English teachers' perceptions of feedback practices under emergency remote teaching conditions: Insights for future use

Toni Mäkipää[1] & Anna Soltyska[2] [1]

[1] University of Helsinki, Finland

[2] Ruhr University Bochum, Germany

The purpose of this paper was to investigate how English teachers employed formative assessment practices in emergency remote teaching (ERT) during the COVID-19 pandemic, and their perceived training needs in formative assessment. The data were collected from a questionnaire and interviews with teachers from Finland (n = 33) and Germany (n = 91). The results suggest that the teachers employed feedback, self-assessment and peer assessment in diverse ways, such as using them along with portfolios and online presentations. However, many teachers faced challenges with the technical aspects of providing feedback. The teachers also mentioned several areas in which they needed further training, such as identifying students' needs and learning new techniques and formats of formative assessment. The results pave the way for reshaping formative assessment practices in post-pandemic language education.

Keywords: formative assessment, emergency remote teaching, emergency remote assessment, feedback, language assessment literacy

Introduction

The primary goal of this paper is to explore some of the ways in which English teachers employed formative assessment practices in foreign language emergency remote teaching (ERT), and what training in formative assessment they may still require to better meet their learners' needs. We focused mainly on teacher feedback, peer assessment, and self-assessment. We used a convenience sampling from Germany and

[1] Email address for correspondence: toni.makipaa@helsinki.fi

Finland to recruit participants, as these were the populations that we had ready access to. Furthermore, as these two countries differ in the overall level of digitalization according to DESI (the Digital Economy and Society Index[2]) indicators, but also in the way teacher education is organized and language assessment is taught to prospective teachers and implemented in the schooling system (Mäkipää & Soltyska, 2023), divergent realizations of assessment under ERT conditions could be expected.

For language teachers, understanding and using formative assessment efficiently in the language classroom is a central priority (Mäkipää, 2021; Li & Gu, 2023; Xiao & Yang, 2019). If formative assessment practices are designed and employed with scrupulous attention, the teacher's support of student learning can be markedly better (Black & Wiliam, 2018). From a teacher's perspective, this requires an in-depth understanding of both curriculum standards and theoretical knowledge (Li & Gu, 2023; Xu & Brown, 2016). According to previous studies on the COVID-19 pandemic (e.g., Cooper et al., 2021; Hodges et al., 2020), the shift from on-site to remote testing coincided with a change from single-instance, end-of-course summative exams to multiple-instance cumulative assessment of a more formative character. Therefore, it is essential to investigate formative assessment practices in ERT to develop the pedagogy of remote teaching.

We will first describe the implementation of formative assessment in ERT and analyze teachers' perceptions thereof. In particular, we will investigate how feedback was provided under those exceptional conditions. Then, following their involvement in ERT, we will summarize teachers' self-reported training needs regarding formative assessment, and establish how this experience might impact their future formative assessment practices and routines. The aim of the study is to shed light on the multifaceted character of formative assessment, especially in remote and online settings, and portray advantages and disadvantages of such practices as implemented under trying ERT conditions. Given the scant amount of research on teachers' formative assessment and feedback in online learning (Jiang & Yu, 2021), the study will add valuable insights to this field. It is also intended to contribute to the discussion

[2] The Digital Economy and Society Index (DESI) is a body of data used to analyze Europe's overall digital performance and track the progress of countries within the European Union (EU) in terms of digital competitiveness (https://digital-strategy.ec.europa.eu/en/policies/desi).

on language assessment literacy in the post-ERT era and identify gains from this experience.

Literature review

The following section briefly presents the concept of formative assessment and shows how the use of digital technology, which was prevalent during school closures during the COVID-19 pandemic, could impact feasibility and quality of this kind of assessment.

Formative and alternative assessment

While the term "formative assessment" refers to a function and purpose of assessment, when talking about "alternative assessment" the focus is placed on the type of activities employed as assessment. In other words, assessment activities can be both of formative and alternative character. According to Bachman and Palmer (1996), formative assessment refers to situations in which "students guide their own subsequent learning, and teachers modify their teaching methods and materials so as to make them more appropriate for students' needs, interests, and capabilities" (p. 98). This term is used to denote assessment activities and tasks of a planned and ongoing character, which are aimed at monitoring student learning rather than measuring their progress or providing evidence for its absence. Formative assessment typically incorporates self- and peer assessment, portfolio assessment and formative feedback provided by teachers and might be realized within a classroom setting or beyond it. According to Gu (2021), a typical formative assessment situation entails several components: clarifying the learning objectives, eliciting evidence of students' understanding or learning, interpreting the elicited information against the learning objectives or assessment criteria, providing the students in question with the feedback on the basis of this interpretation and taking some follow-up action which is aimed at improving learning and which can be performed by the students themselves or their teacher. These assessment components are of a sequential and interactive nature and are aimed to bring the learners closer to their learning targets through numerous iterations of the assessment process.

Alternative assessment, also referred to as classroom-based, qualitative, informal, or performance-based assessment, is a procedure "gathered over a period of time", of a

character "less formal than traditional ones" and typically with a "formative function, low-stakes consequences and beneficial washback" (Alderson & Banerjee, 2001, p. 228). It can be understood as an alternative to closed items, (semi-)standardized formal testing administered at a given point in time and under time constraints, usually for summative purposes. It emerged from a constructivist approach in education which focuses on learner's active participation and self-regulated involvement in the learning and assessment process, leading to more hands-on practices such as self- and peer assessment. There are various ways in which learning outcomes can be assessed alternatively, such as interviews, analysis of written work (e.g., stories, essays, reports, blogs, diaries) and diverse feedback activities, for example, direct observations feedback, paired conversations feedback, group interviews feedback, self-evaluation feedback, peer evaluation/feedback, written tasks feedback, student portfolio feedback (including digital portfolios), and interviews (Inbar-Lourie, 2013). This wide array of assessment techniques, which frequently resemble real-life situations of language use outside the classroom, emphasize assessing the processes as well as products of learning, a characteristic that distinguishes alternative assessment activities from conventional ones (Herman et al., 1992, Janisch et al., 2007, Yusop et al., 2022).

Further features of alternative assessment focus on its action-oriented character, as students are typically asked to perform, create, or produce something. The tasks used for assessment represent meaningful instructional activities and are ideally embedded in real-world settings (Herman et al, 1992, p. 6). The objective of alternative assessment is thus to tap higher-level thinking and problem-solving skills which assume more than one right answer and rely on human judgement for scoring, rather than machines. This assumption is, however, no longer fully accurate as AI-based tools are already capable of making an accurate judgement (cf. Link & Koltovskaia, 2023, Mizumoto & Eguchi, 2023) and providing targeted feedback on a learner's unique language performance (Escalante et al., 2023). As a result, alternative assessment, which is deeply embedded in classroom teaching and places on teachers the responsibility for choosing adequate assessment instruments and criteria, requires from them new instructional and assessment roles. Unlike conventional assessment, alternative assessment is claimed to be "more directly useful to stakeholders" (Janisch et al., 2007, p. 222) as it involves observation over a period in contrast to a one-time

snapshot of performance, its results are typically instantly accessible, and teachers and learners have been involved in co-constructing the assessment process (Janisch et al., 2007, p. 223–224).

The role of feedback in formative assessment

The definitions of feedback have evolved significantly during the last few decades. The first definitions conceptualized feedback as a gift from the teacher or the peer for the student, but the contemporary definitions emphasize the importance of the student's active role (Winstone et al., 2022). In a nutshell, feedback refers to students interpreting and using the information provided to enhance their assignments, work, and strategies (Carless & Boud, 2018). Furthermore, the quantity of feedback is not the principal factor in effective feedback processes (Hattie & Clarke, 2019).

As highlighted in a proliferation of studies, feedback is intrinsic to all learning (Dawson et al., 2023b; Smith et al., 2016; Yang et al., 2023). Students also value language teachers' feedback comments (Mäkipää, 2021; Virkkunen & Toivola, 2023; Xiao & Yang, 2019). However, the results on what type of feedback best enhances language learning are inconsistent (Yang et al., 2022). Despite these conflicting findings, Wiliam (2013) argues that it is meaningless to discuss what type of feedback is the most effective as the same feedback might result in different reactions from different students. Therefore, interpreting feedback lies at the heart of effective feedback process, not merely the act of giving feedback (Wiliam, 2013). Although no simple formula or trick guarantees enhancement of learning, many scholars emphasize that feedback processes are a two-way street; both the teacher and the student need to be working actively with feedback (Dawson et al., 2023b). In essence, instead of being passive, the student should engage with the comments, reflect on them and ponder how to revise a piece of work or learning strategies (Black & Wiliam, 2018; Carless & Boud, 2018). The key notion is impact: teacher feedback needs to provoke an impact on the student (Wiliam, 2013) and move learning forward (Black & Wiliam, 2018). However, previous research (e.g., Jónsson et al., 2018; Mäkipää, 2021) suggests that students and teachers occasionally perceive feedback processes differently. Therefore, more research on feedback is warranted.

In addition to promoting language learning (Suzuki et al., 2019; Xiao & Yang, 2019; Yang et al., 2022), teacher feedback is also indispensable for other reasons. In the

classroom, effective teacher feedback is a powerful enhancer of self-regulated learning, learning to learn and life-long learning (Hattie & Clarke, 2019; Smith et al., 2016; Yang et al., 2023). Additionally, teacher feedback is crucial from the perspective of responsive pedagogy; teachers are required to educate students for the future, in which the role of teacher feedback is underscored (Smith et al., 2016). In terms of assessment literacy, providing instructive feedback is an essential skill for an assessment-literate teacher (Xu & Brown, 2016), and teachers find it important to provide feedback in language learning (Li & Gu, 2023; Yang & May, 2023). However, in remote language teaching providing feedback might be challenging due to the absence of body language and face-to-face contact (Guerra, 2023).

Self- and peer assessment

In their definition of self-assessment, Yan and Brown (2017, p. 1248) emphasize students' active role in seeking feedback, which is in harmony with the notion of feedback literacy (Carless & Boud, 2018). They also explicitly state the purpose of self-assessment (identify one's strengths and weaknesses), which is often lacking in other definitions (Andrade, 2019). In substance, the purposes of self-assessment reflect two objectives: summative (grading one's leaning outcomes) and formative (monitoring learning) (Butler, 2023a). To foster learning, it is advisable to focus more on the formative purpose of self-assessment, in other words, adjustment, correction, and verbalization of learning (Andrade, 2018, 2019; Butler, 2023b). Overall, the aim of this form of assessment is to "move students away from dependence on the teacher toward independence in the power to guide their own learning" (Black, 2010, p. 362). It requires familiarizing students with aims that are to be attained and evaluation criteria that are applied to measure this attainment. It is hoped that by so doing, students will internalize assessment criteria and also learn to formulate actionable feedback for their peers worded in accessible language.

Typically, self-assessment is implemented in the form of a questionnaire completed by students themselves (MacBeath et al., 2000). In practice, students habitually compare their work or learning against a set of criteria. However, Nicol (2021) argues that making comparisons is a pervasive process, in which students use prior experiences, materials, or even peers for reflection. Inasmuch as merely using a list of criteria is not enough for enhancing learning, teachers should encourage their students to compare

their work against external sources of information, such as peers and teachers (Nicol, 2021).

In technology-assisted learning, self-feedback can be formed on the basis of audio or video recordings which enable students to analyze and evaluate their own performance, e.g., against a recorded exemplar performance (Dawson et al., 2023a). Another instrument for technology-enhanced self-assessment is the e-portfolio (e.g., Mahara[3], European Language Portfolio[4]), which is a digital repository of a student's work, typically collected in a data cloud hosted within or independent of a Learning Management System (LMS). Like their conventional paper-based counterparts, e-portfolios involve creation, curation and showcasing learning products, and enable sharing them with a specific audience such as teachers, tutors and peers, and inviting their feedback. In contrast to traditional portfolios, sharing content as it is created, i.e., in the process of learning, and interacting with the audience is easier and can be further facilitated by technology when these take place through websites, blogs or social media.

Many terms regarding peer assessment are employed in the literature, sometimes even interchangeably. For example, other recurrent terms are peer evaluation, peer feedback, and peer grading (Double et al., 2020). A common feature in these terms is the word *peer*. Peers who provide insights and comments on each other's work and performance are typically at similar or identical education levels (Double et al., 2020). According to Topping (2018, p. 1), peer assessment refers to "an arrangement for learners to consider and specify the level, value, or quality of a product or performance of other equal-status learners, then learn further by giving elaborated feedback to and discussing their appraisals with those who were assessed to achieve a negotiated agreed outcome." Peer assessment can be summative (grading the work), formative (providing suggestions on improving the work) or a combination of both (Topping, 2018). It is recommended that peer assessment focuses more on the formative aspect (Topping, 2018). In recent years, peer assessment has become more widespread in teaching (Li et al., 2020; Topping, 2017), including language teaching (Härmälä & Marjanen, 2022).

[3] https://mahara.org
[4] https://www.coe.int/en/web/portfolio

Student-to-student peer assessment encompasses a variety of activities in which learners are encouraged to evaluate the work of their fellow students and provide feedback on their performance and outcomes of their work. When used for formative purposes, peer feedback typically involves constructive comments on draft work, which is subsequently revised and improved before submission, if relevant (Topping, 2018). Using technology can significantly enhance peer assessment in language classrooms and beyond, for example by means of co-writing through shared documents, creating and reacting to interactive blogs or discussion forums. Peer feedback software and tools can be used within purpose-built features of learning platforms (e.g., the Workshop function of Moodle described later in this article) or through existing social networking sites, if their use is encouraged and not prohibited for data protection reasons in specific educational contexts. In fact, students value online peer feedback more than hand-written, paper-based comments, both in terms of feasibility and efficiency of use, as Ho's study on English as a foreign language writing (2015) and McCarthy's study on first-year university students (2016) confirm.

Emergency remote teaching and assessment

As of March 2020, unparalleled measures were undertaken to ensure a safe learning environment for every student. The unprecedented transition to teaching and learning away from classrooms as a result of the COVID-19 pandemic in early 2020 and the months that followed gave rise to the concept of emergency remote teaching (ERT) (Hodges et al., 2020), which is juxtaposed with planned and intentional online learning. In contrast to online education, ERT is temporary in nature, implies a simple and time-efficient setup of teaching and learning activities, and emphasizes reliable access for all stakeholders in the process, with no aspirations to high quality, but rather the intention to return to the previously adopted mode once the emergency has passed. Although characterized in essence as a "temporary solution to an immediate problem" (Bozkurt & Sharma, 2020, p. ii), the provision of educational services continued to be constrained by pandemic conditions and so the temporary nature of the phenomenon dissipated. According to Hill's four-phase model of education response to COVID-19 (Hill, 2020), the initial phase of rapid transition to remote teaching and learning was immediately followed by the phase of "(re)adding the basics", i.e., early institutional support for learner and teacher communities, mainly of a technical nature. These two phases could be called ERT proper. The stage that followed was referred to by Hill as

"extended transition during continued turmoil" and involved preparing all stakeholders for whatever conditions in which they have to teach and learn. Finally, the "emerging new normal" phase, with the establishment of sustainable infrastructure and regulations, commenced in 2021 and continued for varying lengths of time in different parts of the world. These diverse conditions throughout the evolution of the pandemic between 2020 and 2021 and beyond need to be considered when analyzing teachers' assessment choices and perceptions thereof.

Teachers' views on ERT and assessment practices in foreign language education under those trying conditions were studied among others by Bruce and Stakounis (2021, 2022), Drucker and Fleischhauer (2021), Kılıçkaya (2023), Mäkipää (2023), Mäkipää and Soltyska (2023), Sevimel-Sahin, (2023), and Zhang et al. (2021). These studies included a variety of observations on challenging and rewarding aspects thereof. English teachers were not prepared for online assessment as they had not received training in it during their teacher education (Mäkipää & Soltyska, 2023). Still, despite the lack of training and unprecedented nature of the pandemic, language teachers were able to use multifaceted assessment practices to assess their students' proficiency (Drucker & Fleischhauer, 2021; Mäkipää, 2023). English teachers mainly determined the assessment practices based on their own experiences, reflections, and the local policy (Zhang et al., 2021). However, in some contexts, English teachers were not allowed to decide themselves what assessment practices they could employ (Kılıçkaya, 2023; Mäkipää & Soltyska, 2023). In spite of the common understanding of the primary role of the didactics of language learning in shaping teaching and assessment decisions (Drucker & Fleischhauer, 2021), practical aspects of providing instruction and assessment came to the fore. Dealing with technology which the teachers were often unfamiliar with, ensuring assessment security and counteracting academic misconduct in remote conditions, and making ad-hoc assessment-related decisions during the transition from in-person to remote mode of delivery without prior training were among the challenges identified in various language teaching settings (Bruce & Stakounis, 2021, 2022; Drucker & Fleischhauer, 2021; Kılıçkaya, 2023; Mäkipää & Soltyska, 2023; Sevimel-Sahin, 2023; Zhang et al., 2021). Of particular relevance to this paper, the term emergency remote assessment (ERA) was introduced by Bruce and Stakounis (2022, p. 34) to denote "emergency assessment policies, procedures and instruments adopted in the context of emergency remote teaching." As prior

studies on the implementation of formative assessment in foreign language ERT are scarce, the novelty of this study relates to examining ERA from the perspective of teacher feedback, peer feedback, and self-assessment in language education.

Methodology

This article is guided by three research questions: (1) How was formative assessment implemented in ERT? (2) How was feedback used in ERT? (3) What are English teachers' self-reported training needs regarding formative assessment? Even though formative assessment also entails feedback, the aim of the second question is to deepen our understanding of feedback practices. The following section summarizes the type of data that was used in the study and describes the way it was analyzed. Furthermore, the participants of the study are briefly identified. For further details of the respondents and other data collected in the project, see Mäkipää and Soltyska (2023).

Context of the study

The two researchers involved in the project are based in Finland and Germany – two countries which differ not only in terms of their teacher education systems but also in the progress of digitalization in various areas of public life, including education. Anecdotal evidence related to teaching and assessing under ERT in both countries, as well as the analysis of the Digital Economy and Society Index (DESI) indicators for the year 2021 implied that there might be some discrepancy between teachers in Finland and Germany. While Finland was leading the ranking of 29 countries based on DESI 2022 (European Commission, 2022), Germany ranked thirteenth. This progressing digitalization gap between the two economies and societies gave rise to analyzing teachers' assessment skills for digital contexts (for more details of the analysis of DESI indicators for Finland and Germany and differences in teacher education, see Mäkipää and Soltyska, 2023).

Data collection and analysis

This study is part of a larger project, which focuses on English teachers' assessment literacy. As it draws on a subset of the data pertaining to formative assessment, this study at hand contributes to the literature by examining formative assessment practices, feedback practices in particular, in ERT from English teachers' perspective.

This study is a mixed-methods cross-sectional study. It was conducted between October 2021 and March 2022 and relied on data collected by means of an online questionnaire and structured interviews. The questionnaire was designed in line with Dörnyei and Taguchi (2010) and included 25 closed and open-ended items categorized into six topical groups: 1) General information, 2) Assessment-related teacher education, 3) Assessment-related knowledge, abilities and confidence, 4) Experience with emergency online teaching and assessment, 5) Professional development in the field of language assessment and 6) Concluding questions. The questionnaire followed the example of previous language assessment literacy surveys by Vogt and Tsagari (2014), Kremmel and Harding (2020), and Zhang et al. (2021). Its draft was commented on by nine experts in assessment literacy from Finland, Germany and the United Kingdom, and improved accordingly. The final version used for the data collection is attached in the Appendix. The questionnaire was administered in Qualtrics and disseminated among English teachers in Finland and Germany via regional teacher organizations and social media. This resulted in a non-representative snowball sample (n = 124).

The semi-structured interviews (Dörnyei, 2007) were conducted online via Zoom with those respondents to the online questionnaire who voluntarily offered to be interviewed by one of the researchers. Each interview (n = 27) followed a predefined structure consisting of 22 questions or prompts within four thematic blocks: 1) online assessment in general and assessment in ERT in particular, 2) providing feedback in ERT and other contexts, 3) self- and peer assessment in ERT and other contexts, and 4) culture-related aspects of assessment. The interview protocol is attached in the Appendix. All interviews were conducted in English and lasted between 13 and 59 minutes, with an average length 34 minutes. With the participants' permission, the interviews were recorded, transcribed and anonymized. Each of the two researchers independently coded the responses thematically and the codes were then compared and streamlined to an agreed interpretation (Cornish et al., 2014). The analysis was inductive as the researchers did not work with any pre-defined codes but derived them from the analysis of the interviewees' responses in the light of the three research questions. In the course of several streamlining sessions, both researchers agreed on a focal aspect of each response and assigned each answer a code. These codes included, e.g., areas of challenge when providing feedback (time pressure, workload, students'

involvement, technical problems), teachers' training needs (practical aspects, feedback skills, understanding students' perspective), and teachers' views, e.g., on the amount of feedback they provide (positive, negative, unsure). Contrary to the advice expressed by Hodges et al. (2020) not to compare a face-to-face course with its online version, some interview questions targeted differences in formative assessment procedures implemented prior to and during ERT conditions. This interview design was chosen to elicit respondents' reflections triggered by experiencing these two, often significantly different, modes of creating and implementing formative assessment.

This paper is based on a selection of data from both sources. The quantitative data were selected from the sections of the questionnaire focusing on the respondents' previous formal training in assessment and their actual formative assessment practices before and during the period of emergency remote teaching. The data were analyzed using descriptive statistics. We calculated the distribution of scores across the options on the scale in terms of percentages. The qualitative data, mainly derived from Sections 2 and 3 of the semi-structured interviews, offer insight into the nature of formative assessment practices as perceived and adopted by the respondents. Furthermore, two open-ended items from the questionnaire were included in the analysis of the qualitative data and analyzed inductively.

Participants

The participants in this study consist of 124 teachers of English as a foreign language from Germany (n = 91) and Finland (n = 33). However, nine teachers did not answer all the questions relevant to this paper. Therefore, the data from these participants were excluded from the analysis. The two countries in focus of the analysis were selected due to their divergent levels of digitalization and different teacher training paths including assessment practices. Furthermore, the group was a convenience sample chosen due to the authors' current place of work. The German teachers worked mostly in higher education (85%), while most of their Finnish colleagues worked in secondary education (67%). However, some teachers worked simultaneously at more than one level. Regarding teaching experience, most of the teachers were experienced as they had worked for over ten years (71% in Germany, 63% in Finland). Most participants were also female (67% in Germany, 76% in Finland). Regarding their formal training in formative assessment, 59% of the German teachers and 79% of the

Finnish teachers had received either some or extensive training in formative assessment during their teacher education.

In addition to the questionnaire responses, we interviewed 27 teachers: 26 from Germany and one from Finland. Most of the German teachers were female (65%) and worked in higher education (81%). Furthermore, they were experienced as most teachers had more than ten years of work experience (77%). Regarding the teacher from Finland, she worked in higher education and had six to ten years of work experience. It is important to note that more teachers from Germany than Finland reacted to our invitation to participate in the interviews, which led to such an unbalanced sample.

Results

The following section presents the results of quantitative and qualitative analysis. First, teachers' familiarity with various assessment purposes and the use of assessment for particular purposes are shown on the basis of quantitative data derived from the questionnaire. Then, based on the responses from the structured interviews, teachers' use of formative assessment in (ERT) is described with a special focus on various challenges related to providing feedback remotely and engaging students in self- and peer assessment under ERT conditions. Finally, teachers' self-reported training needs with regard to various forms of formative assessment are presented. Whenever information from the structured interviews is quoted, the data have been processed anonymously, and each teacher has been assigned a number (T1 to T27).

Teachers' familiarity with assessment purposes and the implementation of assessment methods

We investigated how familiar English teachers were with the following types of assessment purposes and their application**.** The responses are presented collectively for Germany and Finland in Table 1 (n = 115).

Table 1. English teachers' perceptions of their familiarity of assessment purposes (n = 115)

		Not familiar at all	Slightly familiar	Familiar	Very familiar
1.	To evaluate progress in language learning	3%	11%	48%	38%
2.	To evaluate achievement in language learning	3%	12%	45%	40%
3.	To diagnose learners' strengths and weaknesses	2%	13%	42%	43%
4.	To motivate student learning	3%	10%	40%	47%
5.	To provide feedback	2%	5%	35%	58%

The familiarity of the items is discussed by merging the percentages of "familiar" and "very familiar." The results presented in Table 1 clearly show that most of the teachers were familiar with the purposes of assessment listed above. They were most familiar with providing feedback (93%) and least familiar with evaluating achievement in language learning (85%) and diagnosing learners' strengths and weaknesses (85%).

The aim of the next item of the questionnaire was to discover how familiar the teachers were with the following types of alternative assessments and their application. The results are shown in Table 2.

Table 2. English teachers' overall familiarity with alternative assessment (n = 115)

	Not familiar at all	Slightly familiar	Familiar	Very familiar
Portfolio assessment	6%	20%	37%	37%
Self-assessment	2%	23%	44%	31%
Peer assessment	3%	19%	48%	30%

As the table illustrates, most teachers were familiar with these assessment methods (portfolio, 74%; self-assessment, 75%; and peer assessment, 78%), and no major differences between various types of alternative assessment were detected.

Forms of formative assessment implemented during ERT

The interviewees described their assessment practices in ERT without explicitly identifying their summative or formative purpose. The most prevalent forms of assessment were written exams (taken either online under exam conditions or offline, the so-called open book exams), written assignments, oral presentations via videoconferencing and portfolio assessment (for details, see Mäkipää & Soltyska,

2023). However, for reasons of a practical and didactic nature, many of these assessment practices took place at several points in time throughout the semester, and were regularly followed by teachers' feedback, which points at their formative character. Furthermore, an enhanced use of automated feedback features in various Learning Management Systems (LMS) used at the respondents' institutions (e.g., H5P[5] activities in Moodle) was reported and emphasized as an emerging common component of formative assessment.

Ways of providing feedback in remote teaching

When asked how they provided feedback to their learners in remote teaching, most interviewees admitted that they had provided feedback in writing, usually on students' oral or written performance. Written feedback on students' texts was given in the form of annotations, individualized or pre-written comments or correction codes. As a rule, communication related to feedback happened remotely, i.e., the texts were submitted by email or via LMS, and oral performance, most typically presentations, was delivered in videoconferences; the feedback that followed was communicated digitally, too. All these communication measures were in line with school and university closures and local restrictions on in-person contacts in Finland and Germany aimed at containing the COVID-19 pandemic at the time.

The second most common way of providing feedback in remote conditions was oral feedback, typically provided during video conferences held either with an individual or with several students at a time. This practice was commonly used for speaking activities such as presentations, but also for written production and other skills. If several students were attending a videoconference simultaneously, breakout rooms were sometimes used to allow for some privacy when communicating feedback. A special type of feedback provided orally was mentioned by several interviewees: instead of writing up their comments, teachers made voice recordings of their observations and suggestions related to students' essays or presentations and uploaded these audio files to LMS so that the students could listen to them at a time of their choice.

[5] H5P is the abbreviation for HTML5 package, which can be used to create interactive content for the web and Moodle (e.g. quizzes, timelines, memory games, drag-and-drop tasks, picture puzzles and cloze tests) whereby users typically obtain instant automated feedback on the accuracy of their responses.

Further ways of providing feedback included using pre-defined grading sheets, often embedded in LMS, which listed detailed assessment criteria and documented students' fulfillment thereof. Interestingly, some respondents pointed at using automated feedback provided by AI-based applications, e.g., within LMS such as feedback on discussion forum entries or interactive H5P activities in Moodle, the experience which they clearly linked to the transition to online assessment during the pandemic.

As far as the differences in providing feedback in in-person and online/remote teaching are concerned, almost half of the respondents noticed no major differences between the two modes. However, opinions differed as to what was possible or not possible in remote teaching. On the one hand, providing feedback remotely enabled personal and individualized feedback encounters (e.g., in breakout rooms during videoconferences), with no other students accidentally overhearing comments directed at another student. As far as the level of students' engagement is concerned, some teachers observed that students were more likely to ask for feedback in online environments and they generally learned quickly to actively participate in feedback conversations. Others, however, pointed at reduced interest in interaction with a teacher and as a result, the superficial character of feedback provided online. What was also missing in remote settings in teachers' views was the spontaneity, as feedback was often deferred to avoid interruptions in online communication, and intensity of interaction, as feedback givers and recipients were detached by computer screens. As one of the respondents noticed "the teacher's demeanor is already a form of feedback, and that component was almost non-existent in online teaching" (T3).

Teachers' views on the qualities of meaningful feedback

The most common feature of meaningful feedback as seen by the respondents is that it should enable improvement and encourage further learning. This finding was illustrated by the two excerpts from the participants below:

> "Feedback needs to be a place where they [students] can grow from and improve whatever skill it is they are working on" (T24),
>
> "Feedback at the best of times should be a starting point for future learning" (T5).

Furthermore, feedback should be focused on actual target language use – as one of the interviewees observed, "Meaningful feedback … is something that the students can actually use to improve if they have something similar in the future" (T23). It should also be balanced by including the strengths and weaknesses of a given performance that is being commented on. To be actionable, feedback must be clear and understandable, which is often dependent on a given context: a student in question, their needs and their feedback reception capacity, including language proficiency. In the eyes of T1, "Feedback is only meaningful if you know that the student can make something out of it and use it for their learning" (T1). To be efficient, feedback should be initiated by a student and as such, be voluntary and not imposed by a teacher, as according to one teacher "They [students] have to have the desire and the willingness to change" (T13), and it should require interaction between the feedback provider and recipient, because as T5 sees it "a one-way sort of thing… you hand it over and there is very little engagement… is not very useful" (T5). It should also be actionable and include practicable suggestions for improvement. The recommendation voiced by one of the interviewees was to "Tell [the students] what is wrong and why it's wrong and give [them] the tools…, 'cause without the tools to fix it, the feedback is meaningless" (T13).

When asked whether they thought students understand the feedback they received, most interviewees responded affirmatively (21 positive responses against only one negative; three interviewees said it depends) and provided arguments to support their perspectives. Some teachers felt that the sign of improvement could be attributed to the action of feedback. Other teachers provided examples of how they ensured the feedback was understood (e.g., providing feedback in students' L1, illustrating feedback with numerous examples, inviting questions on the feedback received, and offering to talk in person about the written feedback). Among the interviewees who were undecided whether their feedback was understood, the main reasons cited were students' language skills that were insufficient for engaging with the feedback, cultural differences, or their inability to cope with negative feedback.

Challenges in providing feedback

The major challenges in providing feedback that the respondents faced referred to time constraints, students' engagement with feedback, and the technical aspects of

providing feedback remotely. Teachers often felt overwhelmed by the volume of feedback activity expected of them, both in terms of the number of submissions to be commented on and the depth of comments and suggestions. As one of the respondents observed, the real challenge to be faced was "How do we manage giving good, meaningful feedback without burning ourselves out? That is to me the hardest thing." (T25).

This type of pressure, coupled with missing appreciation on the part of some students or no noticeable further engagement with the feedback delivered (e.g., evidence of students not having read the feedback they received or no progress despite clear suggestions for improvement), was perceived as another challenge. In a similar vein, dealing with students who see feedback conversations as merely an opportunity to negotiate their grade as a testing experience proved difficult and unrewarding. Focusing on cultural differences that might affect the process of giving and receiving feedback was seen as another challenge. Some teachers stressed special attention should be paid to being respectful and polite as well as wording one's critical comments carefully as members of other cultures and speakers of other languages might interpret feedback comments in a manner that was different from that intended. Finally, with a clear reference to providing feedback in remote settings, some teachers mentioned technical problems e.g., with the software and hardware used for communicating feedback to their students as well as technical constraints (e.g., limited accessibility to some tools among students or limitations of the software in use within a given institution). Here further limitations of feedbacking in remote settings were named as well:

> "You can't sneak into a room, [as] your presence is automatically noticed when you enter the [breakout] room. When you are in person, you can surreptitiously observe the students and make mental notes 'Huh, yeah, they got this, oh, that's gotten better, oh, we need to talk about this later…'. That really doesn't work online, the observation part." (T13).

Irrespective of the mode of providing feedback, most respondents felt that they give enough feedback to their students. Thirteen out of 27 interviewees responded positively to the question "Do you feel you give enough feedback to your students?", whereas only six gave a negative response. The remaining eight people were unsure whether the amount of feedback they provide to their students is sufficient. Interestingly, some instructors in fact said they give too much feedback and others

related the amount of feedback provided to the size of their course groups and actual feedback needs of course participants in question.

Teachers' views on self- and peer assessment in remote teaching

As self-assessment is an inherent part of formative assessment and as such should also be maintained in remote instruction, the interviewees were asked whether they encouraged their students to use self-assessments when teaching online and if so, how they did it. Most respondents (20 out of 27) responded positively and only four said they did not encourage this form of assessment in remote teaching at all. However, the way self-assessment was conducted and methods of incorporating it in online settings differed among the respondents. Some teachers asked their learners to study relevant level descriptors and level-specific *can-do* statements from the Common European Framework of Reference (CEFR) and to reflect on their actual language skills. In a similar vein, others familiarized their students with task-specific assessment criteria and asked for a self-evaluation of a given piece of work on the basis of these criteria.

Other teachers used various reflection tasks (e.g., an informal conversation about how students feel about their performance on a given activity), and self-assessment activities offered by textbooks (e.g., selecting smileys or traffic lights to indicate one's self-reported success rate). However, self-assessment was also understood as asking students to work on their own and to either receive a score (and feedback) automatically and instantly generated by an LMS or check their answers against an answer key or a model text when working with a printed textbook. This was done in the hope of developing students' self-awareness and promoting their autonomy, according to the teachers who encouraged students' involvement in these activities.

Similarly, almost all interviewees confirmed that they encouraged their students to engage in peer assessments in remote teaching (23 positive answers against three negative ones). In most cases, peer assessment was implemented to evaluate writing or presentation skills and took place in breakout sessions during video conferences. Some teachers used peer feedback or workshop function in an LMS[6], which helped to ensure fair allocation of workload and anonymity. Not only did many respondents

[6] The workshop feature of the Language Management System "Moodle" is a peer assessment activity whereby students add submissions which are then distributed amongst their peers for assessment based on a grading scale specified by the teacher.

describe how peer assessment was implemented in their online classrooms, but they also named several advantages of this procedure taking place online: breakout sessions offered more closed and secure environments in which students did not feel intimidated to share and receive feedback; collaborative writing tools enabled the exchange of ideas on each other's contributions and their evaluation against a set of criteria; and for a teacher or assessor, it was easier to focus their attention when following a presentation which could be seen on the screen of their own computer. This resulted in a higher quality of notes and accuracy of subsequent comments.

However, several challenges were reported regarding encouraging students to conduct self- and peer assessment remotely. These can be categorized as referring to students' attitudes to assessment or students' insufficient assessment literacy, and teachers' perceptions of self- and peer assessment. Another category of challenges related to pre-requisites for successful self- and peer assessment which are often associated with external conditions and not referring to personal traits of an assessing person. As one of the respondents put it, "The challenge is to draw the student into the purpose [of self- and peer assessment] so that they understand what the point is… It's just because you measure the pig more often, [it] doesn't mean that the pork tastes any sweeter." (T21).

Table 3 summarizes the common challenges related to these two types of assessment and uses teachers' sample responses to illustrate the difficulties that teachers are confronted with.

Table 3. Challenges that teachers faced when encouraging students to use self-assessment and peer assessment

Challenges related to students' attitudes or lacking assessment literacy	**Challenges related to teachers' attitudes or views on self- / peer assessment**	**Challenges related to external pre-requisites of successful self- / peer assessment**
• Students don't take self- / peer assessment seriously. • Students don't see the point in self- / peer assessment. • Students are not equipped to assess all aspects of language. • Students are reluctant to give negative feedback. • Students are not comfortable taking part in self- / peer assessment. • Students don't feel confident to give feedback to others. • Students overrate / underrate themselves in self-assessment. • Students are not able to thoroughly reflect on their own or peers' performance.	• Teachers are afraid that students see self- / peer assessment as offloading teachers' responsibility. • Teachers do not trust that students are doing the right thing. • Teachers perceive self- / peer assessment activities as time-consuming. • Teachers are overwhelmed by the organization and management of self- / peer assessment activities. • Teachers find it difficult to encourage students to speak up.	• Tasks for self- / peer assessment are not precisely defined and clearly described; the instructions from teachers are not always understandable and exhaustive. • Tools for self- / peer assessment that could facilitate the process are not available. • Not everyone involved in peer assessment is aware of data protection regulations and follows them accordingly, which might restrict some processes.

Teachers' perceived training needs with regard to providing feedback

The respondents suggested several aspects of providing feedback that they would like to receive training on. Table 4 summarizes interviewees' self-reported training needs categorized under four headings: feedback skills among teachers, understanding students' feedback needs, practical aspects of providing feedback, mostly related to time and efficiency, and other aspects of feedback literacy.

Table 4. Teachers' self-reported training needs with regard to providing feedback

Teachers' feedback skills	Understanding the students' perspective	Practical guidelines	Other aspects
• How to provide feedback spontaneously • How to give culturally sensitive feedback • How to make feedback concise – what to include and what not in one's feedback • How to word your feedback to avoid misinterpretations • What correction codes to use • How to strike the balance between positive and negative feedback • How to ask the right questions in feedback conversations	• How to learn what kind of feedback the students want • How to react to students' feedback needs • How to motivate the students to make good use of feedback • How do students react when receiving feedback and what to do about it • How to better involve students in giving me feedback	• How to give appropriate feedback in an efficient way • How to create a good feedback rubric without spending too much time on it • How to give expert feedback quickly • How to select an appropriate form of (formative) assessment • What are the strengths and weaknesses of various forms of formative assessment for particular contexts and target groups	• What effects does feedback have on students according to research • How to conduct feedback conversations • How do other teachers shape their assessment and what do they think about it

Teachers' self-perceived needs with regard to self- and peer assessment

With regard to the training on self- and peer assessment, the most commonly expressed wish was to exchange ideas and experience and learn from other practitioners how they design and implement such assessment activities in their contexts. More specifically, the interviewees would like to learn about different (new) ways, formats, techniques of shaping self- and peer assessment, including these available in online settings. Furthermore, several interviewees also mentioned the need of learning about new findings from research to inform their classroom self- and peer assessment practices.

Discussion

This study was motivated by limited research on ERA, in particular on the use of formative assessment practices in remote teaching conditions. Understanding how a

typical formative assessment situation may be affected by a lack of personal contact between its stakeholders and computer-mediated communication is crucial for adjustments of assessment activities in post-pandemic conditions and beyond. Furthermore, previous studies of formative assessment have mostly focused on students' perceptions, with limited research on teachers' views. Hence, the results of this study advance our understanding of formative assessment in ERT from teachers' perspectives and offer useful recommendations for future teacher training in this regard.

The first research question focused on exploring the implementation of formative assessment as a part of ERA from the teachers' perspective. The respondents described various types of assessment which they used with their students, including written online and offline exams, oral presentations via videoconferencing, portfolio assessment and automated feedback provided by LMS tools. Without explicitly calling these assessment practices formative, the respondents clearly referred to their formative nature when describing the monitoring function of assessment tasks (e.g., "it's not for a grade, but so that I get an idea how much they understood", T22), their ongoing character (e.g., "Moodle tests every week", T26), collaborative nature (e.g., "they were allowed to work together online", T23; and "I ask them to have peer assessment before I give them feedback", T25) and often extensive feedback provided by the teacher. Most of the teachers also indicated that they had used both self- and peer assessment in ERT. Combining both self- and peer assessment in teaching is highly recommended as feedback from various sources allows for triangulation (Panadero & Lipnevich, 2022), which is also welcomed by students (Burner, 2023, p. 21). This suggests that these teachers support their students' learning in multifaceted ways.

As alternative assessment typically calls for teachers to accept new instructional and assessment roles, it is perceived as a challenge nevertheless and independent of surrounding conditions in which learning, teaching and assessment take place. It is thus understandable that most respondents in our study encountered challenges in applying formative assessment under ERA conditions, even if they had positive experiences with this type of assessment in in-person settings, an observation shared by other studies (e.g., Panadero et al., 2022, Senel & Senel, 2021). The challenges identified by the respondents referred mainly to the increased workload which they

had to manage in a limited time, teachers' lack of familiarity with technology they were asked to use for teaching and assessing remotely, technical difficulties which affected communication in technology-mediated conditions, and dealing with students' engagement in formative assessment practices. These findings are in line with previous studies on ERT (Bruce & Stakounis, 2021; Kılıçkaya, 2023; Sevimel-Sahin, 2023). However, the fact that the respondents did not mention insufficient test security and in broader terms issues of academic integrity in the list of challenges pertaining to formative assessment may be interpreted as a reinforcement of its monitoring and forming character rather than a summative and grade-awarding function. This might be a valuable lesson for post-pandemic technology-mediated assessment. Future decisions about proctoring online exams, using safe exam browsers and designing timed assessments should be guided by the formative or summative character of particular instances of assessment.

The second research question focused on providing feedback. Teachers evaluated feedback communicated digitally as both advantageous and disadvantageous and pointed at various methods of providing it in remote conditions. Even though it was characterized by reduced interpersonal interaction, with no reliance on extra-linguistic channels such as facial expression or body language (Guerra, 2023), and typically with no capacity for follow-up questions, it also enabled reflection (if communicated asynchronously) and revision of one's feedback (from the teacher's perspective) and re-reading the comments and engaging with them (from the students' perspective). Instantaneous feedback provided on the spot in in-person settings does not bear these features. Furthermore, digital feedback offered a wide variety of communication modes, which feedback providers and recipients could experiment with and subsequently choose the mode that is most efficient for them. While some respondents found using digital annotations, correction codes, pick-and-mix comment systems to be more time-efficient than previously used laborious hand corrections, others admitted that they had needed time to become skilled at using these methods. The same applied to providing recorded oral feedback – for some it was quicker to record a voice message than to formulate a text; for others the written medium was favored. These findings might encourage teachers to implement various modes of communicating feedback to their learners in future if technology-mediated communication e.g., through LMS is sustained in their environment. However, the

modality and particular tool of providing feedback might not be the most decisive factor, as it is the content of feedback that students regard as the most important aspect in formative assessment (Mäkipää & Hildén, 2021).

Interestingly, the amount of feedback that teachers provided during the pandemic seems not to have suffered due to the transition from in-person to remote instruction. Nevertheless, the responses and additional explanations obtained from the interviewees suggest that the teachers are rather critical of themselves and feel that there is "always room for more" (T16) with regard to providing feedback as one of the interviewees put it in a rather self-deprecating way. At the same time, the respondents are also aware of limitations beyond their control such as group size, students' expectations and familiarity with feedback tools and procedures imposed on teachers through their organizations. However, as the amount of feedback is not the most important factor in the implementation of feedback practices (Hattie & Clarke, 2019), more weight should be put on the impact of the feedback (i.e., to what extent students make sense of it, how they engage with it, and how they use it to enhance their learning). Although no single characteristic of feedback automatically enhances learning (Wiliam, 2013), this study indicates that carefully wording one's feedback and ensuring the recipient understands it in the intended way is crucial. This is even more important in multi-cultural groups and when communicating feedback remotely or asynchronously, particularly when no immediate correction or rectification is possible.

As shown by this study, teachers' awareness of impactful feedback is mirrored by their concern about how to motivate and sufficiently equip their students to engage efficiently in self- and peer assessment. This finding, supported by the data on overall familiarity with these forms of formative assessment, clearly indicates the need for professional training in this field, irrespective of the mode of assessment.

This connects to the study's third research question, which concerns English teachers' self-reported training needs regarding formative assessment. Overall, the quantitative results indicate that the respondents were familiar with several formative assessment practices. However, the teachers irrespective of their previous formal education and current place of work pointed out several topics for training needs concerning feedback, such as practical aspects related to time and efficiency, feedbacking skills among teachers, and understanding students' feedback needs. Given that using

formative assessment efficiently in teaching is a key aspect of assessment literacy (Xu & Brown, 2016), future training in feedback should consider these themes to increase teachers' feedback skills and to support their development of assessment literacy. Furthermore, considering that language teachers find it essential to provide feedback in language learning (Li & Gu, 2023; Yang & May, 2023), they require up-to-date, research-based training in feedback to implement it comprehensively into their language teaching. In this respect the experience of providing feedback under ERA conditions seems to have "highlighted gaps in assessment knowledge which existed before the shift to digital assessment" (Thai et al, 2023, p. 178). However, the shift to online teaching seems to have resulted in increased reflection on the role of formative assessment, higher recognition of the importance of feedback and awareness of diverse challenges involved (Jiang & Yu, 2021).

Despite the novelty of the study and possible applicability of the results, it is important to acknowledge some limitations. The study focused only on English teachers' perceptions, but formative assessment is learner-centered, so other stakeholders, such as learners, and parents, need to be included in the study, to provide a more comprehensive picture of formative assessment practices in foreign language ERT. Furthermore, teachers' assessment practices, including formative assessment, have evolved over the course of the ERT, as have their views. The study was conducted between October 2021 and March 2022 and includes respondents' recollections on remote assessment practices until then. Consequently, as it is not always possible to pinpoint specific assessment practices to a particular phase of ERT (cf. Hill 2020) and external variables which might have affected the choice and implementation of assessments, the results of this study cannot be generalized as being representative of the whole duration of pandemic-restricted teaching. Moreover, due to the imbalance of the datasets (e.g., the number of teachers from both countries and the level they teach at), the results are tentative and might be skewed given several differences between the two educational systems and the role that assessment plays in both of them.

Conclusion

Focusing on ERT, this study offers insights into the formative assessment practices of teachers of English at various levels of education during the COVID-19 pandemic and

contributes to the discussion on language assessment literacy in the increasingly digitalized world. In particular, the findings increase the understanding of formative assessment and feedback from the teachers' perspective and point to the training needs of teachers in this regard. The examples of assessment practices reported in this study, along with their strengths and weaknesses and the different types of challenges associated with them, can serve as a source of inspiration and encouragement for practitioners and policymakers in the field of language education.

Teachers' awareness of the importance of feedback for their learners' progress, together with their self-reported need for further training in the area of formative assessment, gives hope that this type of assessment can be further implemented and developed in the future. However, according to Burner (2023, p. 24), "concerted coordination of policy, professional development, and practice in classrooms and schools" is essential to make formative assessment or assessment for learning happen. With this in mind, even though putting into effect any finding from this study seems worthwhile, the recommendations should be considered in a wider context of adjusting formative assessment practices in post-pandemic conditions.

Author disclosures

The authors declare no conflict of interest.

The authors contributed equally to the research, analysis, and writing of this article.

ORCID iDs

Toni Mäkipää https://orcid.org/0000-0003-3598-8840

Anna Soltyska https://orcid.org/0000-0003-4587-6113

References

Alderson, J. C. & Banerjee, J. (2001). Language testing and assessment, part 1. *Language Teaching*, *34*(4), 213–236. https://doi.org/10.1017/S0261444800014464

Andrade, H. L. (2018). Feedback in the context of self-assessment. In A. A. Lipnevich & J. K. Smith (Eds.), *The Cambridge handbook of instructional feedback* (pp. 376–408). Cambridge University Press. https://doi.org/10.1017/9781316832134.019

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education, 4*(87), 1–13. https://doi.org/10.3389/feduc.2019.00087

Bachman, L. & Palmer, A. (1996). *Language testing in practice*. Oxford University Press.

Black, P. (2010). Formative Assessment. In P. Peterson, E. Baker & B. McGraw (Eds.), *International Encyclopedia of Education* (3rd ed., pp. 359–364). Elsevier Science.

Black, P. & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education: Principles, Policy & Practice, (6),* 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Bozkurt, A. & Sharma., R. (2020). Emergency remote teaching in a time of global crisis due to corona virus pandemic. *Asian Journal of Distance Education, 15*(1), 1–6. https://doi.org/10.5281/zenodo.3778083

Bruce, E. & Stakounis, H. (2021). *The impact of Covid-19 on the UK EAP sector during the initial six months of the pandemic. BALEAP-funded report*. https://www.baleap.org/wp-content/uploads/2021/06/BALEAP-Report-Covid-and-EAP-May-2021.pdf

Bruce, E. & Stakounis, H. (2022). Emergency remote assessment (ERA) narratives from the UK English for academic purposes (EAP) sector. In K. Sadeghi (Ed.), *Technology-assisted language assessment in diverse contexts. Lessons from the transition to online testing during COVID-19* (pp. 34–53). Routledge. https://doi.org/10.4324/9781003221463

Burner, T. (2023). A review of formative language assessment research and implications for practitioners. In S.W. Chong & H. Reinders (Eds.), *Innovation in learning-oriented language assessment. New language learning and teaching environments* (pp. 13–26). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-18950-0_2

Butler, Y. G. (2023a). Self-assessment in second language learning. *Language Teaching,* 1–15. https://doi.org/10.1017/S0261444822000489

Butler, Y. G. (2023b). Expanding the role of self-assessment: From assessing to learning English. In D. Valente & D. Xerri (Eds.), *Innovative practices in early English language education* (pp. 191–210). Palgrave Macmillan.

Carless, D. & Boud, D. (2018). The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*, *43*(8), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Cooper, A., Timmons, K. & MacGregor, S. (2021). Exploring how Ontario teachers adapted to learn-at-home initiatives during COVID-19: Blending technological and pedagogical expertise in a time of growling inequities. *Journal of Teaching and Learning, 15*(2), 81–101.

https://doi.org/10.22329/jtl.v15i2.6726

Cornish, F., Gillespie, A. & Zittoun, T. (2014). Collaborative analysis of qualitative data. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 79–93). SAGE Publications Ltd. https://doi.org/10.4135/9781446282243

Dawson, P., Henderson, M., Ryan, T., Mahoney, P., Boud, D., Phillips, M. & Molloy, E. (2023a). Technology and feedback design. In J. M. Spector, B.B. Lockee & M.D. Childress (Eds.), *Learning, design, and technology. An international compendium of theory, research, practice, and policy* (pp. 695–739). Springer.

Dawson, P., Yan, Z., Lipnevich, A., Tai, J., Boud, D. & Mahoney, P. (2023b). Measuring what learners do in feedback: the feedback literacy behavior scale. *Assessment & Evaluation in Higher Education*, 1–15. https://doi.org/10.1080/02602938.2023.2240983

Dörnyei, Z. (2007). *Research methods in applied linguistics.* Oxford University Press.

Dörnyei, Z. & Taguchi, N. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). Routledge.

Double, K. S., McGrane, J. A. & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, *32*, 481–509.

https://doi.org/10.1007/s10648-019-09510-3

Drucker, J. D. & Fleischhauer, K. (2021). Language Pedagogy in a pandemic: The shift to online instruction at a German university during the COVID-19 crisis. *Journal of Pedagogical Research, 5*(1), 172–187. https://doi.org/10.33902/JPR.2021167474

Escalante, J., Pack, A. & Barrett, A. (2023). AI-generated feedback on writing: insights into efficacy and ENL student preference. *International Journal of Educational Technology in Higher Education 20,* 57 (2023). https://doi.org/10.1186/s41239-023-00425-2

European Commission. (2022). Digital economy and society index (DESI) 2021. *Thematic chapters*. https://digital-strategy.ec.europa.eu/en/library/digital-economy-and-society-index-desi-2022

Gu, P. Y. (2021). An argument-based framework for validating formative assessment in the classroom. *Frontiers in Education, Volume 6*. https://doi.org/10.3389/feduc.2021.605999

Guerra, J. (2023). And now what (Part 2)? Language teaching: what did we learn and what remains after the lockdowns? *The American Journal of Social Science and Education Innovations, 5*(6), 18–25. https://doi.org/10.37547/tajssei/Volume05Issue06-05

Hattie, J. & Clarke, S. (2019). *Visible learning: Feedback*. Routledge.

Herman, J. L., Aschbacher, P. R. & Winters, L. (1992). *A practical guide to alternative assessment*. Association for Supervision and Curriculum Development.

Hill, P. (2020). *Revised outlook for higher ed's online response to COVID-19*. https://onedtech.philhillaa.com/p/revised-outlook-for-higher-eds-online-response-to-covid-19

Ho, M.C. (2015). The effects of face-to-face and computer-mediated peer review on EFL writers' comments and revisions. *Australasian Journal of Educational Technology, 31*(1), 1–15. https://doi.org/10.1080/10494820.2014.994532

Hodges, C., Moore, S., Lockee, B., Trust, T. & Bond, A. (2020). *The difference between emergency remote teaching and online learning*.

https://er.educause.edu/articles/2020/3/the-difference-between-emergency-remote-teaching-and-online-learning

Härmälä, M. & Marjanen, J. (2022). *Englantia koronapandemian aikaan. A-englannin osaaminen 9. luokan lopussa keväällä 2021 [English during the corona pandemic. Learning outcomes of A level English at the end of the 9th grade in spring 2021].* Kansallinen koulutuksen arviointikeskus.

Inbar-Lourie, O. (2013). Guest editorial to the special issue on language assessment literacy. *Language Testing 30*(3), 301–307. https://doi.org/10.1177/0265532213480126

Janisch, C., Liu, X. & Akrofi, A. (2007). Implementing alternative assessment: Opportunities and obstacles. *The Educational Forum, 71*(3), 221–230.

Jiang, L. & Yu, S. (2021). Understanding changes in EFL teachers' feedback practice during COVID-19: Implications for teacher feedback literacy at a time of crisis. *The Asia-Pacific Education Researcher, 30*(6), 509–518.

Jónsson, Í. R., Smith, K. & Geirsdóttir, G. (2018). Shared language of feedback and assessment. Perception of teachers and students in three Icelandic secondary schools. *Studies in Educational Evaluation, 56*, 52–58. https://doi.org/10.1016/j.stueduc.2017.11.003

Kılıçkaya, F. (2023). Transition to online assessment: Opportunities and challenges for language lecturers in the EFL tertiary context. In M.-M. Suárez & W. M. El-Henawy (Eds.), *Optimizing online English language learning and teaching* (pp. 153–170). Springer.

Kremmel, B. & Harding, L. (2020). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, *17*(1), 100–120. https://doi.org/10.1080/15434303.2019.1674855

Li, J. & Gu, P. Y. (2023). Developing classroom-based formative assessment literacy: An EFL teacher's journey. *Chinese Journal of Applied Linguistics*, *46*(2), 198–218. https://doi.org/10.1515/CJAL-2023-0204

Li, H., Xiong, Y., Hunter, C. V., Guo, X. & Tywoniw, R. (2020). Does peer assessment promote student learning? A meta-analysis. *Assessment & Evaluation in Higher Education*, *45*(2), 193–211.

https://doi.org/10.1080/02602938.2019.1620679

Link, S. & Koltovskaia, S. (2023). Automated Scoring of Writing. In: O. Kruse et al. *Digital Writing Technologies in Higher Education*. Springer, Cham. https://doi.org/10.1007/978-3-031-36033-6_21

MacBeath, J., Schratz, M., Meuret, D. & Jakobsen, L. (2000). *Self-Evaluation in European Schools: A Story of Change*. Routledge/Falmer.

McCarthy, J. (2016). Global learning partnerships in the Café: Peer feedback as a formative assessment tool for animation students. *Interactive Learning Environments, 24(*6), 1298–1318.

Mizumoto, A. & Eguchi, M. (2023). Exploring the potential of using an AI language model for automated essay scoring. *Research Methods in Applied Linguistics, 2*(2023) 100050. https://doi.org/10.1016/j.rmal.2023.100050.

Mäkipää, T. (2021). Students' and teachers' perceptions of self-assessment and teacher feedback in foreign language teaching in general upper secondary education – A case study in Finland. *Cogent Education, 8*(1), 1–19. https://doi.org/10.1080/2331186X.2021.1978622

Mäkipää, T. (2023). Students' perceptions of teacher assessment practices in foreign language emergency remote teaching during Covid-19 in Finland. *Studies in Language Assessment, 12*(1), 1–25. https://doi.org/10.58379/AHRY2151

Mäkipää T., & Hildén, R. (2021). What kind of feedback is perceived as encouraging by Finnish general upper secondary school students? *Education Sciences, 11*(12), 1–15. https://doi.org/10.3390/educsci11010012

Mäkipää, T., & Soltyska, A. (2023). "We don't know what we don't know": Classroom assessment literacy for remote and online contexts – Insights from an inter-country comparative study. In L. Gokturk-Saglam & E. Sevgi-Sole (Eds.), *Emerging practices for online language assessment, exams, evaluation and feedback* (pp. 184–208). IGI Global.

Nicol, D. (2021). The power of internal feedback: Exploiting natural comparison processes. *Assessment & Evaluation in Higher Education*, *46*(5), 756–778. https://doi.org/10.1080/02602938.2020.1823314

Panadero, E., Fraile, J., Pinedo, L., Rodríguez-Hernández, C. & Díez, F. (2022). Changes in classroom assessment practices during emergency remote

teaching due toCOVID-19. *Assessment in Education: Principles, Policy & Practice, 29*(3), 361–382. https://doi.org/10.1080/0969594X.2022.2067123

Panadero, E. & Lipnevich, A. A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements. *Educational Research Review, 35*, 100416. https://doi.org/10.1016/j.edurev.2021.100416

Senel, S. & Senel, H. C. (2021). Remote assessment in higher education during COVID-19 pandemic. *International Journal of Assessment Tools in Education, 8*(2), 181–199. https://doi.org/10.21449/ijate.820140

Sevimel-Sahin, A. (2023). Academic integrity in online foreign language assessment: What does current research tell us? In D. Köksal, N. K. Ulutas & S. Arslan (Eds.), *Handbook of Research on Perspectives in Language Assessment* (pp. 306–328). IGI Global.

Smith, K., Gamlem, S. M., Sandal, A. K. & Engelsen, K. S. (2016). Educating for the future: A conceptual framework of responsive pedagogy. *Cogent Education*, *3*(1), 1227021. http://dx.doi.org/10.1080/2331186X.2016.1227021

Suzuki, W., Nassaji, H. & Sato, K. (2019). The effects of feedback explicitness and type of target structure on accuracy in revision and new pieces of writing. *System*, *81*, 135–145. https://doi.org/10.1016/j.system.2018.12.017

Thai, T., Sheehan, S., Nguyen, Q. T. N., Nguyen, T. T. P., Nguyen, Y. T. Q., Tran, H. T. T., Nguyen, C. T., Nguyen, H. Q. & Bui, S. (2023). Digital assessment literacy: English assessment practice in Vietnam. In A. L. Gokturk-Saglam & E. Sevgi-Sole (Eds.), *Emerging Practices for Online Language Assessment, Exams, Evaluation, and Feedback* (pp. 162–183). IGI Global.

Topping, K. J. (2017). Peer assessment: Learning by judging and discussing the work of other learners. *Interdisciplinary Education and Psychology*, *1*(1), 1–17. https://doi.org/10.31532/InterdiscipEducPsychol.1.1.007

Topping, K. J. (2018). *Using peer assessment to inspire reflection and learning*. Routledge. https://doi.org/10.1558/wap.v7i2-3.26457

Virkkunen, P. & Toivola, M. (2023). Lukio-opiskelijoiden käsityksiä vieraan kielen ääntämisen opetuksessa saadusta palautteesta [High school students' perceptions of feedback on foreign language pronunciation]. In T. Mäkipää, R. Hilden & A. Huhta (Eds.), *Kielenoppimista tukeva arviointi* [Assessment for

supporting language learning] (pp. 40–59). AFinLA-teema. https://doi.org/10.30660/afinla.121255

Vogt, K. & Tsagari, D. (2014). Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly*, *11*(4), 374–402. https://doi.org/10.1080/15434303.2014.960046

Wiliam, D. (2013). Assessment: The bridge between teaching and learning. *Voices from the Middle*, *21*(2), 15–20.

Winstone, N., Boud, D., Dawson, P. & Heron, M. (2022). From feedback-as-information to feedback-as-process: a linguistic analysis of the feedback literature. *Assessment & Evaluation in Higher Education*, *47*(2), 213–230. https://doi.org/10.1080/02602938.2021.1902467

Xiao, Y. & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System*, *81*, 39–49. https://doi.org/10.1016/j.system.2019.01.004

Xu, Y. & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, *58*, 149–162. https://doi.org/10.1016/j.tate.2016.05.010

Yan, Z. & Brown, G. T. L. (2017). A cyclical self-assessment process: towards a model of how students engage in self-assessment. *Assessment & Evaluation in Higher Education*, *42*(8), 1247–1262. https://doi.org/10.1080/02602938.2016.1260091

Yang, L. F., Liu, Y. & Xu, Z. (2022). Examining the effects of self-regulated learning-based teacher feedback on English-as-a-foreign-language learners' self-regulated writing strategies and writing performance. *Frontiers in Psychology*, *13*, 1–16. https://doi.org/10.3389/fpsyg.2022.1027266

Yang, L. F., Zhang, L. J. & Dixon, H. R. (2023). Understanding the impact of teacher feedback on EFL students' use of self-regulated writing strategies. *Journal of Second Language Writing*, *60*, 101015. https://doi.org/10.1016/j.jslw.2023.101015

Yang, R. & May, L. (2023). Understanding and implementing learning-oriented assessment: Evidence from three Chinese junior high school EFL teachers. *Chinese Journal of Applied Linguistics*, *46*(2), 308–325. https://doi.org/10.1515/CJAL-2023-0210

Yusop, F. D., Firdaus, A. & Syed Omar, S. N. S. (2022.) Introduction to alternative assessment In F. D. Yusop & A. Firdaus (Eds.) *Alternative assessments in Malaysian higher education voices from the field* (pp. 1–10). Springer.

Zhang, C., Yan, X. & Wang, J. (2021). EFL teachers' online assessment practices during the COVID-19 pandemic: Changes and mediating factors. *The Asia-Pacific Education Researcher*, *30*, 499–507. https://doi.org/10.1007/s40299-021-00589-3

Appendix

Section 1: General information

1. Place of work: Are you a teacher in
   - Finland
   - Germany

2. Place of work: What type of school/institution do you teach at? (Multiple responses are possible)

Answers for teachers in Finland:
   - Primary school
   - Lower secondary school (grades 7-9)
   - Upper secondary school (grades 10-12)
   - Vocational school
   - Higher education
   - Other (please specify what type of school or institution you teach at)

Answers for teachers in Germany:
   - Primary school
   - Secondary school (grades 5-10/13)
   - Vocational school
   - Higher education
   - Other (please specify what type of school or institution you teach at)

3. What subject(s) do you teach apart from English?

4. Employment conditions: What kind of employment contract do you currently have:
   - Permanent contract
   - Contract for a limited period of time incl. substitute teacher
   - No employment contract, but freelance or occasional substitute work for a school or institution

5. Place of living: What is the population of the place (city, town, municipality) you live in?
   - under 10,000 inhabitants
   - 10,000 – 50,000 inhabitants
   - 50,000 – 100,000 inhabitants
   - 100,000 – 500,000 inhabitants
   - 500,000 – 1,000,000 inhabitants
   - over 1,000,000 inhabitants

6. Teaching experience: How much teaching experience do you have?
   - less than 1 year
   - 1 to 5 years
   - 6 to 10 years
   - 11 to 20 years

- more than 20 years

7. Age: How old are you?
   - 20-29 years
   - 30-39 years
   - 40-49 years
   - 50-59 years
   - 60 years or older

8. Gender:
   - Male
   - Female
   - Other
   - Prefer not to say

Section 2: Assessment-related teacher education

9. During your formal teacher education, how much training did you receive in the following assessment-related fields:

	no training	some training (the topic was mentioned, but not thoroughly explored/explained)	extensive training (the topic was thoroughly explored/explained)	I don't remember how much training I received
purposes of language assessment				
principles of good assessment				
formative assessment				
summative assessment				
forms of alternative assessments				
assessment-related terminology				
providing meaningful feedback				
creating rating scales				
using rating scales				
interpreting test results				
designing online assessment				

10. If you received thorough training dedicated to any specific aspect of assessment mentioned above, please let us know the details of this training you remember (title/focus, length, context, etc.).

Section 3: Assessment-related knowledge, abilities and confidence

11. How familiar are you with the following terms and their application?

	I don't know the term at all	I have a basic understanding of the term	I have a working understanding of the term	I am confident I understand and can apply the term
reliability				
validity				
fairness				

12. How familiar are you with the following types of assessment purposes and their application?

	not at all familiar with this type	slightly familiar – I have a basic understanding of this type	familiar - I have a working understanding of this type	very familiar - I am confident in my understanding and ability to apply this type
to evaluate progress in language learning				
to evaluate achievement in language learning				
to diagnose learners' strengths and weaknesses				
to motivate student learning				
to provide feedback				

13. How familiar are you with the following types of alternative assessments and their application?

	not at all familiar with this type	slightly familiar – I have a basic understanding of this type	familiar - I have a working understanding of this type	very familiar - I am confident in my understanding and ability to apply this type
portfolio assessment				
self-assessment				
peer assessment				

Section 4: Experience with emergency online teaching and assessment

14. Has the Covid-19 pandemic affected your assessment practice(s)? If so, how?
15. To what extent do you agree with the following statements:

	I don't agree at all	I somewhat disagree	I somewhat agree	I fully agree	I have no opinion

The changes in my assessment practices have been initiated by the institution/school (e.g. not to lower students' grades).					
I have made substantial changes to my assessment practices during remote teaching.					
The changes that I have introduced to my assessment practices are based on theory and research rather than impromptu decisions.					
I have received some external support (training and/or resources) from my institution/school or from a regional/national organisation when reshaping assessment for remote teaching.					
Overall, I am open-minded towards developing new assessment methods.					

Section 5: Professional development in the field of language assessment

16. How much training do you feel you need in the following domains?

	I need no training	I need some training (several hours)	I need extensive training (several days)
Choosing assessment procedures for your learners			
Designing online tests/assessment			
Preparing classroom tests/assessment			
Designing and implementing rating scales			
Giving feedback to students based on information from various forms of assessment			
Using informal, continuous, non-test type of assessment			
Integrating self-assessment in your class/course			
Integrating peer assessment in your class/course			
Using portfolio assessment			
Using the Common European Framework of Reference (CEFR) for assessment purposes			

17. How much training do you feel you need in the following domains?

	I need no training	I need some training (several hours)	I need extensive training (several days)
Assessing receptive skills (e.g., reading/ listening)			
Assessing productive skills (e.g., speaking/ writing)			
Assessing microlinguistic aspects (e.g., grammar/vocabulary)			
Assessing integrated language skills incl. language mediation (e.g., reading-into-writing)			
Assessing aspects of the target-language cultures (e.g., intercultural competence)			

18. How much training do you feel you need in the following domains?

	I need no training	I need some training (several hours)	I need extensive training (several days)
Establishing the reliability of an assessment			
Determining fairness in assessment			
Establishing the validity of an assessment			
Using statistics to analyse assessment quality			
Designing assessment for multicultural learner groups			

19. What type(s) of training would you be interested in attending and how extensive would you like it to be (teaching unit/TU equals 45 mins)?

	2 TU to half day	one full day (approx. 8 TU)	two full days (approx. 16 TU	an extensive training programme of at least one week and 40 TU	other length	I'm not interested in this type of training
Online training						
Face-to-face training						
Blended format (online and face-to-face)						

20. Would you be interested in attending training if it is (multiple answers possible):

- ...only provided within your regular working hours
- ...offered after hours i.e. beyond your regular working hours
- ...offered after hours i.e. beyond your regular working hours and you receive remuneration for training duration
- ...free of charge

21. Please share with us your comments on possible training formats and content that you would like to attend and suggest further conditions that would be relevant for your decision to participate (or not) in such training.

Section 6: Open questions (the answers can be given in English or German/Finnish language)

22. What aspect(s) of assessment do you feel you are good at and confident in? ...
23. What do you find most challenging about language assessment in your context? ...
24. If you would like to clarify any of your answers or comment on something that was not asked earlier, you can do it here. ...
25. Would you be available for a follow-up interview on assessment-related topics which will be conducted online via Zoom in 2022? Interviews will be carried out in German/Finnish or English. If you want to participate, please leave your e-mail address and/or phone number.

Semi-structured interview protocols

Section 1: Online assessment/assessment in remote teaching

1. What kinds of assessment practices have you used in emergency remote teaching?
2. How did you choose your assessment instruments in remote teaching?
3. What makes a good online test or online assessment instrument?
4. How much did you know about assessment in remote teaching before the pandemic?
5. What do you wish you had learnt or had been taught about (online) assessment during your studies? What did your studies not prepare you for in terms of (online) assessment?
6. What was most challenging and most rewarding about assessment in remote teaching?
7. What kind of training would you like to receive on assessment in ERT?

Section 2: Providing feedback

8. How did you provide feedback to your learners in remote teaching?
9. How do you usually give feedback (oral/written/non-verbal) to students in contact teaching?
10. Do you give enough feedback to your students? Please specify.
11. What does meaningful feedback mean to you?
12. What kinds of challenges do you face in providing feedback?
13. What kind of training would you like to receive on providing feedback?

Section 3: Self and peer assessment

14. Do you encourage your students to use self-assessments in remote teaching? If so, how?
15. Do you encourage your students to use peer assessments in remote teaching? If so, how?
16. What challenges do you face as a teacher when encouraging students to use these practices?
17. Do you think your students understand feedback (teachers' or peers')? Please specify.
18. What kind of training would you like to receive on self and peer assessment?

Section 4: Culture-related aspects of assessment

19. Do you teach multicultural classes? If yes, (how) does it influence your assessment practices?
20. Do you design your assessment with multicultural backgrounds of your learners in mind? Please specify.
21. Do you think assessment literacy training should include assessing multicultural classes? What would you like to learn on this topic?

22. Is there anything else you would like to add?