

4

PhD revisited: Formative assessment of writing in English

A school-based study of perceptions, practices and transformations¹

TONY BURNER

University of South-Eastern Norway

ABSTRACT This chapter reports from a doctoral study (Burner, 2016a) that explored teacher and student perceptions and practices of formative assessment (FA) in English writing classes. Four English teachers and their students ($N=100$) took part in the study. The assessment situation was analyzed using mixed methods before a plan for intervention cycles was made continuously throughout a school year. The main results, their implications for teaching English in Norway, and further research will be discussed in this chapter.

KEYWORDS formative assessment | writing assessment | portfolio assessment | classroom interventions

-
1. The chapter presents the overall results of a doctoral study (Burner, 2016a), focusing specifically on its implications for the teaching of English in Norway. This is an article-based thesis, with three published articles (Burner, 2014, 2015, 2016b). The thesis in its entirety can be found here: <https://brage.bibsys.no/xmlui/handle/11250/2379861>

INTRODUCTION

Some of the questions teachers need to reflect on are: What happens prior to and after assessment practices? What are the roles of the students in the different phases of assessment? And how can assessment of student texts promote learning? Politicians and the Ministry of Education and Research have shown a special interest in the field of assessment after the introduction of the current national curriculum in 2006 (Norwegian Ministry of Education and Research, 2006, 2013). Seminars and workshops have been held for teachers and teacher educators on formative assessment (FA), commonly known as assessment *for* learning. Undoubtedly, the increased international research base on FA has influenced the national context. The research gap addressed in the doctoral study reported from here concerned FA in English writing classes in Norway, the views and practices of both teachers and students, and how their views and practices might change over time when writing portfolios as a new formative tool is introduced. Portfolio assessment was used as a tool in this study to enhance FA perceptions and practices. A definition of portfolio assessment is offered by Johnson, Mims-Cox and Doyle-Nichols (2010, p. 5), wherein a portfolio:

- ▶ is a collection of work that has been compiled over a period of time
- ▶ is organized to assess competencies in a given standard, goal, or objective and focus on how well the learner has achieved in that area
- ▶ makes learning concrete and visible
- ▶ is evidence of knowledge (what the student knows), skills (what the student is able to do), and dispositions (reveal the student's attitudes, beliefs, or values)
- ▶ includes higher levels of thinking through the use of inquiry (a process of collecting, sorting, selecting, describing, analyzing, and evaluating) and reflection (questioning and sorting of the selected work; questioning how he or she can improve personal practice)

Hamp-Lyons and Condon (2000) suggest three main categories that define writing portfolios, namely collection of texts over a longer period of time, reflection on those texts, and selection of some of them for final assessment.

Thus, the primary research question for the doctoral study was: In what ways can portfolio as a formative tool influence teachers' and students' perceptions and practices of formative assessment?

In order to study the influences of portfolio assessment, the base-line situation had to be investigated to understand how teachers and students perceive and practice FA in English classes. In addition, the formative potentials of the tool used to enhance

FA, i.e. the writing portfolio, had to be reviewed. Finally, a plan for interventions was made together with the teachers at the end of the school year 2012–2013 and the very beginning of the next school year, 2013–2014. The aim of the interventions was to make use of assessment situations to enhance student learning. The interventions were guided by principles from portfolio assessment (Johnson et al., 2010).

THEORY

Black and Wiliam (1998) acknowledge in their seminal work that FA does not have a “tightly defined and widely accepted meaning” (p. 7). They point out that it consists of several classroom practices: self- and peer assessment, learning strategies, goal orientation, effective tasks, useful feedback, and so forth. Their definition of FA is “all those activities undertaken by teachers, and/or by their students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (Black & Wiliam, 1998, pp. 7–8). The expression *assessment for learning* (AfL) was coined by the Assessment Reform Group (ARG, 1989–2010) to stress the learning dimension since not all assessments, for example mini-summative assessments, lead to further learning (Stobart, 2008).

It was particularly the Assessment Reform Group’s work (ARG, 2002) that effectively put FA on the educational agenda after its decline by 1995 (Black & Wiliam, 2003). Their definition of FA is “The process of seeking and interpreting evidence for use by learners and their teachers, to identify where the learners are in their learning, where they need to go and how best to get there” (ARG, 2002). Wiliam (2006) provides a short and concise definition of FA: “An assessment of a student is formative if it shapes that student’s learning” (p. 284). What Wiliam (2006) lacks in his definition is the type of shaping that teacher and student learning go through. Arguably, shaping that is detrimental to a student’s learning cannot be characterized as formative. Wiliam’s (2006) main concern is that FA has to be integrated as a part of teachers’ daily teaching: “Tools for formative assessment will only improve formative assessment practices if teachers can integrate them into their regular classroom activities. In other words, the task of improving formative assessment is substantially, if not mainly, about teacher professional development” (p. 287). In a follow-up article, Black and Wiliam (2009) add to their definition that this information, or evidence, is used “to make decisions about the next steps in instruction that are likely to be better, or better founded, than the decisions they [teachers and/or students] would have taken in the absence of the evidence that was elicited” (p. 9). Teachers were the focal point of carrying out the interventions in this doctoral study, specifically using portfolios to enhance FA of writing.

Taking into consideration FA's documented learning benefits (Black & Wiliam, 1998), including in foreign language learning contexts (Ross, 2005), the question is then how it is perceived and practiced in specific learning contexts, and in what ways teachers and students take advantage of that information or evidence to enhance teaching and learning. According to Lee (2011), self- and peer assessment, where students assess their own and each other's works, formative feedback through multiple drafting/text revision, conferencing and portfolios are ways of realizing FA in the writing classroom. Such classroom activities allow reflection, interaction, and opportunities to return to one's text and improve it. A review of research on second or additional language (L2) writing and response to L2 writing shows that some scholars contend that L2 writing can and should draw on insights from L1 writing, whereas others claim that L2 writing is inherently different from L1 writing and should be considered separately (Ferris, 2003). Zamel (1985) belongs to the former category and suggests a process-oriented approach to L2 writing. This type of approach to writing includes assessment procedures that correspond to the process nature of writing, for example portfolio assessment (Weigle, 2007). Portfolio is a useful tool for assessment in light of the development described above due to its formative benefits in second and foreign language writing classes.

REVIEW

The research that is relevant for this study is to be found in the interface between formative assessment, writing assessment and portfolio assessment. Internationally, as pointed out by Abedi (2010), there is little research on FA of writing in English as a foreign language (EFL) compared to the vast amount of research on summative assessment of writing in EFL. Previous research into EFL writing has been analytical, focusing mainly on error corrections and their possible effects on students' writing (cf. Hyland's and Ferris's classic studies). Some exceptions are Lee and her colleague's research, shedding some light on FA of writing in EFL (Lee, 2011; Lee & Coniam, 2013). They maintain that the focus of assessment of writing in EFL has been retrospective, i.e. the assessment has served mainly summative purposes, not formative. In addition, research on FA has concentrated mainly on teacher views (Broadfoot & Black, 2004; Brookhart, 2001; Cowie, 2005). Given the dialectic nature of FA in which feedback informs both learning and teaching activities, the present doctoral study added to the research on student perspectives and the relations between their perspectives and teachers' perspectives.

Nationally, research reports (e.g., Sandvik et al., 2012) have investigated assessment in four subjects, amongst others English. The main findings were that teachers are predominantly positive toward the new emphasis on FA, that teachers' subject knowledge seems to be an important factor for their assessment competency, and that the lower grades (grades 1–7) have reached furthest in FA practices, such as self-assessment and student involvement. Empirical research on formative uses of writing assessment has been mainly conducted from a pedagogical perspective in contexts of higher education and in secondary schools (Baird, Hopfenbeck, Newton, Stobart & Steen-Utheim, 2014). Furthermore, there has been some assessment research conducted in L1 contexts, and in foreign languages such as German (e.g., Sandvik, 2011). Hasselgreen, Drew, and Sørheim (2012) collected some of the most important research into foreign languages in Norway, mostly for English. From the collection, it is evident that writing research has been confined to orthography and focus on errors, whereas assessment research has been looking at the use of the Common European Framework (CEFR, 2001) in assessing writing and oral exams in 10th grade. Other assessment research in English in Norway revolves around summative assessment (Reisjø, 2006; Thorenfeldt, 2005; Yildiz, 2011). The present study focused on FA of writing in English.

METHODOLOGY

Cultural-Historical Activity Theory (CHAT), which is rooted in a sociocultural approach to learning and development (Vygotsky 1978, 1986), was the methodological choice for the study. CHAT pays attention to historicity and the current situation, to the individual and the collective system. According to Wardekker (2000), CHAT is a systematic approach to analyzing and developing formative assessment. Pryor and Crossouard (2008) place the theory of formative assessment within CHAT, and claim that it is a useful tool of analysis due to its simultaneous problematization of agency, in this case students and teachers enacting formative assessment.

SAMPLE

For the purpose of the study, it was important to select a school that had FA as one of the main developmental themes, school leaders that support teachers in their professional development, and English teachers who were interested in working with FA but met certain challenges during their teaching practice. The selected

school met all these criteria. Four of the teachers (T1–T4) and their students ($N=100$) gave their consent to participate, two classes in 8th grade, and two in 9th grade. Two of the teachers, T1 and T3, were very experienced, both having taught for around 30 years each. The two others, T2 and T4, were less experienced, having taught for 1–3 years each. They were all female and fulfilled the formal criterion for being qualified English teachers, meaning that they had completed at least 60 ECTS of English as part of their teacher education (equal to one full year study). The students were 14–15 years old, and 12 of these were also selected for interviews, according to their self-reported and teacher-reported level of proficiency in English, i.e. low proficiency (grades 1–2), medium proficiency (grades 3–4), and high proficiency (grades 5–6).

RESEARCH DESIGN AND DATA COLLECTION METHODS

During the period of data collection, the teachers had the opportunity to discuss and reflect upon existing practices and to experiment with new assessment practices, before reflecting again and evaluating the new practices. Through work with a process-oriented assessment practice, the students were given the chance to work with the teacher’s feedback to improve their texts and assess their own progress when doing so. Figure 4.1 illustrates the research process in its three phases: historical and situational analyses, intervention, and evaluation.

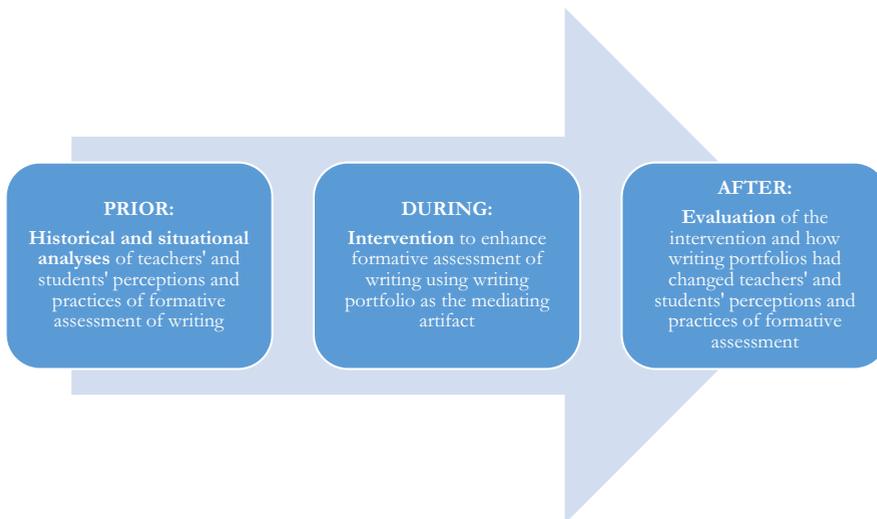


FIGURE 4.1. The intervention and research process.

Data were collected prior to (Phase 1), during (Phase 2), and after (Phase 3) the intervention period. The aim of the data collection was to understand the assessment situation in English writing classes before using portfolios to enhance FA of writing, and to understand how portfolios could influence teachers' and students' perceptions and practices of assessment. As part of the intervention, the researcher led workshops with the teachers. The formative interventions draw on Engeström's activity theory (1987). His activity system was used to find any contradictions in teachers' and students' enactment of formative assessment.

Intervention research aims at facilitating change through formative interventions (Engeström & Sannino, 2010). *Phase 1:* Prior to the intervention period, student questionnaires were handed out to all the students, semi-structured focus group interviews with students from all the four classes and semi-structured one-to-one interviews with all the four teachers were carried out, and observations of writing assessment classes were conducted. The objective was to understand the participants' perceptions and practices of FA of writing. *Phase 2:* During the intervention period, focus group interviews with the same students and continuous observations of writing assessment classes were carried out in order to understand how perceptions and practices of FA of writing would be influenced by using writing portfolios. *Phase 3:* The same procedures for data collection as in the beginning of the study were applied after a period of intervention in three of the classes. The primary sources of data were the interviews and questionnaires, whereas classroom observations were a secondary source of data used to validate findings from the primary sources. One class dropped out of the study half way through due to one teacher's long-term sick leave. Table 4.1 shows the data collection methods used in the various phases of the study.

TABLE 4.1. An overview of the data collection methods

Phases	Data collection	Sample	Duration
Phase 1. Prior (May – Sept 2012)	Observations	4 classes	4–5 hours
Phase 1	Questionnaire	Students, 4 classes, $N=100$	–
Phase 1	Focus group interviews	3 students selected from 4 classes, $N=12$	131 minutes
Phase 1	One-to-one interviews	Teachers T1–T4, $N=4$	162 minutes
Phase 2. During (Oct 2012 – Apr 2013): Intervention	Observations	3 classes	18–19 hours

Phases	Data collection	Sample	Duration
Phase 2	Focus group interviews	3 students selected from 3 classes, $N=9$	66 minutes
Phase 3. After (May – June 2013)	Observations	3 classes	3-4 hours
Phase 3	Focus group interviews	3 students selected from 3 classes, $N=9$	72 minutes
Phase 3	One-to-one interviews	Teachers T1–T3, $N=3$	156 minutes
Phase 3	Questionnaire	Students, 3 classes, $N=70$	–

STUDENT QUESTIONNAIRES

Based on readings of the research literature on FA of writing (see theoretical overview), indicators that were thought to shed light on aspects of FA of writing were identified. The questionnaire consisted of four parts: background variables (gender, class, grade level, language background), a set of items related to student perceptions of FA of writing, a set of items related to students' self-perceived practices of FA of writing, and a set of items related to their preferences of the same practices. The questionnaire was handed out to 8th and 9th grade students of English before ($N=100$) and after ($N=70$) the intervention. For the piloting of the questionnaire, it was given to all the students in 9th and 10th grade ($N=174$). None of these students would participate in the study later. The questionnaire was also handed out to students of English at another randomly selected school ($N=145$). The pilot testing led to changes of items that proved to be unclear for some students. On the part concerning perceptions, students could respond to five items on a 5-point Likert scale (1 = very little extent to 5 = very great extent). The binary items, where the informants could respond with yes/no, concerned students' self-perceived practices and preferences. All items loaded on a single factor with item loadings ranging from .671 to .361. The lowest loading, which was slightly $<.40$, was not discarded since it could theoretically be justified (Ringdal, 2007, p. 293) to be an important part of the construct "formative assessment of writing". Factor analyses with the final data supported a one-factor solution. The loadings from the pre-intervention data ranged from .470 to .796. The loadings from the post-intervention data ranged from .517 to .691. Reliability, expressed through Cronbach's alpha, was found to be .71 for the pre-intervention data and .74 for the post-intervention data, which is a fairly good reliability measure (Ringdal, 2007). Cross-

tabulation analyses were used to reveal any interesting relationships between students' responses and their background. To understand the difference between pre and post intervention, inferential statistics were used, i.e. *t*-tests for the Likert scale items and chi-square tests for the binary items.

TEACHER AND STUDENT INTERVIEWS

One-to-one interviews with the teachers and focus group interviews with 12 students were conducted (see Table 4.1). I considered focus group interviews to be a suitable way of discussing the topics with them. Focus group interviews open up for different views to be discussed (Kvale & Brinkmann, 2009). An interview guide was used for the interviews. All the interviews were recorded. They were semi-structured (Kvale & Brinkmann, 2009), meaning that I had some topics and questions beforehand, related to FA of writing. At the same time, I was open for other relevant issues to be mentioned by the informants. I transcribed the interviews, becoming acquainted with the data and learning a lot about my own style of interviewing (Kvale & Brinkmann, 2009). The constant comparative method of analysis inspired the analyses of the interview transcripts (Creswell, 2013). The first interview transcript was coded together with a colleague, also a PhD candidate who was conducting research on FA, to ensure higher reliability than would be achieved alone.

CLASSROOM OBSERVATIONS

Classroom observations were conducted at the beginning, during and at the end of the study. I was conscious of distinguishing between what I observed and my comments on what I observed (Table 4.2).

TABLE 4.2. An extract from classroom observation notes.

Time and place	Topic	What I see	Comments
September 10, 2012 11:50–12:50; T2's class	Teacher feedback on student letters	Students look at teacher's corrected version of text. Around half of the students have the text in front of them. Some have not, and some have not handed in.	Students have problems seeing the connections between teacher's codes and her corrections.

My role as an observer was the non-participant/observer as participant (Creswell, 2013). I needed to have a participative role as well, where I could achieve a close inside perspective without being one of the participants. In total, I observed English writing classes for 27 hours. The observation notes were used to validate findings from the other data collection methods, i.e. observing to what extent the informants enacted formative assessment the way they described their practices in the interviews.

FINDINGS

In this chapter the findings from the doctoral study will be separated into two areas: those from the base-line study prior to the interventions (Phase 1), where student and teacher perceptions and practices of FA of writing in English were investigated, and those from the intervention study, where writing portfolios were used as a tool to enhance FA.

THE BASE-LINE STUDY: PERCEPTIONS AND PRACTICES

The data from the questionnaires ($N=100$) provided a broad picture of students' responses to how they perceive FA of writing, which aspects they appreciate, and which aspects they find challenging. They reveal that students are positive to their learning outcome and reflective work in English writing classes. A majority of them appreciate text revision. Furthermore, they appreciate being involved in assessment practices, despite claiming that they are not being involved currently. On the other hand, they have different views on teacher feedback and grades. They would appreciate teacher modeling of quality texts. Finally, they claimed they conduct self-assessment, but reported that they do not learn much from it. There was consequently a gap between students' experiences of self-assessment and involvement in assessment practices on the one hand, and their perceptions of them on the other (see Tables 4.3 and 4.4).

The data from the student interviews showed that students have the impression that grades interfere with feedback, sometimes merely because their teacher had told them so. They appreciate feedback, but underscored that they would like to receive more positive feedback, as illustrated by the following quote: "The feedback is too much about the negative things. [The teacher] could write about what was good too" (average performing boy, 9th grade). As for text revision, they said that it is not a common activity in class. In other words, they appreciate revising their texts, but are not offered the opportunity to do so. Moreover, they claimed

they are not involved in assessment practices, and some of them revealed a rather simplistic understanding of what student involvement could mean. The students admitted that they do not follow up feedback on their texts: “I don’t do anything. I don’t think about it. If I’ve made mistakes on a test, then I don’t think about them for the next test” (high-performing girl, 9th grade). The main reason seems to be the negative form of the feedback, but also that they do not always understand the content of the feedback. Classroom observations concurred with the interview data on these points. Moreover, the students preferred a more coherent English subject where various topics are followed up and worked on more thoroughly. Other factors they mentioned which they found particularly challenging with writing assessment in English were grammar and that English is a *foreign* language to them.

In the data from the teacher interviews, the teachers claimed that grades interfere with assessment and that they use assessment rubrics to standardize their assessment practices. They were unsure whether students follow up their feedback or not – “I’m a little unsure about the extent to which the students read the feedback [...] very often they make the same mistakes” (T4, 9th grade) – and frequently consider whether they should provide feedback in Norwegian or in English. They believe writing assessment is more challenging than oral assessment. They emphasized the importance of self-assessment, but at the same time made it clear that 8th graders are not mature enough to reflect upon their own learning. Students were described as possessing inherent characteristics that make them either more capable or less capable in following up the teacher’s feedback. Furthermore, the teachers acknowledged the value of feedback and expressed a research-based understanding of what useful feedback should be like: “we remind each other that the focus should be the way forward for the students” (T3, 9th grade). When asked about challenges in assessment that can be related to the nature of the subject of English, the teachers pointed out four factors: lack of time, wide-ranging subject, subjectivity in assessing students’ work, and the gap between students when it comes to their knowledge of the subject.

THE INTERVENTION STUDY: TRANSFORMATIONS

Table 4.3 shows that students rated feedback from the teacher (Q1), reflection on their own writing (Q4) and teacher modeling of good texts (Q5) higher after the intervention. T-values ranged from $t = -.41$ ($df = 167$, $p = .68$) for Q5 to $t = -1.51$ ($df = 167$, $p = .13$) for Q1, thus not revealing any significant differences in students’ view on aspects of FA of writing before and after the intervention.

TABLE 4.3. Mean and standard deviation for Q1–Q5 before and after intervention.

	Before (M) N=100	After (M) N=70	Before (SD) N=100	After (SD) N=70
Q1. To what extent would you say feedback from the teacher helps you improve your English?	3.49	3.71	.896	.980
Q2. To what extent would you say grades from the teacher help you improve your English?	3.52	3.41	.925	1.110
Q3. To what extent would you say you learn to express yourself in written English?	3.65	3.47	.837	1.003
Q4. To what extent do you think about how you work with a text?	3.63	3.70	1.026	.998
Q5. To what extent would you say the teacher helps you with understanding what a good text is?	3.57	3.63	1.032	.887

Table 4.4 shows that grades on texts (Q2a), text revision after receiving feedback (Q4a) and student involvement in assessment practices (Q6a) revealed the biggest differences between the student scores on assessment practices before and after the intervention period.

TABLE 4.4. Students' perceptions of FA of writing in English before and after intervention.

	Before "No"% (N)	After "No"% (N)	Before "Yes"% (N)	After "Yes"% (N)
Q1a. The teacher gives me both grade and feedback on my texts	66.3 (65)	80.0 (56)	33.7 (33)	20.0 (14)
Q1b. I learn best by receiving both grade and feedback on my texts	20.6 (20)	20.3 (14)	79.4 (77)	79.7 (55)
Q2a. The teacher gives me only grades on my texts	87.8 (86)	100	12.2 (12)	0
Q2b. I learn best by receiving only grades on my texts	84.5 (82)	80.0 (56)	15.5 (15)	20.0 (14)

	Before “No”% (N)	After “No”% (N)	Before “Yes”% (N)	After “Yes”% (N)
Q3a. The teacher gives me only feedback on my texts	15.3 (15)	14.3 (10)	84.7 (83)	85.7 (60)
Q3b. I learn best by receiving only feedback on my texts	55.8 (53)	67.1 (47)	44.2 (42)	32.9 (23)
Q4a. I get the opportunity to work more with a text I have received feedback on	25.0 (24)	2.9 (2)	75.0 (72)	97.1 (68)
Q4b. I learn best by working more with a text I have received feedback on	24.2 (24)	17.6 (12)	75.8 (75)	82.4 (56)
Q5a. I assess some of my own texts	33.0 (31)	25.7 (18)	67.0 (63)	74.3 (52)
Q5b. I learn best by assessing some of my own texts	63.9 (62)	55.1 (38)	36.1 (35)	44.9 (31)
Q6a. I take part in deciding the process of assessment	80.0 (68)	60.9 (42)	20.0 (17)	39.1 (27)
Q6b. I learn best by taking part in deciding the process of assessment	43.3 (39)	38.6 (27)	56.7 (51)	61.4 (43)

After the intervention, the students agreed unanimously that grades were downplayed in the teacher’s assessment of their writing (Q2a). The majority of the students (97%) agreed that they were given the chance to revise their texts (Q4a). And finally, the number of students who experienced that they were more involved in assessment practices had doubled (Q6a). The statistical analyses on all the binary items confirmed this by revealing that Q2a ($\chi^2 = 9.23$, $df = 1$, $p < .01$), Q4a ($\chi^2 = 15.03$, $df = 1$, $p < .001$) and Q6a ($\chi^2 = 6.83$, $df = 1$, $p < .05$) were rated more often *yes* after the intervention than before the intervention. However, no significant differences between the data collected before and after the intervention were observed for the student belief items (Q1b–Q6b), i.e. which assessment practices they believe they learn from.

The student interviews and classroom observations were more positive toward the feedback provided to them on their texts in the final interviews (Phase 3) than in the mid-term interviews (Phase 2). Another finding was that low-performing students tended to prefer grades on all texts, as did also some of the average performing students. However, the high-performing ones were more inclined to adapt to FA practices where the focus was on providing useful feedback and downplay-

ing grades. The high-performing ones also preferred written feedback on written texts, whereas the low-performing ones tended to value oral feedback on written texts. However, all the students were positive toward revising their texts, which was also supported by the quantitative data. They said text revision can be time-consuming, but that it is useful because they learn from it. The high-performing students were more positive toward writing reflective logs than the low-performing ones. Self-assessment was a positive and demanding experience for most of the students. Peer assessment was a double-edged sword in that some perceived it as a good experience they learnt from, whereas others thought it was confronting having peers assess their text. High-performing students were more skeptical, saying they preferred that the teacher was the sole assessor of their texts. There were no visible changes in the feedback practices in the first term, but in the second term students experienced clearer, more selective, and more positive feedback on their texts.

The data from the teacher interviews indicated that portfolios made students work more with writing and writing assessment at school compared to before the intervention. Spending more time on writing at school opened up for more interaction between teachers and students about writing and writing assessment, as illustrated by the following notes from classroom observations:

T2 has prepared the lesson carefully. She has given her students a week to improve their texts according to the feedback they receive, before they upload their text on the Learning Management System and write a reflective log. I ask T2 whether she has changed anything for today with regards to the workshop we have had with the rest of the teacher team or with regards to the last lesson she had with her students. She says “I have tried to limit myself” [in giving students too much feedback].

T2 limits herself to a couple of things regarding grammar, content, and structure of the student texts. She writes “language, content, structure” on the blackboard. Under grammar, she writes “verbs, capital letter, spelling”. Under content, she writes “the wh-words, timeline”. Under structure, she writes “word order, paragraph”. Then T2 goes on to explain the writing process. She tells the students that the writing process means “revise, revise, revise”.

By adopting a portfolio approach, the teachers tried peer assessment of written texts with their students, which was new to all the participating teachers. One of the teachers claimed that peer assessment is as important as the teacher’s assessment of student texts. They all clearly saw the benefits of peer assessment by see-

ing how involved their students became in discussing writing, and how proficient they were in giving each other feedback. Text revision was something none of the teachers had tried before. However, all of them agreed that students benefited from revising their texts when they were given the chance to revise. The teachers agreed that their assessment practices became clearer by using writing portfolio as a tool. Apart from technical issues, the teachers mentioned the comprehensive nature of the subject as challenging. In middle school, English has been allocated two hours of instruction a week and requires that teachers assess their students in both written and oral communication.

DISCUSSION: CONTRIBUTIONS TO THE ENGLISH DIDACTICS FIELD

The present doctoral study reviewed the formative tool used in the interventions, writing portfolios. Furthermore, it shed light on teacher and student perceptions and practices of FA in English writing classes through a base-line study. Finally, processes of change when using portfolio as a tool to enhance FA in English writing classes were investigated through an intervention study. In the following, theoretical, empirical and methodological contributions to the English didactics field will be presented.

THEORETICAL CONTRIBUTIONS

The review study of portfolios in second and foreign language writing classes is a theoretical contribution to the research field. In addition to systematizing writing portfolios' potential formative benefits, the review calls for more research from primary and secondary schools, and more research where classroom observations are used to validate findings. The age of the students in this study also adds new knowledge to the research literature, since most studies on the benefits of portfolio assessment have been conducted in higher education. Portfolio assessment has clear formative potentials in English lessons and should be used more as a tool for writing.

EMPIRICAL CONTRIBUTIONS

In the base-line study, significant gaps were identified in how the informants perceive and act on FA of writing. The gaps need to be addressed in order to make FA effective and meaningful. Notably, teachers' FA practices were not necessarily clear for the students. The conclusion is that there needs to be more interaction and

transparency in the classroom about FA and the writing process, and that teachers' knowledge of research into FA has to be tried out and evaluated with students' experiences of those practices in mind. Compared to first language learning, English in Norway has less hours of instruction. In addition, teachers tend to have low expectations of their students, partly because they underestimate students' abilities. These two factors combined could explain why portfolios are not used more extensively in English compared to first language classrooms, despite the formative benefits. The intervention part of the doctoral study proved to be a challenge in significantly transforming students' beliefs and preferences in relation to FA. Nevertheless, the students in the interviews showed appreciation of the changes (see Lee 2011; Lee & Coniam, 2013), and the large sample of students showed significant changes in their self-reported FA practices.

The practices that significantly changed were related to student involvement, text revision, and the downplaying of grades. The first two were practices that the teachers were reluctant to, whereas the downplaying of grades was something the teachers practiced but not all the students noticed or appreciated. The practice of student text revision was appreciated by the teachers in this study after having tried it, and the students believed text revision is useful but time-consuming (cf. Lee & Coniam, 2013). There were, however, no significant changes in student self-reported practices concerning self-assessment. This study shows that English teachers should not underestimate students' willingness and effort to take part in formative cycles of writing assessment.

Finally, the study reveals that despite what educational authorities believe, in-service courses and seminars on FA are not sufficient for making assessment work. School-based processes of change need to take place in order to provide teachers with the opportunity to act on formative tools that may enhance perceptions and practices formatively.

METHODOLOGICAL CONTRIBUTIONS

Black and Wiliam (1998) assert that research that merely interrogates existing practice can probably do little more than confirming discouraging practices. Consequently, they claim, "To be productive therefore, research has to be linked with a program of intervention" (p. 12). The present study interrogated existing practice and linked the research with a program of intervention, studied through the lens of CHAT, which is a methodological contribution within the field of FA. In most school subjects, however, the appropriate tools to induce change are lacking. Black and Wiliam (2003) claim that tools to work with FA have to be subject-spe-

cific. Writing portfolios will be somewhat different in the school subject History compared to English, or even in a first language context such as Norwegian in Norway compared to English in Norway. Thus, a significant methodological contribution in my doctoral study is the use of a subject-specific tool. As indicated by the empirical findings in this study, FA research will not manifest itself in classroom practices without being mediated by a formative tool (Vygotsky, 1978). A subject-specific tool has to be used in order to mediate what teachers learn at in-service courses and seminars about FA.

One final methodological contribution concerns the methods used in this study. In a recent review of FA research, Black (2015) concludes that there are few studies where observation is used to validate FA practices. The use of mixed methods in the present doctoral study, including classroom observations, has been a methodological contribution in that the methods have complemented and sometimes contradicted each other. The responses regarding involvement in assessment practices from the larger quantitative sample of students complementing the responses from the smaller qualitative sample of students is one example; the responses about the quality of teacher feedback from the teacher interviews contradicting the quantitative student responses about the same topic is another example.

Finally, much research on changes in FA is restricted to shorter periods of time, and often either the student or the teacher perspective is examined. The participants in the present study needed at least a year to get into a transformation phase. Classroom observations validated the self-reported data in that the teachers changed their writing assessment practices, emphasizing the importance of talking more about the recursive nature of writing and by spending more time on clearer and more targeted feedback practices on student texts.

IMPLICATIONS FOR TEACHING ENGLISH

This study points out that there needs to be a change of how teachers and students work with and think about FA of writing in English. Teachers need to provide students with the opportunity to follow up written feedback at school. They can adopt a system where texts are not finished after receiving feedback, but incorporate students' response to the feedback while the teacher is present in the classroom enacting her supervisor role.

Secondly, a portfolio approach in English writing classes seems suitable in order to enhance FA, where students collect their texts in a digital learning management system, reflect on them, the feedback received and their follow-up of the feedback, and at the end of the school year select some of them for final assess-

ment. Peer assessment should be encouraged as part of a portfolio approach to writing, something that saved teachers in this study more time, contrary to what they believed initially.

Thirdly, motivated by a portfolio approach, text revision should be a natural part of English writing classes. English teachers need to consistently give their students the opportunity and encourage them to re-write and re-submit texts by putting the recursive nature of writing at the core of the assessment of written texts.

SUGGESTIONS FOR FUTURE RESEARCH

Since 2016, more research has been conducted within the area of FA in general and FA in English in particular (e.g., Saliu-Abdulahi, Hellekjær, & Hertzberg, 2017). Future research is needed to develop a fuller understanding of the construct of FA of writing, for example the elements that constitute student involvement in assessment practices and the ways students could follow up teacher feedback. Students could be interviewed about their definitions and interpretations of “involvement in assessment practices” and classroom observations could focus on teachers’ ways of involving them. After receiving feedback on texts from teacher and/or peers, their texts could be analyzed to consider the amount of and quality of revisions. The extent to which students use new knowledge and skills in other contexts could be investigated through interviewing and observing one or a few students throughout a school year. Moreover, the data were mainly self-reported in the present study through interviews and questionnaires, and supplemented by classroom observations. A next step could be careful analysis of student texts in order to trace development in their language performance, for example in the reflective statements. There is also a need for more longitudinal studies focusing on one of the most critical parts of interventions, namely the sustainability of changes in student and teacher assessment perceptions and practices a long time after the researcher (or other external collaborator) has left the research field.

REFERENCES

- Abedi, J. (2010). Research and recommendations for formative assessment with English language learners. In H. L., Andrade, & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 181–197). New York: Routledge.
- Assessment Reform Group [ARG]. (2002). *Researched-based principles of assessment for learning to guide classroom practice*. Retrieved from: <http://www.nuffieldfoundation.org/assessment-reform-group>

- Baird, J.-A., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *State of the field review: Assessment and learning*. Oslo/Oxford: Norwegian Knowledge Center for Education/Oxford University Center for Educational Assessment.
- Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice* 22(1), 161–177.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (2003). “In praise of educational research”: Formative assessment. *British Educational Research Journal*, 29(5), 623–637.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5–31.
- Black, P., & Wiliam, D. (2012). The reliability of assessments. In J. Gardner (Ed.), *Assessment and Learning* (pp. 243–263). London: SAGE Publications.
- Broadfoot, P., & Black, P. (2004). Redefining assessment? The first ten years of assessment in education. *Assessment in Education: Principles, Policy & Practice*, 11(1), 7–26.
- Brookhart, S. M. (2001). Successful students’ formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice*, 8(2), 153–169.
- Burner, T. (2014). The potential formative benefits of portfolio assessment in second and foreign language writing contexts: A review of the literature. *Studies in Educational Evaluation*, 43(4), 139–149.
- Burner, T. (2015). Processes of change when using portfolios to enhance formative assessment of writing. *Assessment Matters*. 9(2), 53–79.
- Burner, T. (2016a). *Formative assessment of writing in English: A school-based study of perceptions, practices and transformations* (Doctoral thesis). Trondheim: NTNU. URL link: <https://brage.bibsys.no/xmlui/handle/11250/2379861>
- Burner, T. (2016b). Formative assessment of writing in English as a foreign language. *Scandinavian Journal of Educational Research*, 60(6), 626–648.
- CEFR. (2001). *Common European Framework of Reference for Languages: Learning, teaching assessment*. Council of Europe, Strasbourg: Cambridge University Press.
- Cowie, B. (2005). Pupil commentary on assessment for learning. *The Curriculum Journal*, 16(2), 137–151.
- Creswell, J. W. (2013). *Qualitative inquiry & research design: Choosing among five approaches*, (3rd ed.). Los Angeles: SAGE Publications.
- Engeström, Y. (1987). *Learning by expanding*. Helsinki: Orienta-Konsultit Oy.
- Engeström, Y., & Sannino, A. (2010). Studies of expansive learning: Foundations, findings and future challenges. *Educational Research Review*, 5(1), 1–24.
- Ferris, D. R. (2003). *Response to student writing: Implications for second language students*. New York: Routledge.
- Hamp-Lyons, L., & Condon, W. (2000). *Assessing the portfolio. Principles for practice, theory and research*. Cresskill, NJ: Hampton Press, Inc.
- Hasselgreen, A., Drew, I., & Sørheim, B. (Eds.) (2012). *The Young Language Learner: Research-Based Insights into Teaching and Learning*. Bergen: Fagbokforlaget.
- Johnson, R. S., Mims-Cox, J. S., & Doyle-Nichols, A. (2010). *Developing portfolios in education. A guide to reflection, inquiry, and assessment*. Thousand Oaks: SAGE Publications, Inc.

- Kvale, S., & Brinkmann, S. (2009). *Det kvalitative forskningsintervju*, (2nd ed.). Oslo: Gyldendal Akademisk.
- Lee, I. (2011). Formative assessment in EFL writing: An exploratory case study. *Changing English: Studies in Culture and Education*, 18(1), 99–111.
- Lee, I., & Coniam, D. (2013). Introducing assessment for learning for EFL writing in an assessment of learning examination-driven system in Hong Kong. *Journal of Second Language Writing*, 22(1), 34–50.
- Norwegian Ministry of Education and Research. (2006/2013). *National curriculum for knowledge promotion in primary and secondary education and training*. Oslo: Author.
- Pryor, J., & Crossouard, B. (2008). A socio-cultural theorization of formative assessment. *Oxford Review of Education*, 34(1), 1–20.
- Reisjø, I. A. (2006). *A critical analysis of the final written exam in English in the lower secondary school*. MA thesis in English didactics. Faculty of Education and School Development: University of Oslo.
- Ringdal, K. (2007). *Enhet og mangfold. Samfunnsvitenskapelig forskning og kvantitativ metode*, (2nd ed.). Bergen: Fagbokforlaget.
- Ross, S. J. (2005). The impact of assessment method on foreign language proficiency growth. *Applied Linguistics*, 26(3), 317–342.
- Saliu-Abdulah, D., Hellekjær, G. O., & Hertzberg, F. (2017). Teachers' (Formative) Feedback Practices in EFL Writing Classes in Norway. *Journal of Response to Writing*, 3(1), 31–55.
- Sandvik, L. V. (2011). *Via mål til mening. En studie av skriving og vurderingskultur i grunnskolen tyskundervisning* (Doctoral thesis). Trondheim: Department of Teacher Education.
- Sandvik, L. V., Engvik, G., Fjørtoft, H., Langseth, I. D., Aaslid, B. E., Mordal, S., & Buland, T. (2012). *Vurdering i skolen. Intensjoner og forståelser. Delrapport 1 fra prosjektet Forskning på individuell vurdering i skolen (FIVIS)*. Trondheim: Department of Teacher Education, NTNU.
- Stobart, G. (2008). *Testing Times. The uses and abuses of assessment*. London: Routledge.
- Thorenfeldt, A. E. (2005). *"Unpredictable and full of risks?": An evaluation of the exam assessment in English in the R'94 vocational courses*. MA thesis in English didactics. Faculty of Education and School Development: University of Oslo.
- Torrance, H. (2012). Formative assessment at the crossroads: Conformative, deformativ and transformative assessment. *Oxford Review of Education*, 38(3), 323–342.
- Vygotsky, L. S. (1978). *Mind in society. The development of higher psychological processes*. Cambridge: Harvard University Press.
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge: The MIT Press.
- Wardekker, W. (2000). Criteria for the Quality of Inquiry. *Mind, Culture, and Activity*, 7(4), 259–272.
- Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, 16, 194–209.
- William, D. (2006). Formative assessment: Getting the focus right. *Educational Assessment*, 11(3–4), 283–289.
- Yildiz, L. M. (2011). *English VGI level oral examinations: How are they designed, conducted and assessed?* MA thesis in English didactics. Faculty of Education and School Development: University of Oslo.
- Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19, 79–102.