

An Analysis of the Relationship among Teacher Feedback, Feedforward, and Grade on Swedish University Students' Compositions in English as a Second Language

Monica Karlsson

School of Education, Humanities and Social Sciences
Halmstad University, Sweden

Abstract:

In the present study, with the aim of analyzing the relationship among teacher feedback, feedforward, and grade, the corrections and comments made by four experienced assessors on 187 compositions were under scrutiny. These essays were written by 56 Swedish university students studying English as a second language at three different educational levels. The results reveal that while there were clear links between mid-essay corrections/comments and grades given, the links between mid-essay corrections/comments and end comments were not only comparatively few, but less clear. Moreover, although valued highly in the research literature because of their ability to promote writing skills in an enhanced manner, there were more summative end comments than formative ones. The conclusion was, therefore, drawn that it is quite taxing for assessors, even for experienced ones, to produce connections that involve an alignment among a) mid-essay corrections/comments, b) end comments and c) grade that will, at the same time, promote students' writing skills in accordance with what is suggested by the research literature. The assessors were, however, irrespective of grade given, attuned to the educational level at hand, focusing more on analytic aspects at the two lower levels, while taking a more holistic approach at the highest educational level. This may indicate that offering corrections/comments does not only entail a developmental journey for students, but for teachers too.

Key words: composition writing, feedback, feedforward, formative assessment, L2 English, Swedish University Students, summative assessment, teacher comments, teacher corrections, university level

Cite as: Karlsson, M. (2019). An Analysis of the Relationship among Teacher Feedback, Feedforward, and Grade on Swedish University Students' Compositions in English as a Second Language. *Arab World English Journal*, 10 (3).320.

DOI: <https://dx.doi.org/10.24093/awej/vol10no3.1>

1. Introduction

Many teachers generally seem to think that compositions, in contrast to other types of student production, are comparatively tricky to assess, mark, and grade. Before any teacher sets to work on such a task, several questions, therefore, usually arise. For instance, along what continuums should the various aspects of a learner text be judged? Should only one, several or all elements of the essay be considered, i.e., should a text be approached in a fashion of less-is-more or all-or-nothing? Is a grade given fair in relation to what other students have achieved and received? As if this is not enough, since assessing, marking and grading are not static phenomena, teachers have to adjust their way of working to, for example, what type of essay they are dealing with (e.g., narrative, argumentative), the level of education and proficiency, etc. Moreover, even when all these aspects are taken into account, there always seems to be an element of subjectivity in the final decision.

At the same time, the research literature shows that neither when concerned with correcting nor with commenting do teachers appear to focus on matters that would enhance student development the most. In the former case, many instructors seem to be on a hunt for details rather than in search of an understanding and conveying of the bigger picture. In the latter case, many instructors appear to provide feedback that, while often praise-oriented, is neither what students want nor require to develop their writing competence further (Stefani, 1998). That is, as evidenced in, for instance, Fisher & Frey (2013), much of the work teachers put into supplying comments is not conducive to the process that writing a composition entail.

The present study aims to investigate the relationship among assessing, marking, and grading when working with Second Language (L2) student compositions at university level, while at the same time adhering to the requirements of the course syllabus. The main concern will be the teacher perspective, but to what degree teachers' efforts help promote students' writing skills will also be discussed.

2 Theoretical background

While there are quite a few different assessment types (Council of Europe, 2001), summative versus formative and analytic (also atomistic) versus holistic are of primary interest to the present investigation. These will, therefore, be presented in more detail in Subsection 2.1. In Subsection 2.2, teacher corrections and comments will then be discussed in the light of these assessment types and more, as presented in two case studies.

2.1 Summative versus formative assessment/feedback and analytic versus holistic assessment/feedback

Summative assessment, frequently referred to as *assessment of learning*, and *formative assessment*, commonly referred to as *assessment for learning*, are based on two very different approaches (Lundahl, 2012, p. 485). The former type has developed from behavioristic ways of learning in which it is believed that student achievement is best measured in terms of objective evidence with the help of, for instance, various scoring systems (Shepard, 2000a). The latter type has instead drawn from cognitive, constructivist and sociocultural theories (Shepard, 2000a) according to which a great deal of responsibility for development is put on the students. The aim is here to turn students into autonomous learners, the instructor playing a crucial role in facilitating this goal

(Shepard, 2000b). Within a formative approach to learning, concepts such as ‘zone of proximal development’ (ZPD) (Vygotsky, 1978), and ‘scaffolding’ (Wood, Bruner & Ross, 1976) are, therefore, pivotal. Both focus on what a learner can achieve with the help of a teacher or peer who is more experienced than themselves, and that as such they can provide tailored assistance to guide the learner to the next step in his/her developmental trajectory (Lantolf & Aljaafreh, 1995; Shepard, 2000b). Thus, formative assessment is dynamic, i.e., the composing of a text is a process where a learner always builds on what was previously known (Lantolf & Poehner, 2014; Poehner, 2009; Poehner & Lantolf, 2013), whereas a summative approach regards any written text as a product. Put differently, when formative assessment is implemented, teachers do something *with* their students; when summative assessment is made use of, teachers do something *to* their students (Serafini, 2000/2001). What is more, formative assessment is reciprocal as it does not only supply students with information that will promote their learning, but also empowers teachers to adapt and develop their instruction as they work to meet the learning needs of their students. Moreover, even though summative and formative assessment may seem to serve entirely different purposes, where the former type epitomizes achievement and the latter type propels further progress, it is rather how the assessment is implemented than at what point in time it is given that determines whether it is of a summative or formative nature. Any assessment that offers students input regarding their strengths and weaknesses could thus principally be used for developmental purposes, even summative assessment (Bell & Cowie, 2000).

Over the last few decades, there has been a definite shift away from summative to formative assessment in English language education generally, research most often pointing to the superiority of the latter (Black & Wiliam 1998). Still, most researchers agree that these two different ways of assessing complement each other, so that neither one could be made away with entirely (Skolverket, 2011). Furthermore, although there are comparatively few studies on the use of formative assessment specifically focusing on writing, especially L2 writing research which appears limited to theory (Evans, 2013; Hattie & Timperley, 2007), those that do exist also seem to have observed primarily positive results (e.g. L1 writing – Graham, Herbert & Harris, 2015; Parr & Timperley, 2010, L2 writing – Huang, 2012; Lee & Coniam, 2013).

One crucial part of formative assessment is *feedback*, defined as “information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one’s performance or understanding” (Hattie & Timperley, 2007, p. 81), conveying “direct, usable insights into current performance, based on tangible differences between current performance and hoped for performance” (Wiggins, 1993, p. 182). Lee (2017) considers this type of information to be conceptualized in three main steps: 1) *feedup*, which is supposed to answer the question *where am I going?*, and approached by, for instance, providing course syllabi, other learners’ successful attempts at writing, and various types of scaffolding, 2) *feedback*, which should answer the question *how am I going?*, and approached by providing helpful, diagnostic information from teachers and/or peers that is related to the feedup given in the first step, and 3) *feedforward*, which deals with the question *where to next?*, and approached by implementing the information presented in step two, thus providing learners as well as teachers with new goals.

A great deal of research on feedback in L2 writing has been generated within written corrective feedback (WCF), of which the most important kind generally comes from teachers (Lee,

2017). However, despite the fact that results in this area show that feedback needs to be selective (Ferris, Liu, Sinha & Senna, 2013; Sheen, Wright & Moldawa, 2009), instructors, as hinted at in the introduction to this article, generally appear to comment indiscriminately on all types of errors (Furneaux 2007; Lee, 2004, 2008, 2013). Others seem to prioritize details only, thus losing sight of global trends as they do so. This contrast illustrates the vital distinction between an analytic (or atomistic) assessment approach and a holistic assessment approach, only the latter offering an overall evaluation of a piece of work (Council of Europe, 2001). Advocates of the second approach would set to work with a mindset epitomized in a query such as If a student needs to rethink an entire paragraph, why comment on an awkward sentence in that text part?

Moreover, teachers usually write comments in the form of statements, imperatives, and questions (Ferris, 1997; Sugita, 2006), in connection with which research has shown that teachers are inclined to give imprecise or even negative input (Cumming, 1985; Semke, 1984; Zamel, 1985). An example of this is given in Lee (2017, p. 5) where a comment like “interesting content” as a response to a student draft of a narrative story is clearly not linked to the learning goals of story writing, while “an engaging story opening” (feedback) is. Moreover, a comment like “the story opening is fine, but you could revise it to grab the reader’s attention – e.g., by putting a short dialogue at the beginning”, promotes even further what can be done to develop the story’s narrative appeal (feedforward).

2.2 Case studies of teacher corrections and comments on student essays

In Fritzell (2014), the aim was to investigate teachers’ mindset when assessing and correcting student compositions. As informants, two upper secondary school instructors teaching English within the Swedish school system were included. Both informants were male, one native Swedish teacher, 51 years of age (Sven), and one native British teacher, 40 years of age (Andy). The data gathered consisted of interviews in which open questions were posed. These were followed by think-aloud protocols which took place while the two teachers were asked to try and articulate their line of thought while reading and correcting a piece of learner text. The questions posed were concerned with the teachers’ general attitude to assessing, what their focus is and why they focus on these particular aspects, as well as how and why they correct and assess in specific ways. Both instructors were informed about the aim of the study as well as the reason for doing the observation and what questions would be asked. Finally, to investigate if there was an agreement between what the teachers said they did and what they really did, the teachers’ answers to these open questions were related to the results of the think-aloud protocols.

In the interviews, both Sven and Andy stated that they had received hardly any training in their teacher education as to how to assess and correct student compositions. The knowledge they currently had of how to approach learner text instead consisted of pieces of information from colleagues and what they had been able to acquire on their own. Fritzell, (2014), therefore, concludes that both informants had largely constructed their own knowledge about how to correct and assess written text, and that this makes it all unquestionably subjective. Furthermore, in the interviews both informants gave examples of corrections that imply that they implement direct corrective feedback (feedback that includes the correct solution), indirect corrective feedback (feedback that points to an error, either by underlining or by writing a note in the margin, but does

not offer a solution), as well as metalinguistic corrective feedback (feedback that only provides a clue as to what is incorrect, using either codes or a full description). At the same time, however, when directly asked about specific ways of correcting, both Sven and Andy denied the use of some of them. Fritzell (2014) interprets this to mean that both teachers based their corrections on impression (i.e., a subjective estimation only) as well as guided judgment (i.e., impression + activity of assessment) (Council of Europe, 2001, p. 189). In his interview, Andy further states that “it all really depends on the task” what he focuses on (2014, p. 29), and Sven adds that “sometimes you have to overlook some errors when it is a complex and demanding task” (2014, p. 29). Sven even mentions that he puts on different types of glasses depending on the task at hand and that his ‘linguistic error glasses’ are not his favourites, nor the first ones he puts on.

Sven generally also seems less comfortable with assessing and correcting than Andy does, especially product-oriented tasks such as the one that he was given. In the interviews, he claims to be less interested in atomistic aspects. Instead he assesses and corrects to promote student development. Andy’s answers, on the other hand, portray him as someone methodical with a clear focus who corrects everything and writes formative comments quite frequently. Hence Andy instead assesses and corrects so that students will know where they are in the writing process. This means that, in theory, Sven experiments beyond the analytic base, and instead takes on a holistic approach to essay writing, whereas Andy, product-oriented rather than process-oriented, represents a more analytic approach, giving very much attention to details, but not entirely disregarding the broader picture. This implies that, although both mention communication to be their top priority, Andy is closer to a weak version of Communicative Language Teaching (Spada, 2007) than Sven is. This may be explained by the fact that as a native speaker Andy is possibly more inclined to focus on language errors (i.e., accuracy rather than fluency) than someone who himself is a second language learner.

The results of the think-aloud protocols, however, offer a very different picture. In reality, while holistic matters were commented on, aspects such as spelling, vocabulary, idiomaticity, morphology, grammar, and syntax were undoubtedly considered by both informants, and to a higher degree than what could be understood by their answers in the interviews. One reason for this, Fritzell (2014) argues, may be that it is easier to find surface errors, while holistic aspects are less visible. Moreover, Andy did not offer any formative comments, which, in the interview, he claimed was an essential element of his approach. The few comments Andy made were mostly in the form of recommendations and corrections, as well as a few praises. Sven, on the other hand, was very personal in his comments and offered friendly advice as well as a great many praise-oriented comments. These findings, Fritzell proposes, could indicate that when asked about how they go about assessing and correcting compositions, teachers talk about what they would like to do, and not what they actually do. Accordingly, Sven would like to approach a piece of writing in a holistic formative manner in which the writing of an essay is seen as a process, and Andy would like to give plenty of feedforward comments, supporting learners during their writing development. For Sven, the culprit why he is not able to realize his aims may be down to time, as this is the one thing he returns to repeatedly in the interview. That is, if the primary target is the language part, there will be no time for other aspects. Still, the fact that Sven, who is ten years older than Andy and, therefore, has ten more years of experience in his bag, discusses and thus apparently considers the bigger picture more than Andy does, may also indicate that a teacher

progresses from an atomistic to a holistic approach over time, and comes to see writing as a process rather than being product-oriented, eventually focusing on positives/strengths rather than negatives/weaknesses. Such development would indicate that assessment of learner texts is not only a journey for students, but teachers too.

In Pandey & Magin (2002), the qualitative feedback on 102 essays (between 1000 and 1200 words long) written by first language (L1) and L2 students taking an introductory course in anatomy at the University of New South Wales was analyzed. The assessors, i.e. teachers as well as fellow students, were presented with six criteria: 1) addressing the topic, 2) evidence of research on the topic, 3) adequacy of discussion, 4) coherence and readability, 5) conclusion/justification, and 6) referencing. They were told that their comments should convey constructive feedback on how the essays could be improved in relation to these criteria. The peer assessors were randomly given essays of students belonging to other tutorial groups than their own, while the five teachers assessed the essays that belonged to their respective group.

The findings of the study show that there was peer-teacher agreement in the relative frequency of all six categories implemented, 'coherence and readability' being the most common feedback category (30/37%) and 'addressing the topic' attracting the least number of comments (9/10%). However, the peer assessors made generally fewer comments and were also involved in considerably more cases in which no comments at all had been offered (25%) than were the teachers (none). The students were also more inclined to mention aspects that were not related to the six criteria, as well as voicing suspicion that (parts of) the essay had been plagiarized. The most striking finding, according to Pandey & Magin (2002), is however the fact that the comments offered by the tutors displayed clear differences in focus, one teacher even having a penchant for specific phrases, reusing them over and over again.

In Topping, Smith, Swanson & Elliot (2000), it is stated that comments made by assessors usually form a reliable indication of on what criteria a composition has been judged. The fact that the teachers in Pandey & Magin's study (2002) foregrounded different aspects may, therefore, cause confusion among students (see also Lea & Street, 1998). One reason for the difference in focal points may, according to Pandey & Magin (2002), be attributed to the heavy load of teaching and marking, one way of alleviating this burden being the reuse of formulaic expressions as mentioned above. Such an approach also implies that there may not be a clear link between the assessment criteria implemented and the comments offered (see also Magin, Helmore & Baker, 2001). The extension of this is, of course, that different teachers may, for the same learner text, come to different conclusions as to grade.

3 The present study

3.1 Research questions addressed

In the present investigation, five main research questions are addressed:

- 1) Quantitatively and qualitatively, what corrections/comments do assessors offer on Swedish university students' compositions in English as a second language?
- 2) Do the corrections/comments relate to the grades given?
- 3) Do the corrections/comments made in the essays align with the end comments?

- 4) To what extent do the end comments help promote students' writing skills?
- 5) Do different assessors focus on different aspects of the essays?

3.2 The students and their compositions

A total number of 56 Swedish full-time university students of English as a single subject participated in the present investigation (1). All of these informants studied English for (at least) three terms (henceforth referred to as English A, B, and C), during each of which they were required to write one composition (2). The material, entirely based on availability, was collected during a four-and-a-half-year period with the requirement that the informants, in order to be included, should have passed all their exams at all three levels. As some students also had to do resits to get a passing grade, as many as 187 compositions make up the data. These are distributed as presented in Table 1:

Table 1. The distribution of compositions in the present study.

	A-level	B-level	C-level	Total
Compositions	60 (4 resits)	64 (8 resits)	63 (7 resits)	187 (19 resits)

3.3 The syllabi (3) and the composition courses based on these syllabi

At the A-level, the composing of the essay was placed within a written proficiency module which constituted 25% of the entire course. At the B-level, the essay was also made part of a written proficiency module, but here this subcourse only constituted 15% of the entire course. At the C-level, the essay was instead placed within a general proficiency module, including both spoken and written proficiency, which made up 15% of the entire course. (However, the C-level involves the writing of a degree project, focusing either on linguistics or literature.) At the A-level, the aims and contents of the relevant module are stated as follow (translated into English by the present author):

Based partly on the texts (4) and partly on a comprehensive grammar book of the English language, the ability to understand and use written English in terms of vocabulary, grammar, and stylistics (5) is developed. Proficiency is considered in terms of translation as well as free production.

At this level, the composition itself only made up 5% of the 25% mentioned above, while a test on general questions on grammar as well as translation sentences constituted the rest of the module. At the B-level, much of the A-level syllabus wording remains, but with a few additions (in bold):

Based partly on the texts and partly on **a continued study** of English grammar, the ability to understand and use written English in terms of vocabulary, grammar, and stylistics **is developed further**. Proficiency is considered in terms of translation as well as free production, **with an emphasis on free production**.

Here the composition constituted 5% of the 15% mentioned above, the rest again focusing on general grammar and translation. At the C-level, finally, the aims and contents of the module containing the composing of the essay are stated as follow:

The module involves continued practice of the ability to understand and use spoken language in different communicative situations. In addition, the ability to understand and use written English in terms of vocabulary, grammar, and stylistics **is developed even further**. Proficiency is considered in terms of translation as well as free production.

Here half of the 15% that made up the spoken/written proficiency module focused on the actual writing of the composition.

At each level, the lecturers, always native speakers of English, were given around ten hours per group to teach composition writing. The students, depending on the instructor, then handed in two to three practice essays on which they received feedback. It is here worth noticing that attendance was only compulsory at the C-level. Furthermore, the composition part was, at all three levels, always given in connection with a comprehensive text course. This included contemporary fictional texts (75%) as well as non-fictional texts (25%) at the A- and B-level, and fiction from the 17th century and onwards as well as drama, poetry and culture studies at the C-level. Topics were chosen accordingly. Examples are *A book that changed my life*, *Creating a bond*, *English as a world language*, *Reading as recognition*, *Studies abroad*, *A literary conflict*, *Women in literature*, *The Victorian Era*, *Childcare in the welfare state* and *A good ending*.

When it comes to the final test at the end of the course, the instructions differed somewhat depending on test level. Whereas first-term students were required to write a 300-word essay, second-term students were told to write at least 400 words, and third-term students to write a minimum of 500 words. Students were not allowed to deviate from the stipulated minimum number of words more than 10%. If they did, they would automatically receive a failing grade.

3.4 The lecturers teaching the composition course and assessing the essays

During the years when the data were collected, the same four, very experienced, lecturers were assigned to teach the composition course, assess the students' practice essays as well as grade their achievements on the final exam. (As the students were allowed to take the practice essays home with them, it is only the final exams that make up the material of the present investigation.) All four are native speakers of English, one male, then in his late 50s, speaking American English, and three females, then from 53 to 67 years of age, speaking British English. All have Masters of Art with a literary orientation. Lastly, while all three female lecturers/assessors have some teacher training, this is not the case with the male lecturer/assessor. Table 2 offers a summary of what has been said above.

Table 2. Information about the lecturers/assessors included in the present investigation.

	Gender	Age (during the 4½ years the data were collected)	Nationality	Degree	Teacher education
Lecturer/Assessor 1	male	56-60	American	Master of Arts/literature	-
Lecturer/Assessor 2	female	63-67	British	Master of Arts/literature	Upper secondary school level
Lecturer/Assessor 3	female	56-60	British	Master of Arts/literature	Upper secondary school level
Lecturer/Assessor 4	female	53-57	British	Master of Arts/literature	Middle school level

3.5 The categorization of mid-essay corrections/comments and final comments

In the present investigation, both mid-essay corrections/comments as well as end comments were considered (6), but they are presented separately. In the former case, four categories were formed: A) spelling, B) vocabulary including idiomaticity, C) grammar/syntax and D) topic, content, structure (including paragraphing), punctuation and referencing (7). Here *direct corrective feedback*, *indirect corrective feedback*, as well as *metalinguistic corrective feedback* were all taken into account (see Subsection 2.2).

Category B) requires a special mentioning. This category does not only include clear errors of usage, but also those uses where the students' choice of word was acceptable, but where the assessor believed there to exist a better, more to-the-point alternative. Additionally, the category includes cases where the word or expression used by the student should not have been incorporated at all, as well as cases where the assessor inserted a word or expression, where none was used to begin with.

It is here worth pointing out that it is the assessor's 'perception' of the essay at hand which is important, as this is what the grades were finally based on. This means that errors that were not noticed were not included in the results. While the opposite would also have been the case, it turned out that whereas there indeed were errors that had gone by unobserved (comparatively few), there were no situations in which something correct had been commented on as being incorrect.

Most importantly, while comments falling into Categories A-C can be classified as more atomistic (i.e., analytic) in character, comments found in Category D take on a more holistic approach to essay writing.

As for the assessors' end comments, which are generally used to identify global trends in students' writing and should, therefore, show alignment with mid-essay corrections/comments, these were observed to fall into five main categories: a) summative comments, b) formative comments, c) comments of praise, d) comments of blame, and e) comments on facts. Those belonging to category e) either dealt with a fact that was incorrect or with a fact that, for some reason, was found intriguing by the assessor who, therefore, added a personal thought. While some of these end comments consisted of several elements of information, thus falling into more than

one category, mid-essay corrections/comments as discussed above generally only consisted of one piece of information each, and were, therefore, consistently placed in one category per correction/comment.

A special note here seems called for in connection with Categories a) and b). End comments, especially on an exam, as is the case in the present investigation, may not be thought of in terms of formative at all. However, as the students wrote compositions at all three levels, the assessors may have taken this into account, wanting to provide feedforward comments. On the other hand, the assessors did not know which of the students would continue to the next educational level, and they may, therefore, have been more reluctant to offer formative than summative comments.

4 Results and discussion

In the present section, the research questions will be addressed. Table 3 offers the results of the quantitative analysis of the mid-essay corrections/comments in relation to the grades given.

Table 3. The results of the quantitative analysis (percentage and average) of the corrections/comments found in the essays.

		A) spelling	B) vocabulary, including idiomaticity	C) grammar and syntax	D) topic, content, structure (including paragraphing), punctuation and referencing
U (fail)	no/total no	32.01% (=153/478)	24.48% (=117/478)	35.15% (=168/478)	8.37% (=40/478)
	average/student	6.95 (=153/22)	5.32 (=117/22)	7.64 (=168/22)	1.82 (=40/22)
G (pass)	no/total no	26.91% (=292/1085)	23.96% (=260/1085)	41.11% (=446/1085)	8.02% (=87/1085)
	average/student	3.21 (292/91)	2.86 (=260/91)	4.90 (=446/91)	0.96 (=87/91)
VG (pass with distinction)	no/total no	27.31% (=130/476)	28.99% (=138/476)	36.13% (=172/476)	7.56% (=36/476)
	average/student	1.83 (=130/71)	1.94 (=138/71)	2.42 (=172/71)	0.51 (=36/71)

Table 3 clearly shows that there is a gradual decrease in the average score from a fail to a pass with distinction in all four categories. The precision with which the assessors awarded grades in relation to the corrections/comments made within the essays must, therefore, on a class level, be deemed quite impressive, yet probably exactly what students expect. This accuracy may partly be explained by the fact that all four lecturers were very experienced assessors, and partly by what is referred to as ‘tacit knowledge’. Coined by Polanyi in 1958, the concept refers to a person’s ‘hidden’ knowledge. More precisely, it refers to an “unwritten, unspoken and hidden vast storehouse of knowledge” based on “emotions, experiences, insights, intuition, observations and internalized information” (Luthra, 2007). It may thus be that it is this tacit knowledge that differentiates between an excellent teacher and, if not a bad one, at least a mediocre one. That is, while some teachers may possess this ‘gift’ for correcting/commenting naturally, others do not. It does, however, seem reasonable to suggest that tacit knowledge may also be developed and enhanced over time, being (partly) dormant to begin with. It should be noted though that teacher training does not appear to be a requirement for such knowledge, as one of the assessors in the

present investigation had none, and that, in this respect, there was no evidence of him working differently than the other three assessors who had teacher training.

It can also be observed that more end comments, here only considered from a quantitative perspective, were made on those essays receiving a fail (81.82% (=18/22)) than those awarded with a pass (20.88% (=19/91)) or a pass with a distinction (21.13% (=15/71)), the assessors here seemingly trying to offer enough support to ensure a passing grade at the next attempt. (The qualitative value of these end comments will be discussed further down.)

There are only five cases which deviate from the general trend described above, these students having received lower grades than what could have been expected based on the mid-essay corrections/comments. However, in three of these cases, the assessors instead offered explanations of the grades awarded in their general comments. This means that there are only two students in the whole material that, based on the corrections and comments made by assessors, could have chosen to argue for a higher grade. All in all, this must, as concluded above, be considered quite an achievement on the part of the assessors.

Tables 4-6 offer detailed pictures (percentage and average score) of what the assessors corrected/commented on within each grade in relation to educational level (A-, B- and C-level).

Table 4. Essays receiving a fail in relation to the three different educational levels.

			A) spelling	B) vocabulary, including idiomaticity	C) grammar and syntax	D) topic, content, structure (including paragraphing), punctuation and referencing
U (fail)	A-level	no/total no	30.77% (=36/117)	30.77% (=36/117)	38.46% (=45/117)	0%
		average/student	6.00 (=36/6)	6.00 (=36/6)	7.50 (=45/6)	0
	B-level	no/total no	40.18% (=88/219)	15.53% (=34/219)	34.25% (=75/219)	10.05% (=22/219)
		average/student	9.78 (=88/9)	3.78 (=34/9)	8.33 (=75/9)	2.44 (=22/9)
	C-level	no/total no	20.42% (=29/142)	33.10% (=47/142)	33.80% (=48/142)	12.68% (=18/142)
		average/student	4.14 (=29/7)	6.71 (=47/7)	6.86 (=48/7)	2.57 (=18/7)

Table 5. Essays receiving a pass in relation to the three different educational levels.

			A) spelling	B) vocabulary, including idiomaticity	C) grammar and syntax	D) topic, content, structure (including paragraphing), punctuation and referencing
	A-level	no/total no	21.92% (=73/333)	23.72% (=79/333)	46.55% (=155/333)	7.81% (=26/333)
		average/student	2.35	2.55	5.00	0.84

G (pass)	B-level	no/total no	(=73/31) 27.86%	(=79/31) 23.70%	(=155/31) 44.27%	(=26/31) 4.17%
		average/student	(=107/384) 3.15	(=91/384) 2.68	(=170/384) 5.00	(=16/384) 0.47
	C-level	no/total no	(=107/34) 31.05%	(=91/34) 22.37%	(=170/34) 35.79%	(=16/34) 10.79%
		average/student	(=118/380) 4.07	(=85/380) 2.93	(=136/380) 4.69	(=41/380) 1.41

Table 6. Essays receiving a pass with distinction in relation to the three different educational levels.

		A) spelling	B) vocabulary, including idiomaticity	C) grammar and syntax	D) topic, content, structure (including paragraphing), punctuation and referencing	
VG (pass with distinction)	A-level	no/total no	25.56% (=34/133)	33.08% (=44/133)	35.34% (=47/133)	6.02% (=8/133)
		average/student	1.48 (=34/23)	1.91 (=44/23)	2.04 (=47/23)	0.35 (=8/23)
	B-level	no/total no	24.34% (=37/152)	26.97% (=41/152)	40.79% (=62/152)	7.89% (=12/152)
		average/student	1.76 (=37/21)	1.95 (=41/21)	2.95 (=62/21)	0.57 (=12/21)
	C-level	no/total no	30.89% (=59/191)	27.75% (=53/191)	32.98% (=63/191)	8.38% (=16/191)
		average/student	2.19 (=59/27)	1.96 (=53/27)	2.33 (=63/27)	0.59 (=16/27)

Of particular interest here is not only the fact that Category D, which deals with more holistic aspects of essay writing, displays the lowest number of corrections/comments within all three grades, but that the percentages (as well as average scores) in this category increase from the lowest to the highest educational level (in bold) within each grade. This increase is incremental along all three levels (A, B and C) for a failing grade as well as pass with distinction. For a pass there is a difference between the highest level and the two lower levels, but not between A- and B-level, which can most likely be explained by the fact that a pass includes a much wider span of achievements than do the lowest and highest grade.

In *From "thought and language" to "thinking for speaking"* (1996), Slobin, in a modified version of the Sapir-Whorf hypothesis, suggests that a person's mother tongue determines how he/she presents linguistic information. More precisely, Slobin claims that specific constraints are put on the way spoken messages are uttered, resulting in *a theory of thinking before speaking*. As a continuation of this line of thought, Strömqvist, Nordqvist & Wengelin (2004) suggest that constraints, but of other types, are also imposed on written production. That is, they put forth *a theory of thinking before writing*, which is developed later than *thinking before speaking* and retains some of the elements specific for speaking for quite some time. Along these lines, *an approach of correcting and commenting*, more specifically *an approach of correcting and commenting student compositions in accordance with educational level* is here proposed, whereby assessors, especially experienced one, seem attuned to what educational level they are grading. Irrespective of what grade the students finally received, it appears that the assessors were more likely to correct/comment on holistic aspects at the higher educational levels, while focusing on atomistic aspects at the lower levels, here overlooking errors they did not believe learners studying at the lower levels are able to take in because of their greater overall complexity. The fact that there indeed were errors of a holistic type to correct/comment on at the lower levels supports this interpretation. (It is also further supported by the fact that most of the cases where the assessors had suggested alternative vocabulary to capture a concept in a more precise manner occurred at the highest educational level, while the majority of the corrections/comments on vocabulary at the lower levels were concerned with true errors.) This tallies with the tentative conclusion drawn in Fritzell (2014, where, at least from a theoretical point of view, the more experienced educator valued a holistic approach more than an analytic one. Since all four lecturers included in the present investigation are very experienced assessors, it may be, as an extension of what has been said above, that at the beginning of their careers they focused more on details, language per se being the primary target, and only thereafter slowly started to develop a more holistic perspective, focusing on aspects that affect more substantial parts of an essay. If that is the case, it most certainly means that assessment does not only involve a developmental journey for students, but teachers too.

Two other explanations are possible for this result. On the one hand, it may be that, since students make fewer errors/mistakes belonging to Categories A to C at the highest level, there is then also more time to focus on holistic aspects. As explained in Subsection 2.2, time was considered a possible issue in Fritzell (2014) too. The lack of time is, however, a less likely explanation here as the essays at the C-level were longer than those written at the A- and B-level (See Subsection 3.3). A second explanation may be that the lecturers did not put emphasis on holistic aspects in the composition course until the C-level when attendance was compulsory and that this approach was mirrored in their corrections/comments. Regrettably, it could not be found out whether this was the case or not.

As discussed above, a clear correlation between mid-essay corrections/comments and grade was observed. We will now turn to explore whether a connection can be seen between mid-essay corrections/comments and end comments, and thus also whether there is a relation between end comments and grade. The reader is here reminded that final comments that were made up of several elements were also categorized accordingly (see Subsection 3.5); thus one and the same comment may be subdivided into several categories. Moreover, links in this respect are of course

only relevant in connection with summative/formative feedback/feedforward, which here was considered to be either 'clear', 'partial', 'not present' or 'incorrect'. Table 7 and 8 present the results of these analyses in relation to grade.

Table 7. Types of end comments in relation to grade.

	Type of end comment				
	a) summative	b) formative	c) praise-oriented	d) blame-oriented	e) factual
U (fail)	14	9	5	5	1
G (pass)	17	2	6	1	5
VG (pass with distinction)	7	-	10	-	2

Table 7 reveals that end comments are not very prolific in the material as a whole and that summative ones form the most frequent type at all three educational levels. Logically though, formative comments are more frequent in connection with those compositions that received a fail (9) than with those that received a pass (only 2) and those that received the highest grade (0). However, as 22 out of the 187 compositions were not awarded a passing grade, it would have been desirable to have been able to observe more end comments that could have helped enhance these specific learners' writing development. Praise-oriented comments (e.g. *very good!* and *well done!*), which as discussed in Subsection 2.1 is unlikely to promote learners' writing skills, are comparatively common all throughout the material, whereas comments expressing blame are most frequent in connection with the lowest grade, where they probably are also most likely to do harm(!). Thus while the four lecturers displayed almost precise accuracy as to the correlation between mid-essay corrections/comments and grade, they do not seem equally adept at providing feedforward input. It must, however, be remembered that the present investigation does not include a discussion of what went on in the composition courses (see Subsection 3.4), where, of course, the students most likely received a great deal of supportive feedforward.

Table 8. Summative and formative end comments related to mid-essay comments.

	Type of end comment		Type of link with mid-essay corrections/comments			
	summative	formative	clear	partial	not present	incorrect
U (fail)	14	9	13	4	5	1
G (pass)	17	2	9	3	6	-
VG (pass with distinction)	7	-	4	-	3	-

Table 8 shows that, while the majority of the end comments are indeed linked to the corrections/comments made in the essays and the grades awarded, especially for those that received a fail, there are also cases where the links are only partially transparent or do not exist at all. Thus, had these end comments been the only type of feedback/feedforward, those students would clearly not have understood on what their grades had been based, and they would consequently have been lost regarding what to work on to improve their writing skills.

Based on what has been said above, it is apparent that it is easier to offer mid-essay corrections/comments and relate them to grade than provide comments that will enable students to move ahead, whether it be from a failing grade to a passing one, or from one educational level to the next. Again it needs to be pointed out that a great many such formative comments may, of course, have been offered during the composition course.

Finally, in contrast to what was found in Pandey & Magin (2002) (see Subsection 2.2), while there indeed were reoccurring phrases used by the present assessors, these seemed warranted in connection with what the students had produced and did not point to a specific assessor being especially focused on one particular aspect, while ignoring the rest.

5 Summing up

In the present study, teacher corrections and comments on 187 compositions written at three different university levels in English as a second language were analyzed. The results reveal clear links between corrections of all four categories considered and grades given, i.e., the higher the grade, the lower the average score in all four categories. The assessors' great experience, as well as their tacit knowledge, were thought to be the main reasons for this precision. In connection with the former, it would, therefore, be interesting to explore what less experienced and recently graduated teachers could achieve.

The results also show that the assessors seem attuned to the educational level at hand. That is, they display a thinking-before-correcting/commenting approach, focusing on more analytic aspects at the lowest educational level and more holistic aspects at the highest educational level. If this is the case, it implies that assessment does not only involve a developmental journey for students but for some teachers too, where they move from a detailed-driven approach to an approach where conveying the bigger picture is of greater importance. Again, making a comparison between experienced and less experienced teachers may be of interest.

Finally, though much needed, according to the research literature, comments made at the end of the compositions were summative rather than formative. However, the links made between mid-essay corrections/comments and end comments, although few, displayed alignment in most, but, regrettably, far from all cases. It thus seems easier to produce clear connections between mid-essay corrections/comments and grade than it is to provide end comments and align mid-essay corrections/comments with end comments, which would help promote students' essay writing skills in an enhanced manner.

Notes:

- (1) 46 of the students are female (82%) and 10 are male (18%). This distribution largely agrees with the one in Thagg Fisher (1985) and Karlsson (2012), both focusing on first-term students of English at two different Swedish universities, making the present data representative in terms of gender.
- (2) The academic year at Swedish universities is made up of two terms.
- (3) Since the collection of the material, the university has made changes in the syllabi.
- (4) Here 'the texts' refer to the fictional books read in another module preceding the written proficiency subcourse.
- (5) Stylistics, usually defined as "...the study and interpretation of texts..." "...in regard to their linguistic and tonal style, where style is the particular variety of language used by different individuals and/or in different situations and settings" (<https://en.wikipedia.org/wiki/Stylistics>), is here used in a much wider sense than normal, seemingly focusing on more holistic aspects of the students' written compositions as captured in Category D) discussed in Subsection 3.5. In fact, only one comment made in the whole material, and, therefore, not discussed further, was concerned with style as defined above.
- (6) Tokens, not types, were considered.
- (7) Referencing is here concerned with errors/mistakes with anaphoric reference, having for example discussed an item in the singular, but referring back as if in the plural, thus often causing confusion on a holistic level.

About the author:

Dr. Monica Karlsson is Associate Professor of English at Halmstad University, Sweden. Her main field of interest is Second Language Acquisition (SLA) in the area of which she has published several research articles and books, the most recent one being *Idiomatic mastery in a first and second language*. <https://orcid.org/0000-0002-4138-2338>

References

- Bell, B. & Cowie, B. (2000). The characteristics of formative assessment in science education. *Science Education* 85: 536-553.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education* 5 (1): 7-74.
- Council of Europe. (2001). *The Common European Framework of reference, learning, teaching, assessment*. [pdf] Electronic version. Cambridge: Cambridge University Press.
- Cumming, A. (1985). Responding to the writing of ESL students. *Highway One* 8: 58-78.
- Evans, C. (2013). Making sense of assessment and feedback in higher education. *Review of Educational Research* 83: 70-120.
- Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly* 31(2): 315-339.
- Ferris, D. R., Liu, H., Sinha, A. & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing* 22(3): 307-329.
- Fisher, D. & Frey, N. (2013). Implementing RTI in high school: A case study. *Journal of Learning Disabilities* 46: 99-114.
- Fritzell, M. (2014). It is in the head. A case study: Two teachers of English at upper secondary level in Sweden discuss and demonstrate how they assess and correct written texts. Halmstad University: Unpublished undergraduate study.
- Furneaux, C., Paran, A. & Fairfax, B. (2007). Teacher stance as reflected in feedback on student writing. An empirical study of secondary school teachers in five countries. *International Review of Applied Linguistics in Language Teaching* 45(1): 69-94.
- Graham, S., Herbert, M. & Harris, K. R. (2015). Formative assessment and writing: A meta-

- analysis. *The Elementary School Journal* 115(4): 523-547.
<https://en.wikipedia.org/wiki/Stylistics>
- Hattie, J. & Timperley, H. (2007). The power of feedback. *Review of Educational Research* 77(1): 81-112.
- Huang, S-C. (2012). Like a ball responding to a striker: Instruction contingent on assessment. *English Teaching: Practice and Critique* 11(4): 99-119.
- Karlsson, M. (2012). *Quantitative and qualitative aspects of advanced students' L1 (Swedish) and L2 (English) knowledge of vocabulary*. Halmstad: Högskolan i Halmstad.
- Lantolf, J. P. & Aljaafreh, A. (1995). Second language learning in the zone of proximal development: A revolutionary experience. *International Journal of Educational Research* 23(7): 619-623.
- Lantolf, J. P. & Poehner, M. E. (2014). *Sociocultural theory and the pedagogical imperatives in L2 education: Vygotskian praxis and the research/practice divide*. New York: Routledge.
- Lea, M. & Street, B. (1998). Student writing in higher education: An academic literacies report. *Studies in Higher Education* 23: 157-172.
- Lee, I. (2004). Error correction in L2 secondary writing classrooms: The case of Hong Kong. *Journal of Second Language Writing* 13(4): 285-312.
- Lee, I. (2008). Understanding teachers' written feedback practices in Hong Kong secondary classrooms. *Journal of Second Language Writing* 17(2): 69-85.
- Lee, I. (2013). Research into practice: Written corrective feedback. *Language Teaching* 46(1): 108-119.
- Lee, I. (2017). *Classroom writing assessment and feedback in L2 school contexts*. Singapore: Springer Verlag.
- 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.
- Lundahl, B. (2012). *Engelsk språkdidaktik*. (3rd ed.) Lund: Studentlitteratur.
- Luthra, V. (2007). BusinessDictionary.com. (online) Available at:
<http://www.businessdictionary.com/definition/tacit-knowledge.html>.
- Magin, D., Helmore, P. & Baker, J. (2001). Assessing students' oral communication skills – which skills? In Pudlowski, Z. (ed.), *Proceedings of the 4th UICEE Annual Conference on Engineering Education* (pp. 237-241). Bangkok: UICEE.
- Pandey, P. & Magin, D. (2002). Feedback from peer and teacher assessments of introductory anatomy essays. *Herdsa*: 512-519.
- Parr, J. M. & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing* 15(2): 68-85.
- Poehner, M. E. (2009). Dynamic assessment as a dialectal framework for classroom activity. Evidence from second language (L2) learners. *Journal of Cognitive Education and Psychology* 8(3): 252-268.
- Poehner, M. E. & Lantolf, J. P. (2013). Bringing the ZPD into the equation: Capturing L2 development during computerized dynamic assessment (C-DA). *Language Teaching Research* 17(3): 323-342.
- Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. Chicago: University of Chicago Press.

- Sheen, Y., Wright, D. & Moldawa, A. (2009). Differential effects of focused and unfocused written correction on the use of grammatical forms by adult ESL learners. *System* 37(4): 556-569.
- Shepard, L. A. (2000a). *The role of classroom assessment in teaching and learning*. (CSE Technical Report 517). Los Angeles: University of California, National Center Research on Evaluation, Standards, and Student Testing.
- Shepard, L. A. (2000b). The role of assessment in a learning culture. *Educational Researcher* 29(7): 4-14.
- Skolverket. (2011). *Kunskapsbedömning i skolan – praxis, begrepp, problem och möjligheter*. Stockholm: Fritzes.
- Slobin, D. I. (1996). From "thought and language" to "thinking for speaking". In Gumperz, J. & Levinson, S. (eds.), *Rethinking linguistic relativity* (pp. 70-96). Cambridge, UK: Cambridge University Press.
- Strömqvist, S., Nordqvist, S. & Wengelin, Å. (2004). Writing the frog story – developmental and cross-modal perspectives. In Strömqvist, S. & Verhoeven, L. (eds.), *Relating events in narrative – across languages, cultures and genres* (pp. 359-394). Mahwah, N. J.: Lawrence Earlbaum Associates.
- Thagg Fisher, U. (1985). *The sweet sound of concord. A study of Swedish learners' concord problems in English*. Malmö: Liber.
- Topping, K., Smith, E., Swanson, I. & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education* 25: 149-169.
- Semke, H. D. (1984). Effects of the red pen. *Foreign Language Annals* 17: 195-202.
- Serafini, F. (2000/2001). Three paradigms of assessment: Measurement, procedure, and inquiry. *The Reading Teacher* 54(4): 384-393.
- Spada, N. (2007). Communicative language teaching: Current status and future prospects, In Cummins, J. & Davison, C. (eds.) *International handbook of English language teaching*, Vol. 15, pp. 271-288. US: Springer.
- Stefani, L. (1998). Assessment in partnership with learners. *Assessment and Evaluation in Higher Education* 23: 339-350.
- Sugita, Y. (2006). The impact of teachers' comment types on students' revision. *ELT Journal* 60(1): 34-41.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Wiggins, G. (1993). *Assessing student performance*. San Francisco: Jossey-Bass.
- Wood, D., Bruner, J. & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry* 17: 89-100.
- Zamel, V. (1985). Responding to students' writing. *TESOL Quarterly* 19: 79-101.