Finnish Compulsory School English and Swedish Teachers' Classroom Assessment Practices

Marita Härmälä

Finnish Education Evaluation Centre, Finland

Abstract

The study explores Finnish compulsory school language teachers' formative and summative assessment practices. It aims at finding out what formative and summative assessment modes and tools the teachers use and what kind of evidence they collect for students' final grades. The data were gathered through questionnaires administered in national evaluations of learning outcomes. The results indicate that teachers use multiple tools, i.e., scores, information on a task's objective, oral feedback, and model performances. They listen to students paired discussions and talk with students about their progress. Students also assess their own skills. Teachers' feedback on exams was information on a task's objective, oral feedback, and model performances. When deciding on final grades, teachers use as evidence written tests and students' working in lessons. However, there were some differences between languages and syllabi. Teachers gave the final grades based on students' language proficiency at a given moment stressing the curricula's content areas related to language proficiency. At the same time, teachers indicated to be basing their grades on averages in all the different content areas. The study has implications for teachers' assessment literary training and for exploring more in detail the factors impacting students' final grades.

Keywords: *Classroom Assessment, Compulsory Education, English and Swedish, Assessment Practice*

Introduction

Skills in assessment are an important part of a teacher's professional competence. By teachers' assessment literacy was first referred to "the ability to understand, analyze and apply information on student performance to improve instruction" (Stiggins, 1991) and thereby the quality of instruction. For a long time, teacher assessment referred above all to summative assessment at the end of a program or a course. Today, the scope of teacher assessment has expanded to encompass various kinds of skills, tools and methods needed in classroom

assessment. To better capture the multi-faceted and versatile content of teacher assessment competence, the concept of teacher assessment literacy has been introduced (Tsagari & Vogt, 2017).

In teacher education, the amount and content of training in teacher assessment literacy varies but very often, even the practicing teachers feel that they would need more skills and knowledge on issues around assessment. More training would be needed in the alternatives in assessment methods (Brown & Hudson, 1998) but also in learning the principles of valid and reliable assessment practices.

Student assessment has lately raised lots of discussion in Finland since a new chapter on assessment was added to the Finnish National Core Curriculum (hence NCC) for basic education (FNAE, 2020). There are two major revisions in the new text. The first is that worded criteria for the marks 5, 7, 8 and 9 (on a scale from 4 to 10) are now defined and described for each subject. Second, the new text makes a clear distinction between summative and formative assessment. The aim of the renewal was to stress the use of multiple techniques in student assessment (see Brown & Hudson, 1998) and to unify the assessment practices in different schools, between teachers and across different subjects. In several evaluations of learning outcomes (e.g., Härmälä & Marjanen, 2022; Härmälä & Marjanen, 2023), it has been found that there is considerable variation between schools and students in the final grades. This means that in some schools, the requirements for different grades are higher than in others. This variability increases inequity between students when they apply for further studies (Hildén et al., 2016).

The article explores classroom assessment practices in Swedish and English lessons in Finnish compulsory education. The aim is to investigate what kinds of classroom assessment practices the teachers use, how often they use them and if there are differences between languages and syllabi. There are few studies on language teachers' classroom assessment practices (Saeed et al., 2018) and they are especially scarce in Finnish compulsory education. More research is needed i.e., on the procedures the teachers use when deciding on students' final grades as well as on what kinds of formative assessment practices the teachers use in language classroom to promote learning.

The article starts with a brief review of some basic considerations on classroom assessment inspired by the pivotal work of James Dean Brown, which I then illustrate with examples taken from the Finnish national evaluations of learning outcomes. After presenting the results of the study, I discuss the implications and suggest topics for further studies.

Literature Review

There are two families of testing (Brown, 2012): standardized assessment or norm-referenced testing and classroom assessment or criterion-referenced testing. In standardized assessment, the content tested is very general, and the examinees abilities are interpreted in relation to all other examinees to be able to make i.e., admission or placement decisions. In contrast, classroom assessment is very specific as it focuses on what has been taught during a course or a program and it is mainly used by teachers to determine what the students have learnt and what still needs to be practiced.

Classroom assessment has three main functions depending on what kind of information on student performance the teacher wants to collect and for what purpose. A distinction between summative and formative assessment was first drawn by Michael Scriven in late 60s (Scriven, 1967). According to Scriven, evaluation has two roles: the formative, where the evaluator helps to improve the program, and the summative where the evaluator is determining the worth of the program. There is also a third type of classroom assessment, that is, diagnostic assessment, which aims at providing detailed information on students' learning difficulties and on how to overcome them (Saeed et al., 2018).

Student assessment is often conceptualized through its three main functions: assessment *of* learning, *for* learning and *as* learning (Earl, 2013). Assessment of learning refers to summative assessment, which has as its aim to gather evidence of a student's learning against a predefined standard or objective, i.e., the curriculum objectives. The results of summative assessment may be used for various purposes like certification and admission. Summative assessments are therefore high stakes for the students as the decisions made on basis of them may affect their future lives, i.e., access to further studies.

Assessment for/as learning may be seen as synonymous to formative assessment. Their main function is to give students feedback that enables them to learn more efficiently. This diagnostic information helps the teachers to focus the teaching and to improve teaching materials and learning tasks when needed. Assessment as learning occurs when students assess their own learning. Assessment as learning is a very efficient way of involving students directly in the assessment process (Brown & Hudson, 1998).

In sum, summative assessment refers to tests or exams whereas formative assessment, or formative feedback, is kind of an activity or several activities that the teacher does in classroom continuously and sometimes not even explicitly noticing it. As Brown puts it:

> assessment *activities* are different from tests in that they are not easily distinguishable from other classroom activities because they are thoroughly integrated into the language teaching and learning processes… [they] do not stand alone as different, threatening, or interruptive…. [and they] are different from ordinary classroom activities in that they provide a way of … giving *feedback*. (Brown, 2013, p. 334)

Feedback is according to Brown (2019) "one of the teacher's most powerful tools for shaping how students approach the learning process."

There are various modes and tools for giving feedback (see more in Brown, 2019). For example, feedback may be given in writing or orally and it is delivered usually by teacher, peer, group of students or by a student her/himself. Teacher feedback may take many forms such as the form of a score, notes in the margin and oral comments and focus on various matters such as grammar and vocabulary accuracy, fluency, task completion and getting meaning across (Brown, 2013). It is important that the feedback is given often, immediately, it is clear and constructive and linked to the objectives of the course or the curriculum. To avoid is to focus the feedback entirely on language details like mistakes of grammar (Brown, 2019) and use such methods that do not give teachers extra work in addition to what they are already doing (Brown, 2009). It is also important to use multiple sources of information as every method has its

strengths and weaknesses and only one method does not fit for all purposes (Brown & Hudson, 1998).

For the feedback to be encouraging it needs to focus on the content, i.e., how the student may improve his/her own work (Mäkipää & Hildén, 2021). Mäkipää and Hildén (2021) found that the Finnish general upper secondary school students (n = 160) did not perceive teacher feedback in foreign languages as an intrinsic part of teacher assessment practices. Teacher feedback was appreciated but it should not be on the learning process. The impact that the feedback has on learning, depends on how the learner responds to it (Pollari, 2017). Although students of the Pollari 2017 study were primarily content with their feedback, they wanted more guiding feedback, i.e., more feed forward. Students also wanted more personalised feedback as well as feedback that takes place during the learning process, and not only after it. Pollari (2017) also found that feedback should be more differentiated to better support and empower the students in their EFL learning.

The Finnish Education Evaluation Centre (hence FINEEC) conducted in 2019 (Atjonen et al., 2019) an evaluation where the purpose was to study what kind of experiences and perceptions school principals, teachers, learners and guardians have on the methods, practices and culture of assessment. The results indicated that the main factors contributing to teachers' assessment activities were the established assessment practices and information gained about the learners during learning situations as well as the school community's collegial interaction and discussion practices. The teachers believed they were able to support the learners to recognise and develop their own learning methods, but not as able to help learners to compare their competence to the objectives or assessment criteria. The most frequently used assessment methods according to the qualitative and quantitative data gathered were summative and individual-based. The teachers' feedback practices were perceived to be effective, and the importance of encouraging feedback was stressed. Subject grades were usually based on the objectives and assessment criteria, although learners' performance was also compared to the rest of the group. The formation of grades in the final assessment was an imprecise combination of how the objectives had been achieved on average, the learner's competence at the time of the final assessment, and other factors. Differences were detected in the assessment criteria between schools of different sizes, school subjects, teachers' experience and between lower and upper classes in basic education.

Research Questions
For the purposes of this article, the following research questions have been designed:
RQ1: What modes and tools of feedback are used in classrooms and are there differences between languages and syllabi?
RQ2: What evidence do language teachers gather for students' summative assessment and are there differences between languages and syllabi?

For RQ1, the feedback is categorized as in Brown (2019) in (a) modes such as teacher feedback, self-feedback, peer feedback or combinations of them and in (b) tools such as written or oral comments, feedback lectures to the whole class, class discussions and so forth. For RQ2, the rubrics from Atjonen et al. 2019 have been used as they illustrate well the guidelines of the

NCC and also the principles that the Finnish language teachers frequently use when deciding on their students' final grades.

The data were gathered from three national evaluations of learning outcomes in languages, that is, the evaluation in long syllabus English in 2018 and the evaluation of middle-long and long syllabus Swedish in 2022. The study is an independent part of a larger study designed to evaluate students' learning outcomes in English (Härmälä et al., 2019) and Swedish (Härmälä & Marjanen, 2023). In the final reports of these evaluations, teachers' assessment practices have been described per language and per syllabus. A motivation for this study is, therefore, to go a step further and get a more detailed picture of the Finnish language teachers' assessment practices in compulsory education.

Context of the Study

The context of the current study is the Finnish compulsory education, which is comprised since 2021 of basic education and upper secondary education. The lower level of basic education (classes 1 to 6) is initiated at the age of 7 and the higher level (classes 7 to 9) continues until the student is 15/16 years old. After basic education, the student chooses between general upper secondary education or vocational education and training. Both these continue until the student is 18 years old. Compulsory education, including private schools, in Finland is free of charge. As the context of the study is basic education, I first briefly describe what is being prescribed about student assessment in the National Core Curricula for basic education (FNAE, 2016). The National Core Curriculum for Basic Education (hence NCC) is the national guiding document, which describes in detail i.e., the content and objectives of different subjects in accordance with the Basic Education Act (1998/628). The current NCC is from 2014 and it has been partly updated in 2016. Based on these national guidelines, local education providers are then required to design their own curricula where local requirements and context needs (Brown, 2008), such as i.e., students' linguistic background, may be better taken into account. Consequently, all the schools of each education provider need to follow the local curricula, which has all the potential to be more stakeholder-friendly (Brown, 2008) than the very general and abstract NCC. There are no school inspection system nor national exams in Finland but instead a decentralized system based on trust between the different stakeholders.

In the NCC 2016, student assessment is prescribed to have two tasks: first, to guide and encourage studying and to develop students' self-assessment skills; and second, to define to what extent a student has achieved the objectives set for different school subjects. The first task refers explicitly to formative assessment and the second to summative, end of educational stage assessment. Student assessment in the NCC focuses not only on learning, knowledge, and skills, but also on working and behaviour. To note that how the student behaves does not impact the grade of an individual subject.

According to NCC, formative assessment is part of teaching, and it helps a student to understand his/her own learning and to recognize his/her strengths and weaknesses (see also Brown & Hudson, 1998). Diagnostic information about students' performance on the objectives can help a teacher to decide where to focus his/her teaching to attain the objectives of the course or the curriculum (Brown & Hudson, 2002). Self-assessment and peer assessment are part of formative assessment in the NCC. Formative assessment does not require formal

documenting and it may take various forms such as checklists, self-assessment profiles, portfolios, free feedback, journals, video recordings and presenting the outcomes in small groups. Verbal feedback connected to verbal description is often considered better than just a number (see also Brown & Hudson, 2002, p. 49).

Summative assessment describes how well and to what extent a student has attained the objectives set in the NCC. In foreign languages, the Common European Framework of Reference with its update Companion Volume (Council of Europe 2001 and 2020) are used to indicate the proficiency levels targeted for different languages and for different grades. For example, in long syllabus English, there are in total 10 objectives, five of which are directly connected to language proficiency. The objectives are: Growing into cultural diversity and language awareness (1−3), Language learning skills (4−5) and Evolving language proficiency 6−10). Achieving a higher level of competence in one objective may compensate for failed or poorer performance in another. The assessment of the working skills that a student has demonstrated is included in the final assessment and in the given final grade for studies in the A syllabus in English (FNAE, 2020).

As stated in the NCC, summative assessment is done at the end of every school year, at the transition point between the lower and higher level of compulsory school (grade 6) and at the final phase of basic education (grade 9). It is done in relation to the objectives of NCC. Summative assessment is carried out by the teacher(s) who has been teaching a student and the teacher needs to gather all the evidence having an impact on the summative assessment. The guiding principles for assessment are equity, transparency, co-operation, participation, and planning. All assessment in basic education should be systematic, use versatile methods and be based on objectives and criteria that take students' age into account.

The attainment of the objectives set in the NCC is evaluated in the national evaluations of learning outcomes conducted by FINEEC. These evaluations are sample-based, and they are done in math and L1 every 4th year and in other subjects more seldom. During 2018−2022, the focus of the evaluations has been on national languages (Finnish for the Swedish-speaking and Swedish for the Finnish-speaking students) and the long syllabus English. English is the most widely studied foreign language in the Finnish compulsory education and nearly all students take it as their first foreign language at first grade. Swedish is the second national language and nearly all the Finnish-speaking students are required to start studying it at 6th grade as a middle-long syllabus. Students may choose Swedish also according to the objectives of a long syllabus language, which means that they start studying it at 1st or at 5th grade in addition to English or instead of it. The percentage of students taking long syllabus Swedish has been declining for decades and is now about 5,6 % according to the statistics of the Federation of Foreign Language Teachers in Finland SUKOL (Tilastotietoa kielivalinnoista - Suomen kieltenopettajien liitto ry (sukol.fi).

The results of the learning outcomes evaluations are used mainly for national education policy making decisions, but also locally to develop teaching. For the students, the evaluations are low stakes as the results of the evaluations have no serious implications on their lives (see also Brown, 2015). However, the teachers are encouraged to use the results as additional evidence for their final grades. Approximately 10 % of the age group participates in the evaluations, which may limit the usability of the results at local level.

Data and Method

In this article, I report the teachers' assessment practices with reference to formative and summative assessment. The data come from three national evaluations of learning outcomes administered in 2018 (English) and 2022 (Swedish). In connection to the evaluation, the teachers answered a questionnaire inquiring i.e., on their assessment practices, their ways of giving formative feedback and perceptions of students' self- and peer-evaluation practices. The data were gathered among 7th and 9th graders (students aged 12/12 and 15/16 years) as the evaluations of learning outcomes are usually administered either at the transition point between lower and higher levels compulsory education or at the end of it The English data are from long syllabus and the Swedish data from both long and middle-long syllabi. The sample included both small and larger schools and it was regionally representative (Table 1).

Table 1

Number of Teachers, Schools and Students Participating in the Evaluations in Swedish and English

	Number of teachers	Number of schools*	Number of students
Long syllabus English (7th)	243	130	5 021
Long syllabus Swedish (9th)	30	50	1 439
Middle-long syllabus Swedish (9th)	70	94	4 260
Total	338	274	10 720

Some schools may be the same in English and Swedish.

The schools in the data represent approximately 40 % of all the schools giving compulsory education in Finland. In 2022, there were in basic education 495 schools with grades 1 to 9 and 188 schools with grades 7 to 9 and (https://stat.fi/julkaisu/cl8n08k372so70cw1yrr6z3vr) and the trend is towards larger schools especially in the towns of Southern Finland.

The teacher data consist of teachers in both English and Swedish. In practice all Finnish language teachers have a Master's degree in Teaching and Learning, which include practical teaching training in schools. The language teachers are usually qualified to teach two languages, in this study most of them were teaching English or/and Swedish and possibly some other foreign language.

In all the three evaluations, there were common questions on the frequency of the classroom assessment practices, which the teachers answered by using a 5-point scale where the extremes were *never – almost always* or writing a short answer to an open-ended question. The questions focused on students' self-assessment, peer feedback, and teachers' ways of giving feedback to students in classroom. There were also questions inquiring on the principles the teachers used in their summative assessment. The answering scale for these statements was a 5-point Likert scale ranging from *strongly disagree* to *strongly agree*. The teachers' answers to the statements are presented as means by language and syllabi. The open-ended questions on the ways of giving the students feedback were analysed qualitatively and described without counting percentages.

The content of the statements is based on Atjonen et al. (2019) and on a national evaluation of learning outcomes in foreign languages and long syllabus Swedish in 2013 (Härmälä et al., 2014; Hildén & Rautopuro, 2014).

Results

The results of the study are presented by answering directly the two research questions.

RQ₁: What are the most frequent modes and tools of giving formative feedback and are there differences between languages and syllabi?

The first research question aimed at finding out the most frequent modes of giving formative feedback in classroom. In addition, it was of interest to see if there are differences between languages and syllabi. The results are displayed in Figure 1. In the figure, values 1−1,49 stand for *never*; 1,50−2,49 to *only seldom*; 2,50−3,49 to *sometimes*; 3,50−4,49 to *often*; 4,50−5,00 to *nearly always*.

Figure 1

Frequency of Different Feedback Modes by Language and Syllabi
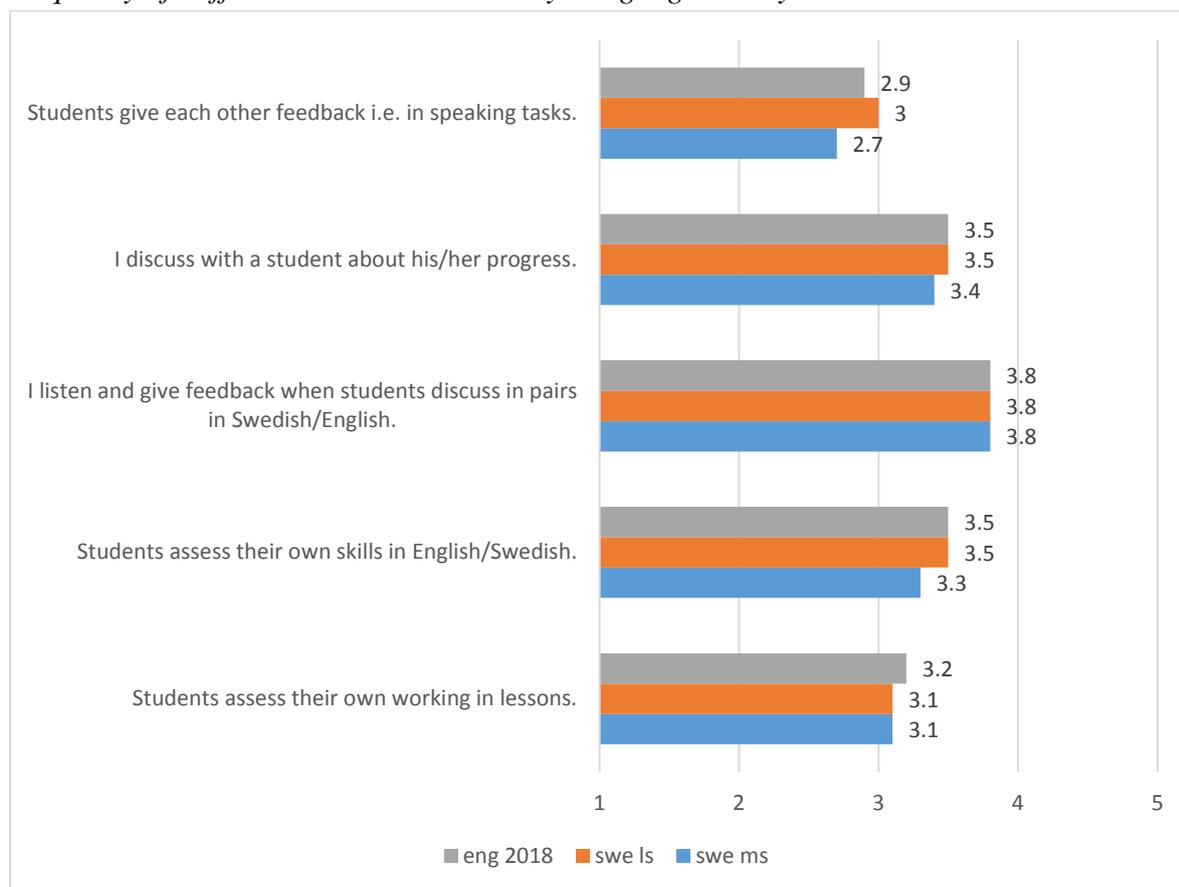
As the Figure 1 shows, there are only slight or no differences between languages and syllabi. The most frequent mode of giving feedback is that the teacher listens to students' pair discussions in the target language and gives them feedback. In both languages and syllabi, this kind of activity is done often on average. Another frequent mode of teacher feedback is that the teacher discusses with individual students about their progress. This is done on average often, in middle-long Swedish syllabus sometimes. As it comes to self- and peer feedback, the students of long syllabi English and Swedish assess their own skills on average often and again in the middle-syllabus Swedish, only sometimes according to the teachers' perceptions. The students assess their own learning in lessons and give feedback to other students i.e., in

speaking tasks. Interestingly, if teachers' and students' answers are compared with each other, it shows that, for example,, in middle-long syllabus Swedish (Härmälä & Marjanen, 2023), the teachers' perceptions about the frequency of giving the students feedback is slightly (approx. 0.6 units) more positive than the students' perceptions. This may be partly because of the fact that even if the teacher feels that s/he gives often feedback to the students, i.e. in oral discussions, it may be that an individual student feels that the feedback is not expressly targeted to him/her and in that way finds it not as useful as the teacher may think.

The teachers were also asked what explicit tools they usually use to give formative feedback to their students on assignments and exams. The means of the teachers' answers are displayed in figure 2. The values 1−1,49 stand for *never*; 1,50−2,49 to *only seldom*; 2,50−3,49 to *sometimes*; 3,50−4,49 to *often*; 4,50−5,00 to *nearly always*.

Figure 2

Teachers' Formative Feedback to Students on Assignments and Exams
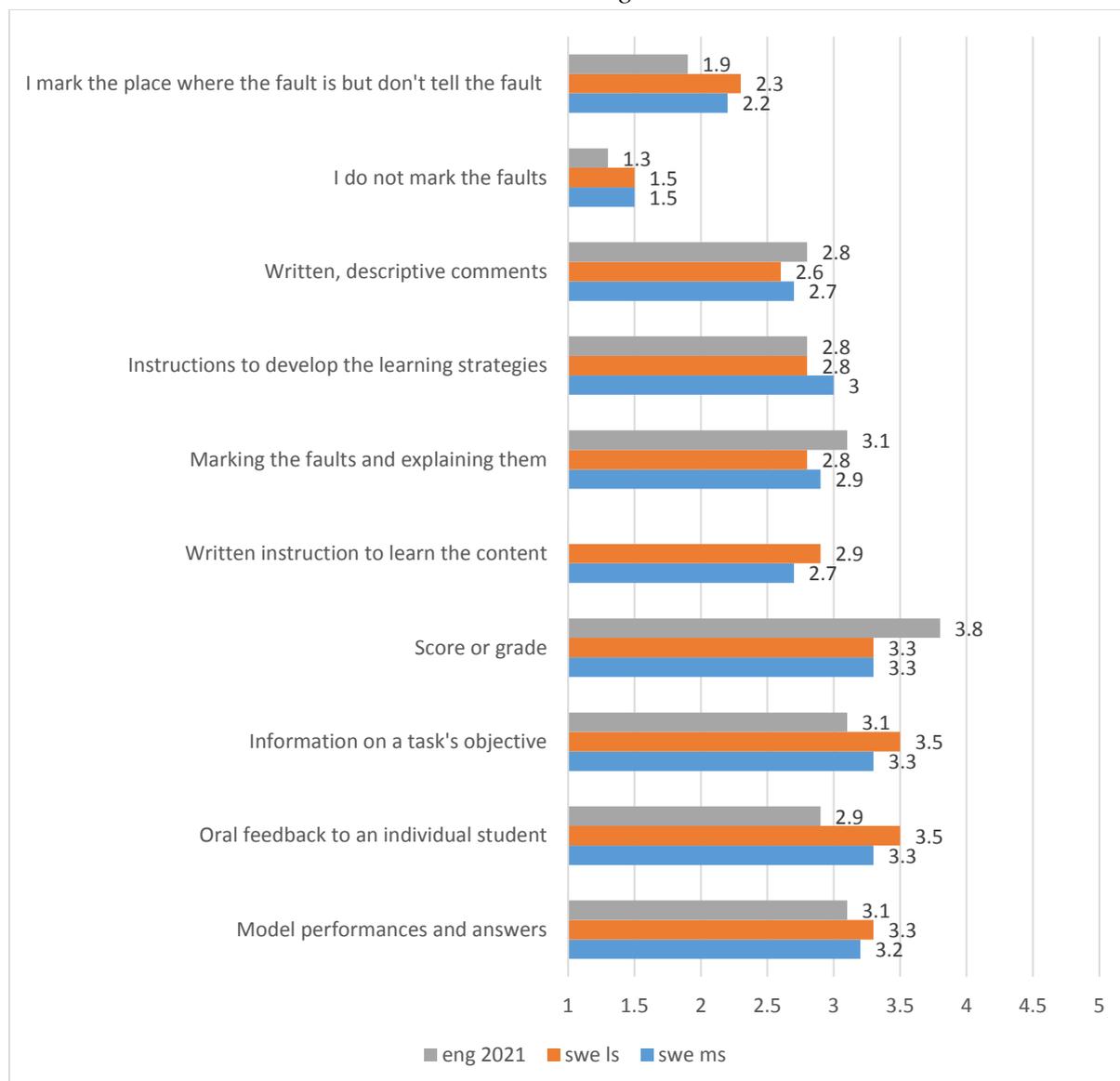
Figure 2 shows that the teachers in both languages use the above specified tools of giving formative feedback on average sometimes. The most frequent tools they use are giving scores, information on a task's objective and model performances and model answers. Teachers also mark the faults that the students have made and tell the students what the fault is. In middle-long syllabus Swedish, the teachers specified in the open-ended questions that they write their written feedback in school's digital application, which allows the feedback to be read also by parents. When giving formative feedback, some teachers use icons, stickers, and thumb-ups to encourage and motivate their students.

There are some interesting differences between languages and syllabi when it comes to the tools teachers use to give feedback. In long syllabus English, for example, teachers give scores or grades to their students on average often whereas in Swedish the average is sometimes. In Swedish, there are also some differences between the syllabi: in long syllabus Swedish, the teachers give oral feedback and information on a tasks' objective often but in middle-long syllabus sometimes. The difference might be partly because the class sizes are smaller in long syllabus Swedish than in the middle-long syllabus, which gives the teachers more opportunities to observe and follow each individual student better. Another interesting result in Figure 2 is that giving oral feedback to students is a bit more frequent in Swedish than in English. The reasons for this may only be speculated: is it because of the smaller class sizes and more heterogeneous students in Swedish? Or is it because there is more need to motivate the students in Swedish than in English and try to encourage them to study further?

In sum, the results indicate that an average Finnish language teacher does not give his/her student feedback so often. However, as the results above are averages, there are in the data also teachers who encourage and give feedback to their students often or almost every lesson but at the same time others who give feedback only rarely. Of course, the feedback must be timely and purposeful, but in general, the results suggest that there is a place for improvement, for example, in using more dynamic feedback that adapts to learner's abilities (see Leontjev, 2016).

RQ2: What evidence do teachers gather for students' summative assessment and are there differences between languages and syllabi?

The second research question aimed at finding out what kind of evidence the teachers gather for deciding on the students' final grade of a course or at the final phase of the basic education. Information on this was gathered through 13 statements that the teachers answered on a 5-point scale ranging from 1 = *strongly disagree* to 5 = *strongly agree*, the value 3 standing for *I do not have a clear perception about the matter.*

Figure 3

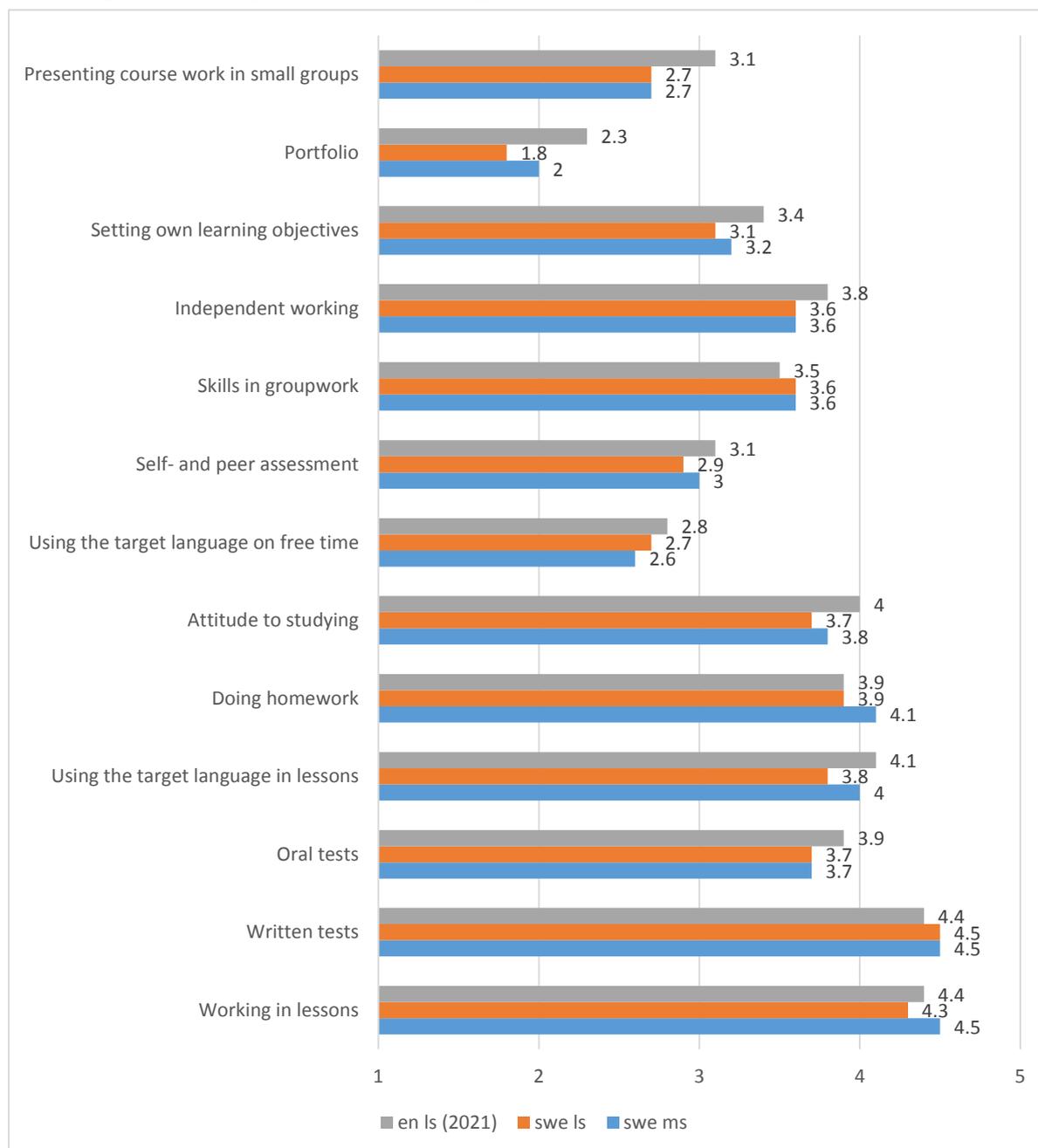*Student Evidence the Teachers Gather to Decide on Course Grades*

Figure 3 shows that two most important types of evidence for final grades are written tests and working in lessons. This is not surprising as students' active participation in lessons is included in the NCC, and it has traditionally been used by teachers as evidence of students learning. In addition, students' attitude to studying, doing homework, using the target language in lessons and oral tests are important sources of evidence for teachers. About the use of the target language the teachers' perceptions vary and on average the teachers do not have a clear perception on how to include it in the summative assessment. In NCC, the use of the European Language Portfolio (ELP) has been strongly encouraged but despite of this the teachers do not use it as evidence for their assessment. This may be partly because the teachers think that it

would mean extra work to implement the portfolio to the everyday teaching and learning activities. However, portfolios have several advantages if used properly (Brown & Hudson 1998): they would focus learners' attention on learning processes, help to motivate students, increase students' involvement in the learning process. For the teacher, portfolios could provide according to Brown & Hudson (1998) a clearer picture of students' language growth and permit the assessment of the multiple dimensions of language learning.

In sum, the evidence gathered for teachers' final grades was quite traditional in the Finnish context and emphasized written tests. The result is not surprising as there is a strong tradition in the Finnish compulsory education to rely on publication companies' printed or digitalized subject books, which include ready-made teaching materials and exams to be used in classroom. Consequently, (nearly) all schools use the same materials, which at least in theory also could guarantee equal learning outcomes. The schoolbooks designers are usually teachers of the subject who have experience in teaching the targeted age group, but, as it goes without saying, the book designed by a group of teachers is always only one possible view on what is important to be taught and learned in foreign language classrooms.

Partly due to the availability of ready-made teaching material, language teachers in basic and general upper secondary education do not often use more authentic materials taken i.e., from the internet, or everyday texts from students' lives, which could at least in some languages, such as Swedish, contribute to increasing students' motivation to study the language and make the students more aware of the affordances to learn a language around them. In learning English this problem is solved smoothly as the English language is a natural part of the young people's life also outside the classroom (see more in Härmälä & Marjanen, 2023). Therefore, it is a relatively depressing result to notice in Figure 3 that using the second national language, Swedish, outside school does not impact the teachers' final assessment even though extra-mural language use is stated in the NCC as being part of a course grade. One reason for this could be that teachers feel they lack national guidelines on how to do this in practice. For the Swedish language, including extra-mural language use in student assessment might also cause problems of equity as Swedish is used in Finland mainly in Southern, South-West and North-West parts of the country. Because of this, all students do not have equal opportunities to use and hear Swedish in their everyday life without making an effort.

To conclude, there are some slight differences between languages and syllabi on what kind of evidence the teachers gather when they decide on students' course grades. Teachers of English seem to utilize somewhat more versatile methods but the differences to Swedish are not significant. One possible reason for English teachers' slightly more varied methods may be that the proficiency level of the students is in English little higher than in long and middle-long Swedish, which allows for using different kinds of methods and tools.

How do the teachers then decide on what evidence to include on the final grades? Figure 4 summarizes the teachers' answers to a question inquiring on their grounds for giving the students final marks. The scale is from *strongly disagree* to *strongly agree*, value 3 standing for *not having a clear perception*.

When giving students final grades at the end of 9th grade, the teachers indicated that they base their grading on the students' language proficiency at a given moment and do not, for example, count an average of the students' earlier marks. Relying on the evidence demonstrated at a given moment was strongest in the long syllabus Swedish where it was just above the limit to always. Another interesting finding was that the teachers stressed the content areas related to language proficiency more than to the other, non-directly language related contents in the NCC. Here again, long syllabus Swedish stressed the most language proficiency. Teachers' perceptions whether they consider also other than directly to language proficiency related factors was not clear as the teachers indicated that they used them on average sometimes. Assessment discussions with parents and students were organized sometimes.

Figure 4

Basis for Giving Final Grades According to the Teachers
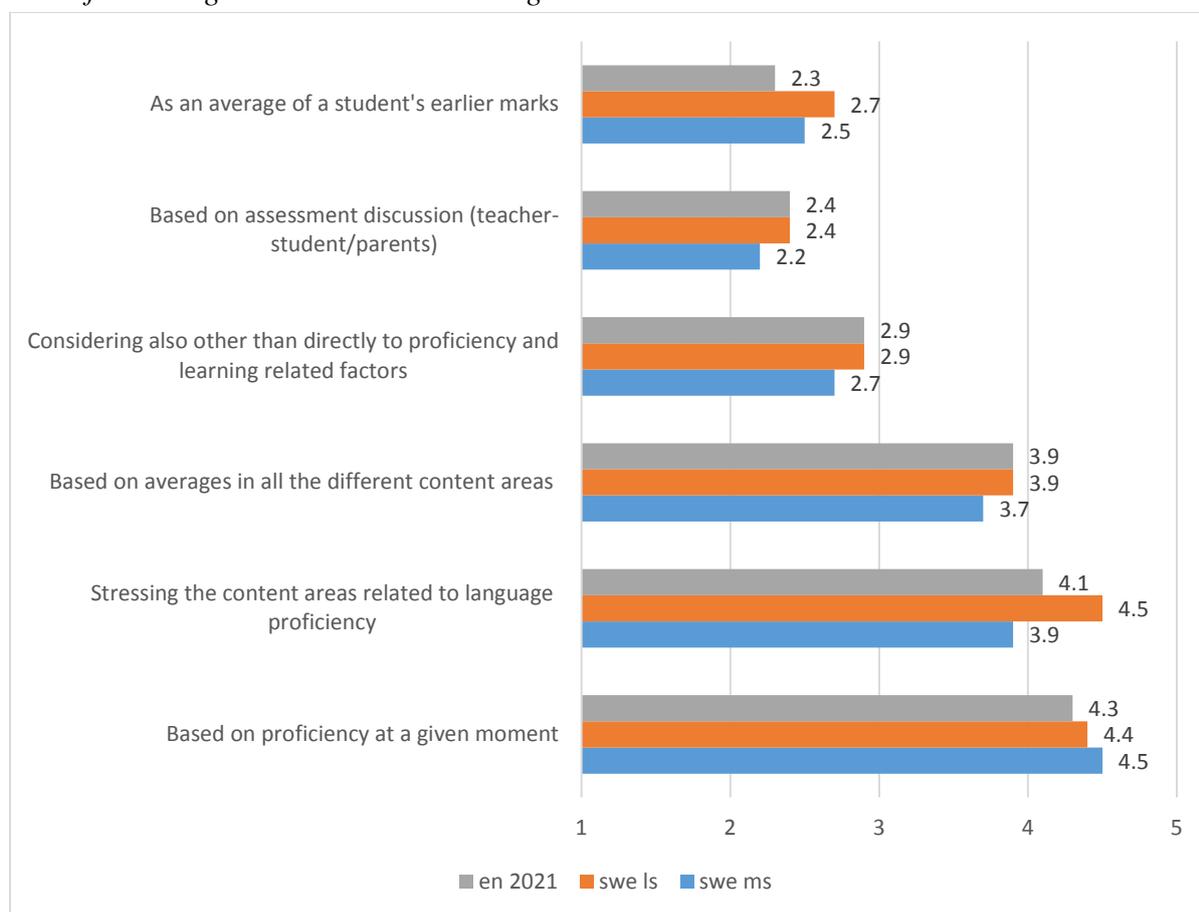
A point to be made in Figure 4 is that in the NCC, it is explicitly stated that none of the 9/10 content areas should be stressed but all be of equal weight. The teachers certainly know this, but the results of the study show that they are not quite convinced that in the final assessment also other than directly to language proficiency related factors should be of equal worth.In teachers' open-ended answers uncertainty and a need for training came clearly up as the following quotes from English teachers well illustrate:

> *"It was useful to practice the use of the assessment criteria also in relation to my own work. I think here especially about the skills of speaking and writing. I have not received any training in them. I would need more training."*
>
> *"The grading instructions were extremely clear and the model answers helped. Anyhow, I noticed that I have no routine at all to do this kind of assessment and I would certainly need some further training about that, too."*

Additionally, some basic principles of student assessment were unfamiliar to some of the teachers. This was illustrated by answers where some teachers wondered why they had to assess other teachers' students and how they could utilize the test results also for diagnostic purposes.

> *"This is not useful for my work as I could not assess my own students."*
>
> *"I cannot use this as part of my course assessment as the results tell me about the students' skills at the beginning of 7th grade."* [1]

Some teachers understood how they could use the results in motivating their students to study.

> *"The students are interested in knowing their own scores even though they do not wish them to impact on their course grades. They think it would be unfair."*
>
> *"In a way this made it clear what needs to be studied under grades 7 to 9 to make the results better."*

Discussion

This article explored Finnish language teachers' perceptions on classroom assessment and in particular on formative and summative assessment. The data were gathered in connection of three national evaluations of learning outcomes where 338 compulsory schoolteachers of English or Swedish participated with their students. The teachers came from 273 schools around Finland, and they had on average 10-15 years' experience in teaching their subject. The aim of the study was, first, to investigate what different modes and tools (Brown, 2019) were used in classrooms to give formative feedback and how often they were utilized. A second aim was to investigate what kind of evidence the teachers gathered to decide on the summative assessment of their students at the end of a course or at the final stage of the compulsory basic education. For both the questions, the differences between languages and syllabi were explored.

The results indicate that relative traditional modes and tools to give formative feedback were used in classrooms, such as teacher listening to students paired discussions in the target language and discussing with the student about his/her progress. Students' self-assessment was

---

[1] The answer is from the first phase of the English evaluation, when the students had just ended the lower level of basic education.

also used frequently. According to the teachers, they give formative feedback to their students often. When giving formative feedback, the teachers use several tools in addition to scores or marks. The most frequent of them were information on a task's objective, oral feedback and model performances and answers. The teachers usually mark the errors that the students have made. When giving course grades, the teachers rely mostly on written exams and students' working in class. In final summative assessment, the teachers tend to stress the objectives directly related to language proficiency even though there was some hesitation whether also other than directly to language related evidence should be considered as stated in the NCC. In all the assessment practices, there were small differences in the averages across languages and syllabi but, on average, the teachers' perceptions were relatively homogenous.

The study has implications for language teachers' assessment literacy training. Even though the average results were quite satisfactory, there are teachers who would benefit from more training, for example., in criterion-referenced assessment and in using the worded criteria for the different school grades. For some teachers even the difference between formative and summative assessment was not clear. This applies especially to documenting the formative assessment evidence. In the NCC, no documentation is required but according to many teachers, it is the parents who wish to see the evidence gathered to understand the motivation to their child's final grade.

Another implication of the study is the need for gathering more evidence on the formation of the students' final grades, for example, by interviewing the teachers. The national evaluations of learning outcomes have shown that in English and Swedish, for example, there is on average an almost two school grades' difference between a student's result in the national evaluation and the teachers' final grades at school (Hildén et al., 2016; Härmälä & Marjanen, 2023). It would be extremely important to know what this other evidence is to make the grades more transparent. The variation between schools is also quite large: in some schools the students get higher grades for the same skills than in others. The issue of equity comes even more urgent now when the worded criteria for different grades have been introduced to all subjects in basic education. This sets totally new kinds of assessment literacy requirements for the teachers.

A third implication of the study is that as Brown and Hudson 1998 suggest, there is still a need for alternatives in assessment. As the heterogeneity of the students in classroom is constantly increasing (even in Finland) it becomes even more important to use sufficiently versatile modes and tools of both formative and summative assessment to capture the skills and knowledge of each individual student in a systematic and fair way. To do this, teachers need to be equipped with sufficient skills in assessment literacy in concordance with the needs of their local teaching context.

Conclusion

Teachers are today strongly encouraged to use multiple tools and modes to gather evidence on their students' skills and knowledge during the whole learning process. Depending on the formative or summative purpose of the assessment, different tools and modes to assess are chosen. However, as our study clearly demonstrated, the assessment practices that the language

teachers used were quite traditional and relatively similar across languages and syllabi. In the final assessment, contents directly related to language competence were emphasized.

Further studies are needed to clarify in what way teachers are already able to take into account the increasing heterogeneity of their students in their assessment practices. How could teachers, for example, incorporate students' extra-mural language use in school assessment and in that way allow for even more versatile alternatives in assessment.

ORCID

🆔 https://orcid.org/0000-0003-1601-3095

Competing Interests

No, there are no conflicting interests.

Rights and Permissions

Open Access

References

Atjonen, P., Laivamaa, H., Levonen, A., Orell, S., Saari, M., Sulonen, K., Tamm, M., Kamppi, P., Rumpu, N. & Hietala, R. (2019). *Että tietää missä on menossa. Oppimisen ja osaamisen arviointi perusopetuksessa ja lukiokoulutuksessa.* Kansallinen koulutuksen arviointikeskus. (Julkaisut 7:2019). https://www.karvi.fi/fi/julkaisut/etta-tietaa-missa-menossa-oppimisen-ja-osaamisen-arviointi-perusopetuksessa-ja-lukiokoulutuksessa

Basic Education Act 1998 (FIN).

Brown, J., D. (2008). Testing-context analysis: Assessment is just another part of language curriculum development. *Language Assessment Quarterly, 5(4)*, 275-312. https://doi.org/10.1080/15434300802457455

Brown, J. D. (2009). Language curriculum development: Mistakes were made, problems faced, and lessons learned. *Second Language Studies*, *28(1)*, 85−105.

Brown, J. D. (2012). Choosing the right type of assessment. In C. Coombe, S. J. Stoynoff, P. Davidson, & B. O'Sullivan (eds.), *The Cambridge guide to language assessment*, (pp. 133-139). Cambridge University Press.

Brown, J. D. (ed.) (2013). *New ways of classroom assessment, revised.* Alexandria, VA: Teachers of English to Speakers of Other Languages.

Brown, J. D. (2015). Language testing: The state of the art (An online interview with James Dean Brown). *International Journal of Language Studies*, *9*(4), 133−143.

Brown, J. D. (2019). Assessment feedback. *The Journal of Asia TEFL*, *16(*1), 334–344. http://dx.doi.org/10.18823/asiatefl.2019.16.1.22.334

Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly 32*(4), 653−675. https://doi.org/10.2307/3587999

Brown, J. D. & T. Hudson (2002). *Criterion-referenced language testing.* Cambridge University Press.

Council of Europe (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment.* https://rm.coe.int/16802fc1bf

Marita Härmälä

Council of Europe (2020). *Common European Framework of Reference for Languages: learning, teaching, assessment. Companion volume*. https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

Earl, L., M. (2013). *Assessment as learning. Using classroom assessment to maximize student learning (2nd ed.)*. Sage.

Finnish National Agency for Education (2016). *National Core Curricula for Basic Education. Publications 5*.

Finnish National Agency for Education (2020). *Päättöarvioinnin kriteerit*. https://www.oph.fi/sites/default/files/documents/Perusopetuksen%20p%C3%A4%C3%A4tt%C3%B6arvioinnin%20kriteerit%2031.12.2020_1.pdf

Härmälä, M. & Marjanen, J. (2023). *B1-ruotsin oppimistulokset perusopetuksen päättövaiheessa 2022*. Kansallinen koulutuksen arviointikeskus (Julkaisut 10:2023). https://karvi.fi/wp-content/uploads/2023/03/KARVI_1023.pdf

Härmälä, M. & Marjanen, J. (2022). *Englantia koronapandemian aikaan – A-englannin osaaminen 9. luokan lopussa keväällä 2021*. Kansallinen koulutuksen arviointikeskus (Julkaisut 22:2022). https://karvi. fi/wp-content/uploads/2022/09/KARVI_2222.pdf

Härmälä, M., Huhtanen, M., Puukko, M. & Marjanen, J. (2019). *A-englannin oppimistulokset 7. vuosiluokan alussa 2018*. Kansallinen koulutuksen arviointikeskus (Julkaisut 13:2019). https://karvi.fi/wp-content/uploads/2019/05/KARVI_1319.pdf

Härmälä, M., Huhtanen, M. & Puukko, M. (2014). *Englannin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Kansallinen koulutuksen arviointikeskus (Julkaisut 2:2014). https://karvi.fi/wp-content/uploads/2014/10/KARVI_0214.pdf

Hildén, R. & Rautopuro, J. (2014). *Ruotsin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Kansallinen koulutuksen arviointikeskus (Julkaisut 1:2014). https://karvi.fi/publication/ruotsin-kielen-oppimaaran-oppimistulokset-perusopetuksen-paattovaiheessa-2013/

Hildén, R., Rautopuro, J. & Ouakrim-Soivio, N. (2016). Kaikille ansionsa mukaan? Perusopetuksen päättöarvioinnin yhdenvertaisuus Suomessa. *Kasvatus 47*(4), 342–357.

Leontjev, D. (2016*). ICAnDoiT: the impact of computerised adaptive corrective feedback on L2 English learners*. [Doctoral dissertation, University of Jyväskylä]. https://jyx.jyu.fi/handle/123456789/49346#

Mäkipää, T. & Hildén, R. (2021). What kind of feedback is perceived as encouraging by Finnish general upper secondary school students? *Education Sciences, 11*(1), 12. https://doi.org/10.3390/educsci11010012

Pollari, P. (2017). To feed back or to feed forward? Students' experiences of and responses to feedback in a Finnish EFL classroom. *Apples. Journal Applied Language Studies 2017, 11*, 11–33. https://doi.org/10.17011/apples/urn.201708073429

Saeed, M., Tahir, H. & Latif, I. (2018). Teachers' perceptions about the use of classroom assessment techniques in elementary and secondary schools. *Bulletin of Education and Research*, *40*(1), 115−130.

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (eds.), *Perspectives of Curriculum Evaluation* (pp. 39-83). Rand McNally.

Stiggins, R. J. (1991). *Assessment literacy*, 72(7), 534–539.

Tsagari, D. & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment, 6(1), 41−63*. https://doi.org/10.58379/UHIX9883