**Standards as a Tool for Teaching and Assessing
Cross-curricular Writing**

Lars Sigfred Evensen

Professor of Applied Linguistics
Department of Language and Literature
Norwegian University of Technology and Science

Kjell Lars Berge

Professor of Textual Studies

Department of Linguistics and Scandinavian Languages

University of Oslo

Ragnar Thygesen

Professor of Special Needs Education

Department of Education

University of Agder

Synnove Matre

Professor of Nordic Language and Literature

South Trondelag University College

Randi Solheim

Associate Professor of Nordic Language and Literature

South Trondelag University College

**Acknowledgement:**

The study was funded by The Research Council of Norway and South Trondelag University College

**Abstract**

The Berge et al. article in this volume presents the functional construct of writing that underlies summative and formative assessment of writing as a key competency in Norway. A functional construct implies that specific acts of writing and their purposes constrain what is a relevant selection among the semiotic resources that writing generally affords.

In this article we present the specific criteria that are currently being introduced in Norwegian teaching and assessment of writing, as well as selected aspects of their development. The article builds on an assumption that assessment criteria have such educational importance that even their origins, intellectual trajectories and underpinnings should be given attention in educational research. In this context, the article presents elements of a rare approach, in that national ‘norms of expected proficiency’ at politically predefined educational grade levels have been grounded in sustained collaboration with experienced teachers of writing across the curriculum, and may thus be viewed as yet inofficial ‘standards’.

In a first step, a combination of existing curricula and literature review of writer development were used to tentatively draft a first set of criteria for the grades included in a 2005 national test of writing (grades 4, 7, 10 and 11). In a second step, such criteria were developed through an iterative, long-term process where initial criteria were confronted with the judgments of experienced teachers. Through ‘think aloud’ assessment interviews, pairs of teachers across Norway were asked to assess specific cases of students’ writings and voice criteria for their judgments, both within and across a series of domains.

In a third step, interview transcripts were used to search for criteria used by several pairs of locally situated teachers across geographically distributed schools. Criteria thus identified, were pooled into a refined set of ‘national standards’ that were subsequently tested out in everyday classroom contexts. On the basis of this confrontation with educational reality, the set has been further refined to form the version presented in this article*.*

The Norwegian case raises a range of issues relating to curriculum development, ‘standards’ and educational sustainability.

**Keywords**:

writing assessment, standards, norms of expectation, curriculum development, educational sustainability

**Background**
Assessment and teaching criteria carry educational and political importance. For this reason, they are normally linked to official curricula that are defined within a top-down political and educational framework. One case of this is the curriculum of South Australia, where an achievement standard in literacy is presented in general terms for each of the school years F-10. The general standard is then exemplified by student portfolios, as in <http://www.australiancurriculum.edu.au/english/Curriculum/F-10#level4>. Internationally, criteria of assessment are increasingly more often developed under a search for national or state level standards of expectation. Such standards have appeared in countries as different as Australia, the U.S.A. and New Zealand.

In most states within the U.S.A., Common Core State Standards ([www.corestandards.org/the-standards](http://www.corestandards.org/the-standards)) are being introduced for ‘foundational writing skills’ since 2010 (Troia & Olinghouse, 2013; Graham & Harris, 2013), with associated tests appearing during 2014- 2015. Two of the issues raised in both of these studies are to what extent these standards are evidence-based and to what extent they may inadvertently lead to a narrowing conception of writing. Graham & Harris (2013, p. 31) also point to a lack of helpful specificity in the current standards for writing, and they point out that they should be seen as educated guesses, for lack of a refined understanding of how writing actually develops in different sub-domains. We believe that our study reported in this article is a much called for supplement in relation to both of these issues.

Within such a context, New Zealand is an ironically special case, following an ‘idiosyncratic approach’ (Thrupp, 2013). In that country, an educationally sensitive academic effort originally led to a set of social practice-oriented criteria for teaching and assessment ([www.asttle.org](http://www.asttle.org)) developed in collaboration with primary school representatives, and with political support to the extent that the criteria were in fact made part of the national curriculum. With a qualitative change in political leadership in 2008, however, a top down edited version of these criteria (NZ Ministry of Education, 2010; Crooks, 2011) was given a new role as official standards, resulting in concerns about ‘teaching to the test’ and a narrowing conception of literacy (cf. Dickie & Shuker, 2014). While the basis for assessment previously had been a set of grade-independent levels of achievement, the current official standards imply using grade-specific criteria (Smaill, 2013).While local results were not previously reported, such reports are now required, with an impact on the role of the assessment (Thrupp, 2013, p. 100).

The trend toward increasing use of state or national standards is related to an international trend towards accountability and documented learning outcomes (Hamp-Lyons, 2011; Smaill, 2013). Within this context, an effect of outcome specification has been documented. 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 standardised 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. Improved outcomes, however, strongly depends upon prolonged teacher learning through social moderation supported by experts, as pointed out by Smaill (2013) discussing the case of New Zealand.

In Norway, educational writing is viewed as a ‘key competency’ (OECD, 2003; Rychen & Salganic, 2003) within the 1-13 curriculum (<http://www.regjeringen.no/en/dep/kd/>). For each school subject, the national curriculum from 2006 is organised as a set of field-specific learning objectives aligned with cross-curricular goals.

On the basis of this curriculum, the Norwegian government in 2008 decided to develop a new assessment system for primary school writing, at two grade levels (5 and 8). This emerging system has a series of challenging characteristics:

* It is primarily intended to support learning across the curriculum. While signaling a low-stakes assessment context, it implies that even teachers who may not have traditionally regarded teaching of writing as a part of their school subject, will be expected to take part in assessing their pupils’ writing proficiency.
* It will be administered locally, with no national reporting system. This emphasises the “non-test” character of assessment in this case, but again implies that it may be carried out by teachers who may not have specific expertise in writing or the assessment of writing
local results may be compared to an available national average, based on sampled pretests graded by a national panel of experienced teachers given extended in-service training in writing assessment for both local and national assessors, a scale of assessment familiar from national tests is used
* The assessment is introduced even at a grade level (5) with no previous formal assessment in writing.

The task of developing this system was given to an interdisciplinary group of researchers headed by the first author. In completing such a task, the above characteristics imply a set of constraints that were decisive for the ensuing development of a specific assessment design. Most importantly, an initial attempt to introduce a national achievement test of writing across the curriculum in 2005 had revealed that there was a lack of shared understanding of what level of proficiency could reasonably be expected at specific grade levels (in that case grades 4 and 7). A first set of assessment criteria had been developed on the basis of existing curricula and a literature review of studies of writer developments, but this set proved insufficient in that the criteria were neither specific nor comprehensive enough (Fasting et al., 2009). Thus, inter-rater reliability coefficients were consistently low, even for domains that might assumedly be easy to assess, like Spelling and Punctuation.

In a situation without politically defined standards or well established teacher norms of expectation it was decided to establish a twin to the national assessment development project, a research project where such norms were developed and tested in collaboration with experienced teachers of writing - The NORMs project (1). This decision may seem paradoxical, however, granted the poor results in the 2005 tests, and raises several issues to be addressed in the Discussion section below.

**Alternative design strategies**

***Authoritarian or minimalist?***

In the Norwegian situation, it might seem obvious to develop an assessment design that might yield high inter-rater reliability coefficients with minimal training. In the international research literature, different options are available for doing so. The approach taken in New Zealand and the U.S. (Glaswell & Brown, 2006) has been to appoint a dominant moderator who can settle all disagreement by virtue of his/her professional authority. This approach raises serious questions about sustainability in local educational situations, however, where no dominant moderator may be available (or even wanted).

A more recent British approach (Pollitt, 2012) implies reducing the assessment challenge by asking assessors to judge two student scripts holistically on the basis of one simple question only: Which one is the better? Summing such comparative judgments yields high reliability coefficients combined with post hoc possibilities for scale development. It also confirms results like the importance of using assessors with recent teaching experience from the grade(s) in question. Still, holistic scores will not yield much information that is useful for formative purposes (cf. Hattie & Timperley, 2007). The Norwegian research group also realised that an aim of content validity would have been hard to achieve within this paradigm. Writing is a multidimensional construct (Berge et al., this volume), which may not be represented well in holistic assessment. Put in other terms, the findings reported in Fasting et al. (2009) and Matre et al. (2011), indicate that a likely outcome from a strictly reliability-oriented strategy would have been to inadvertently perpetuate a narrow perception of the construct - writing as a key competency across the curriculum.

So, none of these alternatives seemed advisable.

***Top down or bottom up?***In a situation with no top down alternatives available, the twin NORMs research project chose a more bottom up-oriented strategy, to actively involve the primary target group – experienced teachers. The aim of this project’s first stage was to jointly define standards for multidimensional aspects of writing in terms of specific criteria for goal achievement, aligned with the writing construct presented in Berge et al. (this volume).

*Developing standards in collaboration with experienced teachers*
The research group first asked teachers at a set of 10 geographically distributed primary schools to participate in discussions of pupils’ writing (cf. Matre et al., 2011). At these schools, experienced teachers at grades 4 and 7 were invited to informal interview sessions to explicate which norms of expectation might be reasonable toward the end of these grades, based on their accumulated professional experience. For these sessions the teachers were invited to bring recent texts from their own classes, and the researchers similarly brought a joint set of texts selected from the 2005 national test of writing (cf. above), as an independent frame of reference.

Teachers at each grade were then interviewed, mostly in pairs, about the specific qualities of the texts they had brought to the session, as well as the texts provided by the researchers:

* which aspects of the individual text at hand would they view as relevant for assessment?
* which aspects might illustrate what was AT their expected level; ABOVE this level or BELOW this level?

In each of these audiotaped ‘assess-aloud’ sessions, the researchers would pose subsequent follow-up questions to clarify, or to ask questions about construct dimensions (Berge et al., this volume) that were not initially mentioned by the teachers.

After these sessions, transcripts were made and investigated to search for assessment criteria that might be used at more than one school. A few excerpts from interview transcripts are presented here to illustrate just one shared criterion that was located during this initial stage of analysis. The example related to the presence and status of a heading or title in a text:

In one of the grade 7 interviews teacher A noted about a piece that it ‘has a heading that catches you’. This phrase seemed to indicate a communicative function. Further in the interview, however, she referred to the same piece that it ‘has a heading’, a descriptor that this time occurred in a specified context of formal concerns (including margins and space). In the same interview, teacher B stated about a different text from the same grade that it has a heading that ‘tells me what I am going to read about’. At a different school, teacher C observed, however, about a different text from the same grade that ‘it has a suitable heading’. On the basis of a number of similar formulations across schools, it was eventually decided that the underlying expectation was one focusing on the writer-reader relation (viewed as a contents preview).

A second example concerns textual aspects of syntax (one aspect of Language use). Several grade 4 teachers observed that a common characteristic of texts rated below their expectations followed an unmarked main clause structure. From a researcher’s point of view this phenomenon is termed theme repetition. Within ‘teacherese’ this observation would frequently be formulated as ‘beginning all sentences in the same way’. The alternative, however, characterising texts that were considered to be ‘as expected’ would have some variation in how consequtive sentences would begin. From a researcher’s point of view this alternative would be termed theme progression. This principle of variation was written into the standards, formulated in ‘teacherese’.

The resulting, preliminary set of criteria was sorted by the research group according to domains of assessment (see below) and then discussed critically within a national panel of raters (see Jølle, 2014, 2015) before the set was tested out at a new empirical set of eight schools, sampled for stage two of the NORMS project. During this test term, the draft criteria were integrated in ordinary teaching, in order to test their sustainability across complex everyday local contexts.

On the basis of these tests, a number of provisional adjustments were made to the initial set of criteria, before all participating teachers were gathered *post hoc* for a national seminar toward the end of term to test the set while assessing a number of texts collected from the participating schools, working in pairs representing different schools (cf. Smaill, 2013). At the end of this seminar, the revised set of criteria was up for critical, plenary discussion. As an example of specific issues discussed during this session, an issue of text organisation from the grade four data was discussed with a view toward paragraphing. The participating teachers at this grade agreed that it seemed reasonable to expect one underlying pattern of global coherence (introduction- main body-ending/coda), but would they also expect that this underlying pattern should be typographically signaled by paragraphing? The joint discussion revealed considerable hesitation, but ended with a recommendation to include a criterion of paragraphing for this particular phenomenon, but for no others at this grade.

After the seminar, audio transcripts from seminar discussions were investigated by the research group during summer, for Beta intervention during the following school year, carried out at yet a new set of 20 nationally distributed primary schools. This finalised version is presented below. [[1]](#footnote-1)

#### ***Assessment domains and strategy***As pointed out earlier, a holistic strategy would not serve a formative purpose of assessment well. For this reason we chose a more complex approach that might yield more specific information. Since Diederich (1974) it has been recognised that even holistic scoring has an underlying factor structure where domains like contents, grammar and orthography play a role. Since then, however, a functional understanding of writing has developed (cf. Berge et al., this volume), with the implication that the writer-reader relationship should be included as a central domain. Furthermore, the current Norwegian curriculum emphasises the multi-modal nature of writing. Granted these developments, the following domains were included:

#### Communication (the writer-reader relationship)

#### Contents

#### Text organisation

#### Language usage (lexicon, syntax and style)

#### Orthography (and inflectional morphology)

#### Punctuation

#### Use of the written medium (handwriting and use of multimodal resources).

To link multidimensional assessment to the Wheel of Writing construct, a primary trait approach was taken, in which assessment is carried out with a perspectival focus on one pre-specified aspect of the writing. In our case this approach implied that student texts would be assessed as viewed through a functional lens of one specified act of writing, being combined with one specified purpose for writing – both to be indicated in each assignment.

***The assessment scale***

In Norwegian comprehensive school, adapted education for all children is required. It was thus an important premise for our choice of assessment scale to focus on pupils who may not benefit well enough from ordinary tuition - either because they suffer from learning difficulties or because they are gifted to such an extent that the regular education programme may not be satisfactory for them. In addition, we wanted the scale to contribute to performance profiles based on central domains of writing for each pupil: ‘In which areas of the assessed writing domains does Ann or John perform well, and in which areas do they show problems?’ A five-level scale, with the phrasing ‘as can be expected from most pupils at this grade’ as the mid category, was chosen. This mid category thus equals the standards presented in Figure 1 below.

For each domain, a text is thus marked as being at one of the following levels of mastery:

* Level of mastery 1*: Very low level of mastery within the domain* (implying that the pupil is *‘at risk level’*),
* Level of mastery 2: *Low level of mastery*,
* Level of mastery 3: *As to be expected from most pupils after 4 or 7 years of schooling* (implying that the pupil is at a *‘functional level’* andable to cope with daily demands in school),
* Level of mastery 4: *High level of mastery*, or
* Level of mastery 5: *Very high level of mastery within the domain* (implying that the pupil is at a *‘challenge level’*)

FIGURE 1: The assessment domains

# Assessment domain 1: Communication

This domain taps how well the writer communicates through the text granted a specified act of writing

## Grade 5

The writer is expected to:

* take a clear position as a writer
* address one or more known readers in a suitable way
* use a title that guides the reader in a relevant way

## Grade 8

The writer is expected to:

* take flexible positions as a writer, and be able to vary among various relevant positions
* address known as well as unknown readers in a suitable way

# Assessment domain 2: Contents

This domain taps whether the topic is dealt with in a relevant and elaborated way

## Grade 5

The writer is expected to:

* present his/her own impressions, experiences, thoughts and/or opinions
* present relevant content elements derived from conceptions, dimensions of experience and/or knowledge familiar to the reader(s)

## Grade 8

The writer is expected to:

* present and elaborate his/her conceptions, experiences, thoughts and/or opinions, as well as those of others
* present and elaborate on content elements that are topically relevant, e.g. to a subject field
* adjust the amount of content relative to the topic

# Assessment domain 3: Text organisation

This domain taps the overall composition of the text, the coherence between individual parts and within each part

## Grade 5

The writer is expected to:

* use some relevant principles of composition (temporal or thematic sequence, etc.)
* use an introduction, a main part and an ending
* create thematic cohesion within the various parts of the text
* create textual cohesion by connectors (*or*, *but*, *because* etc.)

## Grade 8

The writer is expected to:

* use a variety of ways of structuring the text
* structure the text in a purposeful way (e.g. genre)
* use paragraphs as an organising principle
* create cohesion by a variety of connectors

# Assessment domain 4: Language use

This domain taps choice of words, sentence-structure and style

## Grade 5

The writer is expected to:

* use comprehensive declarative, interrogative- and imperative sentences
* use elaborated nominal phrases
* demonstrate some variation at the beginning of sentences
* use a relevant and varied vocabulary, including terms relevant to school-subjects
* include some idiomatic expressions where appropriate
* use direct and indirect speech where relevant

## Grade 8

The writer is expected to:

* build complex and varied sentences
* use a relevant, varied and precise vocabulary, including discipline-specific terms
* use an appropriate tone
* use various idiomatic expressions, where appropriate

# Assessment domain 5: Orthography (with morphology)

This domain taps to what degree the writer masters the orthography as well as morphology related to spelling

## Grade 5

The writer is expected to:

* show control of phonographic spelling
* integrate composite words[[2]](#footnote-2)
* master uppercase letters in proper names, at the beginning of a sentence and after a full stop
* demonstrate control of regular and irregular pronouns and interrogatives
* master double consonants in high-frequency words
* master the morphological system

## Grade 8

The writer is expected to:

* master orthographic spelling

# Assessment domain 6: Punctuation

This domain taps whether the writer masters formal rules of punctuation

## Grade 5

The writer is expected to:

* demonstrate appropriate use of full stops, exclamation marks and question marks
* use commas between items in a list
* use a comma before ‘but’
* mark direct speech by en-dash or colon and quotation marks

## Grade 8

The writer is expected to:

* master the use of colons, parentheses and hyphens
* use a comma between independent clauses
* use a comma after sentence-initial subordinate clauses

# Assessment domain 7: Use of the written medium

This domain taps handwriting and graphic lay-out

### Grade 5

The writer is expected to:

* create texts with a simple lay-out which are easy to follow
* make use of various means of expression, like verbal language, drawings, pictures and symbols
* mark spaces between words clearly
* use and follow margins
* delineate the heading graphically
* mark paragraphs graphically after an introduction and before an ending
* create aesthetically appealing texts
* use cursive handwriting
* use a consistent handwriting
* use a legible handwriting

### Grade 8

The writer is expected to:

* use a clear and suitable lay-out
* delineate paragraphs
* combine multimodal means of expression in order to create cohesion and meaning
* use digital resources for varied communicative and aesthetic purposes

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***The assessment process***

In classroom work in the NORMS project, these norms were integrated first as a general basis for teaching and then as a basis for summative assessment. Formulated as assessment criteria they have been written into the assessment form as a set of support questions for each domain, directing the teacher’ attention towards core indicators of qualities of the text at hand, cf. the assessment form (Figure 2). In addition, annotated benchmark compositions for the grade are placed at the teacher’s disposal. These texts illustrate what kind of achievement can be expected of most students, and what is *much* better than – or *much* below – expected performance.

–––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––

FIGURE 2: The assessment form



In summative classroom assessment the norms were used as criteria to create a multi-faceted picture of where a pupil (or aggregated for a group) is located at the moment of assessment, on the basis of a draft. This picture may be illustrated by a ‘performance profile’, where the assessment scale forms the y-axis and the assessment domains the x-axis, as illustrated in Figure 3 below. In this figure the performance expected at the grade will be marked as a dotted line at Mastery level 3 across domains.

–––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––

FIGURE 3: A performance profile



It should be noted that such performance profiles can hardly be interpreted as a traditional mark, as might justifiably be feared. First of all, they are not summed into a total score, but only presented to teachers as a didactic tool for providing a basis for ‘feed-forward’ in relation to a proposed next step in specific pupils’ learning trajectories. Secondly, they are not reported outside of tutoring contexts. Thirdly, marks are not at all used at lower grade levels in the Norwegian school system. This important difference between marks and performance profiles needs to be illustrated.

The assessment approach presupposes that the pupils write (at least) two versions of their text. The first version is assessed summatively by the teacher, focusing on the question *How am I doing?* The teacher then turns his/her attention to learning-supportive guidance and informs the pupil of the overriding goals for his/her writing development in answer to the question *Where am I going?* He/she talks to the pupil about the progress made so far, and they agree on which specific domain the pupil should focus on. An important aspect of this process is to translate the norms into individual learning objectives for the pupil’s writing progress, taking into account his/her learning trajectory up to the point of formative assessment. Furthermore, the teacher suggests and discusses appropriate working strategies with the pupil in order to reach the objectives. No more than one or two domains should be chosen at each session. This type of feed-forward supervision answers the question *Where to next?*

In many classroom cases the assessment of single domains may be characterised by some uncertainty. Let us suppose that ‘Sally’ meets several of the grade 8 Text organisation criteria for M3 that are relevant for the assignment, but lacks thematic or pragmatic paragraphing. In this process, her teacher might then translate this specific criterion into a specific learning objective for her. As a bridge to formative assessment, her performance is not only presented within a multi-domain profile, but the set of assessment criteria may suggest specific aspects where targeted effort may similarly lift a text to the next level within a specified domain.

Such a formative dialogue may lead to a revised text version subsequently being assessed for just the trait(s) included in the previous ‘feed-forward’ (Hattie & Timperley, 1997). Within the national assessment system, teachers also have annotated benchmark texts for each scale level at their disposal to this end (cf. the section below).

Thygesen & Eggen (2015) discuss details of forming such a bridge between summative and formative assessment in class.

**Discussion**

This article presents an internationally rare empirical approach where experienced teachers have been deeply involved, as either trained members of a national panel of raters or as teachers at intervention schools within the twin NORMs research project. If successful, such an approach may open new ways of thinking about curriculum development. Still, the international potential of a Norwegian case remains to be discussed critically.

One issue relates to imposing an elaborate writing construct (Berge et al., this volume) that may not be totally familiar to all teachers. At this point, however, the feedback from participants has been quite unanimously positive (cf. Matre & Solheim, 2015): Rather than viewing this construct as an imposed burden, they generally view it as a revelation of new potentials for writing instruction and assessment. One aspect of this revelation is the construct’s potential for pointing to domain-specific achievement even among low achievers, which makes it possible to avoid the trap of negative feedback only, prototypically in the form of ‘You must improve your spelling’. In this context, multi-faceted feedback will work to keep up the motivation for both pupils and their teachers.

This article has so far treated ‘norms of expected proficiency’ and ‘standards’ as more or less synonymous terms, even if standards carry an official weight that a specified norm set does not. So: May norms developed in cooperation between researchers and experienced teachers at all become official standards?

In Norway, the developing norms of expected proficiency have served more than the NORMs research project that they have largely been specified within, since they have come to form a core element also in the recently launched official assessment system for writing as a key competency.

As a core element in the still developing assessment system, the 2012 specified norms of expectation informed a governmental process leading to the 2013 revision of the national curriculum, in which progression across the learning trajectory had previously been less than systematic. In this new national context, norms of expectancy for specified grades became a resource for specifying learning objectives. Even more importantly, this approach was used across the curriculum, even outside language curricula.

A second example comes from the current Ludvigsen commission on the Norwegian school for the future (<https://blogg.regjeringen.no/fremtidensskole/mandate-in-English/>). In the recent main report from this commission, the construct behind the Norms project is referred to as one illustration of their recommended research-based way ahead (NOU 2015, 8: 29). Similarly, the norms presented in this article are referred to as exemplary for developing a much needed interpretive community among teachers (NOU 2015, 8: 65, 83).

A third example comes from schools across the country. There is a considerable ‘leakage’ currently going on at grassroot levels, as the norms are found to be of help in an increasing number of everyday classrooms. This leakage is supported by an increasing number of in-service courses and practitioner conferences.

In these local contexts a further development taking place after the first national large scale pilot of the new assessment system in 2012 is influential. The scripts from this pilot were first judged by a national panel of raters (cf. Jølle, 2014, 2015), using the norms set presented in this article, but their judgments formed a basis for detailed investigations of scripts that were judged as belonging to each of the five mastery levels by two independent pairs of raters.

In these investigations, carried out by the National Centre for Writing Education and Research ([www.skrivesenteret.no/om-oss/english/](http://www.skrivesenteret.no/om-oss/english/)), such scripts were for each level searched for de facto descriptors that were later synthesised into generalized level descriptions for each assessment domain at each mastery level. The sum of these analyses formed the basis for a 5 x 6 matrix of *assessment domain-by-assessment level* specifications (2).

The above examples indicate a process where inofficial project norms may incrementally be acquiring a status as *de facto* national standards. If this is in fact the case, there are even more issues to be raised, however. A first issue relates to the intellectual tenability of the existing version of the norms set.

***The tenability of standards***In presenting examples from the teacher interview sessions, it was noted earlier in this article that some assessment domains were initially better represented than others. This raises an issue of ‘tacit knowing’ (Polanyi, 1966). Polanyi noted that perceptually complex phenomena might not be easily verbalised. One example is when you recognise the face or the voice of a friend or a neighbour in some crowded foreign place. What is it specifically that you then recognise? Many have interpreted Polanyi’s observation to imply that such phenomena simply *cannot* be made consciously accessible. But this was not his position. He held that a set of interacting perceptual dimensions might potentially be singled out, described and eventually combined into explicit formulation. The phantom drawings of specialised forensic investigators is one case in point.

As viewed from the first alternative interpretation above, the empirical approach taken in the NORMs project (with its follow-up questions on dimensions not initially volunteered by teachers) would appear to be less than tenable; it could even be seen as imposing a foreign structure on very loosely structured quality judgements. Still, it was observed that several perceptions would in fact frequently be verbalised on the basis of such follow-ups. As viewed from a different angle, then, the domains suggested by the project’s intellectual basis were in fact (more remotely recognisable) parts of teacher observations while assessing. One case in point might be the different functions served by a text’s heading noted earlier in this article, which are no integral part of every teacher’s vocabulary.

In studying interview audiotapes, Matre & Solheim (2015) found that in most cases the teachers did lack a metalanguage to address such cases in a precise way. This finding was initially surprising to the research group, but one institutional fact sheds post hoc light on it. In Norwegian teacher education, most teachers working in primary grades do not specialise in school subjects, but work with a range of these in their classes. And the NORMs project invited all teachers at the grades involved to participate, across the curriculum. Such an invitation certainly implies a need for prolonged dialogue among teachers of different orientations.

At a deeper level of tenability, however, it should be noted that the alignment between the underlying writing construct and the selected domains of assessment is less than clear.

This issue is a question of the applicability of one set of criteria across acts and purposes of writing. Put differently: To what extent may a generic norm set serve teachers who assess a single act of writing? One obvious counter-argument is the observation that not all criteria are applicable granted a specific act of writing underpinning a specific writing task. For this reason, the initial versions of the norms were specified for specific acts of writing. When these versions were tested out in a twin project on a national panel of trained raters working for the national Directorate of Education and Training, however, it appeared that teacher raters struggled with the intellectual task of using partly separate criteria for assessing different acts of writing. And on this basis the Directorate instructed that the norm set should be formulated as a generic set for use with this panel. One possible solution to the tension implied by this instruction would be to specify in the assessment guide accompanying each task a number of criteria that might not easily apply for this task. In the NORMS project, however, teachers would develop their own tasks. In this context, it was decided to include this kind of information in the project’s white book for making tasks.

In a recently published follow-up study, Matre & Solheim (2015) document the tensions implied by such an approach. Assessment criteria may indeed be used mechanically, they found, as teachers across the curriculum test out the generic approach. Still, their study similarly documents that teachers will gradually develop a more context-sensitive practice in their use of the norms as assessment criteria.

Still, the above observation also raises a fundamental issue of sustainability.

***Educational sustainability?***Using primary trait assessment in a functional construct may introduce an intellectual complexity that mirrors the complexity of real life phenomena like writing. In the Norwegian case, this complexity has currently been politically placed on the shoulders of ordinary teachers across the curriculum. In this situation, it is likely that the standards may be translated into directives that may be applied without proper contextual sensitivity. Even if a mechanical use of the criteria becomes gradually more infrequent with increasing practice (as documented in Matre & Solheim, 2015, see above), this problem is likely to remain during the initial years of gradual implementation. The assessment system was formally launched in 2014, for free formative use across schools, with a set of writing assignments related to two acts of writing at each of grades 5 and 8. There are still no studies available, however, of how the system is being interpreted and used across the country.

A combination of summative and formative assessment approaches is still very much an unexplored issue in Norway. Within the NORMs project, teachers have been inspired to focus both on the summative assessment of a pupil’s text and the succeeding formative phase. In general the participant teachers have seemed comfortable applying the norms, formulated as guiding questions on the assessment form, for their summative assessment of a text. Thus, the norms that constitute the reference for the assessment are applicable and valid when it comes to assessment *of* learning.

However, making use of the norms of expected writing proficiency as a core element of the teacher’s formative assessment, i.e. assessment for learning, seems to represent a challenge in general. The UK’s Assessment reform group <http://www.tlrp.org/pub/documents/assessment.pdf> comments that assessment typically is separated from teaching, and that assessment involves only marking and feeding back grades to pupils. Such assessment has increasingly been used to sum up learning; in other words it has a summative rather than a formative purpose. For the pupils, the marks they receive may tell them about their success or failure; however, it tells little about how to make progress towards further learning.

The challenge emerges when the *combination* of the summative and formative approaches becomes the issue. Within our project this transitional process implies that, firstly, the teachers have to make use of the specified assessment criteria as a crucial part of their summative assessment of the text. Secondly, they must transform these norms into individually adapted, specified objectives for the pupil’s further learning progress. Not only must the norms be formulated as objectives related to a text; the teacher must also take his/her pupil’s strengths and weaknesses into account when suggesting the objectives (i.e. ipsative assessment). As for this transitional process, the situation is characterised by an interrelationship between the pupil, the teacher and the text. This situation presupposes that the aims of the summative and formative purposes are seen as mutually dependent. The predefined purposes are substituted by a practice that calls for a bridge between the summative and the formative. This approach seems to be advantageous for the pupils in that it directly affects their learning in a precise and specific way.

Despite the challenges, this study seems to indicate that empirically anchored norms constituting the reference for both summative and formative assessment may also be applicable and valid when it comes to assessment *as* learning. It is our claim that the approach is educationally valid, mainly because the standards are empirically anchored in the accumulated experience of teachers across our country. If this claim holds, the gap between predefined norms and the ipsative application of them can be bridged, granted an emerging interpretive community among teachers, across their local learning environments. This emergence should in our view be seen as a long-term investment in curriculum development.

**Notes**

1. *Developing National Standards for the Assessment of Writing: A Tool for Teaching and Learning*. Homepage: http://norm.skrivesenteret.no
2. In the national assessment system only five acts of writing and six assessment domains are included. The act of writing *To interact* and the assessment domain of Use of the written medium are excluded for reasons that fall beyond the scope of this article.

The authors are members of an interdisciplinary team taking part in a research project entitled *Developing National Standards for the Assessment of Writing: A Tool for Teaching and Learning*, funded by The Norwegian Research Council and South Trondelag University College. The theoretical models and instruments presented in this chapter have been collaboratively developed by the group members. In alphabetical order the team consists of: Jannicke Ohrem Bakke, Vestfold and Buskerud University College; Kjell Lars Berge, University of Oslo; Lars S. Evensen, Norwegian University of Science and Technology; Sindre Dagsland, South Trondelag University College, Trine Gedde-Dahl, University College of Oslo and Akershus; Synnøve Matre (project leader), South Trondelag University College; Hildegunn Otnes, Norwegian University of Science and Technology; Randi Solheim, South Trondelag University College; Ragnar Thygesen, University of Agder; Gustaf Skar, South Trondelag University College.

**References**

Berge, K.L., Evensen, L.S. & Thygesen, R. (this volume). The Wheel of Writing. A model of the Writing Domain for the Teaching and Assessing of Writing as a Key Competency.

Crooks, T. (2011). National Standards revisited. *New Zealand Principal*, *26*, 8-9.

Mansell, W. & James, M. (2009). Assessment for Learning – Beyond the Black Box. In: ARG assessment reform group, Assessment in schools Fit for purpose? A Commentary by the Teaching and Learning Research Programme. July.

Dickie, J. & Shuker, M.J. (2014). Ben 10, superheroes and princesses: Primary teachers‘ views of popular culture and school literacy. *Literacy*, 48, *1*.

Diederich, P. (1974). *Measuring growth in English*. Urbana, Ill: National Council of Teachers of English.

Doherty, R.W. & Hilberg, R.S. (2007). Standards for effective pedagogy, classroom organization, English proficiency, and student achievement. *The Journal of Educational Research*, *101*, 24-34.

Fasting, R. B., Thygesen, R., Berge, K. L., Evensen, L. S. & Vagle, W. (2009). National Assessment of Writing. A Tool for Teaching and Learning. *Scandinavian Journal of Educational Research*, Vol. 53, *6*, 617 – 632.

Glasswell, K., & Brown, G.T.L. (2003). *Accuracy in the scoring of writing:*

*Study in large-scale scoring of asTTle writing assessments.* asTTle Technical Report 26.

University of Auckland/Ministry of Education.

Gorman, T.P., Purves, A.C., & Degenhart, R.E. (1988). *The IEA study of written composition. I: The international writing tasks and scoring scales.* International studies in educational achievement. Vol.5.Oxford: Pergamon Press.

Graham, S. & K.R. Harris (2013). Common Core State Standards, Writing and Students with LD: Recommendations. *Learning Disabilites Research & Practice*, *28*, 28-37.

Hamp-Lyons, L. (2011). Writing assessment: Shifting issues, new tools, enduring questions. *Assessing Writing*, *16*, 3-5.

Hattie, J. & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, *77*.

Hillocks, G. (2002). *The Testing Trap. How State Writing Assessments Control Learning.* New York: Teachers College Press.

<http://www.minedu.govt.nz>. Homepage for The Ministry of Education, New Zealand.

Jølle, L. (2014). Pair assessment of pupil writing: A dialogic approach for studying the development of rater competence, *Assessing Writing*, *14*, 3—24.

Jølle, L. (2015). Rater strategies for reaching agreemen t on pupil text quality. *Assessment in Education: Principles, Policy and Practice*, DOI:10.1080/0969594X.2015.1034087

Matre, S.; Berge, K.L.; Evensen, L.S.; Fasting, R.B.; Solheim, R. & Thygesen, R. (2011). *Developing National Standards for the Teaching and Assessment of Writing. Rapport frå forprosjekt Utdanning2020.* Trondheim: Nasjonalt senter for skriveopplæring og skriveforskning.

Matre, S. & R. Solheim (2015). Writing education and assessment in Norway: Towards shared understanding, shared language and shared responsibility. L1—Educational Studies in Language and Literature, *15*, 1 - 34.

Ministry of Education (2009). *The New Zealand Curriculum Reading and Writing Standards for Years 1-8.* Wellington: Learning Media.

NOU (2015:8). *Fremtidens skole: Fornyelse av fag og kompetanser*. [A school for the future: Renewing subjects and competencies]. Oslo: DIFI.

OECD (2003). *Definition and selection of competencies: Theoretical and conceptual foundations (DeSeCo)*. Summary of the final report: Key competencies for a successful life and a well-functioning society. Paris: OECD.

Parr, J. (2014). Discussant at WRAB: Norwegian Symposium *Developing National Standards for Writing*. Unpubl. manuscript.

Polanyi, M. (1966). *The Tacit Dimension*. London: Routledge and Keegan Paul.

Pollitt, A. (2012). The method of Adaptive Comparative Judgment. *Assessment in Education: Principles, Policy & Practice*, *19*, 281-300. DOI: 10.1080/0969594X.2012.665354.

Purves, A.C. (Ed.) (1992). *The IEA study of written composition. II: Education and performance in fourteen countries.* Vol. 6*.* Oxford: Pergamon Press.

Quality Counts 2006. National assessment of educational progress. *Education Week*, *23*, 33.

Rychen, D.S. & Salganik, L.H. (Eds.) (2003). *Key competencies for a successful life and a well-functioning society*. Göttingen: Hogrefe& Huber Publishing.

Smaill, E. (2013). Moderating New Zealand’s National Standards: teacher learning and assessment outcomes. *Assessment in Education: Principles, Policy & Practice*, *20*, 250—265.

Thrupp, M. (2013). National Standards for student achievement: Is New Zealand’s idiosyncratic approach any better? *Australian Journal of Language and Literacy*, *36*, 99-110.

Thygesen, R.; Evensen, L.S.; Berge, K.L.; Fasting, R.B.; Vagle, W. & Haanæs, I.R. (2007). *Nasjonale prøver i skriving som grunnleggende ferdighet* (National tests of writing as a key competency. Final Report)*.* Sluttrapport. Nasjonalt senter for leseopplæring og leseforsking, Universitetet i Stavanger.

Thygesen, R. & Eggen, A. B. (2015) Bridging summative and formative assessment. In D. L. Cameron & R. Thygesen (Eds.) *Transitions in the field of special education: Theoretical perspectives and implications for practice.* New York: WAXMANN.

Troia, G.A. & Olinghouse, N.G. (2013). The Common Core State Standards and Evidence-Based Educational Practices: The Case of Writing. *School Psychology Review*, *42*, 343-357.

Whitehouse, C. & A. Pollitt (2012). Using Adaptive Comparative Judgment to Obtain a Highly Reliable Rank Order in Summative Assressment. Guildford: AQA.

1. A further elaborated version is employed by The national panel by assignment of The Norwegian Directorate for Education and Training. [↑](#footnote-ref-1)
2. Norwegian spelling differs from English in that most composite words are written as one word cf. sentence pattern [↑](#footnote-ref-2)