Developing Writing Skills in English Using Content-Specific Computer-Generated Feedback with EssayCritic

Irina Engeness

PhD candidate, Department of Education , University of Oslo, Norway
irina.engeness@iped.uio.no

Anders Mørch

Professor, Department of Education , University of Oslo, Norway
anders.morch@iped.uio.no

ABSTRACT

This paper presents a study of Norwegian Upper Secondary School students' writing process in English with: 1) feedback from an essay critiquing system (EssayCritic) (target class) and 2) feedback from collaborating peers (comparison class). The students in both classes significantly improved their grades. In the target class, the feedback from EssayCritic gave content-specific cues and the students included more ideas in their essays than the students in the comparison class who struggled when giving feedback to each other.

Keywords

Developing Writing Skills in English, Computer Generated Feedback, Cultural Historical Theory, Galperin

UNIVERSITETSFORLAGET

idunn.no
Nordiske tidsskrifter på nett

INTRODUCTION

Development of written communication is one of the competence aims of the Upper Secondary School English subject curriculum in Norway[1] and in common with other countries feedback and assessment for learning (AfL) have been encouraged in Norwegian schools[2] (Bueie, 2015; Gamlem & Smith, 2013; Røyeng, 2010). Feedback has been employed in formative assessment to promote the development of students' writing skills (Black & Wiliam, 2009; Hattie, 1999) and AfL has been defined as a classroom practice that involves dialogue and feedback loops between teachers and peers during subject specific problem solving (Gamlem & Munthe, 2014). New digital technologies open up opportunities by providing automated feedback (Dikli, 2006; Kukich, 2000; Sireci & Rizavi, 2000) and specifically for enhancing the development of students' writing skills in English (Lee, 2007; Mørch et al., 2005; Winerip, 2012).

Although several studies have explored the types and the role of feedback teachers and peers give on students' writing (Yang, et al., 2006; Zhao, 2010), to the best of our knowledge none has examined and compared the writing process in English with computer-generated feedback and peer feedback from collaborating peers. By taking a cultural-historical perspective (Edwards, 2005; Galperin, 1969; Vygotsky, 1980), this study aims to fill this gap. The study was undertaken in an Upper Secondary School in Norway, involving one hundred and twenty five students aged 16–17 who wrote essays on the topic *English as a Global Language*. We examined students' writing processes under two conditions of mediation: with the feedback from EssayCritic (target class) and with the feedback from collaborating peers using a set of available (non-computerized) support tools (comparison class).

Feedback in the writing process

The complexity of writing as a process of discovery, learning and communication through language has been emphasized by several researchers (Graves, 1975, 1994; Murray, 1999). Over forty years ago, Murray identified three phases in the classroom writing process: prewriting, writing and rewriting. Prewriting is everything that takes place before the first draft, writing is the act of producing the first draft and rewriting includes reconsideration of subject, form and audience (Murray, 1972). Along the same lines, Hayes and Flower observed that writers employ the following processes: planning, translating (or sentence generation) and revising (Hayes & Flower, 1986). Later the model was revised (Hayes, 1996) where planning was replaced by reflection, translating by text production and revising by text interpretation. In general, learn-

---

1. See: http://www.udir.no/kl06/eng1-03/Hele/Kompetansemaal/kompetansemal-etter-vg1--studieforberedende-utdanningsprogram-og-vg2---yrkesfaglige-utdanningspro-gram/?lplang=eng
2. See: http://www.udir.no/Vurdering-for-laring/4-prinsipper/Viktige-prinsipper-for-vude-ring/fire-prinsipper/

idunn.no
Nordiske tidsskrifter på nett

ing to write has been defined as a complex developmental process (Graves, 1983) where feedback as active intervention mediates learning within the activity of writing (Graves, 1982; Thompson, 2013).

Previous research (Black & Wiliam, 2009) has conceptualized feedback consisting of five key strategies: clarifying and sharing learning intentions and criteria for success; engineering effective classroom discussions and learning tasks that elicit evidence of student understanding; providing feedback that moves learners forward; activating students as instructional resources for one another and as owners of their own learning. According to Hattie and Timperley (2007), effective feedback must answer three major questions: Where am I going? How am I going? and Where to next? These questions correspond to the notions of feed up, feedback and feed forward (Hattie & Timperley, 2007). Four levels of feedback have been identified: feedback about the task, about the processing of the task, about self-regulation and about the self as a person. It has been argued that feedback on how to complete a task is the most effective, whereas feedback related to praise, reward and punishment is the least effective (Hattie & Timperley, 2007). However, research shows that in English writing classes assessment *of* learning often dominates over assessment *for* learning (Lee, 2007).

Echoing international research, Norwegian studies reveal that feedback on students' writing tends to be general and unspecific, consisting mainly of praise, and consequently lacking information on what to do next (Danielsen, et al., 2009; Furre, et al., 2006; Klette & Hertzberg, 2002).

Previous research in computer-based feedback systems includes automated essays scoring (AES). AES systems assign scores to essays written for educational purposes. The score is dynamically determined by machine learning and statistical techniques based on a set of training examples, which typically range from about twenty to several hundred, dependent on the learning algorithms and the desired precision of the feedback (Dikli, 2006). On the one hand, proponents argue for the success of AES systems in terms of how well they compare with the accuracy and reliability of human evaluation (Chung & O'Neil Jr, 1997; Sireci & Rizavi, 2000). On the other hand, critics have pointed out that AES systems can often be fooled by intentionally gibberish essays giving them higher scores (Kukich, 2000; Winerip, 2012). Our version of AES adopts a "qualitative" approach to presenting feedback by asking the students for specified content to be elaborated in their essays instead of the more common "quantitative" approach of presenting scores along a numerical scale.

Despite the considerable improvement of automated formative feedback systems (Shermis & Burstein, 2003), developers acknowledge that it is infeasible for computers to measure every aesthetic property of writing (Landauer, et al., 2003). In addition, classroom teachers express worries that the type of writing being measured with these systems is mechanistic, formulaic in nature and divorced from real-world contexts. Hence, more research is needed to examine

the context in which the system is used, the content of what is written, and the impact on key stakeholders as part of its integration (Ware, 2011).

Early studies of EssayCritic (Mørch et al., 2005) showed that it is a useful tool to facilitate writing in English, as it supplements teachers' feedback. A study in Hong Kong found no significant difference between the grades achieved by students who used EssayCritic and those who revised their texts themselves (Lee, et al., 2009). A later study of the same system (Lee, et al., 2013) compared two conditions: one group that received feedback from EssayCritic and from the teacher, and another group that received feedback from the teacher only. The essays of the group that received two types of feedback were richer in content than the essays of the other group.

Research in peer assessment indicates its summative and quantitative nature, with a strong reliance on scoring and grading (Falchikov & Goldfinch, 2000); and an emphasis on grammar mistakes and other writing errors (Hansman & Wilson, 1998). Despite peer assessment being increasingly used during group work, the link between peer assessment and collaborative learning has not been the subject of much research (Sluijsmans & Strijbos, 2010). To the best of our knowledge, no studies have compared peer feedback and computer-generated feedback in formative assessment in English composition.

Cultural-historical perspective on the development of writing skills

In exploring feedback in writing processes that include both collaborative (discussion) and individual (text processing) activities, we draw on Vygotskian cultural-historical theory. Vygotsky distinguished between social learning and individual development (Vygotsky, 1981), where the latter requires changes in the psychological functions of the learner (Chaiklin, 2003).

Development according to cultural-historical theory (Claxton, 2007; Edwards, 2015) requires employing the cultural tools central to practices. In the case of writing, learning that leads to such development involves feedback to be used by students in their development as writers. However, Vygotsky did not specify how the particular content of instruction (feedback) is related to development, and how specific qualities of the tools acquired by the child affect development. A cultural-historical scholar Galperin (Galperin, 1969; Haenen, 2001) has greatly extended Vygotsky's arguments about the leading role of instruction in the child's development by specifying the kind of instruction that can play such a role (Stetsenko & Arievitch, 2002). Galperin proposed six phases of socially meaningful activity that have implications for pedagogy (Edwards, 1995) and more specifically for the development of writing skills: (1) *motivation*, (2) *orientation*, (3) *materialized action*, (4) *communicated thinking*, (5) *dialogical thinking*, and (6) *acting mentally* (Haenen, 2001; Rambusch, 2006).

In the initial motivational phase, a learner's attitude and relation to the learning outcomes that have to be achieved is formed. In the second orientation phase,

Galperin identified three types of orientation: (a) incomplete, where media-tional means are identified by learners through trial and error; (b) complete, where learners are informed about all the mediational means necessary to solve a particular problem; and (c) complete, but being constructed by learners based on a general approach. In the third phase of a materialized action, learn-ers interact with material or materialized objects, and over time become less dependent on the material support they give and more aware of the meanings they carry. Speech becomes the main guiding tool in the fourth phase of com-municated thinking. The fifth phase, dialogical thinking, establishes a dialogue of a learner with him or herself, so that the action first carried out externally on material objects is being transformed mentally. In the final phase of acting mentally, an action has become a pure mental act with the focus on the out-come of the action (Haenen, 2001; Rambusch, 2006).

According to Galperin, the orientation phase is of particular importance as it introduces mediational tools to learners that will assist and guide further learn-ing process. Two different types of feedback as meditational tools are in the focus of our study: automated and peer. Galperin's pedagogical phases may inform teachers and researchers about the complexity of the processes involved in a learner's move from, for example, orientation to the potential of advice on how to write, to the ability to act as a writer. Therefore, the quality and appropriateness of the feedback guiding students' writing process would seem to be crucial. From this perspective we address the following research question:

How do different types of feedback assist students in their writing process?

METHOD

Participants and setting

Data was collected during the autumn term 2014 in an upper secondary school in Norway that had AfL as a focal area. One hundred and twenty-five students (aged 16–17) from five classes, five teachers and four researchers participated in the project. The teaching plans that preceded the essay writing activity had been discussed and standardized with the teachers who were involved in the project. Previous to the writing activity, the students had covered the same study topics from the textbook *Passage* (Burgess & Sørhus 2009) as sources of information for their essays. The text of the assignment of the essay had been created in collaboration between five teachers and two researchers (the authors): *"Write an essay on the topic of English as a Global Language: Explore how English was spread around the globe, and present the most important reasons for this development"* (300–400 words).

In the first round, the students in five classes wrote the first version of the essay using only a laptop as a writing tool. The learners did not have access to any source materials. Eleven best essays from three classes were selected to code

for the subthemes of the topic "English as a Global Language" and to train EssayCritic (see section EssayCritic System).

The remaining two classes (48 students) were assigned as the target and comparison class. The students of the target class uploaded the first draft to Essay-Critic and after they had discussed the feedback received from the system with their peers, they produced the second draft. The students of the comparison class read each other's essays and gave advice on how to create the second draft. All the students repeated the process of receiving feedback and revising their essays one more time, which resulted in the production of the third draft handed in to the teachers for final evaluation.

During the interventions, the teachers circulated among the groups and responded to questions. The researchers did not interfere in the group work.

EssayCritic System

EssayCritic (EC3) is a web application that analyzes uploaded essays with the help of a Decision Tree machine-learning algorithm (Quinlan, 1986), synonyms from dictionaries and a lexical database for English WordNet (Fellbaum, 1998) and generates feedback based on the content of the essays. Before using the system, a concept tree representing the essay topic was created. The teachers identified eleven subthemes for the topic by using a chapter in *Passage* (Burgess & Sørhus 2009), and the researchers decomposed each subtheme into simpler concepts, represented by a few phrases or word expressions taken from the students' essays that were selected to train the system, together with the synonyms from the dictionaries and the WordNet lexical database. In doing this, a "model" was created and the process of training the system took approximately one month.

Once a new essay was written and uploaded by a student, the system computed a score for the uploaded essay's similarity with the model for each of the subthemes. If this score was below a threshold value, "critique" was given for missing the corresponding subtheme, and if above the threshold, the subtheme was considered to be a part of the essay, and was highlighted in the essay as a phrase or sentence. Figure 1 shows the screen pictures of EC3 in the praise (leftmost picture) and critique (rightmost picture) modes.
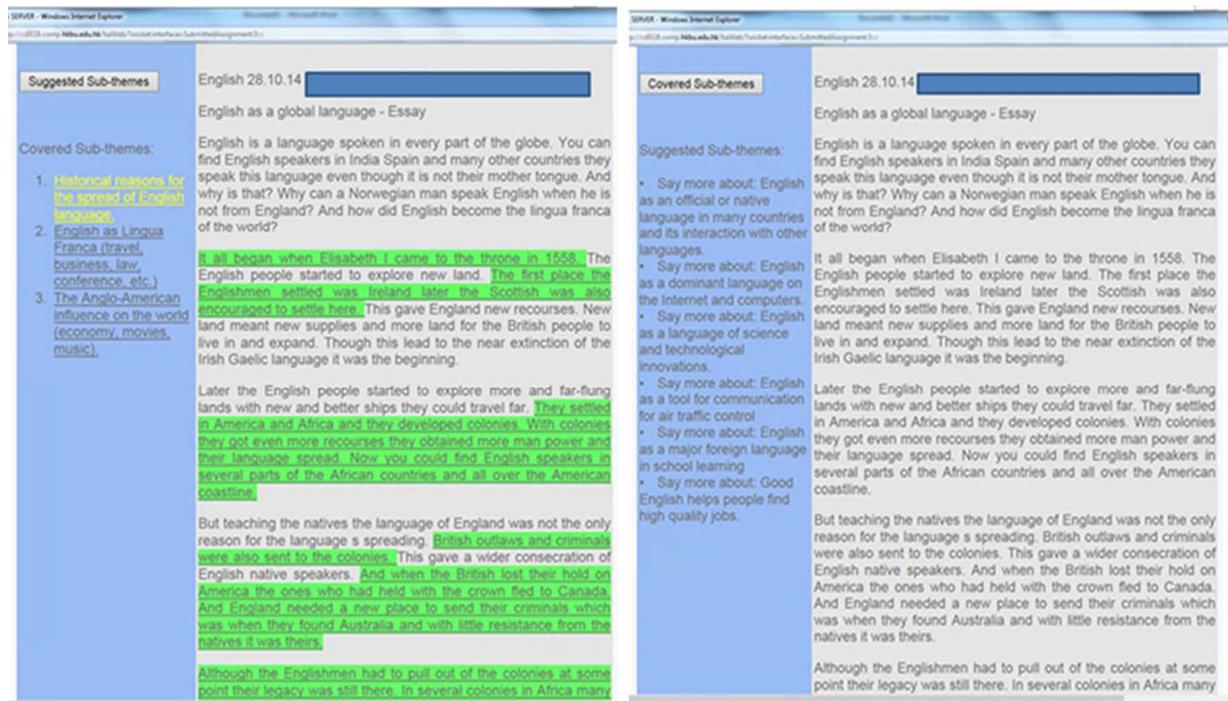
Figure 1. The user interface of EC3: Covered subthemes (left) and suggested sub-themes (right)

Data analysis

Three groups of four students in both classes were video recorded. Another camera followed the teachers. Nine hours of video recordings were transcribed according to the Jeffersonian transcription notations (Atkinson & Heritage, 1999) (Appendix A). Field notes taken during class observations were used to contextualize the data collected (Derry et al., 2010). In total 96 essays, including 24 pre-tests (the first draft) and 24 post-tests (the third drafts) from both classes, were marked on the scale 1–6 by the classroom teachers and an independent English teacher.

To analyze the data we applied *mixed methods* (Creswell, 2012). The paired t-test was used to analyze the improvement between pre- and post-tests and the number of subthemes included in the pre- and post-tests. Cohen's *d* (Cohen, 1992; Field, 2013) was calculated to evaluate the effect size of this improvement. We used the independent t-test (Cohen, et al., 2011; Field, 2013) to calculate: a) statistical difference between the results of the pre- and post-tests and b) statistical difference between the number of subthemes included in the pre- and post-tests.

Two interaction extracts from the target and comparison classes analyzed qualitatively are presented. The groups were selected because the students were verbally active and the chosen extracts represent typical discussions that happened in the observed groups. The analytical procedure was inspired by inter-

action analysis (Jordan & Henderson, 1995). The primary unit of analysis were sequences and turns rather than isolated utterances (Linell, 2009). Once the interaction analysis was completed, the interactions were examined through the analytic lens offered by Galperin's types of orientation and pedagogical phases.

RESULTS

Analysis of pre-test and post-test data

In order to evaluate the writing process, we compare the results of the first (pre-test) and the last (post-test) drafts of students' essays. Forty-eight pre- and post-tests from both classes marked on the scale 1–6 constitute the quantitative data. Our analysis is based on the grades awarded by the independent teacher. The results of the pre- and post-tests are presented in Table 1.

TABLE 1. AVERAGE PRE- AND POST- TEST GRADES IN THE TARGET AND COMPARISON CLASSES AND THE OBSERVED GROUPS.

	Average grade pre-test (essay v.1)	Average grade post-test (essay v.3)	Difference	Paired t-test	Cohen's d
Target class	2.79 M = 2.79 SD = 0.88	4.79 M = 4.49 SD = 1.41	2	t (24) = −7.62 p < .000	d = 2.26
Observed target group	2.75	4.5	1.75		
Comparison class	2.88 M = 2.86 SD = 0.74	4.83 M = 4.83 SD = 0.87	1.95	t (24) = −8.86 p < .000	d = 2.64
Observed comparison group	3.5	4.5	1		
Independent t-test, p	p = .725	p = .903			

The paired t-test (Cohen et al., 2011) shows a significant difference between the average grades of the pre- and post-tests in both classes. Cohen's *d* (Cohen, 1992; Field, 2013) indicates large effect[3] in both classes. The independent t-test (Field, 2013) shows no significant statistical difference[4] between the grades of the pre-tests in the target and comparison classes and no significant statistical difference between the grades of the post-tests in the correspondent classes.

These results triggered our interest to do further analysis and in Table 2 we present the average number of subthemes in the essays in both classes.

---

3. Effect size Cohen's d>0.2 – little effect, Cohen's d>0.5 – middle effect, Cohen's d>0.8 large effect (Cohen, 1992; Field, 2013)
4. p> .05 indicates no significant statistical difference between the means of two samples (Field, 2013)

The paired t-test indicates a significant difference between the number of sub-themes included in the pre- and post-tests in both classes. Cohen's d reflects large effect achieved in both classes. The independent t-test shows significant statistical difference between the number of subthemes in the pre- and post-tests: p= .000. The students in the target class included significantly more sub-themes (almost twice as many) in their post-tests than the students in the comparison class.

TABLE 2. AVERAGE NUMBER OF SUBTHEMES IN THE PRE- AND POST-TESTS IN THE TARGET AND COMPARISON CLASSES AND THE OBSERVED GROUPS

	Average number of subthemes in the pre-tests (essay v.1)	Average number of subthemes in the post-tests (essay v.3)	Difference between the average number of subthemes in the pre- and post-tests	Paired t-test	Cohen's d
Target class	3.25 $M = 3.25$ $SD = 0.79$	7 $M = 7$ $SD = 1.89$	3.75	$t(24) = -9.70$ $p < .000$	$d = 4.72$
Observed target group	4	8.25	4.25		
Comparison class	4.04 $M = 4.04$ $SD = 0.91$	5.92 $M = 5.92$ $SD = 1.14$	1.88	$t(24) = -6.91$ $p < .000$	$d = 2.07$
Observed comparison group	4.5	6	1.5		
Independent t-test, p			$p = .000$		

The students in both classes achieved similar grades on their final drafts. However, the students in the target class included significantly more subthemes than the students in the comparison class. This encouraged us to take a closer look at the learning process and analyze student-student interactions qualitatively.

Analysis of the interactions of the students of the target class

In extract 1, four students are sitting around the table with open laptops. They have uploaded version 1 essays to EC3 and received feedback on the missing subthemes: in Galperin's terms, the learners are in the phase of communicated thinking. The group has just finished discussing the feedback Silje received. It is Thea's turn to share her feedback with the group. The names of the students have been changed for anonymity.

EXTRACT 1

1.	Thea:	We can mention where English is used today. And then I have an air traffic control thing and English as a major foreign language in school.
2.	Silje:	Yes. I have covered that. ((*reading from the screen*)) To be able to read the Internet pages and study, you require at least basic knowledge of English.
3.	Thea:	And to get a job. I think that is why you have to learn English at school or get to know it somehow.
4.	Ida:	I think they speak English in any store. If you go to Spain, for example, and you don't know Spanish and you go to a store and they don't speak English, you can't communicate then.
5.	Thea:	Yes, you have to know English to communicate with others. I was in Germany last summer with the school band and we were at the amusement park and there was a lady behind the counter in the souvenir shop, she hardly knew any English and that was so annoying!
6.	Ida:	The official language, like in France it's like: where can I find the passport desk? Er, er, er ((*making pointing gestures*)) They don't speak a word of English.
7.	Thea:	OK and then I have to say about how English helps people to find high quality jobs. Well, I think high quality jobs are international, in business and in companies. Many companies deal outside the country. My dad, he works with furniture, he has to travel to Asia and Europe. He knows what furniture we are going to sell in Norway and he has to know English very well because he talks to Chinese and Japanese people. And if you are going to be a doctor, many things should be learnt in English.
8.	Ida:	Well, I went to chiropractor, and only the secretary was Norwegian, all the doctors there spoke English; they were Australians and British and Americans.
9.	Thea:	So, if you are going to have a job, you have to have some knowledge.
10.	Ida:	Yes, basically you don't get a good job if you don't speak English.
11.	Thea:	Yes, you have to know some basic English.
12.	Ida:	Yes, at least.
13.	Thea:	Yes, at least.

The extract emphasizes the role of the generated feedback: two subthemes prompted by EC3 trigger a conversation and one of the feedback messages relates to the sentence in Silje's essay (line 2) extended by Thea (line 3). The students engage in the discussion by giving examples from their personal experience, which might serve as a way of making brief feedback more meaningful (lines 4–8). Summarizing their discussion, the girls draw a conclusion that English is very important for working life (lines 9–13).

The discussion may enable the learners to incorporate new ideas in their essays and hence, promote the development of their writing skills. In fact, Thea included the following in her final draft:

"Today we find English all over the globe. In today's society you have to have some basic language to get around. You almost can't visit a website without some knowledge of English, and you can't watch TV without knowing English. We use English everywhere, we are surrounded with it. We use English in science and technological innovations and to get a good

job. We even use English to communicate in air traffic control. English is used all the time and everywhere because the majority of the world's population knows some English."

This shows that the ideas EC3 had prompted appeared twice: 1) when talked about in groups and 2) incorporated as a new text in the essays.

Analysis of the interactions of the students of the comparison class

In extract 2, four students are sitting around the table with open laptops. They have received their first drafts from the teacher who graded them, without giving any feedback: the learners are in Galperin's phase of communicated thinking. Jonas and Tobias have read each other's essays and Jonas is in the middle of giving feedback to Tobias.

EXTRACT 2

1.	Jonas:	If you want, you can write about the upper class situation with learning English.
2.	Tobias:	It isn't coming naturally here, that's the point. And that's about history and I don't want more history.
3.	Jonas:	No, you need other things.
4.	Tobias:	Because ((*takes the assessment rubric*)) I have already five or six sentences about history, I don't think I need more.
5.	Jonas:	It's so difficult not to talk about history because the whole book is about history. Have you talked about music?
6.	Tobias:	Music? I say media, perhaps I should mention music. I say media and entertainment.
7.	Jonas:	You can specify, but that's not needed really.
8.	Tobias:	Then there will be producing of media, music and films and TV series ((*laughing*)) I feel like I want to write entertainment. It feels like it's easier.
9.	Jonas:	Yes, that's more descriptive.
10.	Tobias:	OK. But I feel if I want a better grade, there is something more I have to change.
11.	Jonas:	You decide.
12.	Tobias:	And I don't know what it is. That's kind of annoying.

Having looked through the assessment rubric, Tobias realizes that he has covered the historical reasons for the spread of English, but he is lacking other ideas (line 4). Jonas mentions music (line 5) and Tobias interprets this as "media, music, films and TV series" and generalizes it to "entertainment" (line 8). By eliciting each other's thoughts, the two peers attempt to develop their understanding of how to improve their essays. However, the students are somewhat unsure if they are moving in the right direction (line 12). Tobias includes the following in his final draft:

" …the US has taken the role of a mass producer of media and entertainment in the world. After the Second World War, Europe was in ruins. The

European industry and prosperity was dramatically slowed down, while in the US, the economy grew. Making the US the first and the only international superpower. Now, new Hollywood movies are displayed on the big screen all over the world."

Jonas writes:

" With all new smartphones and streaming possibilities of films and music, we are hearing English more than we used to. With streaming programs like Netflix and Spotify, you can watch Titanic or listen to The Beatles on the bus. The massive information we receive from international news pages also influences us".

These paragraphs reflect the extent to which the two students are able to incorporate the spoken dialogs in their writing. Tobias chooses to write about the Unites States' role in the production of the world's media and entertainment from a historical perspective, whereas Jonas refers to the opportunities offered by streaming. The students' uncertainty about the type of feedback that should be given results in different interpretations of the ideas discussed when transferring them from speech to writing.

The analysis of extract 2 reflects that giving feedback can be difficult for learners. Nevertheless, by using the assessment rubric and exchanging personal ideas, the students are able to provide feedback suggesting potential improvements for their essays.

DISCUSSION

By adopting a cultural-historical perspective, this study explored students' writing processes in English through the exposure to two different types of orientation (see section Cultural-historical perspective on the development of writing skills). In this section we discuss the main findings in relation to Galperin's theory and previous research.

Target class

Galperin's types of orientation (Haenen, 2001; Rambusch, 2006) suggest that the students in the target class were exposed to so-called complete orientation. In other words, the mediational means needed for the learners to receive individual feedback (EC3), elaborate on it (group discussions) and improve their drafts were available and contributed to the significant improvement the students achieved from the pre- to post-tests. The analysis of students' interactions revealed that the discussions in the target group were triggered by the feedback generated by EC3, then elaborated and made sense of by referring to the learners' personal experiences. In this way, collaboration might have contributed to the development of students' understanding of the automated feed-

back, which might have resulted in a greater variety of subthemes included in the final drafts. This potentially indicates the need for incorporating group discussions when students' writing is supported with computer-generated feedback.

These findings reflect the role of EC3 in constantly assisting the students during the writing process by providing individual feedback to all learners simultaneously on multiple occasions and drafts. In doing this, EC3 supplements the teacher's facilitating of students' writing process (Lee et al., 2013). In addition, EC3 assisted the students in their analysis of essays by identifying the missing and covered subthemes, which was challenging for the learners in the comparison class. In this respect, the feedback from EC3 coincides with the five key strategies of formative assessment (Black & Wiliam, 2009): it clarifies and shares learning intentions and criteria for success (feedback on covered and suggested subthemes); engineers effective classroom discussions that elicit evidence of student understanding (initiated group discussions); provides feedback that moves learners forward (suggested subthemes); activates students as instructional resources for one another (initiated group discussions), and as owners of their own learning (individual feedback on multiple drafts and occasions).

On the one hand, EssayCritic is unique in two respects: 1) it utilizes semantic matching using Decision Tree learning algorithms to compute feedback on text-based artifacts and 2) it provides formative feedback which is dynamically computed based on the best essays produced by the community of learners and used to compare individual essays against it, suggesting further improvements in a cycle of write/rewrite activities.

On the other hand, the feedback given by EC3 was framed by the contextual constraints of the textbook Passage (Burgess & Sørhus 2009) and the pedagogical choices of the teachers made when identifying eleven subthemes to train the system (Landauer, et al., 2003; Ware, 2011), which might have narrowed the choice of the subthemes included in the essays. Consequently, the study raises the question of the need for teachers to critically analyze the cost-benefits of using computer-generated feedback as formative assessment in the writing process. In addition, reflecting on the four levels of feedback (Hattie & Timperley, 2007), EssayCritic provides only feedback about the task, but not about the processing of the task. This may be addressed as potential improvements of EssayCritic in further research.

Comparison class

The learners in the comparison class also achieved significant improvement from pre- to post-tests. However, the students' discussions showed that giving feedback was challenging. The learners were unsure about what kind of feedback to give and what tools could assist them. The peers had to identify useful resources when giving feedback, which indicates that they were exposed to the

orientation of the third type (Haenen, 2001; Rambusch, 2006): complete but created by learners based on a general approach. The assessment rubric with the list of the subthemes, the students' essays, the peers and the teacher were the resources available and the students' previous experience in the AfL practice served as a general approach they were to pursue. The learners struggled in their attempt to give feedback: the analysis revealed that the students relied on their general understanding of what had to be improved (Hansman & Wilson, 1998). The assessment rubric, however, was occasionally used as a guiding tool. By reading each other's drafts, the students might have acquired another insight, as one source for ideas to write about. However, students' uncertainty about the type of the feedback that would improve their drafts resulted in different interpretations of the mutually discussed ideas. The quantitative analysis showed that the students in the comparison class included fewer subthemes in their final drafts.

Directions for further research

Our findings suggest that further research is needed to explore: 1) whether EC3 can be modified so that it would provide feedback both on the task and on the processing of the task (Hattie & Timperley, 2007); 2) whether the students exposed to the complete orientation are able to transfer their feedback-giving skills gained working with EC3 to other learning situations without access to EC3, 3) the role of the teacher facilitating students' writing with EC3; and 4) our tentative hypothesis that group discussions based on the feedback given by EC3 are more likely to be incorporated in the text compared to other conditions.

APPENDIX A

Transcript conventions

[ ]	Text in square brackets represents clarifying information
=	Indicates the break and subsequent continuation of a single utterance
?	Rising intonation
:	Indicates prolongation of a sound
(.)	Short pause in the speech
[…]	Utterances removed from the original dialog
-	Single dash in the middle of a word denotes that the speaker interrupts herself
--	Double dash at the end of an utterance indicates that the speaker's utterance is incomplete
((Italics))	Annotation of non-verbal activity
Courier	`Courier   Verbatim reading from screen (typed text)`

REFERENCES

Atkinson, J. M., & Heritage, J. (1999). Transcript Notation – Structures of Social Action: Studies in Conversation Analysis. Aphasiology, 13(4–5), 243–249. doi: http://dx.doi.org/10.1080/026870399402073.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability (formerly: Journal of Personnel Evaluation in Education), 21(1), 5–31. doi: http://dx.doi.org/10.1007/s11092-008-9068-5.

Bueie, A. (2015). Summativ vurdering i formativ drakt–elevperspektiv på tilbakemelding fra heldagsprøver i norsk. Acta Didactica Norge, 9(1), Art. 4, 21 sider. Retrieved from: https://www.journals.uio.no/index.php/adno/article/view/1300

Burgess, R., & Sørhus , T. B. (2009). Passage engelsk vg1 studieforberedende program Cappelen Damm AS.

Chaiklin, S. (2003). The zone of proximal development in Vygotsky's analysis of learning and instruction. In A. Kozulin, B. Gindis, V. S. Ageyev & S. M. Miller Vygotsky's educational theory in cultural context, 1, 39–64. doi: http://dx.doi.org/10.1017/cbo9780511840975.004.

Chung, G. K., & O'Neil Jr, H. F. (1997). Methodological Approaches to Online Scoring of Essays. Retrieved from: http://eric.ed.gov/?id=ED418101

Claxton, G. (2007). Expanding young people's capacity to learn. British Journal of Educational Studies, 55(2), 115–134. doi: http://dx.doi.org/10.1111/j.1467-8527.2007.00369.x.

Cohen, J. (1992). A power primer. Psychological bulletin, 112(1), 155. doi: http://dx.doi.org/10.1037/0033-2909.112.1.155.

Cohen, L., Manion, L., & Morrison, K. (2011). Research methods in education. Milton Park. Abingdon, Oxon,[England]: Routledge.

Creswell, J. W. (2012). Qualitative inquiry and research design: Choosing among five approaches: Sage.

Danielsen, I., Skaalvik, E., Garmannslund, P., & Viblemo, T. (2009). Elevene svarer. Analyse av elevundersøkelsen 2009: Oslo: Utdanningsdirektoratet og Oxford Research.

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., . . . Sherin, M. G. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. The Journal of the Learning Sciences, 19(1), 3–53. doi:http://dx.doi.org/10.1080/10508400903452884.

Dikli, S. (2006). An overview of automated scoring of essays. The Journal of Technology, Learning and Assessment, 5(1). Retrieved from: https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1640

Edwards, A. (1995). Teacher education: Partnerships in pedagogy? Teaching and Teacher education, 11(6), 595–610. doi: http://dx.doi.org/10.1016/0742-051x(95)00015-c.

Edwards, A. (2005). Let's get beyond community and practice: the many meanings of learning by participating. Curriculum Journal, 16(1), 49–65. doi: http://dx.doi.org/10.1080/0958517042000336809.

Edwards, A. (2015). Designing tasks which engage learners with knowledge. In I. Thompson, Designing Tasks in Secondary Education: Enhancing Subject Understanding and Student Engagement, 13–27. Routledge.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. Review of Educational Research, 70(3), 287–322. doi: http://dx.doi.org/10.3102/00346543070003287

Fellbaum, C. (1998). WordNet: Wiley Online Library. Retrieved from: http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal1285/abstract?userIsAuthenticated=false&deniedAccessCustomisedMessage=

Field, A. (2013). Discovering statistics using IBM SPSS statistics: Sage.

Furre, H., Danielsen, I., Stiberg-Jamt, R., & Skaalvik, E. (2006). Analyse av den nasjonale undersøkelsen" Elevundersøkelsen" 2006: Kristiansand: Oxford Research.

Galperin, P. (1969). Stages in the development of mental acts. A handbook of contemporary soviet psychology, 249–273. New York: Basic Books.

Gamlem, S. M., & Munthe, E. (2014). Mapping the quality of feedback to support students' learning in lower secondary classrooms. Cambridge Journal of Education, 44(1), 75–92. doi: http://dx.doi.org/10.1080/0305764x.2013.855171.

Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. Assessment in Education: Principles, Policy & Practice, 20(2), 150–169. doi: http://dx.doi.org/10.1080/0969594x.2012.749212.

Graves, D. H. (1975). An examination of the writing processes of seven year old children. Research in the Teaching of English, 227–241.

Retrieved from: http://www.jstor.org/stable/40170631?seq=1#page_scan_tab_contents

Graves, D. H. (1982). A Case Study Observing the Development of Primary Children's Composing, Spelling, and Motor Behaviors during the Writing Process. Final Report, September 1, 1978-August 31, 1981. Retrieved from: http://files.eric.ed.gov/fulltext/ED218653.pdf

Graves, D. H. (1983). Writing: Teachers and children at work: ERIC. Retrieved from: http://eric.ed.gov/?id=ED234430

Graves, D. H. (1994). A fresh look at writing: ERIC. doi: http://dx.doi.org/10.5860/choice.32-5779.

Haenen, J. (2001). Outlining the teaching–learning process: Piotr Gal'perin's contribution. Learning and Instruction, 11(2), 157–170. doi: http://dx.doi.org/10.1016/s0959-4752(00)00020-7.

Hansman, C. A., & Wilson, A. L. (1998). Teaching writing in community colleges: A situated view of how adults learn to write in computer based writing classrooms. Community College Review, 26(1), 21–42. doi: http://dx.doi.org/10.1177/009155219802600102.

Hattie, J. (1999). Influences on student learning. Inaugural lecture given on August, 2, 1999. Retrieved from: http://xn--www-rp0a.teacherstoolbox.co.uk/downloads/managers/Influencesonstudent.pdf

Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81–112. doi: http://dx.doi.org/10.3102/003465430298487.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. Levy & S. Ransdell The science of writing. Theories, methods, individual differences and applications,1–27. New Jersey: Lawrence Erlbaum Associates. doi: http://dx.doi.org/10.4324/9780203811122.

Hayes, J. R., & Flower, L. S. (1986). Writing research and the writer. American Psychologist, 41(10), 1106. doi: http://dx.doi.org/10.1037/0003-066x.41.10.1106

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. Journal of the Learning Sciences, 4(1), 39–103. doi: http://dx.doi.org/10.1207/s15327809jls0401_2.

Klette, K., & Hertzberg, F. (2002). Klasserommets praksisformer etter Reform 97: Det utdanningsvitenskapelige fakultet, Universitet i Oslo.

Kukich, K. (2000). Beyond automated essay scoring. IEEE intelligent systems and their Applications, 15(5), 22–27. Retrieved from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=889104

Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In Automated essay scoring: A cross-disciplinary perspective, 87–112.

Lee, C., Cheung, W. K. W., Wong, K. C. K., & Lee, F. S. L. (2013). Immediate web-based essay critiquing system feedback and teacher follow-up feedback on young second language learners' writings: an experimental study in a Hong Kong secondary school. Computer Assisted Language Learning, 26(1), 39–60. doi: http://dx.doi.org/10.1080/09588221.2011.630672.

Lee, C., Wong, K. C., Cheung, W. K., & Lee, F. S. (2009). Web-based essay critiquing system and EFL students' writing: A quantitative and qualitative investigation. Computer Assisted Language Learning, 22(1), 57–72. doi: http://dx.doi.org/10.1080/09588220802613807.

Lee, I. (2007). Assessment for learning: Integrating assessment, teaching, and learning in the ESL/EFL writing classroom. Canadian Modern Language Review/La Revue canadienne des langues vivantes, 64(1), 199–213. doi: http://dx.doi.org/10.3138/cmlr.64.1.199.

Linell, P. (2009). Rethinking language, mind, and world dialogically: interactional and contextual theories of human sense-making. Information Age Publishing.

Murray, D. (1972). Teach writing as a process not product. The Leaflet, 71(3), 11–14. Retrieved from: http://www.larue.k12.ky.us/userfiles/1085/Teach%20Reading%20as%20Process%20Not%20Product%20Article.pdf

Murray, D. (1999). Write to learn. Harcourt Brace College Publishers.

Mørch, A., Cheung, W., Wong, K., Liu, J., Lee, C., Lam, M., & Tang, J. (2005). Grounding Collaborative Knowledge Building in Semantics-Based Critiquing. In R. H. Lau, Q. Li, R. Cheung, & W. Liu (Eds.), Advances in Web-Based Learning – ICWL 2005 (Vol. 3583, pp. 244–255): Springer Berlin Heidelberg. Retrieved from: https://telearn.archives-ouvertes.fr/hal-00190025

Quinlan, J. R. (1986). Induction of decision trees. Machine learning, 1(1), 81–106. doi: http://dx.doi.org/10.1007/bf00116251

Rambusch, J. (2006). Situated learning and Galperin's notion of object-oriented activity. Paper presented at the The 28th Annual Conference of the Cognitive Science Society. Retrieved from: http://csjarchive.cogsci.rpi.edu/Proceedings/2006/docs/p1998.pdf

Røyeng, M. G. S. (2010). Elevers bruk av lærerkommentarer: Elevperspektiv på underveisvurdering i skriveopplæringen på ungdomstrinnet. Retrieved from: https://www.duo.uio.no/handle/10852/32366

Shermis, M. D., & Burstein, J. C. (2003). Automated essay scoring: A cross-disciplinary perspective. Routledge. Retrieved from: https://books.google.no/books?hl=no&lr=&id=9qyPAgAAQBAJ&oi=fnd&pg=PP1&dq=Shermis,+M.+D.,+%26+Burstein,+J.+C.+(2003).+Automated+essay+scoring:+A+cross-disciplinary+perspective:&ots=5MQxPKEGc2&sig=ebWoqfbi5XmBII_xjOaMqY1eDHM&redir_esc=y#v=onepage&q&f=false

Sireci, S. G., & Rizavi, S. (2000). Comparing Computerized and Human Scoring of Students' Essays. Retrieved from: http://files.eric.ed.gov/fulltext/ED463324.pdf

Sluijsmans, D. M., & Strijbos, J.-W. (2010). Flexible peer assessment formats to acknowledge individual contributions during (web-based) collaborative learning. In E-collaborative knowledge construction: Learning from computer-supported and virtual environments, 139–161. doi: http://dx.doi.org/10.4018/978-1-61520-729-9.ch008.

Stetsenko, A., & Arievitch, I. (2002). Teaching, learning, and development: A post-Vygotskian perspective. In G. Wells & G. Claxton Learning for life in the 21st century: Sociocultural perspectives on the future of education, 84–96. John Wiley & Sons. doi: http://dx.doi.org/10.1002/9780470753545.ch7.

Thompson, I. (2013). The mediation of learning in the zone of proximal development through a co-constructed writing activity. Research in the Teaching of English, 47(3), 247–276. Retrieved from: http://www.ncte.org/library/nctefiles/resources/journals/rte/0473-feb2013/rte0473mediation.pdf

idunn.no
Nordiske tidsskrifter på nett

Vygotsky, L. S. (1980). Mind in society: The development of higher psychological processes: Harvard university press. doi: http://dx.doi.org/10.2307/1421493.

Vygotsky, L. (1981). The genesis of higher mental functions. The concept of activity in Soviet psychology, 144–188. New York: Sharpe.

Ware, P. (2011). Computer?Generated Feedback on Student Writing. Tesol Quarterly, 45(4), 769–774. doi: http://dx.doi.org/10.5054/tq.2011.272525.

Winerip, M. (2012). Facing a robo-grader? just keep obfuscating mellifluously. New York Times, 22. Retrieved from: http://thelawsofrobotics2013.iankerr.ca/files/2013/09/49-Facing-a-RoboGrader.pdf

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in a Chinese EFL writing class. Journal of second language writing, 15(3), 179–200. doi: http://dx.doi.org/10.1016/j.jslw.2006.09.004.

Zhao, H. (2010). Investigating learners' use and understanding of peer and teacher feedback on writing: A comparative study in a Chinese English writing classroom. Assessing Writing, 15(1), 3–17. doi: http://dx.doi.org/10.1016/j.asw.2010.01.002.