Feedback practices in language classes in Finnish general upper secondary schools

Toni Mäkipää, University of Helsinki

As feedback and formative assessment have a substantial effect on learning, the aim with this paper is to report on a study of the perceptions of Finnish general upper secondary school students of feedback in Swedish and English classes, and to compare how the perceptions differ at language proficiency (CEFR) levels. The data were collected by using a survey and were analysed quantitatively. The results show that several differences occur in Swedish: students with higher proficiency levels find feedback more useful, feel that they receive feedback from teachers, and are more willing to correct their own mistakes. There were no differences in perceptions according to language proficiency levels in English. The results indicate that Swedish teachers should pay more attention to their feedback practices to make sure that they cater for students with different levels of proficiency.

Keywords: assessment, feedback, CEFR, general upper secondary school

1 Introduction

The purpose of this study is to examine the perceptions of Finnish general upper secondary school students about the feedback they receive in Swedish and English language classes. Language proficiency levels are described in terms of the Common European Framework of Reference for Languages (CEFR). The CEFR is an international framework describing language ability and achievement at a range of levels ranging from A1 (basic user) to C2 (proficient user) (Council of Europe, 2001). The National Core Curriculum for general upper secondary education (FNBE, 2015) describes the target levels in each language after completion of general upper secondary school (B1.1 in Swedish and B2.1 in English). Thus, it is of interest to study the perceptions of students of feedback at different CEFR levels to enhance language learning and teaching as well as to differentiate between students. Comparing English and Swedish is also of interest, as they encompass distinct roles in Finnish society; almost 90 per cent of the students agreed that it is useful to know English, particularly in everyday life but also in the workforce. Moreover, about 60 per cent of the students believed that they are good at English (Härmälä, Huhtanen, & Puukko, 2014). In contrast, only 56 per cent of the students believed that they would need Swedish in the workforce, and 54 per cent expressed the view that they would

need Swedish in everyday life. Regarding students' skills, 36 per cent of the students stated that they are good at Swedish (Hildén & Rautopuro, 2014). As these national studies on 9th graders have pointed out, students find English more useful than Swedish, and they assess their skills to be better in English than in Swedish. In addition, according to a national survey on Finns' perceptions of English, people estimate that the role of English in 2027 will be significant in Finland and almost as important as the role of Finnish, whereas Swedish will be needed much less (Leppänen et al., 2011). Given the above, it can be concluded that learning English is almost regarded to be a civic skill in Finland, whereas Swedish is not (Pyykkö, 2017).

In Finland, students usually graduate from general upper secondary school in three years. Students must complete at least 75 courses prior to sitting for the matriculation examination. One course lasts 38 lesson hours, and 47–51 of the courses are compulsory, depending on whether the student undertakes the basic or advanced syllabus in mathematics. In the matriculation examination, students must complete at least four of the possible 39 tests. The only mandatory test is in the student's mother tongue (Finnish or Swedish). Constitutionally speaking, Finland is a bilingual country, with the Swedish-speaking minority making up about five per cent of the total population.

The National Core Curriculum (FNBE, 2015) stipulates that assessment should be diverse. However, Finnish language teachers implement traditional assessment practices in their teaching (Hildén, Härmälä, Rautopuro, Huhtanen, Puukko, & Silverström, 2015; Härmälä, Huhtanen, & Puukko, 2014), and based on a study of 7th graders' skills in English, it seems that language teachers' assessment practices in Finnish schools should be improved (Härmälä, Huhtanen, Puukko, & Marjanen, 2019). Likewise, research indicates that teachers abroad are not aware of current assessment practices and their implications (Bennett, 2011; Taylor, 2009; Vogt & Tsagari, 2014). Moreover, it seems that teachers use more written forms of assessment than oral (Atjonen et al., 2019; Hildén & Rautopuro, 2014). Based on previous research, one may suppose that teachers' assessment practices are somewhat one-sided. Therefore, studying the perceptions of general upper secondary school students of assessment and feedback will enable the development and enhancement of language teaching in Finland. As research elucidates the point of view of students of feedback, the results could be used to strengthen and improve language teachers' assessment practices. Moreover, to the best of author's knowledge, the perceptions of general upper secondary school students of feedback in language teaching at different proficiency levels have not been studied in Finland.

2 Conceptual framework

2.1 Assessment and feedback in language teaching

Usually, a distinction is drawn between *summative* and *formative assessment*: summative assessment is associated with tests and evaluation of outcomes. Students are taught an issue, after which students take a test that will be graded. In contrast, formative assessment seeks to describe how a student has worked, and rarely do students get grades in formative assessment. Formative assessment is used as learning is taking place, whereas summative assessment is

used after learning should have occurred (Harlen, 2007; Sadler, 1989; Tomlinson & Moon, 2013). The aim of formative assessment is to improve learning, whereas summative assessment is most-often used to grade students (OECD, 2005; Cizek, 2010; Harlen, 2007). To differentiate between the core meaning of the assessment types, Stake (2004, p. 17) clarifies that "when the chef tastes the soup, it's formative evaluation, and when the guest tastes the soup, it's summative evaluation." However, the terms *assessment of learning (AoL)* and *assessment for learning (AfL)* are also used instead of summative and formative assessment (Leahy & Wiliam, 2012). Moreover, some scholars also mention *assessment as learning (AaL)*, emphasizing that students learn about themselves as learners through assessment. AaL is usually considered to be an aspect of formative assessment or AfL (Dann, 2014).

Nevertheless, the distinction between summative and formative assessment has been challenged by Newton (2007), who argues that no-one has managed to distinguish them from each other as they belong to different categories, thereby making it problematic to distinguish them within one category. Similarly, Black (2002) argues that there is no neat difference between them, and that they complement each other. Even though scholars have divided opinions on the distinction between formative and summative assessment, it seems that these two types of assessment are interrelated. Even though their primary purposes are opposite, they both contribute to assessment of learning and assessment for learning. Thus, assessment practices should be designed so that formative assessment generates information on assessment of learning, and summative assessment provides information on assessment for learning (Bennett, 2011).

Feedback can be conceptualised as information on the actions of someone with respect to their goals (Wiggins, 2012). With feedback, the aim is "to reduce the gap between where students are and where they should be" (Hattie, 2012a, p. 20). Feedback, as well as formative assessment in general, constitutes one of the more powerful ways to promote learning (e.g. Black & Wiliam, 1998a, 1998b, 2012; Hattie, 2012b; Hattie & Timperley, 2007). The key to the power of feedback is situated "in its double-barrelled approach", namely cognitive and motivational factors. Good feedback explains where the student is in their learning process and what needs to be done next (cognitive factor). When they have understood what to do and why, they gain the feeling of control over their learning process (the motivational factor) (Brookhart, 2017, p. 2).

One of the integral questions regarding teaching efficiently is how teachers can give meaningful and efficient feedback that promotes and enhances learning, performance, and achievement. Several ways have been proposed, such as clarifying the goals. It must be clear what the student's current learning stage is, what goals they set, and students must be provided with enough feedback so that the goal can be achieved (Brookhart, 2012; Hattie, 2012a; Wiggins, 2012). Feedback should address the task at hand, not the student's personality (Brookhart, 2012), and praise should be avoided (Hattie, 2012a; Hattie & Timperley, 2007). Moreover, teachers should ensure that students understand the content of the feedback (Brookhart, 2017; Hattie, 2012a), and that providing students with the feedback is not delayed. Students need to receive feedback as soon as possible for it to have a powerful effect on achievement (Brookhart, 2017; Crooks, 1988; Wiggins, 2012). Feedback must also be tangible and transparent (Wiggins, 2012), as well as differentiated and adjusted to every student's needs (Brookhart, 2012). Additionally, feedback provided within students' zone of

proximal development can promote learning and increase awareness of mistakes (Leontjev, 2014).

Corrective feedback (henceforth referred to as CF) can be defined as "the responses to a learner's non-target-like L2 production." In other words, corrective feedback is given to students in language classes to indicate errors and mistakes (Li, 2010, p. 309). In the literature, the concept of CF has been utilised as an umbrella term encompassing *negative feedback*, *error treatment*, and *error correction* (Sheen, 2011). Regarding written CF, several strategies have been identified. Direct CF involves providing the student with the correct form, whereas in the case of indirect CF, the teacher points out that the student has made a mistake but does not correct it. Meta-linguistic feedback refers to giving students explicit comments about the mistake, for instance with the help of error codes. Moreover, CF can be focused or unfocused: the teacher can either correct all the errors (unfocused) or correct certain error types (focused). The teacher can also reformulate the sentence (Ellis, 2009).

Oral corrective feedback practices can also be categorised into groups. The first of these is *explicit correction,* when the teacher gives the correct form to the student, indicating that the form produced by the student was incorrect. *Recasts* refer to the reformulation of the student's utterance without the error, whereas *meta-linguistic feedback* indicates that there is an error in the student's utterance, but the teacher does not provide the correct form. With respect to *elicitation*, it encompasses three techniques. The teacher repeats the utterance and leaves a blank, indicating that the student should fill it in, or teachers can ask questions, such as "How do you say that?" Teachers can also ask students to reformulate what they have said. *Repetition* involves teachers' repetition of the mistake, usually followed by intonation in order to emphasise the error (Lyster & Ranta 1997). Sheen (2011) divides recasts into two subgroups, didactic recasts (reformulation has a pedagogical purpose) and conversational recasts (reformulation has a communicative purpose). Sheen also points out that explicit correction can be accompanied by metalinguistic explanation.

A large and growing body of literature has investigated the perceptions of students about corrective feedback. Several studies have revealed that students find corrective feedback to be useful (e.g. Ammar & Spada, 2006; Bitchener, 2008; Fathman & Whalley, 1990; Ferris & Roberts, 2001; Li, 2010). However, the relative importance of corrective feedback has been subject to debate: Truscott (1996, 1999a, 1999b, 2007) argues that grammar correction does not facilitate learning, or its contribution to learning is minor. To illustrate this, he points out that it is extremely time-consuming for teachers, and that amount of time could be used better. Moreover, students start to express themselves in a short and simplified manner as they do not want to get corrected, and correction might even discourage students (Truscott, 1996). In contrast, Ferris (1999) is critical of the conclusions that Truscott draws from his findings. She argues that his conclusions are "based on limited, dated, incomplete, and inconclusive evidence" (p. 9), and that the studies in his article are not comparable. In his response to Ferris, Truscott (1999a) emphasises that teachers themselves decide whether they incorporate grammar correction in their classrooms. Moreover, Fazio (2001) did not find any evidence to support the claim that CF promotes learning and accuracy, and she points out that "there are contradictory reports regarding the influence of form-focused feedback on accurate writing in general" (p. 238).

2.2 The characteristics of the Common European Framework of Reference (CEFR) and the proficiency levels

The CEFR is an international framework describing language proficiency at different levels. The Council of Europe (2001, p. 1) defines the term comprehensively:

> The Common European Framework provides a common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, textbooks, etc. across Europe. It describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively. The description also covers the cultural context in which language is set. The Framework also defines levels of proficiency which allow learners' progress to be measured at each stage of learning and on a life-long basis.

The descriptive scheme entails general competencies (knowledge, skills and know-how, existential competence, and ability to learn) as well as communicative competencies (linguistic, pragmatic, sociolinguistic, and sociocultural). Regarding language activity, it is divided into four categories: reception, production, interaction, and mediation (Little, 2007). The CEFR encompasses six levels: A1 (Breakthrough), A2 (Waystage), B1 (Threshold), B2 (Vantage), C1 (Effective Operational Proficiency), and C2 (Mastery). Level A refers to Basic User, Level B refers to Independent User, and Level C refers to Proficient User (Council of Europe, 2001). Table 1 displays some of the characteristics of the levels based on the Council of Europe (2001, p. 24).

Table 1. Characteristics of the CEFR levels.

C2	Can express him/herself very fluently and precisely. Can differentiate between finer shades of meaning.
C1	Can produce clear and well-structured text. Can use language flexibly and effectively.
B2	Can produce clear and detailed text. Can interact fluently with native speakers without strain.
B1	Can produce simple connected text. Can describe hopes and dreams.
A2	Can communicate in simple and routine tasks. Can describe his/her background.
A1	Can introduce him/herself. Can use everyday expressions.

The CEFR has been successful for two main reasons. First, the timing of the publication was right, just at the heart of establishing a common terminology and attempting to link language learning more to real-life situations. Second, the level descriptors are positive in nature (Figueras, 2012). Regarding the CEFR in Finland, Takala (2010) considers it a useful tool for both developing language education and international cooperation. However, the CEFR has also been criticised because it neither makes any reference to specific languages (Little, 2007) nor is it based on empirical research. Instead, it is based on teachers' perceptions (Hulstijn, 2007). In addition, it seems that inadequate definitions, unclear terms, and gaps occur in the CEFR. To illustrate this, one might wonder

if *find* and *locate* are synonymous in the 'Can do' statements (Alderson et al., 2006). This view is supported by Hai & Nhung (2019, p. 50), who mention that some teachers find the CEFR descriptors nonspecific, such as using terms "basic", "short", "simple", and "satisfactory" to describe language proficiency. Moreover, the use of a dictionary is not mentioned at all at lower levels even though learners at an early stage most likely use a dictionary (Alderson et al., 2006). Regarding language teaching, the authors of the CEFR were not explicit about the impact of CEFR on classroom practices (Westhoff, 2007).

In Finnish syllabi, the CEFR was used to create the level division for school purposes. The levels were validated for the previous curriculum by Hildén & Takala (2007). The relationship between the CEFR levels and the Finnish school scales is described in Appendix 1.

2.3 Aims of the Finnish general upper secondary education and the target levels in English and in Swedish

In general upper secondary education, the objective is to enhance transversal general knowledge. During the courses, students start to ponder their places in the world, and they construct their identity and understanding of humanity and worldview. Additionally, students start contemplating their future. The aim is also to promote lifelong learning (FNBE, 2015), which is one of the major skills required in the 21st century (Martin, 2018).

As mentioned in the introduction, the core curriculum emphasises the use of diverse assessment methods in teaching. This is also pointed out both in the previous General Upper Secondary Schools Act (629/1998) that was valid when the material was collected and in the current Act (714/2018). Put differently, teachers should use a variety of assessment methods in their teaching to assess students.

As illustrated in Table 2, after general upper secondary school, students are expected to reach a higher level in English than in Swedish (FNBE, 2015). This is hardly surprising since students usually start learning English in the third grade (A syllabus), whereas starting Swedish does not occur until the sixth grade (B syllabus). However, when the participants of this study started to study Swedish, they started it in the seventh grade, because in 2016, sixth graders started to study Swedish. In this study, the CEFR level was determined by an essay, thus, we are examining the level of text production skills in this article.

Table 2. Target levels at the final phase of general upper secondary education in Swedish and in English.

Language and syllabus	Interaction skills	Text interpretation skills	Text production skills
Swedish, B1	B1.1	B1.1	B1.1
English, A	B2.1	B2.1	B2.1

At the B1.1 level regarding text production skills, students are able to describe the main points and some details of topics that are relevant for their everyday life. Students are also able to use a broad vocabulary that even includes some of the more common idioms. At the B2.1 level, in contrast, students can express themselves about different topics using diverse structures and a broad vocabulary with idioms (FNBE, 2015).

3 Research questions

This article relates to a larger study examining feedback and assessment practices in Finnish general upper secondary schools. The major objective of this study is to investigate Finnish language teachers' assessment practices in general upper secondary education. More specifically, the research task is to investigate the perceptions of students about teachers' feedback practices. To address this task, the major research question is: How do the perceptions of general upper secondary school students about feedback differ at CEFR levels within Swedish and English? To illuminate this question, I will also explore the possible differences between the languages. In my study, feedback refers to teachers' comments about language use, learning process, and work, such as essays and portfolios. The comments refer to reactions to students' erroneous forms, comments on the content of students' essays, remarks on how students are studying in relation to their learning goals, and how fluently students speak the target language.

4 Data and methods

General upper secondary school students answered a web-based survey concerning their perceptions about teachers' feedback practices. The survey included background questions and ten sections with 57 closed-ended questions and one open-ended question. The background questions related to gender, previous course grade, how long students have studied the language, and what students' attitude to the language is. The survey was also used to gather the data for other studies. The survey focused on self-efficacy, correcting mistakes, teachers' formative assessment practices, amount of feedback students want, nature and content of teachers' assessment practices, students' individual needs in receiving feedback, examples of encouraging feedback, and metacognitive awareness. Table 3 displays the theme of the questions, the number of questions in the survey and the scale used.

Table 3. Content of the survey.

theme of the question(s)	number of question(s)	scale
self-efficacy	3	Likert 1-5
correcting mistakes	9	Likert 1-5
teachers' formative assessment practices	22	Likert 1-5
amount of feedback on different issues	1	Likert 1-5
nature and content of teachers' assessment practices	1	Likert 1-5
individual learning needs	3	Likert 1-5
examples of encouraging feedback	1	open-ended question
metacognitive awareness	18	Likert 1-5

To assess the students' CEFR levels, they wrote a short essay prior to answering the survey, and I determined the students' CEFR levels. In Swedish, the length of the essay was between 100 and 130 words, and in English the length was between 150 and 250 words. The topic of the essay was related to the content of the course, such as environment or culture. However, full CEFR exams include other tasks as well, but implementing such an extensive exam was beyond the scope of this study.

For the sake of reliability, every sixth essay from each school was also evaluated by an external rater. In English, 22 essays were evaluated by an external rater, and we agreed on the rating in 16 essays (73 per cent). In Swedish, 13 essays were evaluated by an external rater, and we agreed on 8 essays (62 per cent). Brown, Glasswell, & Harland (2004) found rater agreement of between 70 and 90 per cent in their studies on writing assessment. My results partially accord with those of Brown, Glasswell, & Harland, as the percentage in English was 73 and in Swedish it was 62. Regarding the essays that we did not agree on in English, I rated three essays higher than the other rater who, in turn, rated three essays higher than I. In Swedish, I rated three essays higher than the other rate who, consequently, rated two essays higher than I. This can be explained by individual differences in rater severity that have also occurred in previous studies (e.g. Deygers & Van Gorp, 2015; Holzknecht, Huhta, & Lamprianou, 2018), or even the order in which the essays were rated could have affected the judgement (Barkaoui, 2011; Daly & Dickson-Markman, 1982; Yen, 2016). The external raters were not the students' teachers. Therefore, "some of the motivation and institutional norms that appear in marking for a real exam may have been lacking", which means that no consequences of low reliability would arise (Barkaoui, 2011, p. 290). Another explanation for the lower agreement in Swedish is that familiarisation with the material, which the CEFR manual (Council of Europe, 2009) recommends, was not conducted in Swedish.

The data were collected between November 2018 and March 2019. Altogether 160 English students (57 males, 102 females, 1 did not indicate gender) and 97 Swedish students (35 males, 59 females, 1 did not indicate gender) answered the survey. The students were aged 17 to 19 years. All the participants answered with their own name because some of the participants were also interviewed for another study. The students gave me their written consent to participate in the study and allowing me to use their answers in my study. They also allowed me to combine their names with their answers. The data for this study comprises closed-ended questions with a Likert-type scale from one to five (1 = I strongly disagree, 5 = I strongly agree). As Pollari (2017a) also studied assessment in general upper secondary school language teaching, I used some of her survey questions in my own survey. The surveys for English and Swedish were identical. However, not all students wrote the essay, for some reason. The most common reason was that they were absent from school when the other students wrote the essay, and the teacher did not make them write it later. Also, only six students were at level A2 in English, which is not enough for statistical comparisons. Consequently, the answers from 134 English students and 86 Swedish students were analysed. Five general upper secondary schools participated in the study, three of which are in the Helsinki metropolitan area. Two other participating schools are located outside the metropolitan area, in the Häme and Satakunta regions. Moreover, I conducted a pilot study (N=13) prior to the actual study to ensure that the questions were clear and understandable.

The data were analysed quantitatively using ANOVA, MANOVA, and bivariate correlation (Spearman). If statistically significant differences were found between the groups, post-hoc tests (Tukey) were conducted to determine the levels between which the differences occur. The Mann-Whitney U and the Kruskall-Wallis tests were also used to confirm the analysed test results as not all the dependent variables were evenly distributed.

Based on factor analysis, six factors with Eigenvalues greater than 1 were formed. These variables were formed according different focuses in the survey.

Their items and factor loadings are presented in Appendix 2. Their names, abbreviations, number of items in each factor, and the Cronbach's alpha for each factor were as follows:

1. I want the teacher to correct all the mistakes (TC, 3 items, α = .77).
2. I want to correct all the mistakes myself (SC, 3 items, α = .87).
3. I would like to receive feedback on graded items (FGI, 4 items, α = .83).
4. I would like to receive feedback on the learning process (FLP, 4 items, α = .89).
5. Feedback has helped me in my studying (FIU, 3 items, α = .84).
6. I receive feedback in courses (GF, 5 items, α = .82)

5 Results

The descriptive statistics of the data are introduced in this section. First, the perceptions of general upper secondary school students of feedback practices by gender and the correlations between the variables are presented and discussed. Second, the perceptions by the CEFR levels in Swedish and in English are presented.

Table 4 shows the perceptions of students by gender. In Swedish, girls wanted more teacher correction, feedback on graded items, and feedback on the learning process. In contrast, boys wanted more self-correcting, and they also mentioned that they received more feedback from teachers. Regarding the fifth variable, if feedback has helped students in their studying, girls and boys had the exact same mean. In English, by contrast, the boys had a higher mean in one variable only, namely receiving feedback in courses. One statistically significant difference was found regarding teacher correction in English: girls were significantly more interested in it.

Table 4. Descriptive statistics of the perceptions of the students by gender.

Variables	Girls		Boys			All	
	M	S.D.	M	S.D.	p	M	S.D.
Swedish							
TC	4.35	0.73	4.31	0.81	.843	4.33	0.75
SC	3.56	1.14	3.82	1.02	.295	3.65	1.10
FGI	3.85	0.79	3.66	1.15	.361	3.78	0.93
FLP	2.76	0.99	2.40	1.09	.129	2.63	1.04
FIU	3.41	0.90	3.41	1.08	.999	3.41	0.96
GF	2.89	0.92	3.13	0.90	.247	2.98	0.91
English							
TC	4.32	0.62	3.98	0.98	.045*	4.21	0.77
SC	3.76	1.00	3.45	0.97	.234	3.66	0.99
FGI	3.78	0.78	3.47	0.84	.094	3.67	0.81
FLP	2.73	1.01	2.53	0.95	.134	2.65	1.00
FIU	3.54	0.88	3.33	0.91	.104	3.48	0.90
GF	2.73	0.81	2.89	0.72	.405	2.78	0.78

Note 1: M = Mean, S.D. = Standard deviation, * = p <.05
Note 2: TC = Teacher corrects all the mistakes, SC = Student corrects all the mistakes, FGI = Student would like to receive feedback on graded items, FLP = Student would like to receive feedback on learning process, FIU = Feedback has helped learning, GF = Student receives feedback from teachers.

A correlation analysis (Spearman) was conducted to examine the relationship between the variables. Several correlations are noteworthy in Table 5. In Swedish, the usefulness of feedback (FIU) was correlated with four other variables, whereas the variable on teacher correction (TC) was correlated with three other variables in English. In Swedish, all the variables were correlated with at least two other variables. In contrast, three variables (SC, FGI, and GF) were correlated with only one other variable in English. Moreover, the highest correlation in Swedish was between the willingness to correct one's mistakes (SC) and the usefulness of feedback (FIU), whereas the highest correlation in English was between feedback on graded items (FGI) and feedback on learning process (FLP).

Table 5. Correlations across the sum variables in Swedish and in English.

Swedish	TC	SC	FGI	FLP	FIU	GF
TC	1	.183	.205	.213*	.216*	.061
SC	.183	1	.186	.123	.527**	.221*
FGI	.205	.186	1	.373**	.261*	.074
FLP	.213*	.123	.373**	1	.122	-.002
FIU	.216*	.527**	.261*	.122	1	.409**
GF	.061	.221*	.074	-.002	.409**	1
English						
TC	1	.327**	.261**	.247**	.138	.013
SC	.327**	1	.084	.124	.083	-.028
FGI	.261**	.084	1	.510**	.116	.042
FLP	.247**	.124	.510**	1	.101	.156
FIU	.138	.083	.116	.101	1	.221*
GF	.013	-.028	.042	.156	.221*	1

*Correlation is significant at the 0.05 level (2-tailed).
** Correlation is significant at the 0.01 level (2-tailed).
Note 1: TC = Teacher corrects all the mistakes, SC = Student corrects all the mistakes, FGI = Student would like to receive feedback on graded items, FLP = Student would like to receive feedback on learning process, FIU = Feedback has helped learning, GF = Student receives feedback from teachers.

The research questions focus on the perceptions of general upper secondary school students about teachers' feedback practices in Swedish and English classes, and how the perceptions differ between the levels and the languages. The multivariate result was significant for proficiency level, Wilks' Λ = .636, F = 3.302, df = 2, p = .000, indicating a difference across the levels. Table 6 presents the means of the students at proficiency levels in Swedish and in English classes.

Table 6. The perceptions of students at different language proficiency levels.

	A1		A2		B1		B2		MANOVA	
Variables	M	S.D.	M	S.D.	M	S.D.	M	S.D.	F	p
Swedish										
TC	4.10	0.73	4.29	0.83	4.60	0.48			2.212	.116
SC	3.35	1.19	3.50	1.03	4.24	1.10			4.357	.016*
FGI	3.44	1.06	3.78	0.98	4.06	0.59			2.095	.129
FLP	2.55	1.00	2.65	1.07	2.65	1.02			.057	.944
FIU	2.46	1.02	3.45	0.79	4.05	0.70			17.278	.000*
GF	2.59	0.86	2.96	0.87	3.32	0.97			3.132	.049*

<u>English</u>

TC	4.16	0.70	4.28	0.88	.740	.391
SC	3.54	1.06	3.90	0.81	4.097	.045*
FGI	3.66	0.81	3.69	0.80	.034	.853
FLP	2.62	0.99	2.68	1.01	.096	.757
FIU	3.45	0.87	3.53	0.95	.163	.687
GF	2.77	0.79	2.79	0.78	.018	.892

Note 1: TC = Teacher corrects all the mistakes, SC = Student corrects all the mistakes, FGI = Student would like to receive feedback on graded items, FLP = Student would like to receive feedback on learning process, FIU = Feedback has helped learning, GF = Student receives feedback from teachers.

As portrayed in Table 6, it seems that in Swedish classes, the means increase progressively from the lowest level to the second level to the highest level. In other words, students at higher levels wanted their teachers to correct all the mistakes, were more open to correcting their own mistakes, wanted to receive feedback on items that are graded, felt that teachers' feedback is beneficial for them, and felt that they receive feedback from Swedish teachers. However, the only minor exception is feedback on the learning process: students at the lowest level had the lowest mean, whereas students at levels A2 and B1 had the same mean.

Regarding English, students at the higher level (B2) always had the highest mean. However, there was only one sum variable for which there was a statistically significant difference, namely about correcting one's own mistakes. The p value was .045, which is quite near the boundary of statistically significant difference. However, Levene's test was significant for this sum variable (p=0.032), and this variable was not evenly distributed. Accordingly, the Mann-Whitney Test was conducted to compare levels B1 and B2. The test did not indicate a statistically significant difference (p=0.079).

As not all the variables were evenly distributed, the Kruskall-Wallis Test and the Mann-Whitney Test were run to compare the differences between the groups. These tests confirm the results from MANOVA and indicated the same sum variables in which statistically significant differences occurred.

As demonstrated in Table 5, statistically significant differences were found in three sum variables regarding Swedish: students receive feedback in courses, correcting one's own mistakes, and feedback has helped. Correlation tests (Spearman) were conducted to determine the correlations between the sum variables and the proficiency levels. Using the guide that Geher & Hall (2014) suggest, a weak correlation was found for the first sum variable (0.242). However, a moderate correlation was found between proficiency level and correcting one's own mistakes (0.303) and between proficiency level and the usefulness of feedback (0.509). However, caution must be applied as Geher & Hall (2014, p. 78) emphasise that: "Correlation does not imply causation." For the students studying English, the correlation analysis was not run for the only sum variable where a statistically significant difference occurred, due to the result in the Mann-Whitney Test.

As MANOVA showed that there were statistically significant differences between the CEFR levels in three sum variables in Swedish, post-hoc tests were run to determine the levels between which the differences occur. Tukey's test revealed that differences occurred regarding correcting one's own mistakes

between levels A1 and B1 (p=0.037), and A2 and B1 (p=0.024). Then, differences about the usefulness of feedback occurred between levels A1 and A2 (p=0.000), A1 and B1 (p=0.000), and A2 and B1 (p=0.017). The last sum variable focused on whether students receive feedback from teachers. Differences occurred between levels A1 and B1 (p=0.039).

6 Discussion and conclusions

One of the more striking observations to emerge from the data comparison is that Swedish students seem to receive more feedback from their teachers compared to English students. The means of the English students at levels B1 and B2 were 2.77 and 2.79 respectively, thus, almost the same. In comparison, the means of the Swedish students at levels A1, A2, and B1 were 2.59, 2.96, and 3.32 respectively, and the differences were statistically significant (p=0.049). One explanation is the general proficiency of students in these languages: in general, the level of Swedish is rather low. In 2017, approximately 10 per cent of the boys and 20 per cent of the girls in Swedish had attained the expected level B1 after general upper secondary school. Put differently, most of the students are either at levels A1 or A2 (Juurakko-Paavola & Åberg, 2018). Another study explored the results of the matriculation examination in spring 2012 and discovered that 25-30 per cent of the students in Swedish achieved level B1.1, whereas nearly 70 per cent of the students in English achieved level B2.2 (Juurakko-Paavola & Takala, 2013). Regarding basic education, the means of 13,500 students after the ninth grade were studied, and the mean for the girls in English was 8.15 and for the boys it was 7.93 (on a scale from four to ten). By contrast, the mean for girls in Swedish was 7.90 and that for boys was 7.05 (Pöysä & Kupiainen, 2018). In other words, the grades were better in English than in Swedish. Regarding the matriculation examination, in spring and autumn 2019, the lowest point for obtaining the highest grade in English was 267 out of 299. In contrast, the points in Swedish (B syllabus) were 237 and 243, respectively. Moreover, to pass the English test, students needed 122 points in the spring and 141 in the autumn. To pass the Swedish test, in contrast, students needed 90 points in the spring and in the autumn (YTL, 2020a). Taken together, students in general possess better skills in English than in Swedish, which is noticeable in the means of the grades and in the amount of points needed to pass the test in the matriculation examination or to get the best grade in the test. Therefore, it is possible that Swedish teachers try to encourage their students more with feedback due to their lesser language skills.

Regarding the number of students in the matriculation examination, 18,627 students sat for the Swedish test in 2011, whereas in 2019, the number of students was 14,614, approximately 4,000 students fewer. In contrast, the number of students participating in the English test in 2011 was 39,219, and the number in 2019 was 39,324, approximately the same (YTL, 2020b). Taken together, these statistics show that the number of students participating in the Swedish test has decreased over the course of eight years, whereas there has been no change in the English test. As students' exhibit somewhat negative perceptions of Swedish (Pitkänen & Westinen, 2017), it can be speculated that this also affects the number of students willing to take the Swedish test. Regarding motivation, feedback is one of key aspects of enhancing motivation

(Dörnyei, 2001). Hence, it could conceivably be hypothesised that Swedish teachers emphasize the use of feedback to engage their students more with Swedish and try to enhance students' motivation.

As mentioned in the introduction, students value English as a school subject more than Swedish. This also seems to apply to spare time, as students use more English in their spare time than Swedish (Härmälä, Hildén, & Leontjev, 2016; Luukka et al., 2008). Interestingly, the use of English outside the school context has a moderate impact on English skills (Härmälä, Huhtanen, & Puukko, 2014), whereas the impact on Swedish is extremely weak (Härmälä, Hildén, & Leontjev, 2016). As these studies indicate, students do not use Swedish as much as English outside school, and even if some students do use it, it does not seem to enhance their Swedish skills particularly. Regarding my results, it could therefore be argued that students rely more on teachers' feedback in Swedish than in English, and as a result, Swedish teachers give more feedback than English teachers.

It is apparent from the data that students appreciate teachers' feedback. This finding is consistent with the findings of past studies, including those by Lee (2008), Pollari (2017b), and Yang, Badger, & Yu (2006) to name but a few. Moreover, it seems that students at lower proficiency levels are less interested in receiving feedback in the form of error correcting compared to students at higher levels, albeit the differences are not statistically significant. This reflects the findings of Lee (2008). In general, students seem to want to be corrected by the teacher. This finding is in agreement with Jean & Simard (2011) and Schulz (2001) who found that students preferred to be corrected all the time.

Based on the results, girls prefer error correction from teachers more than boys. One explanation for this is that boys' perceptions of their English skills are somewhat more positive than those of girls (Härmälä, Huhtanen, & Puukko, 2014). Therefore, girls could be more insecure of their skills and want more error correction from teachers. Moreover, Pöysä & Kupiainen (2018) and Kupiainen, Marjanen, & Oukarim-Soivio (2018) raise the issue of the matriculation examination and point out that boys outperform girls in the English test. Indeed, statistics show that 494 boys received the highest grade in the English test in spring 2019, whereas 458 girls received it. Similarly, in autumn, the highest grade was obtained by 522 boys and 405 girls. In contrast, 741 girls and 324 boys failed the test in autumn, and 581 boys and 270 girls failed it in spring (YTL, 2020c). Thus, it can be argued that girls are not as skilful in English as boys and need error correction more.

An interesting finding regarding Swedish is that students at the lowest level feel that teachers' feedback has not helped them (mean 2.46 at level A1), whereas students at the highest level (mean 4.05 at level B1) have the opposite opinion. This may be explained by the nature of feedback: Dinnen & Collopy (2009) found that teachers tend to give positive feedback to strong students and negative feedback to weak students. Compared to positive feedback, negative feedback decreases motivation (Fong, Patall, Vasquez, & Stautberg, 2019). As a result, weak students in Swedish do not necessarily want to use teachers' feedback at all due to its negative nature. However, this must be interpreted with the utmost caution as the results do not reveal whether students have perceived teachers' feedback to be either positive or negative. Another plausible explanation is that Swedish teachers have not been able to target the feedback appropriately with weaker students, as feedback targeted inappropriately is ineffective (Hattie 2012b; Hattie & Timperley, 2007). It is also possible that the

students at level A1 do not perceive Swedish teachers' feedback as being important, as students seem to pay attention to feedback that they consider useful (Leki, 1991). For making students understand the value of feedback, the teacher could explain his or her approach to correcting mistakes because this would increase the usefulness of feedback for the students (Plonsky & Mills, 2006).

One of the more striking observations to emerge from the data comparison are the differences between Swedish and English at the B1 level: one sees immediately in Table 5 that the means in Swedish are always higher, albeit the difference in feedback on the learning process is minimal. It is surprising, for instance, that they have a more positive attitude towards correcting one's own mistakes. However, among the plausible explanations for this finding are the expected proficiency levels after general upper secondary education in these languages. The target level in Swedish is B1.1, whereas the target level in English is B2.1. In other words, the students of this study at level B1 in Swedish are good students as they have reached the target level, whereas English students at level B1 have not yet reached the target level. Therefore, Swedish students might be more willing to correct their mistakes. Interestingly, if we then compare levels B1 in Swedish and B2 in English, which are the target levels in both languages, Swedish still usually has the highest mean, except in feedback on the learning process. Nevertheless, the difference is extremely small. Put differently, Swedish students have received more feedback from teachers, feedback has helped them more, they are more willing to correct their mistakes, they want more feedback on graded items, and they want more error correction from the teacher. It can be hypothesised that the reason for this lies in the discussion above, regarding why students rely more on feedback in Swedish than in English.

Regarding feedback on the learning process itself, the data show that neither Swedish nor English students especially crave it. Nevertheless, this is an alarming result because the learning process is a vital component of learning, and feedback should also be targeted at students' learning (Brookhart, 2012). The analysis showed that students want feedback on graded items, such as exams, written compositions, and oral skills. Also, Marie (2016) found that students appreciate feedback on exams. Nevertheless, as the National Core Curriculum (FNBE, 2015) emphasises life-long learning and self-regulated learning, it is of the utmost importance to give feedback on the learning process itself. The data, however, clearly show that students want feedback on graded items more than feedback on the learning process. Nevertheless, some students are also interested in the learning process, as Pollari (2017b) found that English students want more feedback on study skills. Thus, I argue that the importance of the learning process itself should be emphasised more in language teaching so that it becomes apparent to students that the role of the learning process in learning is significant.

To sum up, this project was undertaken to investigate the perceptions of general upper secondary school students of feedback in language teaching. The present study makes several noteworthy contributions to the current literature. First, feedback has helped Swedish students more, and they also receive more feedback than English students. Second, the opinions of the Swedish students are mixed, whereas the opinions of the English students are more uniform. Third,

students are more interested in feedback on graded items than feedback on the learning process.

The current study reveals only the tip of iceberg about perceptions of feedback at different CEFR levels. Considerably more work will need to be done in this area. First, the sample size should be larger, and familiarisation with the material should be undertaken prior to assessing the levels. Second, peer feedback and self-assessment were excluded from this study. Therefore, more research is needed to assess their connections to the CEFR levels. Third, the results show a moderate correlation between proficiency level and correcting one's own mistakes and the usefulness of feedback in Swedish. Accordingly, more sophisticated analyses could be used to determine the factors affecting the usefulness of feedback, and whether gender, previous course grade, school, or age have any effect on it.

The findings from this study have important implications for future practice. Regarding Swedish, weaker students have not received feedback as much as stronger students, and feedback has been more useful for stronger students than for weaker students. Compared to Swedish students, English students have received less feedback, and feedback has been less useful. Therefore, both Swedish and English teachers should reflect critically on their feedback practices and ponder whether the amount of feedback they give to students is enough, and whether the feedback is differentiated according to students' needs and proficiency.

Acknowledgements

References

Alderson, J. C., Figueras, N., Kuijper, H., Nold, G., Takala, S., & Tardieu, C. (2006). Analysing tests of reading and listening in relation to the Common European Framework of Reference: The experience of the Dutch CEFR construct project. *Language Assessment Quarterly*, 3(1), 3–30.

Ammar, A., & Spada, N. (2006). One size fits all?: Recasts, prompts, and L2 learning. *Studies in Second Language Acquisition*, 28(04), 543–574.

Atjonen, P., Laivamaa, H., Levonen, A., Orell, S., Saari, M., Sulonen, K., Tamm, M., Kamppi, P., Rumpu, N., Hietala, R., & Immonen, J. (2019).*"ETTÄ TIETÄÄ MISSÄ ON MENOSSA" Oppimisen ja osaamisen arviointi perusopetuksessa ja lukiokoulutuksessa* ["So that we know where we stand" Assessment of learning and competence in basic education and general upper secondary education]. Publications of the Finnish Education Evaluation Centre (FINEEC), Finland, 2019:7.

Barkaoui, K. (2011). Effects of marking method and rater experience on ESL essay scores and rater performance. *Assessment in Education: Principles, Policy & Practice*, 18(3), 279–293.

Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25.

Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of Second Language Writing*, 17(2), 102–118.

Black, P. J. (2002). *Testing, friend or foe? : The theory and practice of assessment and testing*. London; Washington: Falmer Press.

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–148.

Brookhart, S. M. (2012). Preventing feedback fizzle. *Educational Leadership*, 70(1), 24–29.

Brookhart, S. M. (2017). *How to give effective feedback to your students*. Alexandria, Va.: ASCD.

Brown, G. T. L., Glasswell, K., & Harland, D. (2004). Accuracy in the scoring of writing: Studies of reliability and validity using a New Zealand writing assessment system. *Assessing Writing*, 9(2), 105–121.

Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In H. L. Andrade, & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 3–17.) New York, NY: Routledge.

Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Council of Europe. (2009). *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): A Manual*. Strasbourgh: Lanuage Policy Division.

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, 58(4), 438–481.

Daly, J. A., & Dickson-Markman, F. (1982). Contrast effects in evaluating essays. *Journal of Educational Measurement*, 19(4), 309–316

Dann, R. (2014). Assessment as learning: Blurring the boundaries of assessment and learning for theory policy, and practice. *Assessment in Education: Principles, Policy & Practice*, 21(2), 149–166.

Deygers, B., & Van Gorp, K. (2015). Determining the scoring validity of a co-constructed CEFR-based rating scale. *Language Testing*, 32(4), 521–541.

Dinnen, J. L. D., & Collopy, R. M. B. (2009). An analysis of feedback given to strong and weak student writers. *Reading Horizons*, 49(3), 239–256.

Dörnyei, Z. (2001). *Motivational strategies in the language classroom.* Cambridge: University Press.

Ellis, R. (2009). A typology of written corrective feedback types. *ELT Journal*, 63(2), 97–107.

Fathman, A., & Whalley, E. (1990). Teacher response to student writing: Focus on form versus content. In B. Kroll (Eds.), *Second language writing: Research insights for the classroom* (pp. 178–190). Cambridge: Cambridge University Press.

Fazio, L. (2001). The effect of corrections and commentaries on the journal writing accuracy of minority- and majority-language students. *Journal of Second Language Writing*, 10(4), 235–249.

Ferris, D. (1999). The case for grammar correction in L2 writing classes: A response to Truscott (1996). *Journal of Second Language Writing*, 8(l), l–l 1.

Ferris, D. R., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing*, 10(3), 161–184.

Figueras, N. (2012). The impact of the CEFR. *ELT Journal*, 66(4), 477–485.

FNBE. *Lukion opetussuunnitelman perusteet 2015* [The National Core Curriculum for General Upper Secondary Schools 2015]. (2015). Helsinki: Finnish National Agency for Education.

Fong, C., Patall, E., Vasquez, A., & Stautberg, S. (2019). A meta-analysis of negative feedback on intrinsic motivation. *Educational Psychology Review*, 31(1), 121–162.

Geher, G., & Hall, S. (2014). *Straightforward statistics: Understanding the tools of research.* Oxford: University Press.

General Upper Secondary Schools Act, 629/1998. https://www.finlex.fi/fi/laki/ajantasa/kumotut/1998/19980629

General Upper Secondary Schools Act, 714/2018. https://www.finlex.fi/fi/laki/ajantasa/2018/20180714

Hai, L. T. T., & Nhung, P. T. H. (2019). Implementing the CEFR at a Vietnamese university – General English language teachers' perceptions. *CEFR Journal – Research and Practice*, 1, 41–59.

Harlen, W. (2007). *Assessment of learning*. London: SAGE.

Hattie, J. (2012a). Know thy impact. *Educational Leadership*, 70(1), 18–23.

Hattie, J. (2012b). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Oxford: Routledge.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.

Hildén, R., Härmälä, M., Rautopuro, J., Huhtanen, M., Puukko, M., & Silverström, C. (2015). *Outcomes of language learning at the end of basic education in 2013*. Helsinki:Finnish Education Evaluation Centre; Finnish National Board of Education.

Hildén, R., & Rautopuro, J. (2014). *Ruotsin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013* [Outcomes of language learning in A level Swedish at the end of basic education in 2013]. Helsinki: Finnish Education Evaluation Centre; Finnish National Board of Education.

Hilden, R., & Takala, S. (2007). Relating descriptors of the Finnish school scale to the CEF overall scales for communicative activities. In A. Koskensalo, J. Smeds, P. Kaikkonen, & V. Kohonen (Eds.), *Foreign languages and multicultural perspectives in the European conte*xt = Fremdsprachen und multikulturelle Perspektiven im europäischen Kontext . Dichtung - Wahrheit - Sprache , vol. Band 9-10 , Lit Verlag, Berlin , pp. 291–300.

Holzknecht, F., Huhta, A., & Lamprianou, I. (2018). Comparing the outcomes of two different approaches to CEFR-based rating of students' writing performances across two European countries. *Assessing Writing*, 37, 57–67.

Hulstijn, J. H. (2007). The Shaky Ground beneath the CEFR: Quantitative and qualitative dimensions of language proficiency. *The Modern Language Journal*, 91(4), 663–667.

Härmälä, M., Hildén, R., Leontjev, D. (2016). Yhdeksäsluokkalaisten opiskelukäytänteet englannin, ruotsin ja ranskan oppitunneilla ja vapaa-ajalla [The study practices of 9th graders in English, Swedish, and French lessons and spare time]. In A. Huhta, & R. Hildén (Eds.), *Kielitaidon arviointitutkimus 2000-luvun Suomessa* [Research on assessment of language proficiency in Finland in the 21st century]. AFinLA-e. Soveltavan kielitieteen tutkimuksia 2016:9, 142–166.

Härmälä, M., Huhtanen, M., & Puukko, M. (2014). *Englannin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013* [Outcomes of language learning in A level English at the end of basic education in 2013]. Helsinki: Finnish Education Evaluation Centre.

Härmälä, M., Huhtanen, M., Puukko, M., & Marjanen, J. (2019). *A-englannin oppimistulokset 7. vuosiluokan alussa 2018* [Outcomes of language learning in A level English at the beginning of the 7th grade in 2018]. Helsinki: Finnish Education Evaluation Centre.

Jean, G., & Simard, D. (2011). Grammar teaching and learning in L2: Necessary but boring? *Foreign Language Annals*, 44(3), 467–494.

Juurakko-Paavola, T., & Takala, S. (2013). *Ylioppilastutkinnon kielikokeiden tulosten sijoittaminen lukion opetussuunnitelman perusteiden taitotasoille* [Placing the results of the language tests of the matriculation examination into the CEFR levels indicated by the national curriculum for general upper secondary education]. Helsinki: Ylioppilastutkintolautakunta.

Juurakko-Paavola, T., & Åberg, A.-M. (2018). Ruotsin kielen osaamisvaatimuksista vapauttaminen korkeakouluissa [Liberalising the requirements in Swedish in higher education]. *Kieli, koulutus ja yhteiskunta*, 9(1).

Kupiainen, S., Marjanen, J., & Ouakrim-Soivio, N. (2018). *Ylioppilas valintojen pyörteissä. Lukio-opinnot, ylioppilastutkinto ja korkeakoulujen opiskelijavalinta* [Student in-between choices. General upper secondary studies, the Matriculation Examination and choice of students for higher education]. Suomen ainedidaktisen tutkimusseuran julkaisuja: Ainedidaktisia tutkimuksia 14.

Leahy, S., & Wiliam, D. (2012). From teachers to schools: Scaling up professional development for formative assessment. In J. Gardner (Eds.), *Assessment and Learning* (2nd ed., pp. 49–71). London: SAGE.

Lee, I. (2008). Student reactions to teacher feedback in two Hong Kong secondary classrooms. *Journal of Second Language Writing*, 17(3), 144–164.

Leki, I. (1991). The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals*, 24(3), 203–218.

Leontjev, D. (2014). The effect of automated adaptive corrective feedback: L2 English questions. *Apples – Journal of Applied Language Studies*, 8(2), 43–66.

Leppänen, S., Pitkänen-Huhta, A., Nikula, T., Kytölä, S., Törmäkangas, T., Nissinen, K., Kääntä, L., Räisänen, T., Laitinen, M., Koskela, H., Lähdesmäki, S., & Jousmäki, H. (2011). *National survey on the English language in Finland: Uses, meanings and attitudes.* Jyväskylä: University of Jyväskylä.

Li, S. (2010). The effectiveness of corrective feedback in SLA: A meta-analysis. *Language Learning*, 60(2), 309–365.

Little, D. (2007). The Common European Framework of References for Languages: Perspectives on the making of supranational language education policy, The *Modern Language Journal*, 91(4), 645–655.

Luukka, M.-R., Pöyhönen, S., Huhta, A., Taalas, P., Tarnanen, M., & Keränen, A. (2008). *Maailma muuttuu – mitä tekee koulu? Äidinkielen ja vieraiden kielten tekstikäytänteet koulussa ja vapaa-ajalla* [The world is changing – what does the school do? Text practices in mother tongue and foreign languages in school and in spare time]. Jyväskylän yliopisto, soveltavan kielentutkimuksen keskus.

Lyster, R., & Ranta, L. (1997). Corrective feedback and learner uptake. *Studies in Second Language Acquisition*, 19(01), 37–66.

Marie, J. A. (2016). Student views on the value of feedback. *Journal of Education and Training Studies*, 4(6), 207–213.

Martin, J. P. (2018). *Skills for the 21st century: Findings and policy lessons from the OECD survey of adult skills*. OECD Education Working Paper, 166(2), 1–33.

Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education: Principles, Policy & Practice*, 14(2), 149–170.

OECD. (2005). *Formative Assessment. Improving learning in secondary classroom.* Paris: OECD.

Pitkänen, V., & Westinen, J. (2017). *Puhut sie ruotsia? Tutkimus suomenkielisten asenteista ruotsin kieleen ja ruotsinkielisiin* [Do you speak Swedish? Study of Finns' attitudes to the Swedish language and Swedish speaking people]. e2 Tutkimus.

Plonsky, L., & Mills, S. V. (2006). An exploratory study of differing perceptions of error correction between a teacher and students: Bridging the gap. *Applied Language Learning*, 16(1), 55–74.

Pollari, P. (2017a). (Dis)empowering assessment? Assessment as experienced by students in their upper secondary school EFL studies. Doctoral Dissertation. University of Jyväskylä, Jyväskylä, Finland.

Pollari, P. (2017b). To feed back or to feed forward? Students' experiences of and responses to feedback in a Finnish EFL classroom. *Apples – Journal of Applied Language Studies*, 11(4), 11–33.

Pyykkö, R. (2017). *Monikielisyys vahvuudeksi: Selvitys Suomen kielivarannon tilasta ja tasosta* (Multilingualism into a strength. A report of the status and levels of language competences in Finland]. Publications of the Ministry of Education and Culture, Finland, 2017:51.

Pöysä, S., & Kupiainen, S. (2018). Tyttöjen ja poikien väliset erot kansainvälisen ja kansallisen arviointitutkimuksen valossa [The differences between boys and girls in light of international and national research on assessment]. In S. Pöysä, & S. Kupiainen (Eds.), *Tytöt ja pojat koulussa. Miten selättää poikien heikko suoriutuminen peruskoulussa?* [Girls and boys in school. How to overcome boys' shortcomings at comprehensive school?]. Valtioneuvoston selvitys- ja tutkimustoimikunnan julkaisusarja 36/2018.

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119–144.

Schulz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar instruction and corrective feedback: USA-Colombia. *Modern Language Journal*, 85(2), 244–258.

Sheen, Y. (2011). *Corrective feedback, individual differences, and second language learning.* Dordrecht: Springer.

Stake, R. E. (2004). *Standards-based & responsive evaluation*. Thousand Oaks, CA: Sage.

Takala, S. (2010). Putting the CEFR to good use: Activities and outcomes in Finland. In J. Mader, & Z. Urkun (Eds.), *Putting the CEFR to good use, Selected articles by the presenters of the IATEFL Testing, Evaluation, and Assessment Special Interest Group (TEA SIG) and EALTA Conference in Barcelona, Spain, 20–30 October 2010*, 96–105.

Taylor, L. (2009). Developing assessment literacy. *Annual Review of Applied Linguistics*, 29, 21–36.

Tomlinson, C. A., & Moon, T. R. (2013). *Assessment and student success in a differentiated classroom*. Alexandria, Virginia.: ASCD.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46 (2), 327–369.

Truscott, J. (1999a). The case for ''The case against grammar correction in L2 writing classes'': A Response to Ferris. *Journal of Second Language Writing*, 8(2), 111–122.

Truscott, J. (1999b). What's wrong with oral grammar correction. Canadian Modern Language Review, 55(4), 437–456.

Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16(4), 255–272.

Vogt, K., & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, 11(4), 374–402.

Westhoff, G. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *Modern Language Journal*, 91(4), 676–679.

Wiggins, G. (2012). Seven keys to effective feedback. *Educational Leadership*, 70(1), 10–16.

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.

Yen, N. T. Q. (2016). Rater consistency in rating L2 learners' writing task. *Journal of Science: Foreign Languages*, 32(2), 75–84.

YTL. (2020a, April 27). *Pisterajat* 2008–2019 [The number of points needed for each grade 2008–2019]. Retrieved from https://www.ylioppilastutkinto.fi/ext/stat/FS2019A2008T3010.pdf

YTL. (2020b, April 27). *Ilmoittautuneet eri kokeisiin tutkintokerroittain 2011–2020* [The number of students enrolling in the tests in each examination 2011–2020]. Retrieved from https://www.ylioppilastutkinto.fi/ext/stat/FS2020A2011T2010.pdf

YTL. (2020c, April 27). *Arvosanajakaumat 2017–2019* [The distribution of grades 2017–2019]. Retrieved from https://www.ylioppilastutkinto.fi/ext/stat/FS2019A2017T4002.pdf

Appendices

Appendix 1. The relationship between the CEFR levels and the Finnish school scale (Hildén & Takala, 2007).

CEFR level	Finnish school scale
A1 Breakthrough	A1.1 First stage of elementary proficiency
	A1.2 Developing elementary proficiency
	A1.3 Functional elementary proficiency
A2 Waystage	A2.1 First stage of basic proficiency
A2+ Strong Waystage	A2.2 Developing basic proficiency
B1 Treshold *[sic]*	B1.1 Functional basic proficiency
B1+ Strong Treshold *[sic]*	B1.2 Fluent basic proficiency
B2 Vantage	B2.1 First stage of independent proficiency
B2+ Strong Vantage	B2.2 Functional independent proficiency

Appendix 2: The sum variables.

The sum variables, their items, and their loadings

1. I want the teacher to correct all the mistakes (α = .77).

- I want the teacher to correct the grammar mistakes I make when I speak English/Swedish. .933
- I want the teacher to correct the pronunciation mistakes I make when I speak English/Swedish. .792
- I want the teacher to correct the mistakes I make when I write essays or do other written tasks in English/Swedish. .470

2. I want to correct all the mistakes myself (α = .87).
- I want to correct myself the grammar mistakes I make with the teachers' feedback when I speak English/Swedish. .905
- I want to correct myself the pronunciation mistakes I make with the teachers' feedback when I speak English/Swedish. .879
- I want to correct myself the mistakes I make with teachers' feedback when I write essays or do other written tasks in English/Swedish. .743

3. I would like to receive feedback on graded items (items, α = .83).
- I would like to receive feedback on essays. .957
- I would like to receive feedback on exams. .828
- I would like to receive feedback on written tasks. .713
- I would like to receive feedback on word tests. .448

4. I would like to receive feedback on the learning process (α = .89).
- I would like to receive feedback on my study skills. 1.015
- I would like to receive feedback on the learning process. .858
- I would like to receive feedback on how I work regarding the learning goals. .792
- I would like to receive feedback on my work during the lessons. .578

5. Feedback has helped me in my studying (α = .84).
- Teachers' feedback has helped me enhance my language skills. .869
- Teachers' feedback has helped and guided my studies. .806
- Teachers' feedback has motivated me. .716

6. I receive feedback in courses (α = .82).
- I regularly receive feedback on my studies. .781
- The teacher writes enough feedback, for instance at the end of an essay. .758
- I receive enough feedback on my learning during the courses so that I can affect my studying during the course. .648
- I receive personal feedback on my studies from the teacher. .641
- When I am working, the teacher gives me feedback on how I can improve my work. .630