payments, accounting and reporting).

Assessment research in Norway is limited in volume. Only a handful of senior researchers are currently qualified for work in the area of writing assessment, most of whom will take part in the present project. In this situation it is strategically important to use our project to recruit promising researchers. Within the project group, there are three associate professors currently in a qualifying process. Postdoctoral resources will be shared among these. In the main stage, two doctoral projects are planned. The doctoral students will each have a stay with one of our international partners.

#### **3.2 Relevance to society**

Assessment is a politically debated issue in Norway. Since the "Bondevik 2" government from the period 2002–2005, a national system of objectives-oriented curricula, with cross-curricular key competencies like writing, is connected to a system of national tests. In this situation it is important that even politicians are given access to empirically valid information about current educational reality and its challenges. One illustrative case in point is relevant in this respect. When the national test of writing in 2005 showed poor reliability, this was ascribed to the quality of the test, whereas writing assessment researchers know that it was bound to take strategic, long-term effort to develop a set of shared norms in a system where systematic assessment had that far not been practiced.

#### **3.3 Environmental perspectives**

The proposed project will not affect the external environment.

#### **3.4 Ethical aspects**

The proposed project will be carried out according to the detailed ethical standards of the *British Association of Applied Linguistics* (BAAL 2006). All participation within cases will be based on voluntary participation and anonymity, and all available data will be digitized only after anonymization and removal of indexical information that may lead to participant identification.

#### **3.5 Gender equality and gender perspectives**

It has been documented in the QAL project (Berge et al 2005) that there are significant gender differences in learning outcomes in the field of writing. Girls typically write better than boys across narrative genres. This difference, however, diminishes significantly across expository and argumentative genres (Vagle 2005). Taking gender into account may counteract the documented gender bias in Norwegian elementary school teaching of writing. For this reason, it would be

possible to collect gender information as a background variable. The field study design makes this accessible while still acting in accordance with research ethical principles.

The project group has equal gender representation, and will be directed by two female senior researchers.

## **4 Communication with users and utilization of results**

### **4.1 Communication with users**

Both school leaders and teachers will be directly involved in the project. Results will be used as a basis for developing new courses in Norwegian teacher education and school leadership. Empirical results, resulting standards and sample pupil texts will be distributed through the National Centre for Writing Education and Research, as well as through national channels like *Nordisk lærerutdanningskonferanse*; *Bedre skole*, *Skoleledernytt*, *Norskklæringen* and *Utdanning*. In co-operation with the centre, research based material relevant for teachers and teachers' education will be developed and published.

### **4.2 Dissemination plan**

Toward the end of the project period the research group will organize an international conference with our international partners, where results will be presented and published as an international volume. The project will be presented at Norwegian and Nordic conferences, as well as international conferences organized by *Assessment in Education – Europe*, and conferences such as *Writing Research across the Curriculum*. Presentations and publications from the project will appear through the EARLI SIG Writing/SIG Assessment conference series and internationally published book volumes originating from that series. The following international journals will be relevant: *Assessing Writing*, *Reading and Writing Quarterly*; *Reading and Writing*, *European Journal of Educational Research*, *Written Communication*, *Assessment in Education* and *Educational Assessment*.

## **Selected references**

- BAAL (2006) Recommendations on good practice in applied linguistics. <http://www.baal.org.uk/aboutgoodpracticefull.pdf>. Downloaded January 27, 2008.
- Berge, K. L. (1996) *Norskensensorenens tekstnormer og doxa (Textual norms and doxa of NLI assessors)*. Dr.art.-avhandling. NTNU: Trondheim.
- Berge, K. L. (2002) Hidden Norms in Assessment of Students' Exam Essays in Norwegian Upper Secondary Schools. *Written Communication*. 19(4). 537–571.
- Berge, K. L.; Evensen, L. S.; Hertzberg, F. & Vagle, W. (2005) *Ungdommers skrivekompetanse (The writing competency of teenagers)*, Vols 1 and 2. Oslo: Scandinavian University Press.
- Biesta, G. (2008) Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*(21), 33–46.
- Black, P. & William, D. (2001) *Inside the black box: Raising standards through classroom assessment*. London: Kings's College, London School of Education.
- Breland, H. et al (1987) *Assessing Writing Skill. Research Monograph No. 11*. New York: College Entrance Examination Board.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of Assessment in Education. *Assessment in Education*, 11(1), 7–27.
- Brown G., Glasswell, K. & Harland, D. (2004) Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing* 9, 105–111.
- CCCC 2009 *Writing Assessment: A Position Statement*. (<http://www.ncte.org/cccc/resources/positions/writingassessment>, retrieved 11/02/2011)
- Cizek, G. J., Bunch, M. B. & Koons, H. (2004) Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, Issue 25, 31 – 50.
- Coffman, W. E. (1971) Essay examination. In R L Thorndike (Ed.) *Educational Measurement*. Washington DC: American Council of Education, 71–302.
- DeSeCo (2000) *Definition and Selection of Key Competencies. A contribution to the OECD program definition and selection of competencies: Theoretical and conceptual foundations*. INES General Assembly 2000.
- Diderich, P., French, J., & Carlton, S. (1961) Factors in the judgment of written composition. *Research Bulletin RB61–15*. Princeton NJ: ETS.

- Dobsen, S., Eggen, A. B., & Smith, K. (2009) *Vurdering, prinsipper og praksis. Nye perspektiver på elev- og læringsvurdering*. Oslo: Gyldendal akademisk.
- Doherty, R. W. & Hilberg, R. S. (2007) Standards for effective pedagogy, classroom organization, English proficiency, and student achievement. *The Journal of Educational Research*, September/October, Vol. 101, Issue 1, 24–34.
- Fasting, R. B. & Thygesen, R. (2009) National Assessment of Writing Proficiency Among Norwegian Students in Compulsory Schools. *Scandinavian Journal of Educational Research*, 53(6), 617–637.
- Gorman, T. P., Purves, A. C. & Degenhart, R. E. (1988) *The IEA study of written composition I: The international writing tasks and scoring scales*. Oxford: Pergamon Press.
- Hamp-Lyons, L. (2011) Writing assessment: Shifting Issues, new tools, enduring questions. *Assessing Writing* 16, 3–5.
- Hillocks, G. jr. (2002) *The Testing Trap. How State Writing Assessments Control Learning*. Teachers College, Columbia University: New York & London.
- James, M.; Black, P.; McCormick, R.; Pedder, D.; & Wiliam, D. (2006) Learning How to Learn, in Classrooms, Schools and Networks: aims, design and analysis. *Research Papers in Education*, 21(2), 101–118.
- Matre, S.; Berge, K.L.; Evensen, L.S.; Fasting, R.; Solheim, R. & Thygesen, R. (2011) *Developing national standards for the teaching and assessment of writing*. Rapport frå forprosjekt Utdanning2020. Trondheim: HiST, Skrivesenteret.
- Messick, S. (1989) Validity. In Linn, R. L. (eds) *Educational measurement*, London: Collier Macmillan, 12–103.
- Meyer, L. H.; McLure, J.; Walkey, F.; Weir, K. F. & McKenzie, L. (2009) Secondary student motivation orientations and standards-based achievement outcomes. *British Journal of Educational Psychology*, Issue 79, 273–293.
- Ministry of Education (2009) *The New Zealand Curriculum. Reading and Writing Standards*. Wellington.
- Quality Counts 2006. National assessment of educational progress. *Education Week*, April/May 2006, Vol. 23, Issue 5, 33.
- Reys, R.; Reys, B.; Lapan, R.; Holliday, G. & Wasman, D. (2003) Assessing the impact of standards-based middle grades mathematics curriculum materials on student achievement. *Journal for Research in Mathematics Education*, January, Vol. 34 Issue 1, 74.
- Rupp, A. A. & Lesaux, N. K. (2006) Meeting expectations? An empirical investigation of a standards-based assessment of reading comprehension. *Educational Evaluation and Policy Analysis*, Issue 28, 315–333.
- Scherer, M. (2001) How and why standards can improve student achievement. A conversation with Robert J. Marzano. *Educational Leadership*, September.
- Smidt, J. (ed) (2010) *Skriving i alle fag – innsyn og utspill* (Writing across the curriculum). Trondheim: Tapir.
- Snow, R. E. (1974) Representative and quasi-representative designs for research on teaching. *Review of Educational Research*, 44(3), 265–291.
- Thygesen, R.; Evensen, L. S.; Berge, K. L.; Fasting, R. B.; Vagle, W. & Haanæs, I. R. (2007) *Nasjonale prøver i skrivning som grunnleggende ferdighet. Sluttrapport*. Nasjonalt senter for leseopplæring og leseforskning, Universitetet I Stavanger.
- Timperley, H., & Parr, J. (2010) *Weaving Evidence, Inquiry and Standards to Build Better Schools*. Wellington: NZCER press.

## Note

- (1) In relation to the distinction in classical norm theory between constitutive and regulative norms (Searle 1969, Berge 1990), the development of norms for writing assessment concerns both. Constitutive norms decide what kind of writing construct is used as the basis for assessment. These were explicated in the pilot study. Regulative norms concern how writing is assessed in an authentic writing assessment situation, i.e. which criteria are used to place pupils' texts at different proficiency levels. Such norms will be primary in the main stage study.