Introducing teachers to new semiotic tools for writing instruction and writing assessment: consequences for students' writing proficiency

Kjell Lars Berge, Gustaf Skar, Synnøve Matre, Randi Solheim, Lars Evensen, Hildegunn Otnes & Ragnar Thygesen

Abstract

This article reports consequences for student writing quality based on a long-term professional learning project. Project teachers, representing all school subjects in grades 3–7, were presented with a writing construct, 'Wheel of Writing', and norms of expectation for writing proficiency. Participating teachers used the writing construct and norms as a basis for writing instruction and writing assessment. The project was conducted in 24 schools across Norway. 3088 students from 20 project schools participated. 233 students from 4 schools were used as a comparison group. The investigation showed that students in primary school improved their writing quality significantly. Students in lower secondary school did not. However, there was substantial variation in writing quality effects between schools, classes, and individual students. For instance at a number of schools, project students from lower secondary school improved their writing quality significantly. The article discusses potential explanations of the effects.

Keywords: Writing, writing assessment, writing instruction, construct definition, effects of writing intervention

1. Introduction

This article reports consequences for student writing quality based on a long-term Norwegian professional learning project, the NORM-project. During the project teachers in grades 3–7 were presented with a novel construct definition of writing, the so-called 'Wheel of Writing' (Berge, Evensen, & Thygesen, 2016) as well as norms of expectation (or 'standards') for writing proficiency (Evensen, Berge, Thygesen, Matre, & Solheim, 2016).1 The background of the project was the latest school reform in Norway, which introduced writing as one of five key competencies, to be taught and assessed across the curriculum (Norwegian Directorate for Education & Training, 2007). Outcomes from the OECD-project Defining and Selecting Key Competencies (Organisation for Economic Co-operation and Development, 2005) were influential. All students, whether attending compulsory grades 1–10 or non-compulsory upper secondary school (grade 11–13), were to receive subject specific and subject relevant training in five key competencies, namely writing, reading, ICT, mathematics and oral competence. Hence, all teachers, regardless of subject were to focus on these competencies.

The school reform acknowledged the need for teachers across subjects to teach *disciplinary literacy* (Berge, 2005; Shanahan & Shanahan, 2008), i.e. the kind of literacy competence needed for students to meaningfully engage with and participate in school subjects using writing as a mediating tool (c.f. Russell & Yañez, 2003). However, the new curricula did not offer any tools for the instruction and assessment of writing, and, thus, the integration of writing as a key competency was left to the individual teacher.

With the reform, a tension arose between the intended curriculum, and the one teachers came to enact. An evaluation indicated that six years after the reform, teachers were still uncertain about how to teach *disciplinary literacy*, or even if it was the responsibility for anyone outside the language arts subjects/mother tongue education (Aasen et al., 2012). In turn, this indicates that students were given few opportunities to learn how to use writing as a tool for 'doing' different subjects.

The NORM-project (Matre & Solheim, 2015) tried to resolve this kind of tension by facilitating professional development through a long-term professional learning project, targeting teachers in all subjects in grades 3–7. A premise for the project was that all teachers, in fact, *might* be teachers of writing if they have tools to re-contextualise their subject from a writing perspective. A major challenge then is to introduce semiotic tools (Säljö, 2000; Wertsch, 1991) in writing instruction and assessment that might be effective.

The nature of writing instruction and assessment has changed dramatically since the 1970s (Galbraith & Rijlaarsdam, 1999). The question of what is to be regarded as best practice when teaching writing has received increased attention. Several interventions have shown to be successful. In an early meta-analysis Hillocks (1986) investigated the consequences of different foci in writing intervention and found the 'inquire' approach, to provide the best results. In this approach, students use writing systematically and functionally to achieve communicative goals (c.f. Hillocks, 1999). Meta-analyses by Graham and colleagues (Graham, McKeown, Kiuhara, & Harris, 2012) have documented positive consequences for students texts' writing quality, computed as large mean 'effects' associated with both so-called 'strategy instruction' ($ES = 1.17$), peer assistance ($ES = 0.89$), and feedback from adults ($ES = 0.80$). For example, teaching students specific writing strategies, such as planning and mastering a writing process, has shown to be highly beneficial for both older and younger students (Fidalgo, Torrance, & García, 2008; Glaser & Brunstein, 2007; Harris, Graham, & Mason, 2006; Tracy, Reid, & Graham, 2009), and for students with disabilities (McConnell, Little, & Martin, 2015). Engaging students in 'cross-age tutoring', i.e. peer assistance, has been indicated to be of academic benefit for grade 4 students, and an appreciated practice by younger students (Paquette, 2008), and in a study investigating effects of adult modelled feedback the treatment group students outperformed non-treatment students (Lumbelli, Paoletti, & Frausin, 1999). Moreover, other modes and foci of writing instruction that in average do not yield high effect sizes have been suggested to be successful in individual studies. For example, teaching of grammar is usually deemed to be ineffective (Graham & Perin, 2007), but other studies report positive effects (Myhill, Jones, Lines, & Watson, 2012).

Alongside specific writing interventions, different professional development programmes targeting writing instruction have also proven to be successful (McCarthey & Geoghegan, 2016). For example, facilitating teachers to implement process writing approaches associated with the National Writing Project in the US has shown to be very valuable, as indicated by increase in student proficiency (Pritchard & Marshall, 1994). Also, a large national New Zeeland project aiming at facilitating teachers to interpret assessment data to provide better targeted writing instructions resulted in significant increases in student writing scores (Parr & Timperley, 2010; Parr, Timperley, Reddish, Jesson, & Adams, 2007). In Norway, examples of professional development through a 'bottom-up design', i.e. teacher initiated professional development, have indicated 'fruitful collaboration across subject areas', to cater for implementation of writing as a key competency (Hertzberg & Roe, 2015).

To date, then, several modes and foci of instruction, as well as types of professional development programmes, indicate effective strategies that are indeed evidence based. However, when introducing writing as a key competency across subjects, these modes and focuses are not readily available for all teachers; they all rest on in-depth knowledge of writing and writing development. As the NORM-project aimed to resolve tension between intended and enacted curriculum, it was decided that it was necessary to go beyond modes and foci

of writing instruction. Therefore, the project was set up to (1) facilitate an explicit as well as theoretically comprehensive understanding of what writing could be, (2) to develop resources (incl. norms of expectation) and strategies for teachers to appropriate this understanding in practice, and (3) to use this writing definition as a semiotic mediating tool for construction of writing instruction, writing tasks and writing assessments.

For a period of two years, between 2012 and 2014, teachers and students from 20 schools (grades 3–7) across Norway, participated in a programme aimed at introducing a novel construct definition of writing – depicted in the model *Wheel of Writing* (Berge et al., 2016) – and norms of expectations for writing proficiency (Evensen et al., 2016).2 Teachers were also facilitated in their work of using the Wheel of Writing and the norms as tools for instructing and assessing disciplinary literacy.

The project was set up as a repeated measures field study, including teachers and students from comparison schools. In this article, we investigate the possible results, computed as effect sizes, on student writing proficiency of introducing these new semiotic tools across and within different school subjects. We will answer the following research question:

> *Will a semiotic re-contextualisation of school subjects using the Wheel of Writing and associated norms of expectation as a basis for teaching and assessing writing, make students better writers?*

2. Methods

2.1. Research design and participants

The project was set up as a repeated measures field study, and was conducted between October 2012 and June 2014 in 24 schools across Norway. Given the size of the project, the schools were selected to reflect the school variation in Norway, as realistically as possible. Thus schools from different parts of Norway was included, as well as schools from rural and urban environments. These schools also matched the use of the two norms of written Norwegian,3 as well as students with oral Norwegian as first language (L1) and other first oral languages (L2). The gender distribution was also taken into account (see Berge & Skar, 2015). Among the 24 schools, 20 were assigned to the project mode, rendering a total of 3,088 student and 500 teacher participants. 4 schools were assigned to a comparison mode, rendering 233 student participants. Pre- and posttest-data for both project groups and comparison groups were collected in the fall of 2012 and in October 2014, about 5 months after the project was ended.

Table 1. Participating 300-students and 600-students.

	300 project group	300 project group	600 project group	600 project group	N
All participants (*n*)	75	54	78	58	265
Girls (*n*)	45	23	40	31	139
Boys (*n*)	30	31	38	27	126
Age[a]	7.8	7.8	10.8	10.8	

In this article, we will present analyses regarding students' writing proficiency of 265

students entering the project in grade 3 or in grade 6. These students were randomly selected from the large pool of students. The number of selected students was based on available economical resources for assessment of students' scripts. These students will be referred to as 300-students and 600-students, respectively. The students in the project group participated indirectly in the project, receiving writing instruction and writing assessment from teachers who had learned about the Wheel of Writing and the norms of expectation, both of which are further described below. Characteristics of participating students are described in Table 1.

Due to missing data, not all schools in the project group are represented at both 300- and 600-level. This is presented in more detail in the results section.

2.2. The project

The NORM-project had four phases:

- First, the conception of writing was developed into a theoretically based, explicit construct definition, The Wheel of Writing (Berge et al., 2016).
- Second, norms of expectation, or standards for writing proficiency operationalizing the construct, were developed in cooperation with teachers (Matre et al., 2011; Evensen et al., 2016).
- Third, the Wheel of Writing and norms of expectation were introduced to teachers participating in the project.
- Fourth, the Wheel of Writing and norms of expectation were used in project classes as a basis for writing instruction and writing assessment.

2.3. The writing construct and the expectation norms (standards)

The writing construct, the Wheel of Writing (Berge et al., 2016), used in the NORM-project was developed to account for the complexity and variation of writing in different school subjects as well as in writing across the curriculum. The writing construct is constituted by three dimensions to account for what is considered as the intentionality of writing: (1) writing as an act of meaning making, (2) the purpose of writing, and (3) writing as semiotically mediated (Mertz & Parmentier, 1985) by different modalities. The Wheel of Writing consists of six acts with corresponding purposes. The six writing acts are the following: to reflect, to describe, to explore, to imagine, to convince and to interact by means of writing. There are six purposes that may fulfil these acts: identity formation, knowledge organisation, knowledge development, creation of textual worlds, persuasion, and exchanging information. As an example of these default relations between act and purpose in writing, the act 'to describe' may be related to the purpose of 'knowledge organisation and storing', as when we describe an engine with the purpose of defining the qualities of the engine. Of course actual texts may consist of more than one act and purpose. Therefore, the relations between acts and purposes may be arbitrary as well as contingent. For instance the act 'to describe', by default related to the purpose of 'knowledge organisation and storing', may also fulfil the purpose 'persuasion'. A writer may describe some natural phenomenon like global warming and by this description persuade the reader to accept as a fact that the climate crisis is a reality and a threat. The Wheel of Writing is presented in Figure 1.

In the NORM-project the participating teachers used the Wheel of Writing as a mediating tool for defining the writing acts and purposes used when teaching different school subjects as defined in the curriculum. In addition, this specific writing construct was used to create subject relevant writing tasks, aligned with the norms of expectation or standards.4 The dimensions representing the semiotic resources necessary to mediate written communication are presented in Figure 2. In the visual metaphor of the Wheel of Writing, the four resources together represent the wheel's hub:

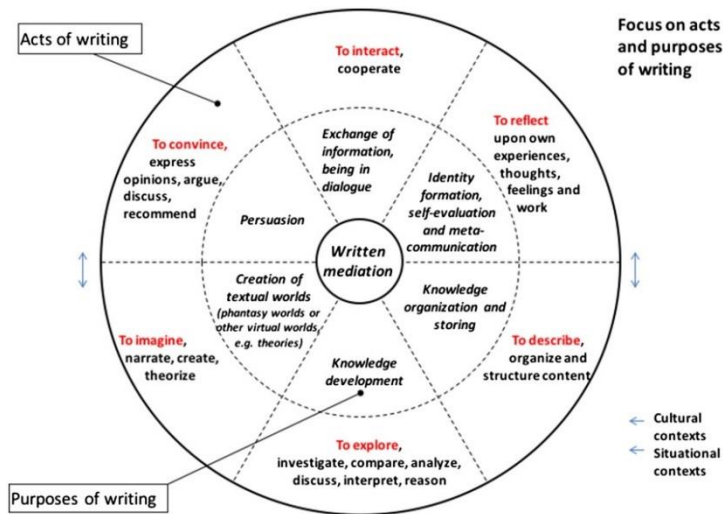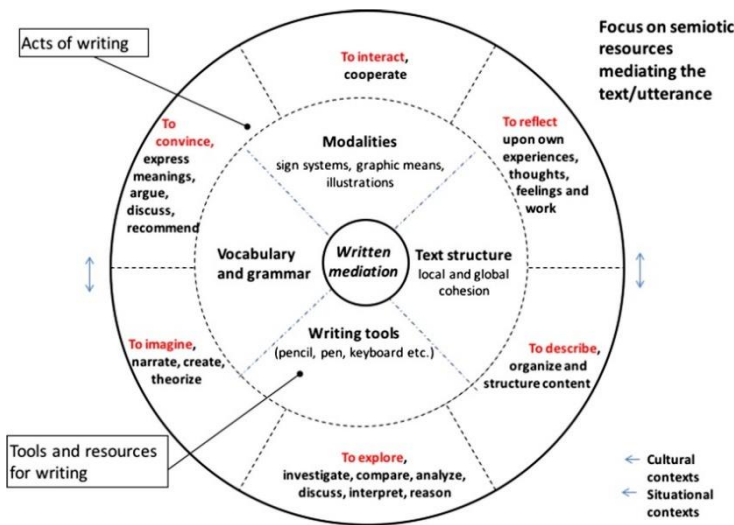Figure 1. The Wheel of Writing: acts and purposes of writing

Figure 2. The Wheel of Writing: semiotic mediation (the tools and resources are in this version of the model covering the purposes).

The four resources enabling written mediation are: writing modalities, material writing tools, vocabulary and grammar (lexico-grammatical resources), and text structure (textual resources). These four dimensions are further differentiated into seven assessment domains, or scales (Evensen et al., 2016), sub-characterised as functional competencies and code competencies:

Functional competencies:
- Communication (mediating the writer–reader relationship in the text)
- Content as mediated in the texts (i.e. topics, discourses)
- Text organisation
- Language use (lexicon, syntax, and style)

Code competencies:
- • Orthography (and inflectional morphology)
- • Punctuation
- • Use of the written medium (handwriting and use of multimodal resources).

When the teachers involved in the NORM-project were teaching and assessing writing according to the seven assessment domains, the assessment was done using scales consisting of five band levels (1–5). The mid-level, band level 3, was defined as 'as can be expected from most pupils after 4 or 7 years of schooling' or put differently, at the start of year 5 or 8. This means that band level 3 was awarded to scripts meeting the norms of expectation. An example of this way of defining the scales is given in Appendix A.

There were one set of descriptors for end of school year 4, and one set of descriptors for end of school year 7. As described in Evensen et al. (2016) the descriptors were developed by deducing possible relevant assessment domains from earlier research (for instance Diederich, 1974; Gorman, Purves, & Degenhart, 1988), and then, in close collaboration with Norwegian compulsory teachers from school years 5 and 8 by explicating expectations of writing proficiency associated with those domains (Matre et al., 2011).

2.4. The Wheel of Writing and standards as semiotic tools

At each project school, teaching and assessment practices were planned in concert between principals, teachers, the projects schools' local coordinators and two members of the research consortium, so that each plan would fit its local learning ecology. The implementation of the project at the schools was understood and interpreted as a field study. During the local integration, three sets of intellectual instruments were central – the construct of writing, the norms and a multidimensional scoring instrument. At each project school, the core semiotic instruments (i.e. the Wheel of Writing and the norms) of the study were in combination used as a practical pedagogical nave. Teachers were instructed to create different writing tasks, so that all acts of writing were instructed and assessed.

As a consequence of the project, the teachers and students should understand the local integration of the writing resources as a part of normal teaching practice anchored in the national curriculum. The participating teachers received systematic training in the writing construct, the norms/standards, writing assessment, and research-based examples of effective writing instruction through several workshops. Consequently, the training involved modelling curriculum relevant writing assignments and the use of assessment information for planning of future instruction. As professionals, the teachers decided on how to integrate the standards and models in their own educational settings. However, the research group regularly visited schools, made observations and offered support between workshops.5 The researchers also cooperated closely with the principals at the schools, and as a consequence, most of the principals were actively involved in the project together with, and in close cooperation with, the teachers.

These steps were followed during the cooperative implementation of the project at the participating schools:

- • Participating teachers were introduced to the Wheel of Writing as the constitutive construct of writing in the project period

- • Each participating teacher at every project school was introduced to the norms of expectation as a tool for instruction, for feedback to students' ideas and drafts and assessment of the students' written texts

- • At each project school the participating teachers were shown how to develop writing prompts based on the writing construct, and relevant teaching goals in the curriculum Each participating teacher at every project school was introduced to research-based

strategies for formative and summative feedback and assessment of writing

- At each project school the participating teachers were introduced to how writing assessment could be performed as a cooperation between teachers, and how a reliable interpretative community could be achieved

- At each project school the participating teachers were introduced to how to plan writing instruction relevant for each writing prompt, differentiating between the students' texts as drafts and as finished texts

- At each project school the participating teachers were introduced to how to plan writing instruction in all school subjects over the whole school year.

2.5. Data collection

Pretest data were collected in October 2012, and posttest data were collected 5 months after the project was ended, in October 2014. On both occasions, students wrote standardised writing tasks, but in low-stakes settings.6 Students were informed that scores would be used only for purposes of research. On the last occasion, the project and comparison group students were given the same task as the one administrated in the Norwegian Sampled-Based Writing Test (NSBWT), a test that is also based on the Wheel of Writing (Skar, Evensen, & Iversen, 2015).7 The NSBWT is administrated to 10- and 13-year olds, which therefore makes the results across groups comparable. NSBWT scores will be used for comparison.

Members of a professional panel8 rated all student scripts. These were rated analytically in six of the seven scales described above, namely communication, content, text organisation, language usage, orthography and punctuation by two raters.9 The analytical rating yielded a sumscore, with a minimum of 12, and maximum of 60. Although all scales were of equal length, there were different scale descriptors for 4th and 7th graders. This means that scores given to 5th graders are incomparable to those of 8th graders. However, in both grades, a score of 36 represents achieving in average band level 3 on each scale, and thus in average meeting the 'norms of expectation' for 4th or 7th grade, respectively. In the analysis below, 36.0 is used as cut-score, for classification of students as at or below the norm.
Two different strategies to secure reliability were taken. In the pretest, most students wrote two tasks. Two raters rated each script individually, after which a consensus decision was formed. The individual ratings were not recorded, and reliability estimates are therefore not available. Each student received a sumscore based on the two tasks.10 In the posttest, students wrote one task, which was rated by two independent pairs of raters. Scores were summed across raters. The reliability on the posttest was acceptable (ICC[1] = .775 for ratings of 300-students, and ICC[1] = .807 for ratings of 600-students).

The data collection procedures reveal an important caveat. The NORM-project built on the premise that any sort of professional learning project must be flexible enough for teachers to be able to adapt it to local circumstances. As suggested elsewhere (Parr et al., 2007, p. 6), the relationship between teacher participation in professional development and student outcome, then, becomes characterised by 'black boxed processes'. In this particular case, with a few exceptions, we do not have direct access to teacher understanding of the tools introduced in the NORM-project, neither do we have directly access to student learning processes at all project schools. This limits our possibilities to fully explain links between student outcomes and teacher participation in the project. However, some of the practice from teachers involved in the NORM-project has been observed. Teachers and principals have been interviewed during and after the project period. Findings suggest altered teacher practice (Matre & Solheim, 2015).

2.6. Data analysis

To answer the research question of writing instruction results, computed as 'effects', we

have used descriptive and inferential statistics. The observed score for each student on each occasion has been used to calculate means, standard deviations and t-statistics for mean differences within and between groups.

Furthermore, we have conducted analyses of covariance (ANCOVA), to adjust for any initial differences between the groups. The ANCOVA was performed in two steps. First preliminary checks were conducted to investigate whether or not there were violations of the basic assumptions (Pallant, 2013). No such violations were found. We also computed the standardised mean effect (Cohen's *d*) for each comparison, using Coe (2000), SPSS 23 (IBM, 2015) and Wiseheart (2014).

3. Results

The results show that on average, and irrespective of group, all students make gains in writing proficiency throughout the project period. 300-students in the project have the largest gain, from a mean score of 30.4, to a mean score of 40.2, resulting in an effect size of $d = 1.54$. Put differently, by the time of the posttest, the average student amongst 300-students in the project group received scores well above mid level on all scales. The gain in the 300-students comparison group, from a mean of 28.0 to a mean of 34.7, equals $d = 1.16$. The average comparison student performed near the mid level on all scales. See also Table 2.

In the 600-students project group, the gain was modest, from a mean of 28.0 to a mean of 34.1, resulting in an effect size of $d = 0.78$. The gain amongst 600-students comparison group was even lower, equalling $d = 0.47$. See also Table 3.

Table 2. 300-students: Means for pretest, and posttest, and analysis of mean difference.

Group	n	M_{pre}	SD_1	M_{post}	SD_2	r	t	p	*ES (d)*
300 proj.	75	30.4	7.1	40,2	8.8	.68	13.0	.000	1.54
300 comp.	54	28.0	5.8	34.7	6.8	.58	8.5	.000	1.16

Table 3. 600-students: Means for pretest, and posttest, and analysis of mean difference.

Group	n	M_1	SD_1	M_2	SD_2	r	t	p	*ES (d)*
600 proj.	78	28.0	6.5	34.1	8.8	.47	6.6	.000	.78
600 comp.	58	27.8	6.9	31.7	8.2	.50	3.8	.000	.47

The comparison of results within each group reveals that 300-students have gained more than 600-students, and that the largest gain is within the 300-students project group. The 300-students on average perform on mid level on all scales, which in turn strongly suggests that this group on an aggregated level meets the norms of expectation. The results also suggest that the average project student from both the 300- and 600-group outperform the average comparison student. This is further explored below.

The initial differences between project and comparison schools were significant at the 10

per cent level for 300-students, $t(157) = 1,74$, $p = .084$, and non-significant for 600-students, $t(155)=-.045$, $p = .964$. However, to control even for these small differences, we conducted an ANCOVA-analysis, using the pretest as a covariate.

As can be seen in table 4, there is a significant difference between the project group and the comparison group amongst 300-students. The adjusted mean score for project students is 39.4, while the result for the comparison group is 35.8. The mean result on the NSBWT ($N = 340$) for the same age group was 33.8.[11] The comparison group of 300-students thus seem to be somewhat stronger writers than might be expected when compared to NSBWT results. The effect size based on the adjustment following the ANCOVA is $d = 0.57$, which indicates a moderate effect (Cohen, 1988), and about half the gain of the comparison group throughout the two-year project (cf. table 2).

Table 4. Mean differences between project and comparison group, 300-students.

Group	n	M	SD	M_{adj}	df	F	p	ES (d)	Pwr
300 proj.	75	40.2	8.8	39.4					
300 comp.	54	34.7	6.8	35.8					
300 proj. > 300 comp.					1.126	10.13	.002	.57	.88

Table 5. Mean differences between project and comparison group, 600-students.

Group	n	M	SD	M_{adj}	df	F	p	ES (d)	Pwr
600 proj.	78	34.0	8.8	34.0					
600 comp-	58	31.7	8.2	31.7					
600 proj. > 600 comp.					1, 133	3.13	.079	.31	.42

We also investigated the difference between project and comparison groups amongst 300-students on sub-group level. All 300-students scoring an average of 1.00–2.49 on the original scale were grouped into a 'below expected level-group' (*n*: project = 36, comparison = 22) and all students scoring 2.50–3.50 on the original scale were grouped into the 'at expected level-group' (*n*: project = 34, comparison = 32).[12] The analysis shows that the difference is greatest among students scoring at expected level on pretest, $F (1, 55) = 8.47$, $p = .005$, with an ANCOVA-adjusted effect size of $d = 0.78$. The effect size for the group scoring below expectation is smaller, $d = 0.51$, following a smaller, but significant mean difference, $F (1, 63) = 4.16$, $p = .05$.

Turning to the 600-students, we see that the overall effect is smaller for this student group. The adjusted mean score for project 600-students are 34.0, while the result for the comparison group is 31.7. The mean result on the NSBWT ($N = 411$) for the same age group was 31.2. The comparison group thus seems to perform as expected. The ANCOVA-based effect size amounts to only $d = 0.31$, which, however, corresponds to two thirds of the gain of the comparison group throughout the two-year project (cf. table 3). Furthermore, the difference between the project group and comparison group is non-significant, $F (1, 133) = 3.13$, $p = .079$. This is congenial with the weak statistical power (.42). See also table 5.

Also for 600-students, we conducted sub-group analyses. As with the 300-students, only two groups were formed, below expected level (n: project = 49, comparison = 35) and at expected level (n: project = 28, comparison = 21). The results resemble those for the 300-students. There is a greater difference between project and comparison groups among students scoring at the expected level. However, in this case it is non-significant, $F (1, 46) = 1.1$, $p = .31$, and the effect size rather modest, $d = 0.31$. For the group scoring below expected level amongst 600-students, the difference between project and comparison was non-significant, $F (1, 81) = 1.48$, $p = .23$, and small, $d = 0.27$. The statistical power was weak, .17 and .22, respectively. As a last sub-group analysis, we compared posttest results within the whole group of schools and between groups of project schools and comparison schools.

Table 6 depicts posttest mean value for each of the participating schools at both grade levels. All school mean values exceeding 36, which represents an average of band-level 3 (i.e. meeting the norms of expectation) are bolded. As can be seen, some mean values are associated with high standard errors, reflecting low n. When reading the table, one must bear in mind that although scripts from 300- and 600-students were rated on scales with the same lengths, these were different scales, meaning that a score of 36 should read meeting expectations for 4th and 7th grade, respectively.

Table 6. Posttest mean values for all participating schools.

		300-students				600-students			
		N	Mean	SD	SE	N	Mean	SD	SE
Project	School 1	11	**43.5**	8.6	2.6	8	**39.8**	8.5	3.0
	School 2	7	**38.9**	5.8	2.2	4	31.3	4.1	2.1
	School 3	5	**41.6**	13.5	6.0	6	**41.8**	5.6	2.3
	School 4	–	–	–	–	3	27.0	11.4	6.6
	School 5	11	**39.6**	8.2	2.5	6	**36.3**	9.9	4.0
	School 6	2	**39.5**	12.0	8.5	2	34.5	0.7	0.5
	School 7	2	**46.5**	5.0	3.5	2	22.0	2.8	2.0
	School 8	1	20.0	n/a	n/a	1	**40.0**	n/a	n/a
	School 9	2	**40.5**	7.8	5.5	3	33.0	8.5	4.9
	School 10	––	–	–	–	6	26.3	15.5	6.3
	School 11	2	33.5	0.7	0.5	3	**37.0**	1.7	1.0
	School 12	5	**45.6**	8.4	3.8	6	34.5	8.9	3.6
	School 13	6	35.3	5.4	2.2	4	**40.3**	7.3	3.6
	School 14	3	**39.7**	6.0	3.5	5	33.4	6.0	2.7

Table 6. Posttest mean values for all participating schools.

		300-students				600-students			
		N	Mean	SD	SE	N	Mean	SD	SE
	School 15	2	**40.0**	17.0	12.0	3	25.0	8.9	5.1
	School 16	4	35.3	3.9	1.9	4	34.3	6.8	3.4
	School 17	1	20.0	n/a	n/a	–	–	–	–
	School 18	1	**39.0**	n/a	n/a	2	35.0	2.8	2.0
	School 19	6	**47.0**	6.7	2.7	5	32.2	3.6	1.6
	School 20	4	**39.3**	9.8	4.9	5	32.8	4.9	2.2
	Total	*75*	*40.2*	*8.8*	*1.0*	*78*	*34.1*	*8.8*	*1.0*
Comparison	School 21	20	35.2	8.0	1.8	26	29.3	8.0	1.6
	School 22	8	33.4	7.9	2.8	10	31.8	8.2	2.6
	School 23	13	34.2	6.3	1.8	14	33.4	9.0	2.4
	School 24	13	35.3	5.0	1.4	8	**36.1**	6.3	2.2
	Total	*54*	*34.7*	*6.8*	*0.9*	*58*	*31.7*	*8.2*	*1.1*

Setting aside somewhat broad confidence intervals, table 6 indicates that there are non-trivial differences within the project group. For example, almost all project schools at 300-level exceeds an average of 36. Of the 18 project schools (2 schools are left out due to missing data), only 5 have an average below 36.0 points. All comparison schools average below 36.0 points. Quite the opposite pattern reveals itself when reading results for the same schools at 600-level. Of 19 project schools (1 school is left out due to missing data), 6 schools have an average exceeding 36. The average student in the remaining 13 schools scores below 36.0, as does the average comparison school student. We can also note some interesting variation within schools. For example, the 300-students in school 19 perform well above 36.0 on average (M = 47.0, 95% CI [41.7, 52.3]), while 600-students in the same school perform below 36.0 on average (M = 32.2, 95% CI [29.1, 35.3]). In the same school, then, young learners outperform the average comparison student with more than 10 raw scores, while the older learners do so with only 0.5 raw scores. Finally, 3 of the 6 schools on 600-level that score above 36.0 on average, do not do so on 300-level. In all, then, 3 project schools score above the cut score both on 300- and 600-levels (School 1, School 3, and School 5).

To further nuance the results at school level, we clustered students from project schools into two groups, students from schools with or above 36.0 points on average (referred to as '300/600 above norm', and students from schools with less than 36.0 points in average (referred to as '300/600 below norm'). The results are shown in table 7 and in table 8.

At the 300-level, there were 13 schools with average scores above 36.0 points. These schools were represented by 61 students having an adjusted mean of 40.5, 1 score point above the 300 project group on average (c.f. table 4). Table 7 shows that the distance to the comparison group is larger than the distance for the whole 300 project group, suggested by a higher effect size (d = .67), equalling 58% gain increase compared to the average gain in the comparison group (c.f. table 2). Fourteen students represent the 5 schools with average below 36.0 points. Although the raw score difference suggests that the comparison school students outperform this group of project students, the difference is not significant.

Turning to 600 students, we see that there were 6 project schools with an average score above 36.0 points. 28 students represented these schools. As can be seen in table 8, the difference between these students and students from comparison schools was quite large. The effect size equals d = .90, almost twice the effect size for comparison schools (c.f. table 3). The 13 schools scoring below the cut score in average are represented by 50 students. There are no significant difference between this group and the comparison group.

Table 7. Comparing 300 schools with average above and below cut score with comparison schools.

Group	n	M	SD	M$_{adj}$	df	F	p	ES (d)	Pwr
300 above norm	61	41,8	8,4	40.5					
300 comparison	54	34,7	6,8	36.2					
300 above > 300 comp					*1.112*	*12.5*	*.001*	*0.67*	*.94*
300 below norm	14	32.9	6.7	35.1					
300 comparison	54	34.7	6.8	34.1					
300 below > 300 comp					*1.65*	*0.4*	*.550*	*−0.16*	*.09*

Table 8. Comparing 600 schools with average above and below cut score with comparison schools.

Group	n	M	SD	M$_{adj}$	df	F	p	ES (d)	Pwr
600 above norm	28	39.3	7.3	38.5					
600 comparison	58	31.7	8.2	32.0					
600 above > 600 comp					*1.83*	*16.6*	*.000*	*.90*	*.98*
600 below norm	50	31.1	8.3						
600 comparison	58	31.7	8.2						
600 below > 600 comp					*1.105*	*.001*	*.980*	*.00*	*.05*

Table 9. Summary of results.

	300-students, *ES*	600-students, *ES*
All project students > comparison students	.57**	.31
Students below cuts-core > comparison students below cut-score	.51*	.27
Students above cuts-core > comparison students above cut-score	.78**	.31
Schools below cut-score > comparison schools	−0.16	.00
Schools above cut-score > comparison schools	.67**	.90**

Summing up, the project has generated some mixed results. On the one hand, 300-students participating in the project made non-trivial gains. The group outperforms the comparison group, whose results, with reference to the NSBWT, are somewhat higher than expected. Moreover, when investigating sub-groups we see that 300-students performing at expected level at the start of the project demonstrate an even larger gain. These students benefit greatly from the project.

4. **Discussion**

The NORM-project represents a comprehensive and conceptually coherent way of locally integrating shared understandings of a writing construct and cross-curricular norms of expectation for writing to promote systematic writing instruction. Unlike the bulk of previous writing intervention studies, the NORM-project did not focus on isolated aspects of writing instruction, but drew on teachers' professionalism and experience to operationalise the consequences of adopting a functional approach to writing. A basic warrant for the project's design has been that the project should be carried out as a co-operation between the researchers behind the project and the participating principals and teachers. Consequently, the project should be integrated in the participating schools' normal teaching practices and habits. The teachers' assumed tacit teaching knowledge and competence was to be respected and considered as a resource. Thus, the project should not be considered as something alien or a brief intermezzo in the schools' everyday learning environment and practices. In this way the project's results documented above should be interpreted and discussed in terms of a cultural and/or ecological environmental approach. This tradition-sensitive cooperative approach towards the schools' traditional learning culture may shed light upon the variation in writing quality effect we have presented above. As documented, the variation occurs at four levels:

First, there is variation in writing quality results between the age groups 300- and 600-students: The NORM-project has developed the 300-students' writing ability with an effect size variation of .57 to .78. Depending on frame of reference, this represents 0.5 years or 1.5 years of schooling (c.f. Hattie, 2009). On aggregated level the results for the 600-students do not suggest any considerable gain. There are no significant differences between project students and comparison students.

The 300-students compared to the 600-students seem to be a more homogenous and apparently at an age when learning and the developing of abilities like writing is more effortless that at a later age. Amongst 600-students, it is more demanding to develop the students' writing abilities. This age group variation has been observed in a number of research

projects. It seems to be a pattern, reported for example in numerous studies in Australia, that effect sizes related to reading and writing instruction are generally 'higher in primary and lower secondary schools […] and lower in upper secondary schools' (Hattie, 2015, our translation). Other explanations may also be relevant, for instance older students may develop their interests more narrowly, in a way that requires more intensively motivational writing instruction.

Second, there is variation in writing quality results between schools. At some schools age does not seem to matter when it comes to improving writing quality. For 13 of the schools there is a significant effect size of .67 among 300-students. For six (6) schools there is a significant effect size of .90 for 600-students. At three of these high achieving schools (School 1, School 3, School 5), covering 30.7% of the project students, there is no difference in results that varies with age group. At one participating school (School 16), the students, regardless of age group, do not improve their writing more than comparison students at all during the project. This important outcome seems to contradict the age difference pattern presented above.

Third there is variation in writing quality results among classes at the same schools. At one school (School 19) 300-students have achieved a very high level of writing quality not matched by their fellow 600-students. At two other schools (School 11, School 13) the opposite pattern may be observed: 600-students at these schools have achieved a very high level of writing quality not matched by their fellow 300-students at the same school. These three schools cover 13.9% of the participating students. The variation between schools, in between classes at the same school may indicate different learning cultures and different learning environments, and possibly teacher effects.

Fourth, there is variation in writing quality results among students both individually and between groups at different levels of ability. We have seen that low achieving students have not improved their writing at a level comparable to other students. This observation contradicts findings in some research projects (Parr et al., 2007), but corroborates findings in others (e.g. Myhill et al., 2012). Another observation is that the writing quality of the students' texts varies with writing events, and writing tasks. Only 37% of the variance amongst 300-students and 23% of the variance amongst 600-students in the low stake posttest in October 2014 may be predicted by results of the pretest in October 2012 (Berge & Skar, 2015). This outcome is in agreement with other writing research (Breland, Camp, Jones, Morris, & Rock, 1987; Gorman et al., 1988). Obviously, students' writing competency is an unstable context-sensitive resource. This result is accordance for comparable research (Bouwer, Béguin, Sanders, & van den Bergh, 2015).

Thus, the answer to the research question presented in the introduction is that the substantial variation in writing quality effects between schools, classes, and individual students, suggests that it is plausible to assume the exposure to new semiotic re-contextualisation of school subjects using the Wheel of Writing and associated norms of expectation as a basis for teaching and assessing writing *does not in itself* lead to writing development.

Still, another finding is that at most project schools the project has contributed to the development of students' writing ability at a quality level they would not have reached if they had not participated in the project. An important discovery in the NORM-project is that the students participating in the NORM-project in comparison to students taking the NSBWT obtained relatively high results on the posttest. The posttest was administrated more than 5 months after they had been exposed to a writing instruction using the writing construct and expectation norms developed by the NORM-project. Most 600-students had also changed classes and schools between the end of the project and the posttest. They thus took the posttest in a completely different learning environment than the writing instruction environment at their old school. The fact that the project schools perform significantly

better on the posttest (i.e. the national writing test) than both comparison school students and Norwegian students in general, implies therefore that the project schools have developed sustainably improved writing competence by participating in the NORM-project. Consequently, the answer to the research question of whether the educational integration of an explicit writing construct and shared explicit standards have specific consequences for the quality of pupils' writing or not, is complex and points to individual and group-related explanations, as well as to how the writing construct and shared explicit expectation norms have been understood, developed, and used as semiotic tools in writing instruction at the different schools and in the different classes. Further studies of how the project has been developed at the schools or classes with high writing quality gains, might provide information to explain these possible consequences. Corresponding studies at schools or classes with no learning effect on students' writing abilities might explain the missing effect. More qualitatively oriented on-going research in the NORM-project (e.g. Matre & Solheim, 2015) indicates some factors that may lead to valid interpretations of the variation in writing quality results:

First, the general learning environment established before and independent of the participation in the project may explain the improved writing quality at some schools. For instance, eight of the high achieving schools (either in one or both age groups) do systematically obtain higher than normal results at the yearly Norwegian national test in reading. Another observation is that even if the students at some of these schools had not achieved more than a normal level at the pretest in October 2012, it did not take long before they performed better than students from most other schools. A group of high achieving schools had already reached a high quality level in the students' texts before participating in the project, indicating that competent writing instruction was established as a habit at schools regardless of teacher, or amongst some teachers at the schools.

Second, most teachers and principals report that they fully support the project. They considered it as a necessary, as well as supplementary, teaching practice that in a relevant way translated the National curriculum to meaningful writing instruction. Still, at three of the low achieving schools the qualities of the local implementation may partly explain the negative results on students' writing. These schools lacked stable and consistent leadership. The principals or other leaders have left the school during the project, leaving it up to the individual teacher to implement a new writing construction. On the contrary, the four high achieving schools regardless of age group, were all characterised with a strong and committed leadership and a distinct engagement, not only from project teachers at the schools but all teachers regardless of whether they participated in the project or not. Therefore, the writing results may not be explained by individual verbal or discursive commitment, but to systematic and practice changing implementation of a new writing instruction at the schools making possible/resulting in a new learning environment. It is only under such conditions that the semiotic tools developed in the NORM-project (i.e. the Writing Wheel and the norms of expectation) may be interpreted and used in a way that in the long run and in stable way improves the text quality of the students' texts.

As mentioned earlier, the research design leaves some of the processes black-boxed. Therefore, future research that might explain the results of the NORM-project should focus on the different writing instruction consequences presented in various meta-analyses (Bangert-Drowns, Hurley, & Wilkinson, 2004; Graham et al., 2012; Graham & Perin, 2007; Koster, Tribushinina, de Jong, & van den Bergh, 2015):

> at a more general level, these findings show that it is advantageous to explicitly and systematically
> teach adolescents the processes and strategies involved in writing (including planning, sentence construction, summarizing, and revising). It is also advantageous for teachers to structure writing by having students work together in an organized fashion, establishing clear

and reachable goals for writing assignments, providing models of what the end product should look like, and engaging students in activities that help them acquire, evaluate, and organize ideas for their writing. (Graham & Perin, 2007, p. 467)

Consequently the results from the NORM-project indicate that the explanation could be a combination of the three following qualities:

(1) How the project in general and writing construct and norms of expectation in special were understood and integrated locally by the school leaders, as well as by the participating teachers.

(2) How writing instruction and writing assessment have been developed and carried out through the project. For instance how the writing prompts have been developed, presented and modelled for the students.

(3) Whether there have been feedbacks (Hattie & Timperley, 2007) to students' writing encouraging writing strategies like planning and revising.

In other words, might the consequences of participating in the NORM-project be explained by the specific qualities of the writing instruction at the project schools, the project classes as well as the project teachers where students have been exposed to writing in the subjects in accordance with the Norwegian curriculum of 2006.

Notes

1. The model and definition of standards are exhaustively presented in (Berge et al., 2016) and (Evensen et al., 2016), respectively.

2. The 'norms of expectation' express performance standards developed for students at the end of school year 4 and at the end of school year 7. The development of these standards is presented in Evensen et al. (2016).

3. There are two official norms of written Norwegian, 'New Norwegian (*nynorsk*)' and 'Book Language (*bokmål*)'. Students learn to write in New Norwegian or Book Language.

4. Teachers created writing tasks according to the local curriculum.

5. Observations of these classroom activities are being analysed in ongoing research.

6. To allow for comparisons across groups, the pretest and posttest dealt with general topics.

7. The NSBWT is a writing test on a nationally representative sample of students in primary and lower secondary school (NSBWT-5 for school year 5 and NSBWT-8 for school year 8). See e.g. Skar and Jølle (in press).

8. The NSBWT panel consists of 80 raters, trained to rate student scripts.

9. These are the six domains included in the NSBWT, and for reasons of comparison, the seventh domain was not included in the NORM-project.

10. Not all students that participated in both pretest and posttest wrote two pretest tasks. Of the 300-student, 36 completed only one task (15 in in the project group, and 21 in the control group). Of the 600-students, 41 completed only one task (32 in the project group and 9 in the control group). To put all scores on the same scale, scores from students writing only one pretest task were doubled.

11. It should be noted that the NSBWT for 10-year olds was scored on slightly different scales. While mid-level was identical to the one used in the NORM-project, all scales had only three band-levels. This caused re-scaling, transforming a NSBWT scale value of 2.0 to 3.0, and NSBWT scale value 3.0 to 5.0, leaving 2.0 and 4.0 as unobserved categories.

12. There were no comparison group students scoring above 3.50 on the pretest.

References

Aasen, P., Møller, J., Rye, E., Ottesen, E., Prøitz, T. S., & Hertzberg, F. (2012). *Kunnskapsløftet som styringsreform - et løft eller et løfte?* [Knowlegde promotion as reform.] Oslo: NIFU/Institutt for lærerutdanning og skoleforskning, University of Oslo.

Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based writingto-learn interventions on academic achievement: a meta-analysis. *Review of Educational Research, 74*, 29–58.

Berge, K. L. (2005). Skriving som grunnleggende ferdighet og som nasjonal prøve - ideologi og strategier [Writing as basic competency and as national test – ideology and strategies]. In A. J. Aasen & S. Nome (Eds.), *Det nye norskfaget* (pp. 1–22). Oslo: Fagbokforlaget.

Berge, K. L., & Skar, G. B. (2015). *Ble elevene bedre skrivere? Intervensjonseffekter på elevers skriveferdigheter og skriveutvikling (Rapport 2)* [Did the students become better writers? Intervention effects on students' writing competency and development of writing]. Trondheim: Høgskolen i Sør-Trøndelag, Avdeling for lærer- og tolkutdanning.

Berge, K. L., Evensen, L. S., & Thygesen, R. (2016). The Wheel of Writing: A model of the writing domain for the teaching and assessing of writing as a key competency. *The Curriculum Journal,* 1–18, doi:10.1080/09585176.2015.1129980

Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing, 32*, 83–100. doi:10.1177/0265532214542994

Breland, H. M., Camp, R., Jones, R. J., Morris, M. M., & Rock, D. A. (1987). *Assessing writing skill*. New York, NY: College Entrance Examination Board.

Coe, R. (2000). *Effect size calculator* [Excel spreadsheet]. Retrieved April 23, 2015, from http://www.cem.org/effect-size-calculator

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Evensen, L. S., Berge, K. L., Thygesen, R., Matre, S., & Solheim, R. (2016). Standards as a tool for teaching and assessing cross-curricular writing. *The Curriculum Journal,* 1–17. doi:10.1080/09585176.2015.1134338

Fidalgo, R., Torrance, M., & García, J.-N. (2008). The long-term effects of strategy-focussed writing instruction for grade six students. *Contemporary Educational Psychology, 33*, 672–693. doi:10.1016/j.cedpsych.2007.09.001.

Galbraith, D., & Rijlaarsdam, G. (1999). Effective strategies for the teaching and learning of writing. *Learning and Instruction, 9*, 93–108. doi:10.1016/S0959-4752(98)00039-5

Glaser, C., & Brunstein, J. C. (2007). Improving fourth-grade students' composition skills: Effects of strategy instruction and self-regulation procedures. *Journal of Educational Psychology, 99*, 297–310. doi:10.1037/0022-0663.99.2.297

Gorman, T. P., Purves, A. C., & Degenhart, R. E. (Eds.). (1988). *The IEA study of written composition I. The international writing tasks and scorings scales*. Oxford: Pergamon Press.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology, 99*, 445–476. doi:10.1037/0022-0663.99.3.445

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*, 879–896. doi:10.1037/a0029185

Harris, K. R., Graham, S., & Mason, L. H. (2006). Improving the writing, knowledge, and motivation of struggling young writers: Effects of self-regulated strategy development with and without peer support. *American Educational Research Journal, 43*, 295–340. Retrieved from http://www.jstor.org/stable/3699421

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* London: Routledge.

Hattie, J. (2015). Synlig læring i dag [Visible learning today]. *Paideia, 5*, 9–22. Retrieved from https://utdanningsforskning.no/artikler/hattie-synlig-laring-i-dag/

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*, 81–112. doi:10.3102/003465430298487.

Hertzberg, F., & Roe, A. (2015). Writing in the content areas: A Norwegian case study. *Reading and Writing.* doi:10.1007/s11145-015-9607-7.

Hillocks, G. (1986). *Research on written composition: New directions for teaching.* Urbana, IL: National Council of Teachers of English.

Hillocks, G. (1999). *Ways of thinking, ways of teaching.* New York, NY: Teachers College Press. IBM (2015). *Spss Statistics (version 23)* [Computer Software]. New York, NY: IBM.

Koster, M., Tribushinina, E., de Jong, P. F., & van den Bergh, H. (2015). Teaching children to write: A meta-analysis of writing intervention research. *Journal of Writing Research, 7*, 249–274. doi:10.17239/jowr-2015.07.02.2

Lumbelli, L., Paoletti, G., & Frausin, T. (1999). Improving the ability to detect comprehension problems: From revising to writing. *Learning and Instruction, 9*, 143–166. doi:10.1016/S0959-4752(98)00041-3

Matre, S., & Solheim, R. (2015). Writing education and assessment in Norway: Towards shared understanding, shared language and shared responsibility. *L1 Educational Studies in Language and Literature, 15*, 1–33. doi:10.17239/L1ESLL-2015.15.01.05

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 Utdanning 2020). Trondheim: Nasjonalt senter for skriveopplæring og skriveforsking.

McCarthey, S. J., & Geoghegan, C. M. (2016). The role of professional development for enhancing writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (2nd ed.). (pp. 330–345). New York, NY: The Guilford Press.

McConnell, A. E., Little, K. S., & Martin, J. E. (2015). Transition planning and writing instruction: The effects of a brief intervention. *British Journal of Special Education, 42*, 87–111. doi:10.1111/1467-8578.12071

Mertz, E., & Parmentier, R. J. (Eds.). (1985). *Semiotic mediation: Sociocultural and psychological perspectives.* Orlando: Academic Press.

Myhill, D. A., Jones, S. M., Lines, H., & Watson, A. (2012). Re-thinking grammar: The impact of embedded grammar teaching on students' writing and students' metalinguistic understanding. *Research Papers in Education, 27*, 139–166. doi:10.1080/02671522.2011.637640

Norwegian Directorate for Education and Training. (2007). *The knowledge promotion.* Retrieved from http://www.udir.no/Stottemeny/English/Curriculum-in-English/_english/Knowledge-promotion–Kunnskapsloftet/

Organisation for Economic Co-operation and Development. (2005). *The definition and selection of key competencies – Executive summary.* Paris: Author. Retrieved from http://www.oecd.org/pisa/35070367.pdf

Pallant, J. (2013). *SPSS survival manual*. Maidenhead: McGraw-Hill.

Paquette, K. R. (2008). Integrating the 6+1 writing traits model with cross-age tutoring: An investigation of elementary students' writing development. *Literacy Research and Instruction, 48*, 28–38. doi:10.1080/19388070802226261

Parr, J. M., & Timperley, H. (2010). Multiple "black boxes": Inquiry into learning within a professional development project. *Improving Schools, 13*, 158–171. doi:10.1177/1365480210375349

Parr, J. M., Timperley, H., Reddish, P., Jesson, R., & Adams, R. (2007). *Literacy professional development project: identifying effective teaching and professional development practices for enhanced student learning*. Wellington, NZ: Ministry of Education. Retrieved from www.minedu.govt.nz/goto/2107

Pritchard, R. J., & Marshall, J. C. (1994). Evaluation of a tiered model for staff development in writing. *Research in the Teaching of English, 28*, 259–285.

Russell, D. R., & Yañez, A. (2003). 'Big picture people rarely become historians': Genre systems and the contradictions of general education. In C. Bazerman & D. R. Russell (Eds.), *Writing selves, writing societies: research from activity perspectives* (pp. 331–362). Fort Collins, CO: WAC Clearinghouse. Retrieved from http://wac.colostate.edu/books/selves_societies

Säljö, R. (2000). *Lärande i praktiken* [Learning in practice]. Stockholm: Prisma.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking contentarea literacy. *Harvard Educational Review, 78*, 40–59. doi:10.17763/haer.78.1.v62444321p602101

Skar, G. B., Evensen, L. S., & Iversen, J. M. (2015). *Læringsstøttende prøver i skriving 2014. Teknisk rapport* [Formative writing tests 2014. Technical report]. Trondheim: Nasjonalt senter for skriveopplæring og skriveforskning.

Skar, G. B., & Jølle, L. (in press). *Teachers as raters: Investigation of a long-term writing assessment program*.

Tracy, B., Reid, R., & Graham, S. (2009). Teaching young students strategies for planning and drafting stories: The impact of self-regulated strategy development. *The Journal of Educational Research, 102*, 323–332. doi:10.3200/JOER.102.5.323-332

Wertsch, J. V. (1991). *Voices of the mind. A sociocultural approach to mediated action*. Cambridge, Mass.: Harvard.

Wiseheart, M. (2014). *Effect size calculator*. Retrieved from http://www.yorku.ca/ncepeda/effectsize.html

Appendix A

LEVEL	Level 1	Level 2	Level 3	Level 4	Level 5
MEANING	Very low level of mastery within the domain	Low level of mastery	**As to be expected from most pupils after 4 or 7 years of schooling**	High level of mastery	Very high level of mastery within the domain
DESCRIPTOR for scale used at school year 4	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.)				
DESCRIPTOR for scale used at school year7	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 organizing principle • Create cohesion by a variety of connectors				

The scale "text organization" with descriptors for school years 4 and 7