Contents

Editorial: Self- and peer assessment as fully fledged assessment methods in learning and teaching, especially in foreign languages (*Věra Janíková, Jana Veličková*) 365

Articles ____ **368**

Kateřina Keplová: Self-assessment, self-evaluation, or self-grading: What's in a name? 368

Stephan Schicker: Teachers' beliefs of pre-service German teachers about student self-assessment 383

Jana Veličková: Developing oral presentation-related self-assessment among prospective teachers of German as a foreign language: Analysis of self-assessment comments 410

Blanka Pojslová: The role and perception of peer and teacher feedback in multiple-draft feedback provision on foreign language learners' writing. . . . 444

Discussion ____ **484**

Martina Šindelářová Skupeňová: The more the merrier? Analysing self-assessment tool set use 484

Obsah

Editorial: Sebehodnocení a vrstevnické hodnocení jako plnohodnotné metody hodnocení ve vyučování a učení, především v oblasti cizích jazyků
(*Věra Janíková, Jana Veličková*) 365

Studie ______ 368

Kateřina Keplová: Sebehodnocení, sebe-evaluace, nebo sebe-známkování: Co je po jméně? 368

Stephan Schicker: Přesvědčení budoucích učitelů německého jazyka o sebehodnocení studentů 383

Jana Veličková: Rozvíjení sebehodnocení zaměřeného na ústní prezentace u budoucích učitelů němčiny jako cizího jazyka: Analýza sebehodnoticích výroků 410

Blanka Pojslová: Role a vnímání vícenásobné vrstevnické a učitelské zpětné vazby při rozvoji cizojazyčného psaní 444

Diskuze ______ 484

Martina Šindelářová Skupeňová: Čím více, tím lépe? Analýza využití sady sebehodnoticích nástrojů 484

Editorial: Self- and peer assessment as fully fledged assessment methods in learning and teaching, especially in foreign languages

At the outset of the thematic scope of this special issue, the question was raised: who should have control over the educational process in the classroom – the teacher or the students as well? This is also related to the question of whether assessing or evaluating the learning process of students solely belongs in the hands of the teacher or whether it can also be entrusted to the learners themselves.

On the one hand, there are proponents of exclusive teacher assessment, who critically view the limited objectivity, reliability, and validity of learner self- and peer assessment, considering these assessment methods only as supplementary (Vollmer, 2007, pp. 368–369). On the other hand, advocates for involving learners in managing and assessing their learning process emphasize the potential of self- and peer assessment for increasing learners' awareness of their learning process and their role within it (Vollmer, 2007, pp. 368–369). This, in turn, contributes to the development of responsibility for their learning (Ross, 1998, p. 2) and the desired learner autonomy (Tassinari, 2010). Furthermore, it can be assumed that the proper implementation of self- as well as peer assessment supports the effectiveness of the learning process (Boud, 2003, pp. 14–15) and has been empirically proven to enhance learning achievement (Hattie, 2018).

Due to the tendency in the educational and didactic discourse over the past 30 years towards considering assessment in education not only as assessment of learning but also as assessment for learning (Brown & Harris, 2013, p. 367) and assessment as learning (McMillan, 2013, pp. 4–6), we can observe that self- as well as peer assessment are gaining well-deserved prominence as fully fledged assessment methods. However, is the pace of increasing interest in these methods and their implementation in educational reality sufficient?

This special issue responds to the growing interest among practising teachers and researchers in self- and peer assessment in the context of foreign language learning and teaching and aims to support it further. The goal is to provide a platform for discussion on the current state of working on and with self- and peer assessment as well as for sharing of experiences with their

development, realization, and evaluation of their impact on foreign language learning and teaching. Considering the articles‘ focus, this issue is primarily intended for educators working in the tertiary education sector, especially those in teacher education.

In the first article, Kateřina Keplová addresses terminological questions and discusses the use of the terms "self-assessment" and "self-evaluation" in the area of teaching and learning a foreign language, taking into account their development in the fields of psychology and pedagogy.

The second text is a study by Stephan Schicker, who investigates the beliefs of Austrian and Czech pre-service teachers of German regarding student self-assessment, as this is a significant factor in whether and how teachers foster students‘ self-assessment skills.

In the third study, Jana Veličková also focuses on prospective teachers of German as a foreign language and their experience in developing their self-assessment skills, which is considered a precondition for effectively fostering self-assessment skills in their future learners. The author examines the characteristics of self-assessment comments provided by prospective teachers during an intervention aimed at developing self-assessment skills.

The next article is by Blanka Pojslová. She discusses the decision-making process regarding how and when to implement computer-mediated peer feedback in the classes. She presents a study that investigates whether incorporating peer feedback as a component of multiple-draft feedback provision, while following best practices in feedback, can contribute to improvements in the quality of learners‘ writing after they have been instructed on academic writing conventions and genre requirements.

In the final contribution of this issue, Martina Šindelářová Skupeňová introduces tools offered to university students at the beginning of the course "English Autonomously". The article explains how these tools are presented to students in introductory sessions, shows how individual students choose to use them in diverse ways, and discusses whether the toolset allows students to approach self-assessment in an individualized and efficient manner.

As evident from the contributions, self- and peer assessment are multifaceted constructs that rightly deserve recognition as fully-fledged assessment methods. We believe that the diverse perspectives on the exploration and

implementation of these methods in the context of foreign language learning and teaching presented in this issue will contribute to the ongoing discussion on self- and peer assessment.

Finally, we extend our gratitude to all the authors, reviewers, and members of the editorial team who played pivotal roles in bringing this special issue of the Journal of the Czech Pedagogical Society to fruition.

guest editors
Věra Janíková, Jana Veličková

References

Boud, D. (2003). *Enhancing learning through self-assessment.* Routledge Farmer.

Brown, G. T., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 367–393). Sage.

Hattie, J. (2018). *Hattie Ranking: 252 Influences And Effect Sizes Related To Student Achievement.* https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/.

McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 3–16). Sage.

Ross, S. (1998). *Self-assessment in second language testing: A meta-analysis and analysis of experimental factors. Language Testing*, *15*(1), 1–20.

Tassinari, M. G. (2010). *Autonomes Fremdsprachenlernen: Komponenten, Kompetenzen, Strategien.* Lang.

Vollmer, H. J. (2007). Leistungsmessung, Leistungskontrolle, Selbstbeurteilung: Überblick. In K.-R. Bausch, H. Christ, & H.-J. Krumm (Eds.), *Handbuch Fremdsprachenunterricht* (pp. 365–370). A. Francke.

Self-assessment, self-evaluation, or self-grading: What's in a name?

Kateřina Keplová

Department of English and American Studies, Faculty of Arts and Philosophy, University of Pardubice

Received 14th November 2022 / final version received 20th June 2023 / / accepted 22nd June 2023

Abstract: The ability of judging one's own performance seems to be an increasingly attractive topic in foreign language learning and teaching research. Although there have been numerous studies confirming the positive impact of such a competence of students at various levels of education on the improvement of the foreign language acquisition process, questions regarding terminology conventions remain and discussions continue.

The aim of this article is, therefore, to respond to some of those questions, namely: What should we call this ability, *self-assessment*? Is *self-assessment* the same as *self-evaluation*? What does it mean to *self-grade*? This article provides an insight into the origins of the terms *self-assessment* and *self-evaluation* and their development in the fields of psychology and pedagogy, before focusing on their use in the area of teaching and learning a foreign language.

Keywords: self-assessment, self-evaluation, self-grading, development of self-assessment

Studies show that developing the competence of self-assessment in students has a positive impact on the overall learning of foreign languages (see, for example, Hattie, 2018). One of the potential shortcomings of the studies on self-assessment appears to be the lack of agreement when it comes to terminology. Based on the conclusions of a recent literature review (Keplová, 2021), it seems that authors assume the readers' knowledge and understanding of the term and, therefore, do not feel the need to provide an explanation. What adds to the confusing situation is that authors of some studies might omit to clarify even the reason for choosing self-assessment as the specific activity to be carried out as part of their research. As Andrade (2019) suggests in her review, the purpose for which teachers and/or researchers use self-assessment activities shapes the procedure of conducting the self-assessment. Not only that, but the purpose may actually determine the terminology, i.e. whether the term *self-assessment* or *self-evaluation* is

https://doi.org/10.5817/PedOr2022-4-368

appropriate. The current situation resembles what Roeser et al. (2006, as cited in Panadero et al., 2016, p. 810) aptly describe in their study of the types of student self-assessment (SSA): "it seems reasonably clear that SSA is in danger of jingle-jangle fallacies in that different kinds of SSA are given the same name, while similar kinds of SSA are sometimes given different names."

This article, therefore, offers an overview of commonly used terms for the area of students making judgements on their own performance. It aims to specify the difference between *self-assessment*, *self-evaluation*, and *self-grading*, in order to define the specific uses of these terms.

The focus of this article being teaching and learning English as a foreign language, the most relevant supporting sciences to take into account when looking for definitions and uses of the terminology in question are psychology and pedagogy. Psychology provides the basis for the understanding of the self and its place in the personal development, including the development of foreign language competence. Pedagogy links the general ideas of language teaching and learning with the area of language assessment and evaluation but also introduces the term *self-grading*.

This article starts with a brief introduction of the terms assessment and evaluation, defining their original meanings and their adaptation in the fields of psychology and pedagogy. Further, the terms self-assessment, self-evaluation and self-grading are specified for the area of foreign language teaching and learning. This is by no means the ultimate guide to selecting the terminology. Nevertheless, it suggests one possible approach to the decision-making process.

1 The origins of *assessment* and *evaluation*

Originally, the meaning of the two terms, *assessment* and *evaluation*, was very similar. The 16th century term assessment originally stood for the "value of property for tax purposes" (Online Etymology Dictionary, n.d.), while the 18th century term *evaluation* meant "the action of appraising" (Online Etymology Dictionary, n.d.). In both cases, the outcome was a numerical estimate. However, these terms changed their meaning in due course and their use became more common outside the taxation and sales field. There are currently several definitions of *assessment* and *evaluation* available. From a general point of view, the online Cambridge Dictionary (n.d.) offers the

following definitions of the two terms: *assessment* is "the process of testing, and making a judgment about, someone's knowledge, ability, skills, etc., or the judgment that is made," while *evaluation* is "the process of judging or calculating the quality, importance, amount, or value of something."

To draw the distinction between *assessing* and *evaluating*, it is important to know what is being judged. Therefore, the terms learning *product* and learning *process* need to be clarified. Rephrasing Spada's (1987, p. 137) definitions of language learning programme product and process to reflect the learner-centred take on the terms, the *product* will be understood here as that which the learner produces in terms of language learning and the *process* as what the learner accomplishes in terms of learning practices and procedures. The product and process of learning form the objects of judgements and, therefore, will eventually determine the terminology to be used.

Narrowing the field of assessment and evaluation by restricting the agent making judgements, *self-assessment* and *self-evaluation* are the processes of students making judgments on their own learning products or their own learning processes. Before the details of further distinction are delved into, however, the use of *self-assessment* and *self-evaluation* from the point of view of psychology as a supporting science is clarified, then the point of pedagogy is taken into account and finally the view of teaching and learning English as a foreign language.

2 Self-assessment and self-evaluation in psychology

The concept of the Self is based on the widely used definition of *I-self* and *Me-self* as presented by William James (1892). James proposed that "the total self of me, being as it were duplex, partly known and partly knower, partly object and partly subject, must have two aspects discriminated in it, of which for shortness we may call one the *Me*, and the other the *I*." (1892, p. 176). *Me*, therefore, is the object and *I* is the subject of attention. James (1892, p. 187) established an equation to describe how a person determines their own worth, which he calls *self-esteem*:

$$\text{Self-esteem} = \frac{\text{Success}}{\text{Pretensions}}$$

The equation illustrates the way we estimate our value, i.e. self-esteem. The pretensions in the equation represent the expectations we set for ourselves. Self-esteem is formed based on the ratio between the success we experience and the expectations we set for ourselves. This means, each individual can influence their self-esteem by setting realistic goals for themselves and succeeding to reach those. The explanation appears to be close to the definitions of assessment and evaluation as already introduced in this article, i.e. valuing or appraising of something. The something, in this case, being the individual carrying out the valuing.

The ideas proposed by James have been expanded on, rather than contested, by symbolic interactionists such as C. D. Cooley and G. H. Mead (Blatný, 2010). Current research still operates with a Self as the subject and a Self as the object, two functions of the *self-system* which are different but not separate. Such research provides the basis for the study of self-regulation and, consequently, self-assessment.

The concept of self-regulation is also taken up by Mareš (2013), who works within the field of pedagogical psychology and sees the concept as a continual characteristic a learner might possess. The recurring theme in Mareš's (2013) account of self-regulation is the importance of aims, which learners set for themselves and, using self-selected strategies, monitor their own progress towards these aims by means of clear criteria. Mareš aptly comments that "[r]esearchers agree that [in terms of developmental changes of learner's self-regulation] it is a lifelong process" (2013, p. 235). This view is clearly in line with James's note of "our self-feeling is in our power" (James, 1892, p. 188).

The recurring theme of monitoring one's own work, i.e. paying close attention to the completed work, and using clear criteria to decide how well the work was completed, is important for understanding how close the terms *self-assessment* and *self-evaluation* actually are, but also how close they appear to be to other terms, such as *self-regulation* and *self-reflection*. Blatný (2010, p. 125) defines self-assessment as "the image a person has of themselves in terms of values and competences." Self-assessment, according to Blatný (2010), is the outcome of the monitoring of own activities and comparing the results, using set personal criteria, within the given social context. To be able to self-assess, a learner first needs to reflect on their work and, based on the purpose of the specific self-assessment, to recognise their

own strong and weak points. Once the learner is able to reflect on their work and can use clear criteria to recognise its worth, i.e. to assess, they are then able to regulate their learning to achieve their goals.

As shown above, the view of psychology is concerned mainly with the competences, the ability to achieve goals rather than the journey towards the goals and the way of acquiring the competences. In terms already defined, this view seems mainly focused on the product of learning rather than the process. This is further supported by a Dörnyei's framework (2009), which describes the various aspects of learners acquiring the abilities of making their own judgements about their own language production.

From the point of view of the psychology of foreign language acquisition, therefore, the process of self-monitoring is central to the framework prepared by Dörnyei (2009). The author builds on the ideas of Markus and Nurius (1986, as cited in Dörnyei, 2009) to define: "Ideal L2 Self, which is the L2-facet of one's ideal self"; "Ought-to L2 Self, which concerns the attributes that one believes one ought to possess to meet expectations and to avoid possible negative outcomes"; and "L2 Learning Experience, which concerns situated, 'executive' motives related to the immediate learning environment and experience" (Dörnyei, 2009, pp. 217–218). Dörnyei works with the concept of the learners monitoring both their learning process and performance to achieve the desired results, which offers a steppingstone to self-assessment and self-evaluation.

3 Self-assessment and self-evaluation in pedagogy

Trying to find the distinction between *assessment* and *evaluation* in pedagogy, or to determine whether there is a distinction, indeed, appears to be topic of some discussion as the terminology is determined by the perceived meaning of the phrase. It is, therefore, important to explore the meaning of *self-assessment* and *self-evaluation* in specific contexts. An early definition by Scriven (1981) actually states that *assessment* is "often used as a synonym for evaluation" (p. 12).

Reviews of pedagogical literature and research of self-assessment frequently include the key terms of *self-assessment* and *self-evaluation*, without further distinction made between the two. Some authors of reviews acknowledge work done on clarifying the distinction between the two terms, for example,

Brown and Harris (2013) admit that "distinguishing between assessment and evaluation has become commonplace in the [assessment for learning] community" (p. 369) but they make it clear that they do not share the same view: "it is our position that there is little merit in creating a dichotomy between assessment and evaluation" (p. 369).

The close relationship between the activity of assessing and evaluating is demonstrated in current research, for example, by Panadero et al. (2016): "Student self-assessment (SSA) most generally involves a wide variety of mechanisms and techniques through which students describe (i.e. assess) and possibly assign merit or worth to (i.e. evaluate) the qualities of their own learning process and products" (p. 804). Referring to a number of studies, Panadero et al. (2016) consider the types of self-assessment and the various methods of including *self-assessment* in the teaching and learning process to determine the correct terminology. Panadero et al. (2016) use the term self-assessment as the central term, the one which refers to the main activity of students' making judgements about their learning process and/or product. The authors provide various typologies of self-assessment, e.g. knowledge interest or student-teacher involvement typologies. They draw on their extensive review to distinguish, among many others, *self-grading*: "upon request by the teacher (e.g., via task) or a system (e.g., via computer) students assess at a surface level and mainly with summative purposes" (Panadero et al., 2016, p. 807). From this point of view, *self-grading* can be seen as simplified self-assessment or, perhaps, a preparatory stage of *self-assessment*, a step before students learn to work with criteria and delve into details of the language produced, i.e. their language product.

For further distinction in the terminology, it is important to consider learner self-regulation, learner autonomy and learner motivation. Each of these areas offers a unique take on the learners' ability to judge their own work and will be looked at in detail.

3.1 Learner self-regulation and self-assessment

Andrade (2010) argues that "self-regulation and self-assessment are complementary processes that can lead to marked improvements in academic achievement and autonomy" (p. 3). According to the author, "self-regulated learning is a dynamic process of striving to meet learning goals by generating, monitoring, and modifying one's own thoughts, feelings, actions and, to some

degree, context" (p. 5). In her later work, Andrade (2019) works with the term self-assessment as the overarching term for the student appraisal of their work. She points to a close link between *self-assessment* and *self-regulated* learning. Andrade's main focus is on the purpose of self-assessment, and she stipulates that for summative purposes, self-grading is suitable as the form of self-assessment for the product of learning (2019, p. 3).

Zimmerman (1998) makes the case that *self-evaluation* is an integral part of self-regulated learning which he sees as an essential ability not only for students but for everyone in their daily life. Zimmerman (1998, p. 83) proposes "a cyclical model of self-regulated learning" in which goal setting and strategic planning lead to strategy implementation and monitoring. This is then followed by strategic outcome monitoring which, in turn, feeds into self-evaluation and monitoring. This stage initiates a new cycle of goal setting and strategic planning and so on.

Zimmerman's choice of *self-evaluation* rather than *self-assessment* may indicate his understanding of *evaluation* as being connected with the process of learning, rather than its product as in the case of Andrade (2019).

3.2 Autonomous learner and self-assessment

The interest in learner autonomy is steadily rising. According to Benson (2013, p. 839), "Autonomy refers […] to a capacity to control important aspects of one's language learning."

Richards (n.d., online) puts even more stress on the initiative of the learner in his definition: "Learner autonomy refers to the principle that learners should take an increasing amount of responsibility for what they learn and how they learn it." He proposes five principles for helping learners to develop autonomy, among which he includes "Encouraging reflection" and suggests the European Language Portfolio as a practical tool for the development of learner autonomy. Such reflection cannot be achieved with the learners' use of set criteria and monitoring of their own work – the principles of self-assessment and self-evaluation.

While Gardner (2000) describes *self-assessment*, a term he uses consistently throughout the paper, as a process which "refers simply to the mode of administration, i.e., assessments which are self-administered," (p. 50), the author provides evidence of the benefits (and pitfalls) of practicing

self-assessment for autonomous learners. Gardner accepts Holec's definition of autonomy as "the ability to take charge of one's learning" (Holec, 1981, as cited in Gardner, 2000, p. 20). In this sense, the focus of judgements broadens to include the product of learning as well as the process.

Autonomous learners are, as hinted above, expected to take responsibility for their learning. To do that, they need to monitor their learning and their progress to be able to identify areas for improvement and plan their further learning. In other words, learners look at the *process* of their learning. However, learners also need to know how well they did in a particular performance and, therefore, they judge the *product* of their learning, too.

3.3 Learning and motivational strategies

Self-assessment and self-evaluation are often referred to as parts of strategies research. The two main areas in which they appear are learning strategies and motivational strategies.

Oxford et al. (1989) suggest that "good language learners manage their own learning process through metacognitive strategies, such as paying attention, consciously searching for practice opportunities, planning for language tasks, self-evaluating, and self-monitoring" (p. 30). The authors refer to *evaluation* and *self-evaluation* without further explanation for their choice of terminology. In her later article, Oxford (1999) continues to refer to *evaluating* as one of the metacognitive learning strategies learners should acquire to progress towards communicative competence. This seems to point to the focus on the *process* of learning.

Strategies for motivating learners are a topic spanning at least two scientific fields: educational psychology and pedagogy. In his research of strategies to motivate learners of foreign languages, Dörnyei (2001) presents the following definition of the strategies: "Motivational strategies are techniques that promote the individual's goal-related behaviour" (Dörnyei, 2001, p. 28). The author lists *self-assessment* as one of the efficient ways to keep language learners motivated and his use of the terms *self-assessment* and *self-evaluation* is very interesting. Dörnyei (2001) often uses *self-assessment* as a noun but *self-evaluation* as an adjective, for example: "Encourage accurate student *self-assessment* by providing various self-evaluation tools" (p. 134). He also refers to students *evaluating* their language performances, for example: "Goals are not only outcomes to shoot for but also standards by

which students can evaluate their own performance and which mark their progress" (p. 82). Dörnyei does not provide an explanation of the difference he sees between the two terms, if any, but from his use of them, it seems that *self-assessment* refers to the whole process of setting the task, completing it, applying criteria of success, and drawing conclusions for further study. *Self-evaluation*, on the other hand, seems to describe the actual act of applying the criteria only.

4 Self-assessment and self-evaluation in teaching and learning English as a foreign language

There are a number of considerations to be taken into account when determining the most suitable terminology to be used in educational research. Not the least of those is the actual identity of the teaching and learning English as a foreign language (TLEFL). Among clarifications of the supporting sciences and their role in forming this identity, Píšová (2011) stresses the need to distinguish between TLEFL and theories of foreign language acquisition (FLA). The author sees FLA as intentional or incidental language acquisition in any context, whereas TLEFL focuses specifically on the language acquisition within the educational context, taking into consideration the teaching and learning processes. The area of FLA does not concern itself with learner assessment or evaluation and is, therefore, irrelevant for this study. However, there are several reliable sources of information regarding the terminology in the TLEFL field.

4.1 Teaching and learning English as a foreign language and self-assessment

A useful explanation of what self-assessment is can be found in Harris and Brown (2018): "self-assessment is a descriptive and evaluative act carried out by the student concerning his or her own work and academic abilities" (Brown & Harris, 2013, as cited in Harris & Brown, 2018, pp. 6–7). Once more, the adjective *evaluative*, used here to describe the activity of judging own work, appears to confirm that *self-assessment* seems to be used synonymously with *self-evaluation*. The authors accept that the definition does not suit all self-assessment opportunities, but they propose it as suitable for learners who are beginning to develop their self-assessment abilities.

A trustworthy and widely used source of terminology for language assessment which should be included in determining the correct use of *self-assessment* is the Common European Framework of Reference for languages (CEFR). The CEFR was created based on numerous studies and discussions among experts in the language learning, teaching, and assessment fields. Both the original CEFR (2001) and the CEFR Companion Volume (2020) use the term *self-assessment* only, making no reference to *self-evaluation*. It is, therefore, reasonable to assume that the descriptors are to be used to judge the language ability of a learner, i.e. the *product* of language learning, rather than the impact of the course the learner might be attending to improve their competence in the language, i.e. the *process* of learning the target language. The availability of self-assessment grids (CEFR CV, 2020, pp. 177–181) may be seen as proof of such understanding.

To compare, Tsagari et al. (2018) discuss *self-assessment* frequently in their *Handbook of Assessment for Language Teacher*. Interestingly, their definition of self-assessment, i.e. "the involvement of learners in assessment procedures" (p. 217) and their definition of assessment, i.e. "Language assessment is the practice of evaluating the extent to which learning and teaching have been successful, focusing on what learners can do with the language, on their strengths rather than their weaknesses" (p. 184) seem to suggest that, for the authors, assessment is the act of evaluating, similar to other works already mentioned. It may also be noteworthy that the authors consider the success of the teaching and learning *process* to be an integral part of the language learners use of the language, i.e. the *product*.

4.2 Language testing and self-assessment

To complete the teaching and learning English as a foreign language picture, it is useful to consider one more area of research: language testing. As the research of teaching and learning foreign languages grew in importance, the development of testing and specific national/international/specialist examinations became important milestones in determining the learners' progress and/or achievement.

The Association of Language Testers in Europe (ALTE), which focuses on the quality of testing foreign and second languages, published a *Multilingual glossary of language testing terms* (further referred to as the *Glossary*) in 1998, to help standardise the use of relevant terminology. Their distinction

between *assessment* and *evaluation* is stated clearly: *assessment* is defined as "the measurement of one or more aspects of language proficiency, by means of some test or procedure" (p. 135). Whereas *evaluation* is the "gathering [of] information with the intention of using it as a basis for decision-making. In language testing, evaluation may focus on the effectiveness or impact of a programme of instruction, examination, or project" (p. 144). In other words, in language testing, *assessment* provides information about the learners' ability to use a specific area of language but *evaluation* informs of how well a programme prepares learners for the test. This would imply, in line with what has already been presented, that *assessment* is the term to be used when referring to the judgement regarding the *product* of language learning, whereas evaluation provides information on the *process*.

In the *Glossary* (1998), the authors also provide definitions for other essential terms within language testing, such as *marking*, with its definition of "assigning a mark to a candidate's response to a test. This may involve professional judgement, or the application of a mark scheme which lists all acceptable answers" (p. 152); or *grading*, defined as "the process of converting test scores or marks into grades" (p. 146).

The whole hierarchy may, therefore, be summed up as follows: Grades are created by converting test scores into a standardised scale. Usually, these are represented by a letter representing the level of quality of language output. The test scores are the result of a marking process, which is completed by language assessors. The marking, or assigning of marks based on a mark scheme, is completed for a specific instance of assessment, for example a written test, which focuses on a selected language feature or area. Several assessments are commonly conducted over a period of time to provide basis for a longer-term process of evaluation of the teaching/learning process. Such evaluation then provides feedback on the learning process.

The *Glossary* (1998) also offers a definition for the key term of this paper, i.e. *self-assessment*, which is here related to the specific field of language learning. It defines self-assessment thusly: "The process by which a student assesses his/her own level of ability, either by taking a test which can be self-administered, or by means of some other device such as a questionnaire or checklist" (p. 162). The use of *self-assessment*, rather than *self-evaluation*, is consistent with the nature of assessment as defined by the *Glossary* (1998)

and points to the act of self-assessment being focused on a specific language performance conducted by a particular test or procedure.

Mirroring the abovementioned hierarchy of grading – marking – assessing – evaluating, it is possible to create a similar hierarchy involving the learners themselves as the judges. Therefore, *self-grading* can be seen as referring to the students' assigning grades to their work (i.e. the product of the language learning), *self-marking* as describing the act of students using a mark scheme to decide which of their responses are correct and which are not, *self-assessment* as pointing to the use of criteria, whether provided by the teacher or co-created with the teacher, to analyse a specific product of language learning a student produced, and finally, *self-evaluation* as the reflection a student conducts to analyse how effective the process of learning was to achieve the student learning goals.

5 Conclusion

In current research of teaching and learning foreign languages, there seems to be little to no consensus on how different *self-assessment* and *self-evaluation* are and when to use each of the terms. This article aimed to provide an insight into how terminology is used in the research of learners making judgements about their own language performance. The various viewpoints of the supporting sciences of psychology and pedagogy have been presented to introduce the terms *self-assessment* and *self-evaluation* in a general way. This then provided the basis for presenting the terms in the field of teaching and learning English as a foreign language, with the inclusion of the specific term *self-grading*.

The understanding of the terms *self-assessment* and *self-evaluation* appears to be synonymous in the field of psychology, although the meaning and use of the terms in pedagogy starts to offer some distinction, especially that between appraising the product or the process of learning. Also, the term *self-grading* appears in the field of pedagogy, though it is not relevant for psychology. Once the research field is narrowed further to the specific area of teaching and learning English as a foreign language, the three terms show a marked distinction in their meaning and use. *Self-assessment* appears to be mostly relevant when students are asked to analyse a *product of their language learning*, such as an essay they produced. They are encouraged to monitor their spoken or written performance and use set criteria to decide

how well they performed in that particular instance. *Self-evaluation* most commonly describes the analysis of the *language learning process*, be it a language course, a school term of language learning, and so on. Students do not consider one particular language performance but rather analyse their approach to learning, the time spent learning in a formal and/or informal way, and other aspects of learning a language. *Self-grading* is the narrowest of the three terms and refers exclusively to students assigning grades to their own language learning product, based on a list of correct responses rather than complex criteria. The term often implies that there is little to no analysis of the learner's work.

As research on self-assessment and self-evaluation continues, it will be interesting to monitor how the use of the terminology develops and how the meanings of the terms become more refined.

References

ALTE members. (1998). *Multilingual glossary of language testing terms.* Cambridge University Press. https://www.alte.org/resources/Documents/9780521658775book_D.pdf

Andrade, H. L. (2010). Students as the definitive source of formative assessment: Academic self-assessment and the self-regulation of learning. In *NERA Conference Proceedings 2010*.

Andrade, H. L. (2019). A critical review of research on student self-assessment. In *Frontiers in Education* (p. 87). Frontiers.

Benson, P. (2013). Learner autonomy. In *TESOL Quarterly, 47*(4), 839–843. https://onlinelibrary.wiley.com/doi/10.1002/tesq.134

Blatný, M. (2010). *Psychologie osobnosti. Hlavní témata, současné přístupy.* Grada.

Brown, G. T. L., & Harris, L. R. (2013). Student Self-Assessment. In J. H. McMillan (Ed.), *The SAGE Handbook of Research on Classroom Assessment* (pp. 367-393).

Cambridge Dictionary. (n.d.). Assessment. In Cambridge Dictionary. Retrieved October 12, 2022, from https://dictionary.cambridge.org/dictionary/english/assessment

Cambridge Dictionary. (n.d.). Evaluation. In *Cambridge Dictionary*. Retrieved on October 12, 2022, from https://dictionary.cambridge.org/dictionary/english/evaluation

Council of Europe. (2001). *Common European framework of reference for languages: Learning, reaching, assessment.* https://rm.coe.int/1680459f97

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment. Companion volume with new descriptors.* ://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4

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

Dörnyei, Z. (2009). *The psychology of second language acquisition.* Oxford University Press.

Gardner, D. (2000). Self-assessment for autonomous language learners. *Links & letters*, (7), 049–60. https://www.researchgate.net/publication/28057303_Self-assessment_for_autonomous_language_learners

Harris, L. R., & Brown, G. T. L. (2018). *Using self-assessment to improve student learning.* Routledge.

Hattie, J. (2018). *Hattie ranking: 252 influences and effect sizes related to student achievement.* https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/

James, W. (1892). *Psychology. Briefer course.* Henry Holt and Company.

Keplová, K. (2021). Sebehodnocení ve studiu angličtiny v terciárním vzdělávání: přehledová studie. In *Výzkum v didaktice cizích jazyků IV* (pp 47–65). Masaryk University.

Mareš, J. (2013). *Pedagogická psychologie.* Portál.

Online Etymology Dictionary. (n.d.). Assessment. In *Online Etymology Dictionary.* Retrieved October 15, 2022, from https://www.etymonline.com/search?q=assessment

Online Etymology Dictionary. (n.d.). Evaluation. In *Online Etymology Dictionary.* Retrieved October 15, 2022, from https://www.etymonline.com/search?q=evaluation

Oxford, R. L. (1999). Relationships between second language learning strategies and language proficiency in the context of learner autonomy and self-regulation. *Revista Canaria de Estudios Ingleses*, *38*, 109–126. https://www.academia.edu/24255872/relationships_between_second_language_learning_strategies_and_language_proficiency_in_the_context_of_learner_autonomy_and_self_regulation

Oxford, R. L., Lavine, R. Z., & Crookall, D. (1989). Language learning strategies, the communicative approach, and their classroom implications. *Foreign Language Annals*, *22*(1), 29–39. https://doi.org/10.1111/j.1944-9720.1989.tb03139.x

Panadero, E., Brown, G. T., & Strijbos, J. W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*, 803–830. https://www.researchgate.net/publication/288179989_the_future_of_student_self-assessment_a_review_of_known_unknowns_and_potential_directions

Píšová, M. (2011). Didaktika cizích jazyků: otázky identity. In *Pedagogická orientace*, *21*(2), 142–155. https://journals.muni.cz/pedor/article/view/868

Scriven, M. (1981). Evaluation thesaurus. Edgepress.

Spada, N. M. (1987). Relationships between instructional differences and learning outcomes: A process-product study of communicative language teaching. *Applied Linguistics*, *8*(2), 137–161. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=55f58290e63981d3bcaaed1ed39e8812d200fdb3

Tsagari, D., Vogt, K., Froehlich, V., Csépes, I., Fekete, A., Green, A., Hamp-Lyons, L., Sifakis, N., & Kordia, S. (2018). *Handbook of assessment for language teacher.* http://taleproject.eu/

Zimmerman, B. J. (1998). Academic studying and the development of personal skill: A self-regulatory perspective. *Educational psychologist*, *33*(2/3), 73–86. https://www.tandfonline.com/doi/abs/10.1080/00461520.1998.9653292?journalCode=hedp20

Author

Kateřina Keplová, University of Pardubice, Faculty of Arts and Philosophy, Department of English and American Studies, Studentská 95, 532 10 Pardubice 2, 166744@muni.cz

Sebehodnocení, sebe-evaluace, nebo sebe-známkování: Co je po jméně?

Abstrakt: Schopnost hodnotit vlastní výkon se zdá být čím dál atraktivnějším tématem výzkumu v oblasti výuky a učení se cizímu jazyku. Přestože existuje řada studií, které potvrzují pozitivní vliv této schopnosti na rozvoj řečových dovedností v procesu osvojování cizího jazyka na různých stupních vzdělání, stále zůstávají otázky ohledně terminologie a diskuse nad správným pojmenováním se stále vedou. Cílem tohoto článku je tudíž pokusit se odpovědět na tyto otázky, a to konkrétně: Jak se tato schopnost nazývá, sebehodnocení? Jsou sebehodnocení a sebe-evaluace to samé? Co znamená sebe-známkování? Tento článek nabízí vhled do původu sebehodnocení a sebe-evaluace a jejich rozvoje v oblasti psychologie a pedagogiky. Následně se zaměřuje na použití těchto termínů v oblasti výuky a učení se cizímu jazyku.

Klíčová slova: sebehodnocení, sebe-evaluace, sebe-známkování, rozvoj sebehodnocení

Teachers' beliefs of pre-service German teachers about student self-assessment

Stephan Schicker

University of Graz, Center for Germans as a Second Language and Language Education

Received 1st November 2022 / final version received 24th April 2023/ accepted 25th April 2023

Abstract: The aim of the research is to investigate beliefs of Austrian and Czech pre-service teachers of German about student self-assessment (SSA). In the first part of the paper important theoretical and empirical findings on and principles of SSA and about teachers' beliefs are discussed. After the description of the research design the data analysis is presented. Results show that only a minority of trainee teachers participating in this survey have experienced SSA as students themselves and that even fewer have been able to implement SSA as teachers in their classroom. Moreover, it was verified that most of the trainee teachers have theoretical knowledge about student self-assessment. If one looks at the statements of the individual pre-service teachers as a whole and assign them to a growth mindset (=self-assessment skills can be learned with suitable training) or a fixed mindset (=self-assessment skills are only mastered by certain particularly reflective students), it is evident that although 43.8% cannot be classified and 9.4% make statements that can be assigned to both mindsets, 28.1% of the prospective teachers can be assigned to a growth mindset and 18.8% to a fixed mindset. Didactically, it would be desirable if it were clearly accentuated that self-assessment skills can be learned through suitable didactic training.

Keywords: student self-assessment, teachers' beliefs, self-regulated learning, German as a second language, German as a foreign language

> *Beim ersten Mal funktioniert es meistens noch nicht so gut, aber nach etwas Übung wird es besser und kann effektiv zur Verbesserung eines Textes beitragen. (AT_3, Pos. 7; translation: It usually doesn't work so well the first time, but after a little practice it gets better and can be effective in improving a text.)*

This quote from a pre-service German teacher from Austria refers to a process, in which learners evaluate their own work or their learning process. Various terms have been established that refer to the assessment of one's own performance by learners. These terms, which may also have different theoretical foundations, include "self-assessment", "self-evaluation" (judgments used for grading), "self-reflection", "self-monitoring" and "reflection" (cf. Ross, 2006, p. 2). In this paper, the term student

https://doi.org/10.5817/PedOr2022-4-383

self-assessment and its abbreviation SSA will be used henceforth to refer to the following process:

> Self-assessment is a process of formative assessment during which students reflect on and evaluate the quality of their work and their learning, judge the degree to which they reflect explicitly stated goals or criteria, identify strengths and weaknesses in their work, and revise accordingly. (Andrade & Du, 2007, p. 160)

Most other definitions of the term have a significant overlap with this definition by Andrade above: Brown and Harris (2013, p. 368), for example, define the term as a "descriptive and evaluative act that the student undertakes in relation to his or her own work and academic skills." According to Panadero et al. (2016, p. 804) the term refers to a "variety of mechanisms and techniques students use to describe (i.e. assess) and potentially assign value (i.e. evaluate) to the qualities of their own learning processes and products." Epstein et al. (2008, p. 5) define SSA for the area of science as the ability "to notice our own actions, curiosity to examine the effects of those actions, and willingness to use those observations to improve behavior and thinking in the future." Even though these quotes make it clear that there is a common ground between the definitions of student self-assessment, the term itself refers to various didactic activities, "such as assigning a happy or sad face to a story just told, estimating the number of correct answers […], using a rubric to identify strengths and weaknesses in one's persuasive essay, writing reflective journal entries, and so on" (Andrade, 2019, p. 1). However, what all these activities have in common, is that they require some type of assessment of one's own performance. SSA can have different purposes in language classes, which can range from raising awareness about language aspects to self-reflections about the current language level or future learning goals. As this article will outline, there is extensive evidence of the benefits of SSA: It can promote metacognitive skills (cf. Siegesmund, 2016), academic performance (cf. Brown & Harris, 2013), learning autonomy (cf. Andrade & Du, 2007) and motivation (cf. Brown & Harris, 2013). Moreover, "from a pedagogical perspective, effective learning can only occur when students have a realistic sense of their own performance so that they can direct their further learning on critical aspects of their learning needs" (Yan et al., 2020, p. 509). Yet, these empirical findings alone do not guarantee that SSA is going to be implemented in the classroom because teachers play a decisive role in facilitating the implementation of didactic concepts. Their beliefs about,

experiences with and attitudes towards SSA are significant factors for the implementation of SSA. Research shows that teachers' beliefs not only have a central function in planning, designing, and managing classrooms (e.g., cf. Kratzmann et al., 2017; Bromme, 2014), but they also determine individual acceptance of (new) didactic concepts (cf. Bredthauer & Engfer, 2018, p. 2) such as SSA. Hence, the purpose of this research is to investigate beliefs that pre-service German teachers have about SSA.

For the scope of this research, two specifications must be made to these remarks. Firstly, the explanations on self-assessments in the theory section, which discuss empirical as well as theoretical findings on SSA, are meant to provide the basis in terms of objective theories of language didactics for comparison with teachers' subjective theories (beliefs) on SSA in the empirical part. Secondly, this paper does not address teacher beliefs on SSA on the whole, but investigates them in language learning and more specifically SSA to written work in language classes[1]. This focus on written work is based on the understanding of SSA as expressed in the quote above that it is ultimately about revision for learner work based on self-assessment. Such a revision is only possible to a limited extent in the case of (oral) utterances of students.

First, important theoretical and empirical findings on and principles of SSA and about teachers' beliefs are presented in Section 1. After the description of the research design of this paper (data collection, data analysis, research questions, etc.) in Section 2, the data analysis about teacher's beliefs of pre-service German teachers is presented (Section 3), before the findings from the analysis are summarized in the concluding section.

1 Theoretical overview

In the following part, the scientific and didactical discourse on the aspects of SSA relevant for this survey are presented in order to introduce objective theories of didactics before subjective theories of the teachers are analyzed and related to these "objective" theories of didactics.

[1] Most of the references made in the following article refer directly to theoretical findings or empirical studies from the field of language learning. If references are also made to other academic disciplines, these mostly concern the didactic concept of SSA in general, so that it can be assumed that they also apply to language learning to a large extent.

1.1 Student self-assessment in language learning: Formative or summative

An important question of SSA is if it should be formative, i.e. feedback that is provided during the learning process so that students can still improve their (written) learning outcomes, or summative, which means that feedback is given after the assessment in the form of a final mark. In the definition of SSA above it is clear that Andrade & Du (2007) perceive SSA as a formative feedback tool. They ground this in its function as a way of providing feedback that then leads to revision or optimization of the learning outcomes:

> Why do we ask students to self-assess? I have long held that self-assessment is feedback […] and that the purpose of feedback is to inform adjustments to processes and products that deepen learning and enhance performance; […] if there is no opportunity for adjustment and correction, self-assessment is almost pointless. (Andrade, 2019, p. 2)

Panadero et al. (2019, p. 147) use a similar argument. They suggest that the concept of SSA should be moved towards self-feedback, "in which the final goal is for students to produce and search for feedback to close the gap between their current and desired performance."

Research (e.g. Tejeiro et al., 2012) shows that summative SSA (especially when the assessment contributes to the final grade) is perceived by students mostly as a tool to give oneself a better grade rather than to really evaluate the qualities of one's own texts. When the purpose of SSA is learning-oriented, the student judgments of their learning outcomes or texts are more consistent with judgments of professors or experts/researchers (cf. Barney et al., 2012; Panadero & Romero, 2014) or teachers (cf. Chang et al., 2012). In summary, it can be stated that if SSA does not play a role in the final grade, the learner's judgment may not always be accurate either, but deliberate distortions in favor of a better grade are avoided and a stronger focus on the learning process seems more likely to be guaranteed.

1.2 The effects of student self-assessment on written skills in language learning

There are numerous studies, especially in the Anglo-American world, which investigate the effectiveness of SSA in relation to (language) teaching and written performance: For a broader understanding of the topic, two meta-analyses of the effects of SSA on learning are presented. Brown & Harris (2013), who included 24 studies in their meta-study, found a median effect

from |d|[2] = 0.40 to 0.45 on academic achievement in general. The meta-analysis of Graham et al. (2015) including 11 studies, which investigated the effects of SSA on writing, yielded an average effect size of |d|= 0.62.

With reference to language learning and written performances, some studies will now be discussed in detail. Andrade et al. (2008) and Andrade & Boulay (2003) conducted quasi-experimental studies to investigate the effect of using rubrics[3] when revising a text. The first study was conducted at the primary level (116 learners) and shows significant effects (|d|= 0.87). The second study (107 learners) was conducted at the secondary level and shows no effect (|d| = 0.00) of training on the text quality of revisions. It is noteworthy that the intervention group in Andrade & Boulay (2003) was only very briefly trained in self-assessing their own texts through "rubrics", which could be an explanation for the outcome. The learners in the intervention group in Andrade (2008), however analyzed a model text and used this model essay to generate a list of criteria that made the model text a well-written text.

Sadler & Good (2006) and Andrade et al. (2010) also reported significant effects (|d| = 0.82 and |d| = 0.66, respectively) for lower secondary level (126 learners) in a quasi-experimental setting and for primary level (162 learners) in a quasi-experimental setting by using "rubrics." Duke (2003), Guastello (2001) and Ross et al. (1999) investigated for different age groups the influence of using rubrics when revising text structure (composition). While Duke (2003) for the upper secondary level (164 learners) and Ross et al. (1999) for the 4th to 6th grades (296 learners) could only prove minimal effects of SSA on text composition (|d| = 0.29 and |d| = 0.20), Guastello (2001) found a significant improvement (|d| = 1.27) for the fourth grade (167 students). Glaser et al. (2010) found rather moderate influences (|d| = 0.38) in a true experimental study at the primary level (105 learners), in which they investigated the effects of self-regulation and assessment training on the writing performance and the self-efficacy of

---

2 The effect sizes used in this article are those indices that are also given in the original publications. In this case, these are Cohen's |d|, Hedge's |g| and the eta2 (η2). These effect sizes can be interpreted as follows:
Cohen's |d| and hedge's |g|: small effect size:|d| <= 0.2; medium size effect: |d| <= 0.5; large effect size: |d| <= 0.8;
eta2 (η2): small effect size =< 0.05; medium size effect =< 0.13; large size effect => 0.14;

3 Andrade & Du (2005, p. 5) define rubrics as "a document that articulates the expectations for an assignment by listing the criteria, or what counts, and describing levels of quality from excellent to poor."

learners. The results of the intervention study of Schicker (2020) confirm the effectiveness of a didactic setting which focuses on SSA in terms of promoting textual assessment skills, revision skills, the learner's argumentative writing skills and increasing the motivation for revision. Depending on the selected rated texts, there is a medium or large effect of the didactic setting on the textual assessment skills of the learners ($\eta 2$ = .12 or $\eta 2$ = .21), as well as on the interrater reliability of the intervention groups (intervention groups posttest: ICC 2 = 0.97, control groups: ICC2 = 0.57). There is a substantial effect on the revision motivation ($\eta 2$ = .24) and revision skills ($\eta 2$ = .23) and a medium effect on the argumentative writing skills ($\eta 2$ = .08).

This above-depicted potential of SSA to promote written language skills is theoretically (and empirically) also explained by its link to foster self-regulated-learning (SRL)[4] skills. This competence of "learning to learn" is closely connected to the ability to assess one's own texts or skills. Brown & Harris (2014, p. 8) even see SSA as an essential component of SRL as self-reflection is an integral part of self-regulated learning. In Brown & Harris (2013) they also proved the connection between SRL and SSA empirically.

1.3 Didactic implications

Research shows that the following didactic premises and aspects are of especially great importance for the success of SSA and its promises. First, studies (cf. Eva & Regehr, 2008) have shown that (formative) SSA is more effective when it is more task-specific rather than generic to a very abstract competency. Hence, it is more effective for learning to give the feedback that the composition of a particular text does not follow standard text type norms than simply stating that one is generally bad at writing. This is certainly also of particular importance with the "growth mindsets"[5] and "fixed mindsets"[6] identified by Dweck (2008) in her psychological studies on motivational aspects of learning. Learners with a growth mindset focus on the learning process and that they can in general acquire (almost all) skills if they try hard enough. Dweck (2008) deals in her work with changing such a "mindset" in

---

4 Zimmerman (2000, p. 14) defines SRL as "self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals."

5 A growth mindset can be defined as "a belief that suggests that one's intelligence can be grown or developed with persistence, effort and a focus on learning" (Ricci, 2013, p. 3).

6 A fixed mindset is "a belief system that suggests that a person has a predetermined amount of intelligence, skills or talents" (Ricci, 2013, p. 3).

the course of pedagogical practices. For learners, concrete and task-specific criteria would make it more transparent in which areas they need to improve their performances or their skills because “students assess their own writing to appraise growth, determine strengths, and identify areas in need of further development” (Graham et al., 2011, p. 11f).

Secondly, research (cf. Andrade & Boulay, 2003; Andrade & Du, 2007; Andrade et al., 2008, 2010; Panadero & Romero, 2014) also indicates that a clear reference to standards or criteria as scaffolds for the learning process is beneficial for the learning process. Most frequently, these specific standards or criteria are given in the forms of rubrics. Jönsson & Panadero (2017, p. 99) define them broadly as “assessment instruments designed to assist in identifying and evaluating qualitative differences in student performance.” More specifically, Andrade (2008, p. 61) outlines that a “rubric is a document that lists criteria and describes varying levels of quality, from excellent to poor, for a specific assignment.”[7] In the context of SSA, research also highlights the importance of rubrics. They can “aid assessors in achieving higher levels of consistency when scoring performance tasks” and they “promote learning and/or improve instruction by making assessment expectations explicit and aiding the feedback process” (Jönsson & Panadero, 2017, p. 99). Looking at the effects of using rubrics, Andrade (2019, p.4) reports an average effect size of small to moderate considering all the studies, which focused on SSA using rubrics compared to control groups.

Jönsson & Panadero (2017, p. 99) outline that the two main difficulties students face when using feedback are that they do not comprehend the feedback or they do not know how they can use the feedback to improve their skills. Rubrics make assessment criteria explicit so that students can understand the feedback. Because of the fact that rubrics include detailed descriptions of student performance, they also have the potential to give students “instructions” on how to use feedback.

[7] This definition indicates that rubrics contain more information than “Kriterienkataloge” (Becker-Mrotzek 2014), which are often used in the context of German as a first, second and foreign language teaching, as rubrics also specify different levels for each criterion with precise descriptions of various levels for achievement.

For their use, Panadero et al. (2016, p. 317f) recommend the following principles:

- Define the criteria by which students assess their work
- Teach students how to apply the criteria
- Give students feedback on their self-assessments
- Give students help in using self-assessment data to improve performance
- Provide sufficient time for revision after self-assessment
- Do not turn self-assessment into self-evaluation by counting it towards a grade.

More generally for feedback, Panadero et al. (2016, p. 321) also highlight that it can be very beneficial if students are involved in developing the assessment criteria. Studies (cf. Sadler & Good, 2006; Andrade et al., 2010) show that students who are involved in formulating criteria for assessment also achieve better results. Jönsson & Panadero (2017, p. 108) add to the following three aspects to these principles. It can be beneficial (1) to use an analytic scoring instrument "so that the aspects to be assessed are explicitly spelled out" and (2) to use various quality levels, "so that the quality sought becomes visible to the students." Moreover, it can be helpful (3) to specify task-levels, "so that rubrics are neither too closely tied to the particular task nor too generic."

This section dealing with didactic principles for the implementation of SSA shows that clear criteria that are comprehensible for learners are of great importance for the implementation of SSA.

1.4 Consistency instead of accuracy

When it comes to measuring the "significance" of SSA, ratings of students are often compared with ratings of teachers or other professionals in terms of correlations to measure their "quality". For this correlation, Andrade (2019) argues that the term consistency is more precise than the term accuracy as there is much evidence that ratings of teachers or other professionals are also unreliable (cf. Brown et al., 2015). Generally, Brown & Harris (2013) reported for a very broad variation of forms of SSA from weak to strong correlations between ratings of students of their own work and external ratings (e.g. from teachers, experts) (ranging from r = 0.20 to 0.80). Research

(cf. Butler, 2018) also indicates that more skilled and more competent learners are more consistent with external evaluators than less experienced learners. Hence, consistency can be improved by more experience in SSA (cf. Lopez and Kossack, 2007) and the use of rubrics (cf. Panadero & Romero, 2014). Additionally, older research also shows that – not surprisingly – the degree of accuracy/consistency of SSA rises with simple and concrete tasks (cf. Bradshaw, 2001). For narrating a story, Kaderavek et al. (2004) were able to verify in the case of formative assessment that older, higher qualified students were more consistent in their judgements than younger, less qualified students. In addition, male students had the tendency of being more likely to overestimate the quality of their works than female students.

When it comes to SSA, Andrade (2019, p. 6) also states that consistency is not the goal of SSA, as the goal of SSA is learning-oriented:

> Many if not most of the articles about the accuracy of self-assessment are grounded in the assumption that accuracy is necessary for self-assessment to be useful, particularly in terms of subsequent studying and revision behaviors. Although it seems obvious that accurate evaluations of their performance positively influence students' study strategy selection, which should produce improvements in achievement, I have not seen relevant research that tests those conjectures.

This section emphasizes that the didactic value of SSA lies less in a consistency of learner judgements with expert judgements but rather in the intensive engagement of learners with their performance or learning process.

1.5 Student perceptions

There are also a number of studies focusing on how students, pupils, and learners perceive SSA (e.g., cf. Micán & Medina, 2017; Bourke, 2014; Ndoye, 2017; van Helvoort, 2012; Siow, 2015). These studies confirm that it is very important for students to understand the purpose of SSA. Bourke (2016) was able to show in her study that younger students often do not understand the purpose of SSA and this leads to the result that the processes of SSA are often insufficiently or poorly executed. In contrast, students in higher education or university students tend to consider SSA to be beneficial and useful for their learning process (cf. Micán & Medina, 2017; Lopez & Kossack, 2007; Bourke, 2014; Ndoye, 2017; van Helvoort, 2012; Siow, 2015). For this context, research (e.g., cf. Bourke, 2014) also suggests that – as already mentioned above – it is additionally beneficial if learners can formulate and develop the criteria for assessment themselves.

1.6 Teachers' beliefs

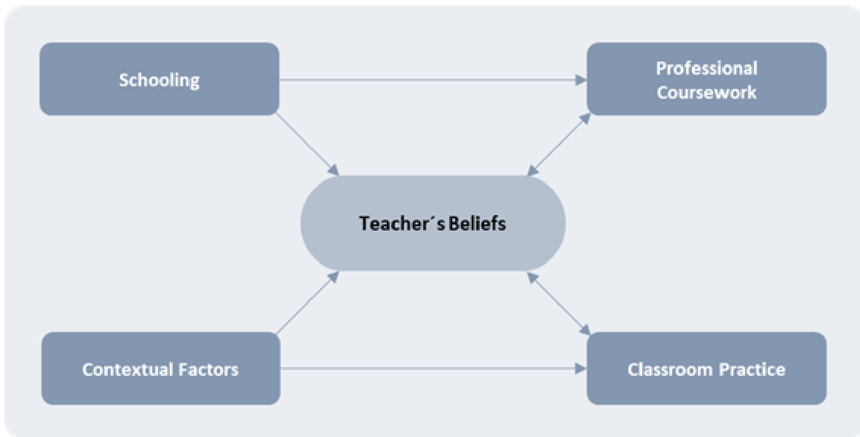
The effectiveness of SSA for language learning in general and for the promotion of writing skills in particular has been discussed above and proven in numerous studies (cf. Section 1.2). There are also numerous studies on how to implement SSA (e.g. cf. Jönsson & Panadero, 2017; Andrade et al., 2008, 2010; see Section 1.4). In language teaching, however, not only scientific theories and empirical findings are vital for didactic choices made in the classroom, but also the beliefs or conceptions of teachers concerning how language is best learned/taught, are crucial. For the implementation and application of SSA concepts in classrooms, it is therefore also significant that teachers subjectively perceive this didactic concept as effective and beneficial.

Before we clarify the connection between SSA and teachers' beliefs, the teachers's beliefs are reviewed in general. Bredthauer & Engfer (2018, p. 3) use the term teachers' beliefs to refer to "teachers' perceptions, attitudes, and internal representations of instruction." Borg (2006, p. 272) defines the term as "an inclusive term referring to the complex, practically-oriented, personalized, and context-sensitive networks of knowledge, thoughts and beliefs that language teachers draw on in their work." Such teachers' beliefs play a central role when it comes to implementing didactic concepts. In fact, most research (cf. e.g., Bredthauer & Engfer, 2018; Kratzmann et al., 2017) from the field of teachers' beliefs is based on the view that these beliefs have a major influence on the practice of teaching. Hence, as "teachers' beliefs guide teachers in understanding educational policies, deciding what is important, and determining what should be done" (Panadero & Brown, 2017, p. 134), it is first necessary to understand and change the beliefs of teachers (about feedback) to alter classroom practices.

The following adapted figure based on Borg (2003, p. 82) shows factors that have the potential to influence teachers' beliefs: It highlights that next to contextual aspects and classroom practice one's own language learning experience and the teacher training itself are important factors.

Figure 1

Teacher's beliefs (own illustration, based on Borg 2003, p. 82)



The present study is located at an important intersection as far as its subjects are concerned. The subjects (pre-service teachers of German) are presumably still strongly influenced by their own language learning experiences during their own school years and they are currently undergoing studies in which they are confronted with objective theories about language learning. Regarding the relationship between one's own language learning experiences in school and scientific theories in teacher education, Haukås (2019, p. 346) notes how "a number of studies show that beliefs that were established prior to language teacher education are resolutely held and that it can be difficult to change students' views."

It is of significance for the research interest of this study that there are already studies on "teachers' beliefs" about SSA in foreign language learning (particular for English) available internationally (e.g., cf. Remesal, 2007; Brown & Harris, 2013; Cephe & Yalcin, 2015; Gebril & Brown, 2020), but a desideratum is still the question of how prospective German teachers in the Czech Republic and Austria view SSA and its didactic potential.

2 Research design

The present study explores the beliefs pre-service teachers have about SSA: All participating Czech German teachers are trained in teaching German as a foreign language, all participating Austrian German teachers are trained in teaching German as a second and first language.

An anonymous self-report questionnaire, consisting of 23 open questions was used as a survey tool. The questionnaire consisted of five big thematic blocks: (a) demographic information, (b) experience in SSA, (c) perceived advantages/disadvantages of SSA, (d) consistency of SSA and (e) received training in SSA. On the questionnaire, the definition of SSA was also provided, as specified introduction of this paper.

Before the survey was carried out, the questionnaire was tested in a pilot study with one prospective teacher, who also conducted the survey. By means of the „thinking aloud" procedure (cf. Schramm, 2018, p. 65) the participant verbalized everything that went through his mind during the survey. With the results of the thinking-aloud protocol, the questionnaire was slightly revised in relation to linguistic aspects and then employed in a seminar and a workshop on feedback at the beginning of the seminar.

The results of the survey were coded in the MAXQDA program and then categorized and evaluated according to qualitative content analysis according to Mayring (2010, p. 60). An inductive approach was taken to the analysis and the category system was adapted several times as part of a cyclical revision process. In the first step, the statements of the students were paraphrased and, in the second step, summarized into categories on a higher level of abstraction. During the analysis, there was an external coder in addition to the researcher. In the first step, both coded the data material independently of each other with the help of the coding guide[8]. In case of discrepancies, coding was made consensually in the second step after a comparison.

A total number of 8 students from the Czech Republic (2 male students, 6 female students) and 23 students from Austria (6 male students and

[8] The two coders first agreed on steps on how to proceed with the coding: These included independently passages relevant to the previously formulated research questions and summarizing them at a higher level of abstraction in a code. In the coding guide, the two coders also collected actual examples for the formulation of codes together from the corpus in advance, so that there was a common basis of understanding of the level of abstraction.

17 female students) participated in the survey. The students studying in the Czech Republic were between 23 and 26 years old (average age: 24,6), they were, on average, in their 9th semester of study and had an average of 17.7 months of practical experience in teaching German classes in school. The students studying in Austria were between 21 and 38 years old (average age: 24,8), they were, on average, in their 8th semester of study and had an average of 7.6 months of practical experience in teaching German in schools.

The questionnaire was used to collect data to answer the following research questions:

- RQ 1: What experience do pre-service teachers have with various aspects of SSA?
- RQ 2: How do pre-service teachers think SSA is best implemented (didactic approach, aims, target group)?
- RQ 3: What advantages and possible problems do pre-service teachers see in SSA?
- RQ 4: Do pre-service teachers consider SSA "accurate" and how do they justify their opinion?
- RQ5: How can the statements of the participants be assigned to the concepts of a growth and fixed mindset?

3 Analysis

Due to the small size of the sample and the fact that there are hardly any systematic differences in the answers of students from the two countries, the evaluation for most questions is presented for both countries together and not separately by country.

RQ1: What experience do pre-service teachers have with various aspects of SSA?

Table 1 shows that the majority of trainee teachers (CZ 75%; AT 52%) in both countries have not used SSA in their classrooms. Interestingly, the second largest group is of those who say they have had experience with SSA, but only in relation to their own work (i.e. for planning a lesson) and not in their own teaching as a teacher. Those students then also state that they have had very positive experiences with SSA in relation to their own work, as this text quote shows:

> Sehr gute Erfahrungen; wichtig für Entwicklung der Lehrpersönlichkeit; hoher Lernfaktor. (AT21, Pos. 7; translation: Very good experience; important for the development of the teaching personality; high learning factor)

Table 1

SSA used in once classroom

	CZ		AT	
Use of SSA	Frequency	%	Frequency	%
no	6	75	13	52
yes, for my own work (lesson planning)	1	12.5	4	16
yes, unspecified (generally for feedback)	1	12.5	3	12
with private tutoring (one-to-one teaching setting)	0	0	2	8
at the end of chapters to reflect on the progress of learning	0	0	1	4
to check prior knowledge	0	0	1	4
yes, for revision	0	0	1	4
Total number	**8**	**100**	**25**	**100**

Some further questions related to whether the student teachers experienced SSA personally as learners. Also, with this question only very small differences can be found between the two countries: Only half of the participating students from the Czech Republic and 42,85 percent of the students from Austria have experienced forms of SSA as learners.

Table 2

Experience with SSA as a student

	CZ		AT	
Experience with SSA as a Student?	Frequency	%	Frequency	%
no	4	50	12	57,14
yes, but it was difficult	1	12,5	4	19,05
yes, in foreign language teaching	1	12,5	2	9,52
yes, at university level	1	12,5	1	4,76
yes, with positive experiences	1	12,5	1	4,76
yes, entrance examination for teacher training	0	0	1	4,76
Total number	**8**	**100**	**21**	**100**

A sub-question to this research question shows what concrete experiences those students who have already used SSA in their teaching have had (Table 3). Consistent with the research findings, students emphasize that criteria as scaffolds are very important for students (cf. Andrade et al., 2008, 2010; Panadero & Romero, 2014), that assessment skills can be increased through training (cf. Schicker, 2020) and that assessing other people's texts seems to be easier for students than assessing their own texts (cf. Fix, 2006, p. 176f).

Table 3

Experience

Experience Using SSA?	Frequency	%
no experience	11	47.83
precise and clear criteria/questions or scaffolds are important	3	13.04
good experience with SSA (unspecified)	3	13.04
self-evaluation-competence increases with experience	2	8.7
to increase the ability of learners to self-reflect	2	8.7
it is easier to evaluate another person's text than to self-evaluate one's own text	2	8.7
Total number	**23**	**100**

RQ2: How do pre-service teachers think SSA is best implemented (didactic approach, aims, target group)?

To answer the second research question, the trainee teachers were first asked how they have or would didactically guide SSA (Table 4).

Table 4

Methodical procedure

Methodical Procedure	Frequency	%
with a questionnaire	7	29.17
categories/criteria were provided	5	20.83
comparing self-evaluation with external evaluation	5	20.83
digital instruments	4	16.67
learning journal	1	4.17
explaining SA	1	4.17
pupils line up according to self-assessment	1	4.17
Total number	**24**	**100**

Similar to the question above, most prospective teachers make statements regarding the didactic procedure that have also been discussed theoretically in the didactic discourse (see Section 1.3) and are empirically examined as effective. These include, in particular, working with concrete criteria or questionnaires, in which reference can also be made to criteria for assessment in the form of questions. Further suggestions of the students concerning the didactic implementation, such as keeping a learning journal or the comparison of external and self-assessment, can also be classified in the didactic discourse as theoretically well-founded and meaningful. A student from Austria makes a practical suggestion for the didactic procedure of comparing self-assessment and peer assessment:

> Ich würde es eventuell im Anschluss an eine schriftliche Übung oder ein Referat machen und die Schüler*innen [SuS] bitten, ihre eigene Leistung in Kategorien einzuschätzen und anschließend im Plenum die Kategorien (unabhängig von jenen der/des SuS) besprechen und um ein konstruktives Feedback bitten. Der/die SuS hat dabei die eigene Bewertung noch im Hinterkopf und kann sich dann an der Fremdeinschätzung orientieren (AT_1, Pos. 5; translation: I would possibly do it after a written exercise or a presentation and ask the students to assess their own performance in categories and then discuss the categories (independently of those of the student) in plenary and ask for constructive feedback. The student still has his/her own evaluation in mind and can then orientate him/herself on the external evaluation.)

It is interesting to note that this suggested didactic approach can also be found in the principles formulated by Panadero et al. (2016, p. 317f, see Section 1.4.).

Another aspect of RQ 2 referred to the learning goals of SSA (Table 5). The trainee teachers listed numerous learning objectives that can be achieved with the help of SSA.

Table 5

Learning goals

Learning Goals	Frequency	%
to assess/reflect one's skills and knowledge	13	35.14
promoting skills for revision	8	21.62
promoting motivation	6	16.22
self-assessment ability is promoted	6	16.22
promoting literacy skills	3	8.11
learning as a process activity becomes visible	1	2.7
Total number	**37**	**100**

In addition to the frequently mentioned learning objective of being able to assess one's own abilities, many prospective teachers locate the central objective of SSA in carrying out revisions or increasing motivation. The importance of SSA for revisions is shown in the modeling of the revision process by Bereiter & Scardamalia (2009). Revision is successful when conspicuities are identified by comparing intention and its realization (=compare), the discrepancy or inadequacies are identified (=diagnose) and only then improvements are made (=operate).

For the mentioned learning goal of enhancing motivation through SSA, studies show that SSA is associated with improved motivation, more engagement, and self-efficacy (cf. Munns & Woodward, 2006; Ross, 2006, p. 6). The answers to this question (learning goal) are also connected with the question whether SSA should be implemented for summative or formative purposes (see Section 1.1). The answers of the majority of the students accentuate the role of SSA as a formative feedback tool as they highlight its function for revision, self-reflection or as part of the learning process.

A related question regarding the implementation of SSA refers to the suitable target groups of SSA (Table 6). The answers to the question about target groups indicate that many trainee teachers consider the assessment ability of learners analogous to a fixed mindset as relatively static and not trainable. There are some answers to this question, such as "more effective from upper school onwards as students are more reflective" (Pilsen_AT21, item 27; German translation: Ab Oberstufe wirksamer, da SuS reflektierter sind), which neglect the aspect that training and experience with SSA also increase learners' ability to assess and improve their knowledge and products through SSA. A student from the Czech Republic points to the aspect of the importance of training self-assessment skills when she writes that she "believes that it is also suitable for younger ones already, but you have to work it out step by step according to the learning level of the students." (Pilsen_CZ5, item 27; translation: "Ich glaube, dass es auch für Kleinere schon geeignet ist, aber man muss es schrittweise nach der Lernstufe der Schülerinnen erarbeiten".)

Table 6

Target group

Target Groups	Frequency	%
secondary level II	14	41.18
for all levels	7	20.59
secondary level (I and II) + higher levels	6	17.65
only in classes with „good" students	3	8.82
only for university students	2	5.88
students need (years of) training in SA	2	5.88
Total number	34	100

RQ3: What advantages and possible problems do pre-service teachers see in SSA?

The third research question is related to what benefits (Table 7) and possible problems (Table 8) prospective teachers see in SSA. With regard to the benefits of SSA, the respondents emphasize, among other things, its importance in promoting the self-reflective skills of learners and learner autonomy (the promotion of learner autonomy and students do not have to depend on the feedback of teachers). In the Anglo-American world, this aspect of „self-assessment" to promote learner autonomy (cf. Andrade & Du,

2007, p. 161) and self-regulated learning (see Section 1.3) is accentuated. One trainee teacher emphasizes this, for example, when she writes that learners can "track their own progress" thanks to SSA (Pilsen_CZ6, item 11; translation: "Sie können ihren Fortschritt selbst verfolgen").

Table 7

Advantages

Advantages SSA	Frequency	%
self-reflection: reflection of one's own learning progress	12	26.67
students learn to assess their own works and skills	9	20
promotion of learner autonomy	8	17.78
documentation of the learning process (for others)	5	11.11
increases self-confidence	5	11.11
authentic feedback (about the skills of students)	2	4.44
the ability to criticize is encouraged	2	4.44
students do not have to depend on feedback from teachers	1	2.22
Promotion of language awareness	1	2.22
Total number	**45**	**100**

The prospective teachers see possible disadvantages or problems in the use of SSA mainly in the fact that the students „misjudge" themselves. As discussed in Section 1.1 and 1.4, Brown et al. (2015, p. 4) address this fear of trainee teachers by highlighting a significant aspect of assessment processes: "Does it matter if students are inaccurate in their self-assessments, as long as they are engaged in thinking about the quality of their work?"

Table 8

Possible problems

Possible Problems	Frequency	%
underestimation or overestimation	10	26.32
mismatch between self-assessment and external assessment	9	23.68
pupils with incorrect SA could be strengthened in this	5	13.16
students have no motivation for SA	4	10.53
students do not take it seriously	3	7.89
shyness/fear to assess themselves	3	7.89
too little experience	2	5.26
institutional frameworks are not suitable for SA	1	2.63
time-consuming	1	2.63
Total number	**38**	**100**

RQ 4: Do pre-service teachers consider SSA accurate and how do they justify their opinion?

When asked whether they consider SSA to be accurate and objective, the relative majority of trainees state that it is on the whole neither accurate nor objective (Table 9). And as stated in Section 1.4. with reference to research findings, a smaller percentage of trainee teachers also state that as empirically proven the consistency of judgment with expert judgment can be increased through more experience and the provision of clear criteria. In a sub-question to this, trainee teachers were also asked in which direction they thought students tended to be wrong in their judgments. Here, most student teachers state that they believe students tend to both overestimate and underestimate themselves (62.5%). A quarter of students believe that students tend to overestimate themselves and 12.5% of students believe that students tend to underestimate themselves.

Table 9

SSA accuracy and objectivity

Is SSA accurate and objective?	Frequency	%
no, it is difficult to assess one's own abilities (subjective)	13	46.43
yes (various other reasons or unspecified)	5	17.86
only if the questions or criteria are clear/precise	4	14.29
only with training in SSA	3	10.71
only with certain (good) classes	2	7.14
only in a limited way	1	3.57
Total number	**28**	**100**

RQ5: How can the statements of the participants be assigned to the concepts of a growth and fixed mindset?

All statements made by the respondents were also examined to determine whether indicators of a growth or fixed mindset could be derived from them. For example, the answer that in "lower school only certain pupils are suitable for SSA" (AT_3, pos. 29-30; translation: "Unterstufe nur bei geeigneten SuS") was seen as an indicator that the trainee teacher sees the ability to assess one's own assignments more as a predetermined skills which cannot be changed by training (=fixed mindset).

When assigning the statements of all the pre-service teachers as a whole to a growth mindset (=self-assessment skills can be learned with suitable training) or a fixed mindset (=self-assessment skills are only mastered by certain particularly reflective students), it is evident that although 43.8% cannot be classified and 9.4% make statements that can be assigned to both mindsets, 28.1% of the prospective teachers can be assigned to a growth mindset and 18.8% to a fixed mindset.

4 Conclusion and limitations

4.1 Conclusion

Summarizing the results of this research, it appears that only a minority of trainee teachers participating in this survey have experienced SSA as students themselves (in total for both countries: 46.4%) and that even fewer have been able to implement SSA as teachers in their classroom (in

total for both countries: 22.25%). However, the answers of the pre-service teachers on how best to implement SSA didactically certainly reflect the current didactic research discourse. The fact that the trainee teachers have theoretical knowledge about SSA is not only evident in their answers to the question of how best to implement SSA in the classroom, but also in the fact that in both countries, 71.6% of the study participants state that they have already learned and heard something about SSA in their teacher training.

The following conclusions can be drawn from the results of the survey. Firstly, although the results show that the vast majority of trainee teachers have learned something about SSA in their teacher training, only a minority have been able to carry out self-assessments themselves as learners at school and in their university studies. This circumstance must ultimately also be taken into account in the didactic design of seminars at the university. As it seems that it still occurs all too often that concepts such as SSA are taught in teacher training but are then not implemented in the didactic design of seminars in university teaching.

Secondly, if one looks at the statements of the individual prospective teachers as a whole and assign them to a growth mindset (self-assessment skills can be learned with suitable training) or a fixed mindset (self-assessment skills are only mastered by certain particularly reflective students), it is evident that 9.4% make statements that can be assigned to both mindsets, and 18.8% to a fixed mindset. Didactically, it would be desirable if it were clearly accentuated that self-assessment skills can be learned through suitable didactic training, and (linguistic) competencies can thus not only be appropriately assessed but also promoted. Two student teachers emphasize this aspect when they write that "nicht nur die LP hat die Aufgabe den Lernfortschritt der SuS festzustellen, sondern die Schüler werden aktiv eingebunden (Pilsen_AT16, Pos. 11; translation: Not only the teacher has the task to determine the learning progress of the pupils, but the pupils are actively involved) and that SSA "macht deutlich, dass man Schreiben nicht einfach ‚kann', sondern ‚lernen' kann" (AT_7, Pos. 12; translation: Makes it clear that one cannot simply "do" writing but can "learn" it).

4.2 Limitations

Due to the qualitative nature of the study and the small, non-representative sample, the results of the study cannot be generalized beyond the current sample. The other major limitation of this study is its self-reported nature. As

the survey investigates teachers' perceptions, responses may reflect despite the anonymity of the survey some elements of social desirability. Perhaps quite different results would emerge were the students of these teachers surveyed or their classrooms observed. For future studies a triangulation of the investigation beliefs and actual classroom practice would be desirable. In addition, an analysis of the teacher training curricula that the students have gone through would be of interest in order to be able to establish points of reference to the concrete statements made by the students and their training.

References

Abdou, N. (2017). Peer/self-assessment and student learning. *The International Journal of Teaching and Learning in Higher Education*, *29*, 255.

Andrade, H. (2008). Self-assessment through rubrics. *Educational Leadership*, *65*(4), 60–63.

Andrade, H. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, 4. https://www.frontiersin.org/articles/10.3389/feduc.2019.00087

Andrade, H., & Boulay, B. (2003). Role of rubric-referenced self-assessment in learning to write. *Journal of Educational Research.* (97), 21–34.

Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment Research & Evaluation*, *10*(3), 5–16.

Andrade, H., & Du, Y. (2007). Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education, 32*(2), 159–181.

Andrade, H., Du, Y., & Mycek, K. (2010). Rubric-referenced self-assessment and middle school students' writing. *Assessment in Education: Principles, Policy & Practice*, *17*(2), 199–214.

Andrade, H., Du, Y., & Wang, X. (2008). Putting rubrics to the test: The effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students' writing. *Educational Measurement: Issues and Practice*, *27*, 3–13.

Barney, S., Khurum, M., Petersen, K., Unterkalmsteiner, M., & Jabangwe, R. (2012). Improving students with rubric-based self-assessment and oral feedback. *IEEE Transactions on Education*, *55*(3), 319–325.

Becker-Mrotzek, M. (2014). Schreibleistungen bewerten und beurteilen. In H. Feilke & T. Pohl (Eds.), *Deutschunterricht in Theorie und Praxis. (DTP.) Handbuch zur Didaktik der deutschen Sprache und Literatur in elf Bänden. 4. Schriftlicher Sprachgebrauch – Texte verfassen* (pp. 501–513). Schneider-Verl. Hohengehren.

Bereiter, C., & Scardamalia, M. (2009). *The psychology of written composition. The psychology of education and instruction.* Routledge.

Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe, and do. *Language Teaching*, *36*(2), 81–109. https://www.cambridge.org/core/article/teacher-cognition-in-language-teaching-a-review-of-research-on-what-language-teachers-think-know-believe-and-do/F6B40C79983C27649FC1157D4023A776.

Borg, S. (2006). *Teacher cognition and language education: Research and practice.* Continuum.

Bourke, R. (2014). Self-assessment in professional programmes within tertiary institutions. *Teaching in Higher Education*, *19*(8), 908–918.

Bradshaw, B. (2001). Do students effectively monitor their comprehension. *Reading Horizons*, *41*(3), 143–154.

Bredthauer, S., & Engfer, H. (2018). *Natürlich ist Mehrsprachigkeit toll! Aber was hat das mit meinem Unterricht zu tun?* https://kups.ub.uni-koeln.de/8092/.

Bromme, R. (2014). *Der Lehrer als Experte. Zur Psychologie des professionellen Wissens.* Waxmann.

Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: Directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, *22*(4), 444–457.

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 367–393). SAGE.

Brown, G. T. L., & Harris, L. R. (2014). The future of self-assessment in classroom practice: Reframing self-assessment as a core competency. *Frontline Learning Research*, *2*(1), 22–30.

Butler, Y. (2018). Young learners' processes and rationales for responding to self-assessment items: Cases of generic can-do and five-point, Likert-type formats. In J. M. Davis, J. M. Norris, M. E. Malone, T. H. McKay, & Y.-A. Son (Eds.), *Useful assessment and evaluation in language education* (pp. 21–40). Georgetown University Press.

Carless, D., Bridges, S. M., Chan, C. K. Y., & Glofcheski, R. (Eds.) (2018). *The enabling power of assessment: Vol. 5. Scaling up assessment for learning in higher education.* Springer Nature Singapore.

Carlsen, C., Dypedahl, M., & Iversen, S. H. (Eds.) (2019). *Teaching and learning English.* Cappelen Damm Akademisk.

Cephe, P. T., & Yalcin, C. G. (2015). Beliefs about foreign language learning: The effects of teacher beliefs on learner beliefs. *The Anthropologist*, *19*(1), 167–173.

Chang, C.-C., Tseng, K.-H., & Lou, S.-J. (2012). A comparative analysis of the consistency and difference among teacher-assessment, student self-assessment and peer-assessment in a web-based portfolio assessment environment for high school students. *Computers & Education*, *58*, 303–320.

Davis, J. M., Norris, J. M., Malone, M. E., McKay, T. H., & Son, Y.-A. (Eds.) (2018). *Useful assessment and evaluation in language education.* Georgetown University Press.

Duke, B. L. (2003). *The influence of using cognitive strategy instruction through writing rubrics on high school students' writing self-efficacy, achievement goal orientation, perceptions of classroom goal structures, self-regulation, and writing achievement.* Norman.

Dweck, C. (2008). *Mindset: The new psychology of success.* Ballantine Books.

Epstein, R., Siegel, D., & Silberman, J. (2008). Self-monitoring in clinical practice: A challenge for medical educators. *The Journal of Continuing Education in the Health Professions*, *28*(1), 5–13. https://doi.org/10.1002/chp.149

Eva, K. W., & Regehr, G. (2008). "I'll never play professional football" and other fallacies of self-assessment. *The Journal of Continuing Education in the Health Professions*, *28*(1), 14–19.

Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education.* Routledge Falmer.

Fitzgerald, J., & Markham, L. (1987). Teaching children about revision in writing. *Cognition and Instruction*, *4*, 3–24.

Fix, M. (2006). *Texte schreiben: Schreibprozesse im Deutschunterricht.* UTB.

Gebril, A., & Brown, G. (2019). Student self-assessment foreign language quality of feedback in TESOL: A learning-oriented assessment perspective. In J. de Dios Martínez (Ed.), *Quality in TESOL and teacher education. From a results culture towards a quality culture* (pp. 140–149). Routledge.

Glaser, C., Kessler, C., Palm, D., & Brunstein, J. (2010). Improving fourth graders' self-regulated writing skills: Specialized and shared effects of process-oriented and outcome- related self-regulation procedures on students' task performance, strategy use, and self-evaluation. *Zeitschrift für Pädagogische Psychologie*, *24*(3), 177–190.

Graham, S., Harris, K. R., & Hebert, M. (2011). *Informing writing: The benefits of formative assessment.* Carnegie Corporation of New York.

Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal*, *115*(4), 523–547. https://doi.org/10.1086/681947

Guastello, E. F. (2001). Parents as partners: Improving children's writing. In *Twenty-third yearbook of the College Reading Association* (pp. 279–293).

Haukås, Å. (2019). Teachers' beliefs about language instruction. In C. Carlsen, M. Dypedahl, & S. H. Iversen (Eds.), *Teaching and learning English* (pp. 343–357). Cappelen Damm Akademisk.

Jönsson, A., Panadero, E. (2017). The Use and Design of Rubrics to Support Assessment for Learning. In: Carless, D., Bridges, S., Chan, C., Glofcheski, R. (eds) Scaling up Assessment for Learning in Higher Education. *The Enabling Power of Assessment*, vol 5. Springer, Singapore. https://doi.org/10.1007/978-981-10-3045-1_7.

Kaderavek, J. N., Gillam, R. B., Ukrainetz, T. A., Justice, L. M., & Eisenberg, S. N. (2004). School-age children's self-assessment of oral narrative production. *Communication Disorders Quarterly*, *26*(1), 37–48.

Kostons, D., van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction*, *22*, 121–132.

Kratzmann, J., Jahreiß, S., Frank, M., & Ertanir, B. (2017). Einstellungen pädagogischer Fachkräfte in Kindertageseinrichtungen zur Mehrsprachigkeit. *Zeitschrift für Erziehungswissenschaft*, *20*(2), 237–258.

Lopez, R., & Kossack, S. (2007). Effects of recurring use of self-assessment in university courses. *The International Journal of Learning: Annual Review*, *12*, 203–216.

Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken.* Beltz Verlagsgruppe.

Micán, A., & Medina, L. (2017). Boosting vocabulary learning through self-assessment in an English language teaching context. *Assessment & Evaluation in Higher Education, 42*(3), 398–414.

Munns, G., & Woodward, H. (2006). Student engagement and student self-assessment: The REAL framework. *Assessment in Education: Principles, Policy & Practice*, *13*(2), 193–213.

Ndoye, A. (2017). Peer/self assessment and student learning. *The International Journal of Teaching and Learning in Higher Education*, 29, 255–269.

Panadero, E., Brown, G., & Courtney, M. (2014). Teachers' reasons for using self-assessment: A survey self-report of Spanish teachers. *Assessment in Education: Principles, Policy & Practice*, *21*(4), 365–383.

Panadero, E., Brown, G., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*, 803–830. https://doi.org/10.1007/s10648-015-9350-2

Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education*, *32*(1), 133–156. www.jstor.org/stable/44951865.

Panadero, E., Jönsson, A., & Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, *22*, 74–98. http://urn.kb.se/resolve?urn=urn:nbn:se:hkr:diva-17143

Panadero, E., Jönsson, A., & Strijbos, J.-W. (2016). Scaffolding self-regulated learning through self-assessment and peer assessment: Guidelines for classroom implementation. In D. Laveault & L. Allal (Eds.), *Assessment for learning: Meeting the challenge of implementation* (pp. 311–326). Springer International Publishing.

Panadero, E., Lipnevich, A., & Broadbent, J. (2019). Turning self-assessment into self-feedback. In M. Henderson, R. Ajjawi, D. Boud, & E. Molloy (Eds.), *The impact of feedback in higher education. Improving assessment outcomes for learners* (pp. 147–163). Palgrave Macmillan.

Panadero, E., & Romero, M. (2014). To rubric or not to rubric? The effects of self-assessment on self-regulation, performance and self-efficacy. *Assessment in Education: Principles, Policy & Practice*, *21*(2), 133–148.

Remesal, A. (2007). Educational reform and primary and secondary teachers' conceptions of assessment: The Spanish instance, building upon Black and Wiliam. *The Curriculum Journal*, *18*(1), 27–38.

Reynolds, C. J., Hill, D. S., Swassing, R. H., & Ward, M. E. (1988). The effects of revision strategy instruction on the writing performance of students with learning disabilities. *Journal of Learning Disabilities*, (21), 540–545.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation 11*, (10). http://pareonline.net/getvn.asp?v=11&n=10

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1999). Effects of self-evaluation training on narrative writing. *Assessing Writing*, *6*, 107–132.

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, *11*(1), 1–31.

Sanchez, C., Atkinson, K. M., Koenka, A. C., Moshontz, H., & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology*, *109*, 1049–1066.

Schicker, S. (2020). *Förderung der Textbeurteilungskompetenz von Lernenden: Eine Interventionsstudie in sprachlich heterogenen Klassen. Sprachliche Bildung Studien: Band 4.* Waxmann.

Schramm, K. (2018). Methoden zur empirischen Erforschung des Schreibens in Deutsch als Zweitsprache. In W. Grießhaber, S. Schmölzer-Eibinger, & H. Roll (Eds.), *DaZ-Handbücher. Schreiben in der Zweitsprache Deutsch. Ein Handbuch* (pp. 33–49).

Siegesmund, A. (2016). Increasing student metacognition and learning through classroom-based learning communities and self-assessment. *Journal of Microbiology & Biology Education*, *17*(2), 204–214.

Siow, L.-F. (2015). Students' perceptions on self- and peer-assessment in enhancing learning experience. *Malaysian Online Journal of Educational Sciences*, *3*, 21–35.

Tejeiro, R. A., Gomez-Vallecillo, J. L., Romero, A. F., Pelegrina, M., Wallace, A., Emberley, E. (2012). Summative self-assessment in higher education: Implications of its counting towards the final mark. *Electronic Journal of Research in Educational Psychology*, *10*(2), 789–812.

van Helvoort, A. J. (2012). How adult students in information studies use a scoring rubric for the development of their information literacy skills. *The Journal of Academic Librarianship*, *38*(3), 165–171. https://www.sciencedirect.com/science/article/pii/S0099133312000596.

Yan, Z., Brown, G. T. L., Lee, J. C.-K., & Qiu, X.-L. (2020). Student self-assessment: Why do they do it? *Educational Psychology*, *40*(4), 509–532.

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, M. Zeidner, & P. R. Pintrich (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press.

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *The educational psychology series. Handbook of metacognition in education* (1st ed., pp. 299–315). Routledge.

Author

Stephan Schicker, University of Graz, Center for German as a Second Language and language education, Universitätsplatz 1/I, 8010 Graz, Austria, e-mail: stephan.schicker@uni-graz.at

Přesvědčení budoucích učitelů německého jazyka o sebehodnocení studentů

Abtrakt: Cílem studie je prozkoumat přesvědčení rakouských a českých budoucích učitelů německého jazyka ohledně sebehodnocení studentů. V první části studie jsou diskutována empirická zjištění a principy týkající se sebehodnocení, a také přesvědčení učitelů. Dále je prezentována metodologie sběru a analýzy dat. Výsledky ukazují, že jen málo učitelů zažilo sebehodnocení, když byli sami studenty, a ještě méně jich využilo sebehodnocení v rámci vlastní praxe. Většina respondentů má teoretické znalosti sebehodnocení. Při analýze výroků budoucích učitelů z hlediska „growth mindset“ (= sebehodnocení jako dovednost, které se lze naučit) a „fixed mindset“ (= sebehodnoticí dovednosti jsou osvojitelné jen velmi reflektivně založenými jedinci) se ukázalo, že 28,1 % budoucích učitelů lze zařadit ke „growth mindset“ a 18,8 % k „fixed mindset“ (43,8 % nešlo zařadit a 9,4 % bylo možné zařadit k oběma). Z didaktického hlediska je žádoucí, aby učitelé vnímali sebehodnocení jak o osvojitelné skrze vhodný trénink.

Klíčová slova: sebehodnocení studentů, přesvědčení učitelů, autoregulované učení, němčina jako cizí jazyk, němčina jako druhý jazyk

Developing oral presentation-related self-assessment among prospective teachers of German as a foreign language: Analysis of self-assessment comments [1]

Jana Veličková

Masaryk university, Faculty of Education, Department of German language and literature

Received 28th May 2023 / final version received 12th February 2024 / / accepted 14th February 2024

Abstract: Learner self-assessment is a significant predictor of learning outcomes (Hattie, 2018). However, it is insufficiently implemented in Czech secondary schools (Czech School Inspectorate, 2021). One of the reasons for this may be the lack of teachers' experience in developing their own self-assessment skills. This paper presents a study framed by a 12-week intervention programme to develop self-assessment skills focused on presenting in German among prospective teachers of German as a foreign language (n=15). The study examined the content (characteristics) of the participants' self-assessment comments collected before and after the intervention program. A total of 25 self-assessment comments were collected using the "Lautes Erinnern" method (13 before the intervention, 12 after the intervention) and analysed using the inductive category formation of the qualitative content analysis (Mayring, 2014). The analysis revealed three main characteristics of the development of self-assessment: increasing evidence in the self-assessment comments, a shift in focus from the predominance of non-language-specific to language-specific assessment, and a shift in focus from mostly negative to also positive aspects of performance. The study concludes with a discussion of the implications for better teacher education that develops their self-assessment skills appropriately.

Keywords: student self-assessment, development of self-assessment skills, content analysis, German as a foreign language, teacher education

Research has shown that student self-assessment increases student motivation (Benson, 2001), contributes to the development of learner autonomy (Tassinari, 2010), and has a positive impact on the quality of

[1] This study is part of the project MUNI/A/1335/2022 – *Research in foreign language didactics*.

https://doi.org/10.5817/PedOr2022-4-410

student learning and learning outcomes (Hattie, 2018; McDonald & Boud, 2003). Furthermore, education should not only provide learners with knowledge and skills but also teach them how to assess and manage their learning so that it (ideally) becomes a lifelong process (Boud, 2003, p. 13).

In many countries, however, student self-assessment is not widely used in the classroom[2], although teachers are encouraged to promote self-assessment by professional frameworks[3] or sometimes even by curriculum and legal documents[4].

Not surprisingly, the lack of guided opportunities for self-assessment in the classroom can prevent learners from adequately developing their self-assessment skills (Apeltauer, 2010, pp. 22–27). Possible reasons for teachers not developing self-assessment may be related to busy lesson plans, a low belief in the effectiveness of self-assessment (e.g., Mäkipää, 2021), or a lack of experience and knowledge of self-assessment implementation (e.g., Volante & Beckett, 2011), as teachers may not be sufficiently trained to promote assessment for learning, not just of learning (McMillan, 2013, p. 5).

Building on the premise articulated by Raya (2014, p. 149), this article argues that prospective teachers need to gain experience in developing their self-assessment skills during teacher education programs. However, in order to design teacher education programmes that enable prospective teachers to (more effectively) implement student self-assessment in their (later) classroom practice, teacher educators need to understand, how the process of developing self-assessment skills occurs in prospective teachers.

Therefore, this article presents an exploratory study focusing on the content of selfassessment comments collected during a facilitated process to develop self-assessment skills. The comments are seen as a manifestation of the development of self-assessment skills.

The present study is based on a specially designed intervention to promote the self-assessment skills of Czech prospective teachers of German as a foreign language in the specific area of giving a short presentation. The aim

---

[2] For the Czech Republic, see Czech School Inspectorate (2021); for Canada, see Hunter et al. (2006); for Finnland see Lasonen (1995).

[3] For the Czech Republic, see *The framework of professional teacher qualities of a foreign language teacher* (Klečková et al., 2019).

[4] For the Czech Republic see e. g. the *Elementary Education Act* 561/2004 Sb. (MŠMT, 2004) and *Framework education programme for elementary education* (MŠMT, 2017).

is to describe the content characteristics of the self-assessment comments collected at the beginning and at the end of this intervention. The results will shed light on how self-assessment is altered by scaffolding and serve as an empirically supported example for working with the development of self-assessment in teacher education.

First, student self-assessment is conceptualized in the context of education and foreign language learning. Then, in section 2, the research design of the study is presented (research question, participants, data collection, data analysis), followed by the analysis of the results (section 3). The paper ends with a discussion of the results and a conclusion.

1 Theoretical framework: Defining self-assessment in an educational and foreign language learning context

In the educational context, self-assessment has received more attention since the 1990s in relation to its conceptualisation as an essential aspect of formative assessment and the assessment-for/as-learning approach (Brown & Harris, 2013, p. 367; McMillan, 2013, pp. 4–6). In foreign language teaching, self-assessment has been of interest since the 1970s, as foreign language didactics has tended towards the constructivist paradigm (Weskamp, 2007).

To conceptualise student self-assessment for research and teaching purposes, authors have created taxonomies, typologies (e.g. Panadero, Brown & Strijbos, 2016; Taras, 2010, among others), or categorisations (Boud & Brew, 1995) of student self-assessment. However, there is no generally accepted definition. In summary, student self-assessment can first be conceptualised as a process of assessing the quality of one's abilities (skills, competences), processes or products related to learning (Andrade, 2018, p. 377), "based on evidence and explicit criteria, for the purpose of doing better work in the future" (Rolheise & Ross, 2001). This process usually takes place cyclically over time in relation to a particular task or performance and involves the use of various self-assessment tools (happy/sad face, rubrics, reflective journals, portfolios etc.; see e. g. Schneider, 1996; Wilkening, 2013) or implementation of self-assessment methods (Dochy et al., 1999, p. 335), practices or techniques such as "self-ratings, self-estimates of performance, and criterion- or rubric-based assessments" (Brown & Harris, 2013, p. 369). These tools or techniques represent the second conceptualisation. Using them, "students

describe (i.e., assess) and possibly assign merit or worth to (i.e., evaluate) the qualities of their own learning" (Panadero et al., 2016, p. 804).

However, self-assessment should not be limited to assessing the quality of learning. A higher level of cognitive engagement in self-assessment involves "deep engagement with the processes affiliated with self-regulation (i.e., goal setting, self-monitoring, and evaluation against valid, objective standards)" (Brown & Harris, 2013, p. 386). These cognitive processes can be referred to as metacognition (Belgrad, 2013, p. 335), i.e. the cognitive essence of self-assessment, thanks to which one can become aware of and reflect on one's own actions (Krykorová, 2010, pp. 27–28).[5] Thus, thirdly, self-assessment can be conceptualised as a self-regulatory ability (Brown & Harris, 2014).

The ambiguity in the conceptualisation of self-assessment is also reflected in the ambiguity in the terminology, as the term self-assessment is sometimes used as a synonym for self-evaluation (Boud, 2003, p. 13). The term self-assessment emphasizes the procedural understanding of self-assessment, is associated with a formative understanding of assessment and "involves students collecting data to evaluate their own progress" (Brown & Harris, 2013, p. 368). In contrast, self-evaluation refers to a summative understanding of self-assessment (Brown & Harris, 2013, p. 369) and can be conceptualised as one of the sub-components or phases of the self-assessment process (McMillan & Hearn, 2008, p. 41).

In this study, self-assessment is understood as a cyclical/iterative process of metacognitive operations that includes awareness of the goals of the activity, focusing on the object of evaluation (monitoring), evaluating the quality of this component, and formulating alterations to improve the quality. Therefore, the term "self-assessment" is used here.

According to a review study by Andrade (2018, pp. 309–401), the process of developing self-assessment in the learning context seems to be an under-researched area, as research tends to focus on the accuracy and consistency of self-assessment. The findings of these studies suggest that problematic self-assessment accuracy or consistency can be eliminated through appropriate and scaffolded self-assessment development (see, for example, Brown & Harris, 2013, p. 384; Ross, 2006; Ross et al., 1998).

[5] Therefore, self-assessment is rightly reffered to as a metacognitive learning strategy by the authors of various classifications (see Janíková, 2007, p. 95–106).

In order to develop the most accurate selfassessment possible, certain factors should be considered. The first factor to mention is the assessment criteria and descriptors, which are important in obtaining the most accurate self-assessment (Andrade & Valtcheva, 2009). The most effective way is to negotiate them directly with the self-assessors, or at least ensure that they understand them properly (Rolheiser & Ross, 2001, p. 7). Predictors of an accurate self-assessment also include learners' prior experiences with (Ross, 1998, p. 17) and perceptions of self-assessment (Brown & Harris, 2013, pp. 383–384), as well as various intrinsic factors such as fear of making mistakes or self-efficacy (Blanche, 1988, pp. 84–85). As for sociocultural factors, the influence of a culture's attitude towards self-criticism and self-praise (Hosseini & Nimehchisalem, 2021, p. 858) should be taken into account. In the context of foreign language learning, another factor that needs to be considered is the influence of learners' L2/L3 proficiency. Individuals with lower levels of language proficiency, and especially younger individuals, are more likely to overestimate themselves (Butler, 2023, p. 44).

The present intervention study takes the above points into account. Its design is also inspired by studies such as Léger (2009) and her self-assessment forms; Chen (2008), who developed a self-assessment through peer feedback; and Gil-Salomov and Benlloch-Dualde (2016), who investigated self-assessment through peer feedback. Peer feedback is an important part of the presented intervention as it provides an additional incentive to reflect on one's own performance. It has also been argued that peer feedback is more acceptable and uses more natural and understandable language than teacher feedback (Black et al., 2004, p. 14). In addition to the feedback recipient, the feedback giver also benefits from the feedback process (Nicol et al., 2014). This is because it leads to a more intensive and deeper processing of the learning process and also serves as a stimulus for reflecting on one's own performance (i.e. for self-assessment) (Grotjahn & Kleppin, 2015, p. 145). Last but not least, the importance of peer feedback is also reflected in the design of some models for the development of self-assessment – for example in the model by Rolheiser and Ross (2001), which was an important starting point for the intervention in this study.

2 Methodology

This study examines the process of developing self-assessment skills of Czech pre-service teachers of German as a foreign language in the context of a specially designed intervention to promote self-assessment skills with a focus on giving a presentation. The aim of the study is to investigate the characteristics of self-assessment comments in the first and last phase of the intervention.

2.1 Sample

The participants were 15 prospective teachers in the second year of a bachelor's degree programme for teaching German as a foreign language at the secondary level (ISCED 2) who attended a one-semester German course at B2 level as part of their studies[6]. A purposive sampling strategy was used. As mentioned in the introduction, the first selection criterion was that they were prospective teachers. The second criterion was the level of German language proficiency. It was assumed that participants with advanced language proficiency (B1+ or higher) would have a deeper understanding of language structure, a higher level of language awareness, and would be better able to provide detailed selfassessments than participants with less advanced language skills.

A total of 17 students took part in the course and 15 of them were included in the study.[7] All participants had passed a language exam at B1+ level in the previous semester and therefore met the language level requirements. German was their second foreign language and they had been learning it for 5–10 years.

The intervention consisted of six phases, but the reported data refer to the first phase (phase 1) and the last phase (phase 5). The first data collection was conducted with 13 participants and the second with 12 participants.[8] Table 1 summarises the experience of all participants with selfassessment.

[6] The language levels are based on the *Common European framework of reference for languages* (see Council of Europe, 2001).

[7] The study did not include one student who had gone abroad during the course and one student who was studying a different programme from the other participants and whose language level could not be verified either.

[8] They are not the same participants - three participants from data collection 1 did not take part in data collection 2 and two participants from data collection 2 did not take part in data collection 1.

Table 1

Previous experience of the participants (n=15) with self-assessment[9]

How often have you experienced self-assessment	All the time	Often	Sometimes	Rarely	Never	I don‘t know
in the context of learning German as a foreign language?	1	1	2	10	1	0
out of the context of learning German as a foreign language?	0	2	3	8	2	0

2.2 Procedure

In line with the stated aim, a qualitative intervention research design according to Krainer and Lerchster (2012) was chosen. The intervention was based on the models of Rolheiser and Ross (2001) and Zimmerman (2002) and followed the logic of the transition from an object of assessment (other-regulated) to an active agent of assessment (self-regulated). This process is divided into six phases (phase 0 – phase 5) reflecting different degrees of dependence on external assessment. The goal was to bring the participants as close as possible to the stage of independence, in which they should already be able to evaluate their performance without external help (independent stage; Oscarson, 1997, cited in Poehner, 2012, p. 612) – see Table 2.

[9] Respondents were asked to tick one option, but not everyone actually ticked one of the answers.

Table 2

Description of the intervention

Phase 0	• Clarification of self-assessment, its relevance for learning	in-class
Phase 1	• Collaborative negotiation between the learners about the (partial) goals of the task and their concretization based on success criteria describing the characteristics of a good presentation and their indicators.[10] • Presentation • Oral self-assessment (data source 1) • Oral peer feedback • Written teacher feedback (added after a few days)	in-class
Phase 2	• Presentation • Oral self-assessment + Peer feedback (receiving and giving to another peer)	in-class
Phase 3	• Presentation • Written self-assessment	in-class
Phase 4	• Presentation • Written self-assessment	in-class
Phase 5	• Setting goals for the presentation (to provide the participants with further internalization of the objective criteria for a successful presentation) • Presentation • Written self-assessment (data source 2)	individually, out-of-class (online)

This 6-phase intervention (phase 0 – phase 5) was implemented in a 12-week (one-semester) German language course for prospective teachers of German as a foreign language. One of the aims of the course was to develop learners' ability to prepare and give a short oral presentation (about 3–5 minutes) on a selected topic. These presentations were the subject of the self-assessment tasks. The course took place for 90 minutes each week. The intervention was realised every one to three weeks and lasted about 30 minutes. The specific intervals between the phases were as follows: phase 0–1 one week, phase 1–2 two weeks, phase 2–3 three weeks, phase 3–4 one week, phase 4–5 two weeks. The intervals resulted from the inclusion of supportive interim phases between phases 1–4. These phases were completed by the participants

[10] Learners formulated success criteria first in small groups and then under the supervision of an experienced assessor (course teacher). The negotiated rubrics, in the form of a mind map, was available to the learners throughout the whole intervention. The categorical system was used during the collaborative evaluation of a video of a foreign person giving a presentation in order to refine the categorical system and ensure a shared understanding of each criterion.

individually online via the Moodle platform. The aim of these phases was to support the development of self-assessment.[11]

The self-assessment and the feedback from peers and the teacher were conducted in Czech, the participants' mother tongue. It can be assumed that a lack of foreign language proficiency would hinder the verbalisation of more complex cognitive content and lead to shorter self-assessment comments of poorer quality, which would impair the validity of the study (Seliger & Shomamy, 1989, p. 170).

2.3 Data collection, research instruments and data sources

The intervention included six phases. However, this text focuses only on the data from phase 1 (the first self-assessment in the intervention) and phase 5 (the last self-assessment in the intervention). The data was collected using forms with open-ended questions, which were answered verbally (data source 1) and in writing (data source 2). In order to anonymise the data, participants marked their self-assessments with unique codes that were assigned to them at the beginning of each data collection. In the oral self-assessment, participants said the code at the beginning of the recording.

The two data sources are:

- Data source 1 (DS 1) is the first self-assessment of the intervention (from phase 1), i.e., before peer and teacher feedback. The research instrument was an oral self-assessment in class, formulated on the basis of the following questions: How did it go? What went well? What could be done better?[12]

---

[11] Participants watched the recording of their presentation and completed an additional self-assessment of this recording. This additional self-assessment was intended to simulate the evaluation of others' performance, which may seem easier than evaluating one's own performance. However, at the same time, the evaluating/assessing others contributes to the development of effective self-assessment (Hattie, 2020, p. 146). Some of the supportive phases also included reflection on peer and teacher feedback.

[12] By adding the two more specific questions, it was assumed that a self-assessment structured with additional questions would contribute to a higher quality and range of self-assessment, similar to the study by Gan and Hattie (2014) on peer feedback.

- Data source 2 (DS 2) is the final self-assessment of the intervention (from phase 5). The research instrument was an out-of-class[13] written self-assessment formulated on the basis of the following questions: How did it go? What went well? What could have been better and how could it be improved?

The oral self-assessments (from data source 1) were recorded using iPads and transcribed as pure verbatim protocols (Mayring, 2014, p. 45).

2.4 Data analysis

A total of 26 self-assessment comments from both data sources (13 from data source 1, 12 from data source 2) were analysed using qualitative content analysis according to Mayring (2010 in German; 2014 in English), in particular with inductive category formation (Mayring, 2014, p. 79). Due to the different nature of data sources 1 and 2 (see section 2.3), each data source was analysed separately, and the comparison of the results is only made at the level of discussion. The aim was to examine the content (characteristics) of the self-assessments from the first and last phase of an intervention to promote self-assessment skills. The coding was carried out using the software MAXQDA. Although the author analysed the data in Czech, the selected excerpts were subsequently translated into English for this article.

3 Results

First, an overview of the assigned codes and the inductively formed categories is given. The categories represent the content (characteristics) of the student self-assessments in the first and last phase of the intervention – see Table 3. This table also provides an overview of the frequency of occurrence of the most common characteristics of the student self-assessments. However, as the categories are not disjunctive, the frequencies only provide a rough overview. Next, the content of the self-assessments from data sources 1 and 2 is described and illustrated with examples from the data. The chapter concludes with a discussion of the results, including the limitations of the study.

[13] Participants recorded their presentations and uploaded the recordings to the Moodle platform. It can therefore not be completely ruled out that they did not play back the recording of their presentation before the self-assessment.

Table 3

Codebook: Description of the categories; description and frequency of the allocated codes

Categories	Description of the categories	Allocated codes	Description of the codes	Frequency in DS 1	Frequency in DS 2
1. Focus	Responses focusing on the whole performance (i.e., presentation) or the sub-aspects of the performance.	1.1. Global assessment	Responses focusing on the whole performance (i.e., presentation).	8	12
		1.2. Specific assessment	Responses focusing on sub-aspects of the performance (fluency, structure, content, grammar, vocabulary etc.).	61	223
2. Quality of the performance	Judgments about how well a task is performed.	2.1. Negative judgments	Judgments expressing negative, low, or undesirable quality of the performance or its sub-aspects.	44	60
		2.2. Positive judgments	Judgments expressing positive or good quality of the performance or its sub-aspects.	9	61
		2.3. Unclear or unspecific judgments	Judgments expressing unclear or unspecific quality of the performance or its sub-aspects.	4	2

Categories	Description of the categories	Allocated codes	Description of the codes	Frequency in DS 1	Frequency in DS 2
3. Reference norms for the quality	Benchmark against which participants compared the quality of their performance/task completion.	3.1. Criterion-referenced norm	Comparing one's performance to established standards represented by criteria (Brookhart, 2013, p. 258) – e.g., fluency, structure, grammar, vocabulary, presentation skills, pronunciation, …	52	100[14]
		3.2. Individual goal-referenced norm	Comparing one's performance to individual goals of the performance or desirable outcomes expected by the participants themselves.	3	
		3.3. Social-referenced norm	Comparing one's performance to that of another person, e.g., a classmate (William, 2013, p. 207).	1	0
		3.4. Self-referenced norm	Comparison of current performance with some of one's own past performances (Brookhart, 2013, p. 258).	1	23
4. Attributions	Explanations of the cause of the positive or negative performance quality.	4.2. Attributions for the negative quality of the performance.	Explanation of the cause of the negative, lower, or undesirable performance quality.	12	14

[14] As the self-assessment task in Phase 5 also includes a request to assess goals, it is not possible to distinguish clearly in the data between criterion-referenced and individual goal-referenced norms. Therefore, both types of norms are reported together.

Categories	Description of the categories	Allocated codes	Description of the codes	Frequency in DS 1	Frequency in DS 2
		4.3. Attributions for the positive or high quality of the performance.	Explanation of what caused the positive or higher performance quality.	3	14
5. Alterations for improving the quality of the performance	Suggestions for alterations that could improve the quality of the performance.	5.1. Presentation-related alterations	Suggestions for alterations that could have contributed to a better quality of the specific performance (presentation) or could contribute to a better quality of the next specific performance (presentation) in the future.	7	14
		5.2. General alterations	Suggestions for alterations that might help improve the quality of general language performance in the future. Expressed in general terms or as commitments.	3	18
6. Task-related responses	Reflection on the contextual aspects of the task outside of the performance (the task, the topic, the preparation, other settings – recording, place – at school / at home, the role of peers).	6.1. Task assignment-related responses	Reflection on the assignment, including the presentation topic.	1	7

Categories	Description of the categories	Allocated codes	Description of the codes	Frequency in DS 1	Frequency in DS 2
		6.2. Task-preparation-related responses	Reflection on the preparation of the presentation.	14	24
		6.3. Task-settings-related responses	Reflection on the task settings: influence of the recording, place – at school / at home, the peers).	0	12
7. Process-related responses	Description and feedback on the effort and processes involved in completing a task.	7.1. Process-related responses	Description and feedback on the processes involved in completing a task.	2	4
		7.2. Effort	Description of the effort involved in completing a task.	1	16

3.1 Characteristics of self-assessment comments at the beginning of the intervention (from data source 1)

The first self-assessment as part of the intervention comprised 7–159 words. The main focus here was on the sub-aspects of their performance (presentation). In these aspects, the participants primarily commented on their quality, which they evaluated by comparison to reference norms. Some participants also formulated attributions about the perceived quality of the sub-aspects and alterations to improve them. The category *focus of the self-assessment* is therefore cross-cutting and is presented separately for each area (quality assessment, attribution, alterations). In addition, some comments related to the task (e. g., the task assignment and the preparation of the presentation) as well as the processes and efforts involved in completing the task.

Quality of the performance

Participants expressed the perceived quality of their performance in the form of judgments about how well the task was performed. These judgments mainly referred to sub-aspects of the performance and not to the overall performance. More often, the judgments commented on non-language-specific aspects such as presentation skills ("Of course, the contact with the listener could be better. And not looking so much at the notes." 1B_09), fluency ("I was actually stuttering." 1B_14), or presentation structure ("The structure could be better." 1B_03). Less evaluated were the language-specific aspects of performance such as grammar ("I know there were definitely some grammatical mistakes." 1B_02) or vocabulary ("I definitely think the vocabulary could be better because I don't think I have enough vocabulary knowledge for this topic." 1B_11).

We can assume that the participants either do not consider the linguistic aspects of performance (grammatical and lexical correctness) to be important or they find it more difficult to evaluate them. Although the participants have language proficiency at the B1+/B2 level, they may not have sufficient knowledge of the language system and the terms used to describe the linguistic phenomena.

The perceived quality of their performance was verbalised by the participants as positive (focusing on strengths) or negative (focusing on weaknesses). Although the self-assessment task included the question "What did I do well?", negative evaluations predominated. The focus of the negative

evaluations was mainly on the overall performance ("My presentation was horrible." 1B_07). For the sub-aspects, the participants tended to give negative evaluations for fluency of the speech ("I stuttered and did not finish my sentences." 1B_10) and presentation skills ("The contact with the listener could of course be better." 1B_09). The positive evaluations occurred only marginally and focused on presentation skills ("What did I do well? […] Greeting you nicely." 1B_07) or the elaboration of the topic ("But yes, I talked about the topic, that was good." 1B_03).

The predominant focus on negative aspects of one's performance is usually discussed in the context of individuals who have a higher level of competence or knowledge and tend to underestimate themselves (Oscarson, 2009; Kruger & Dunning, 1999). Moreover, the focus on mistakes can be seen as a concomitant factor in the development of self-regulation (Keith & Frese, 2005). It can therefore be argued that developing a positive view of one's performance is also essential, as self-efficacy of one's actions forms the basis for the perception of one's self-efficacy in future actions, as well as for the self-assessment itself (McMillan & Hearn, 2008, p. 44). However, given the superficial view of one's own performance described above, it is questionable whether dealing with mistakes in this case can be seen as promoting the effective development of self-regulation.

Reference norms for the quality of the performance

The quality of performance was assessed by comparison with four norms. The criterion-referenced norm was predominant, as reflected in the use of labels for the evaluation criteria – e.g., "What could be better – probably everything, e.g., vocabulary, grammar." (1B_07). Given that the success criteria were negotiated with the participants at the beginning of the intervention, it is not surprising that they refer to them in their self-assessments. Although the criteria seem to be the most comprehensible reference norm, in most cases, the self-assessments remained only at the rather general level of these criteria ("I am sure the structure could be better." 1B_03) and more specific self-assessments were rare. Thus, this initial self-assessment can be seen as a somewhat superficial consideration of one's performance, as it consisted only of mentioning selected or tangible aspects of the performance without delving deeper into the specific evidence.

As this was the first self-assessment conducted, participants may also have been overwhelmed by the complexity of the task. Furthermore, it can be assumed that the participants based their assessment on general impressions rather than specific evidence due to a lack of knowledge in the specific disciplines (such as phonetics, phonology, morphology, syntax, semantics, and pragmatics).

Although they had not been instructed to set task goals, some of the participants also compared their performance with their own expectations (code goal-referenced norm): “Better than what I prepared.” (1B_10).

To a lesser extent, there was also the self-referenced norm (“DSD[15] made me a hundred times more nervous.” 1B_17) and the social-referenced norm (“This girl here was excellent, a hundred times better than me, but she also prepared two weeks in advance.” 1B_12).

The lack of a self-referenced norm is probably related to the fact that this was the very first presentation and the associated self-assessment in the context of the intervention and the participants were therefore not explicitly offered their own similar performance for comparison with the current one.

However, the low incidence of verbalization of social-referenced norms is surprising. Since the self-assessment in this phase was conducted orally in groups of two or three, one might expect participants to compare their performance with that of their classmates. There are two possible interpretations. The self-assessors were likely so focused on their own performance at that moment that a more tangible assessment framework for them was exactly the criteria discussed at the beginning of the intervention, and further comparison was already beyond their current cognitive capacity. An explanation based on the interaction between the individual’s so-called academic self-concept and the social norm is also offered (see Stiensmeier-Pelster & Schöne, 2008, pp. 66–67). The participants were only able to make the comparison in their minds and concluded from the prevailing negative self-assessment of their communication partner(s) that they also had a problem with the task and therefore possibly no longer considered it important or appropriate to articulate the comparison out loud.

[15] DSD stands for the language exam for the certificate *Deutsches Sprachdiplom*.

Attributions

In addition to evaluating the quality of the performance, participants also commented on possible reasons for the perceived quality of the performance. Among these comments, those referring to possible reasons for a lower quality of the performance predominated. The participants attributed the low quality mainly to external factors, in particular, the limited preparation time ("If I had prepared it, it would not be a problem to cover it, but as it is, it's almost on the spot." 1B_12) and the given topic of the presentation. Among the external factors, they also mention poor language skills, especially in speaking: "I think my performance was very poor because I have a problem expressing myself unless it's in writing, and communication is just a big problem for me. I can't express myself, I can't respond that quickly." (1B_22)

Three justifications referred to the positive quality, that participants attributed to good preparation and prepared notes: "Maybe it was better when I didn't have that support, and then when I was just at the end and I hadn't written anything yet, I talked as if I was just thinking about myself and not sticking to what I had written because I was so lost in it." (1B_10).

The predominant justification for the negative quality of the performance could be related to the fact that the self-assessments focus almost exclusively on the negative aspects of the performance. It is interesting to note that when it comes to attributing failures or negative characteristics of their own performance, participants tend to attribute these to external factors. This tendency is referred to as ego-defensive or self-protective attribution and is associated with a reduced willingness to take responsibility for failures or negative consequences of one's actions (see Miller & Ross, 1975; Weiner & Kukla, 1970).

At the same time, attribution to external factors may indicate a preference for causes that are easier to infer (cf. the principle of cognitive economy – Vašátková, 1995, p. 11). Given the cognitively demanding nature of self-assessment, which takes place as self-monitoring during the performance itself, this explanation seems logical. Due to the cognitively demanding nature of self-assessment, participants likely no longer have the cognitive capacity to search for deeper causes for the quality of their performance.

Alterations

In addition to the negative evaluation of performance by pointing out performance weaknesses, only a few participants also formulated suggestions for their improvement. These alterations were more often specific to a presentation format, i.e. they related to how to improve the quality of the presentation given or how to improve the quality of the next presentation. The focus of these alterations was mainly on better structuring the presentation: "Firstly, I would definitely separate the individual parts so that I can formulate this in German. I would separate the fast food and the preparation of food at home and maybe focus a bit more on the disadvantages of eating at home or preparing food at home." (1B_21). Preparing a presentation and working with notes was also mentioned: "And don't look at the papers so much" (1B_09).

On a side note, there were also general alterations, i.e. not related to the presentation format, but to improving language skills in general: "Overall, I should probably learn to communicate better when it comes to oral communication." (1B_22).

The focus of the presentation-related alterations corresponds to the focus of the evaluations, i.e. mainly on non-language-specific aspects. These alterations are relatively specific, so it can be expected that participants are more likely to improve the non-language-specific aspects of their presentation than the language-specific aspects. The alterations that focus on linguistic aspects tend to be very vague, so it can be expected that they are more difficult to implement and less likely to lead to improvements.

Task- and process-related comments

In the initial self-assessment of the intervention, there were also some task-related comments. They mainly concerned the inappropriateness of the presentation topic ("And overall, I found it really difficult to talk about this topic." 1B_01) and the short preparation time ("My presentation was terrible because I didn't have enough time [to prepare]." 1B_07)

One comment described the effort made to complete the task: "I tried to speak slowly." (2B_02). Also only marginally represented were descriptions of the processes involved in completing the task, which can be illustrated by the following comment: "[…] I was just thinking about it and didn't stick to what I had written." (1B_10)

The presence of this marginal category illustrates that although the participants focused primarily on their own performance in their self-assessment, there are indications that they also take into account the situational context (preparation, task, topic) and are able to perceive the procedural level of performance to a certain extent.

3.2 Characteristics of self-assessment comments at the end of the intervention (from data source 2)

The final self-assessment in the intervention was 59–226 words long and contained mainly comments on the sub-aspects of one's performance (presentation) – e.g., their quality and the reference norms for the evaluation, attributions to justify the perceived quality of the performance and suggestions for its improvement (alterations). In a few cases, process- and task-related comments were also found in the data. The cross-sectional category *focus of self-assessment* is again reported as part of the other thematic categories.

Quality of the performance

Most of the comments on the quality of performance related to partial aspects of performance. Predominant were comments on fluency ("I was speaking fluently and I think I managed not to repeat myself." 1H_14), grammar ("I think I managed to minimise my grammatical mistakes today." 1H_13), presentation skills ("I also managed to be on time because my presentation is three and a half minutes long." 1H_20) and the structure of the presentation ("I followed the structure of the presentation." 1H_02). Among the evaluations, both the positive (i.e., referring to the strengths of the performance) and the negative (referring to the weaknesses) are almost equally represented. The focus of the *negative evaluations* is primarily on fluency ("In particular, the omission of the parasitic or filler sounds ("ehm" etc.) could have been better." 1H_21). There were also some negative evaluations of the linguistic aspects of the performance in terms of grammar ("I made a lot of grammar mistakes." 1H_20) and vocabulary ("I didn't know the vocabulary." 1H_15).

The *positive judgments* include comments that focused primarily on the structure ("The presentation had structure. I mentioned the conclusion, advantages, disadvantages, general information and my own opinion." 1H_10). The overall performance was also often rated positively ("I think it

was better than the very first presentation." 1H_15). The linguistic aspects (grammar and vocabulary), on the other hand, were rarely rated positively (e. g. "I used more connecting expressions – einerseits, trotzdem, weder – noch etc." 1H_01).

The high frequency of positive judgments could be related to the use of the self-referenced norm, because when using this norm two thirds of the participants formulated positive judgments, i.e. they focused on the improvement of the quality of their performance compared to the previous presentation(s). The increase in positive self-assessment due to the influence of an individual reference norm is attributed in particular to the fact that it strengthens confidence in one's own abilities, weakens fear of failure and increases motivation (Rheinberg, 1980, cited in Rheinberg, 2008, pp. 183–184). The influence of considering oneself as successful on increasing learners' self-efficacy (beliefs) is also confirmed by various studies conducted in the context of foreign language learning (e.g. Baleghizadeh & Masoun, 2013).

Reference norms for the quality of the performance

When evaluating the quality of the presentation, the participants compared their performance primarily with the *criterion-referenced norm* ("I managed to keep to the structure of the presentation." 1H_03) and with the individual *goal-referenced norm* ("I managed to keep to the structure of the presentation, which was one of my goals." 2H_20). Since the self-assessment task in this phase also contains the request to evaluate goals, it is not possible to make a clear distinction between criterion-referenced and individual goal-referenced norms. Therefore, both types of norms are reported together.

The evaluations of the sub-aspects were often quite specific, i.e. they did not remain at the superficial level of the evaluation criteria negotiated with the participants at the beginning of the intervention. This is illustrated by the following comments – one referred to the presentation skills and preparation of the presentation: "This time I managed to give a long presentation, over 4 minutes. And the preparation also only took 10 minutes, I managed to write in paragraphs." (1H_15), the other on structure: "I think I did it right because I had an introduction in which I introduced the potential audience to the topic, then I explained the advantages, disadvantages and my own point of view and thanked them for their attention, so I think I did all the steps right." (1H_20).

The comments with the poorest evidence were related to grammar. These comments often remained at the level of the assessment criteria discussed at the beginning of the intervention, as in the following comment: “[What could be done better?] The grammar.” (1H_05). A rare evidence-based self-assessment targeting grammar is the comment: “Sometimes I replaced a verb with a noun.” (1H_02)

At this point, we can discuss the possible influence of participation in the intervention, i.e. the repeated reflection on one’s performance together with peer feedback on the performance of others, which may have contributed to the internalisation of the content of the negotiated success criteria and a more detailed view of the performance. As the self-assessment in this phase took place outside the classroom, the unlimited time for writing the self-assessment may also have contributed to a more specific and comprehensive self-assessment. The specific case of grammar-focused self-assessment is discussed in section 4.

Very often there was also a *self-referenced norm*, i.e. participants compared their current performance with previous presentations. Most frequently, participants compared the overall presentation with all previous presentations (“That was my best attempt.” 1H_01) or with a specific presentation (“But I think it was better than the very first presentation.” 1H_15). Individual aspects were only marginally compared with previous presentations (“Relatively fluent presentation compared to my other presentations.” 1H_22).

The increase in the individual reference standard could be related to the repeated performance of the task, in which participants can compare their current performance with a similar previous performance. At the same time, the increase in the individual reference norm seems to have been reflected in an increase in the positive evaluation – see above.

Attributions

The participants formulated the same explanations for the causes of positive (higher) and negative (lower or undesirable) performance quality. They attributed the lower quality of their performance to various factors, with a particular emphasis on task-related circumstances (“[…], but when I started filming, I got quite nervous.” 1H_20) and poorer language skills (“On the other hand, I used listed phrases – that could also be because they were

new to me and I did not know the vocabulary for them, which is why I got stuck." 1H_03). The higher quality was mainly attributed to the task settings, which was related to the fact that the self-assessment was conducted on an optional topic ("The topic was close to my heart and I had something to say about it." 1H_15) and in conditions outside the classroom ("It was relatively easy. When you are alone, you are less nervous and can concentrate better." 1H_03).

In the self-assessment after the intervention, a balanced reflection on possible causes for both the positive and negative quality of one's performance can be recognised. This could indicate a tendency for self-efficacy to improve in similar tasks in the future – i.e., to be aware not only of the aspects that need to be eliminated in order to achieve better quality but also of those that need to be strengthened. Considering that participants attributed both the positive and negative quality of their performance mainly to external factors (situational and task-related conditions), it can be concluded that participants view the quality of their performance to a certain extent as an interplay of coincidences or circumstances beyond their control.

Alterations

Most participants formulated more general, not just presentation-related alterations. These focused mainly on speaking and vocabulary, such as in the following comment:

> Include new words in your vocabulary because they are very familiar vocabulary that you will encounter throughout your life. Practice speaking more, for example by standing in front of a mirror and trying to speak or asking someone if you can try to present in front of him/her. Make pauses when speaking. (1H_02)

There were also alterations in terms of how the presentation could have been improved ("Maybe I could have talked about more areas that are relevant to the topic." 1H_05) or could be improved in the future ("Be more natural. Don't stick too much to predetermined points. Speak fluently." 1H_22). As this example shows, the focus of these alterations was mostly on fluency. Interestingly, although grammar was one of the most negatively evaluated aspects in this phase, there was only one alteration that focused on improving grammar ("I will pay attention to some grammatical phenomena when I write notes, but I'm afraid that when I start speaking, I won't be able to focus on grammar anymore." 1H_20).

Overall, it can be said that both types of alterations (general and presentation-related alterations) are formulated quite clearly and with a fairly high level of evidence, as illustrated, for example, by the following example: "For the presentations to get better, it's probably important to rehearse at least once a week, to record myself, to listen to myself and see if I'm making progress. And I think it then tends to get better." (1H_14).

Both types of alterations are quite specifically formulated and show that participants have gained a deeper understanding of the desired performance and knowledge of specific strategies for improvement. The increase in cognitive capacity through repeated self-assessment, which became more routinized and therefore required less cognitive load, may also have played a role and enabled an enhancement of cognitive processes.

Task-related comments

The task-related comments primarily referred to the preparation of the task and the lack of preparation time ("There could have been a little more time for preparation." 1H_02). The second most common were task *settings-related comments*. They reflected a shift in presentation and self-assessment situation from in-class to out-of-class presentations:

> My performance was definitely influenced by the topic – I was allowed to choose a topic that I enjoyed and was interested in. And also the home environment. I was alone at home; no one was looking at me or listening to me. I wasn't stressed that I might say something wrong. (1H_01)

Finally, the appropriateness of the presentation topic was also discussed:

> I found the task assignment clear and the topic interesting. The holidays in our country and the holidays in Germany are very topical and it's not bad to know something about them, whether there are differences, etc. (1H_02).

Overall, we can see that the participants talk quite a lot about the situational aspects of the performance, including its preparation. So, they do not limit their self-assessment to the presentation itself but perceive the performance in a much broader framework.

Process-related comments

This marginal category consists of comments describing the processes involved in carrying out the task. The main focus here was on the effort involved in completing the task. The *effort* was either only described: “That’s why I only chose the two best-known holidays (Christmas and Easter) and tried to describe them.” 1H_02), or more often its effectiveness (or efficacy) was also evaluated (positively and negatively in equal measure): “I tried to speak slowly and clearly – with occasional stuttering or longer pauses, I succeeded.” 1H_02). Exceptionally, there was also a justification for the efforts made: “In terms of grammatical correctness and word order, I tried not to make my sentences too long so that I wouldn’t get stuck and say stupid things.” 1H_02). Participants reported that they made an effort to elaborate on the topic well (“I tried to explain my point of view.” 1H_05) and to speak fluently (“I paid attention to speak fluently, but sometimes a new idea came to my mind and I wanted to say it, and then I realized I didn’t know one word of the sentence I wanted to say and I got thrown off track.” 1H_15).

Participants mainly described their efforts, which can be interpreted as “the effort is appreciated” or “this is my merit”, depending on the context. Participants also often directly evaluated whether their efforts led to success. This is likely to have a more positive impact on future goal setting and achievement, as participants can refer back to what worked for them and what did not. Describing and evaluating one’s own efforts can therefore be seen as a desirable feature of self-assessment. In the literature, it is linked to the concept of self-efficacy (Bandura, 1997).

4 Discussion

This study aimed to investigate the characteristics of the self-assessment comments from prospective teachers of German as a foreign language in the first and last phase of the intervention for developing self-assessment of speaking (giving a short presentation in German as a foreign language). The results of this study suggested three main characteristics of self-assessment comments that need to be considered in developing self-assessment skills: level of evidence, focus on strengths and weaknesses, and focus on language- and non-language-specific aspects of performance (giving a presentation).

On the one hand, three characteristics changed during the intervention. Firstly, there was a shift from an almost exclusive focus on weaknesses (negative evaluations) to an equal representation of positive and negative evaluations. Secondly, comments on non-language-specific aspects of performance predominated at the beginning of the intervention, while comments at the end also included language-specific aspects (grammar, vocabulary). Thirdly, the evidence in the self-assessment comments changed from poor to stronger over time. These aspects have already been discussed in the results section. On the other hand, one aspect of the self-assessment – the non-language-specific aspect of grammar – did not correspond to this trend. Although the participants frequently commented on grammar, the comments on grammar tended to have poor evidence and be evaluated negatively. Therefore, the self-assessment of grammar is discussed below. The discussion concludes with the limitations of the study.

The finding that the self-assessment of grammar tended to be negative and with poor evidence can primarily be attributed to the fact that the criterion of grammatical correctness in oral production is not so easy to define in terms of what it entails. This raises the question of how (and whether at all) learners should make and develop a self-assessment of grammar when giving a presentation or speaking in general.

There are not many studies on self-assessment of grammar in foreign language learning, which might illustrate the difficulty of this task. One of the few studies was conducted by Nurov (2000) in EFL settings. The study investigated the correlations between grammar-focused teachers' evaluations, students' self-assessments, and a test. The results showed a low correlation between the students' self-assessments and the other types of evaluations. Therefore, the question arises: What accuracy in self-assessment of grammar can be achieved?

Accuracy in self-assessment (not just) of grammar is thorny, claim Brown, Andrade and Chen (2015, p. 445). A research review by these authors suggests that a simple and concrete task and specific and concrete reference criteria promote accuracy. They also note that „more accurate self-assessors tend to be less optimistic than more inaccurate self-assessors" (Brown, Andrade & Chen, 2015, p. 446) and conclude their directions and cautions for research on student self-assessment by asking, "Does it matter if students are inaccurate in their self-assessments, so long as they are engaged in thinking

about the quality of their work?" (p. 445). In answering this question, we can rely on Brown, Andrade and Chen's (2015) argument that effective self-assessment doesn't necessarily have to be only accurate (p. 448).

In our study, self-assessment of grammar had relatively poor evidence, mainly in the form of vague phrases such as "The grammar could be better" or "There were a lot of grammar mistakes." If effective self-assessment serves "the purpose of doing better work" (Rolheiser & Ross, 2001, p. 8) and evidence-based assessment is essential for formative (self-)assessment (Brown & Harris, 2013, p. 368), it is not surprising that insufficient evidence does not serve this purpose of self-assessment well. Nevertheless, where is the desired level of evidence in a self-assessment focusing on the broad area of grammar? In the study by Pereira, Bermúde and Medina (2018), participants used video recordings of their speech to assess grammatical accuracy and range. In particular, they were able to focus on the confidence and clarity of grammatical structures, error-free sentences, and verb forms. However, they faced challenges in widening their grammatical structures. Despite the relatively poor evidence in the self-assessment of grammar, the participants improved their grammatical accuracy. The authors, therefore, conclude that the goal of self-assessment in grammar should not only be to improve awareness of correct grammar use but also to compare individual performance to get a sense of their improvement and support their motivational potential. Regarding the cognitive demands of self-assessment focused on speaking, self-assessment using recordings of one's performance seems to be an essential training tool. When high cognitive demands are combined with a difficult-to-delimit reference level of the criteria to be assessed, such as grammar, learners easily slip into "The grammar could be better." or "There were a lot of mistakes.". Or they do not pay attention to the grammatical level of the production in their self-assessment, as the following statement illustrates: "I pay attention to some grammatical phenomena when I write notes, but I'm afraid that when I start speaking, I won't be able to focus on the grammar anymore." 1H_20).

Limitations

Due to the qualitative research design and small sample size, the study's main limitation is that the results cannot be generalised. The research attempts to compensate for this by providing a greater depth of data.

Another limitation relates to the different forms of self-assessment prompts in each intervention phase. For this reason, data from the first and last phases are reported separately, and a comparison of the results is only discussed.

Finally, it should be noted that self-assessment processes have been examined on the basis of verbalized cognition, which may not correspond to fully realized cognition.

5 Conclusion

The study presented has provided valuable insights into the content of self-assessments, which otherwise often remains hidden, and their characteristics before and after participation in an intervention to develop self-assessment skills.

We can conclude that although there are many manuals for teachers on developing learners' self-assessment, the findings underline the importance of prospective teachers of German as a foreign language gaining experience in developing their own self-assessment skills during teacher education. The findings showed that prospective teachers do not necessarily know how to carry out self-assessment effectively, i.e., making it evidence-based and focusing on different levels of quality (positive and negative) of both language-specific and non-language-specific aspects. It can therefore be assumed that they would not develop this effectively with their students either. However, the completion of the intervention appears to contribute to the effectiveness and therefore validity of the self-assessment skills, thus enhancing the impact on teachers' ability to develop learners' self-assessment skills in their subsequent teaching practice. Based on these findings, the following implications can be drawn, which relate primarily to foreign language teacher education

At a general level of teacher education, the aim should be for future teachers to develop the habit of self-assessment or, more generally, of self-reflection on their actions – whether about their learning or, later, about their teaching. Of course, practice alone is not a sufficient condition. What is important is to develop self-assessment through reflective and structured work with goals and criteria at the outset and on an ongoing basis. Regardless of the area or focus of self-assessment, the goal is to adopt some kind of universal practice: If someone wants to evaluate the quality of his/her actions, he/she

needs criteria and indicators that represent the desired outcome and help him/her to find the evidence in their performance. In addition, he/she should proactively calibrate his/her own self-assessment with external feedback and, for example, work with video recordings of performance that he/she cannot get back to (e.g. an oral speech). Subsequently, working with goal setting, evaluating the effectiveness of one's efforts (attribution), and formulating alterations, and their implications. The intervention-based research design presented can serve as an empirically supported example of working with self-assessment development in teacher education that incorporates these aspects and through which problematic areas of self-assessment development can be identified and further addressed.

At the same time, it should be assumed that the development of self-assessment is highly individualized due to its interaction with various individual-specific variables, which also underlines the importance of an interdisciplinary approach to the development of self-assessment (and research on it). For curriculum development, it would be desirable to link the different approaches to the development of students' and prospective teachers' self-assessment within the pedagogical-psychological, domain-specific, and field-didactic dimensions of the studies and to take a more integrative approach to this topic. While the pedagogical and psychological components of the studies can effectively contribute to the individual-specific level of the self-assessment, the field didactics can contribute to the specifics of the subject of the self-assessment.

Specifically for the learning and teaching of foreign languages, the results suggest that a strong emphasis should be placed on negotiating criteria and then working with them. One particular area is the language-specific aspects of grammar and vocabulary and the associated criteria, where not everything can be covered. However, at least the grammatical phenomena that are addressed at a given language level can be clearly defined. In terms of vocabulary, it is also possible to focus on specific areas relating to the curriculum and the use of associated phrases.

A related point is that prospective teachers should have adequate knowledge of linguistic disciplines such as phonetics, phonology, morphology, syntax, semantics, and pragmatics in order to be able to name individual linguistic structures at an appropriate level of concretisation and performance. At the same time, it should be borne in mind that teachers should be expected to have a different knowledge of terminology and theory than learners.

Based on the above, the process of negotiating criteria can be considered as a research desideratum for an area where further data is needed, as it shows the importance of this phase for the subsequent process of developing effective self-assessment.

I would like to thank to the reviewers, as well as to the guest editor, Prof. Janíková, and Dr. Minaříková, the editor-in-chief of the Pedagogical Orientation journal, for their invaluable contributions and guidance throughout the publication process.

References

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.

Andrade, H., & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, *48*(1), 12–19.

Apeltauer, E. (2010). Lernerautonomie, Lehrerautonomie und Deutsch als Fremdsprache. In Y. Eğit (Hrsg.), *XI. Türkischer Internationaler Germanistik Kongress. Globalisierte Germanistik: Sprache- Literatur-Kultur. Tagungsbeiträge* (pp. 15–34). Ege Üniversitesi.

Baleghizadeh, S., & Masoun, A. (2013). The Effect of Self-Assessment on EFL Learners' Self-Efficacy. *TESL CANADA JOURNAL/REVUE TESL DU CANADA*, *31*(1), 42–58.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W. H. Freeman.

Belgrad, S. F. (2013). Portfolios and E-Portfolios: Student reflection, self-assessment, and goal setting in the learning process. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 331–346). Sage.

Benson, P. (2001). *Teaching and researching autonomy in language learning.* Longman/Pearson Education.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappa*, *86*(1), 8–21.

Blanche, P. (1988). Self-assessment of foreign language skills: Implications for teachers and researchers. *RELC Journal*, *19*(1), 75–93.

Boud, D. (2003). *Enhancing learning through self-assessment.* RoutledgeFarmer.

Boud, D., & Brew, A. (1995). Developing a typology for learner self-assessment practices. *Research and Development in Higher Education*, *18*, 130–135.

Brookhart, S. M. (2013). Grading. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 257–271). Sage.

Brown, G. T., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 367–393). Sage.

Brown, G. T., & Harris, L. R. (2014). The future of self-assessment in classroom practice: reframing self-assessment as a core competency. *Frontline Learning Research*, *3*, 22–30.

Brown, G. T. L., Andrade, H. L., & Chen, F. (2015). Accuracy in student self-assessment: directions and cautions for research. *Assessment in Education: Principles, Policy & Practice*, *22*(4), 444–457.

Butler, Y. G. (2023). Self-assessment in second language learning. *Language Teaching*, *57*(1), 42–56.

Chen, Y. M. (2008). Learning to self-assess oral performance in English: A longitudinal case study. *Language Teaching Research*, *12*(2), 235-262.

Council of Europe (2001). *The common European framework of reference for languages.* Cambridge: Cambridge University Press.

Czech School Inspectorate (2021). *Kvalita vzdělávání v České republice: Výroční zpráva 2020–2021.* Praha: Česká školní inspekce. Available from https://csicr.cz/CSICR/media/Prilohy/2021_p%c5%99%c3%adlohy/Dokumenty/VZ_CSI_2021_e-verze_22_11.pdf (01.11.2022)

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education, 24*(3), 331–350.

Gan, M. J. S., & Hattie, J. (2014). Prompting secondary students' use of criteria, feedback specificity and feedback levels during an investigative task. *Instructional Science*, *42*(6), 861–878.

Grotjahn, R., & Kleppin, K. (2015). *Prüfen, Testen, Evaluieren.* Klett-Langenscheidt.

Hattie, J. (2018). *Hattie Ranking: 252 Influences And Effect Sizes Related To Student Achievement.* https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/.

Hattie, J. (2020). *Lernen sichtbar machen für Lehrpersonen.* Schneider Verlag Hohengehren.

Hosseini, M., & Nimehchisalem, V. (2021). Self-assessment in English language teaching and learning in the current decade (2010-2020): A systematic review. *Open Journal of Modern Linguistics*, *11*, 854-872.

Hunter, D., Mayenga, C., & Gambell, T. (2006). Classroom assessment tools and uses: Canadian English teachers' practices for writing. *Assessing Writing*, *11*(1), 42–65.

Janíková, V. (2007). *Autonomní učení a lexikální strategie při osvojování cizích jazyků.* MU.

Keith, N., & Frese, M. (2005). Self-regulation in error management training: Emotion control and metacognition as mediators of performance effects. *Journal of Applied Psychology*, *90*(4), 677–691.

Klečková, G., & S., Hanušová et al. (2019). *Rámec profesních kvalit učitele cizího jazyka.* NÚV.

Krainer, L., & Lerchster, R. E. (Hrsg.) (2012). *Interventionsforschung.* Band 1: Paradigmen, Methoden, Reflexionen. Springer.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121–1134.

Krykorková, H. (2010). Autoregulace a metakognice v kontextu psychologie učení. In K. Hrbáčková et al. (Eds.), *Kognitivní a nonkognitivní determinanty rozvoje autoregulace učení studentů* (pp. 27–44). Paido.

Lasonen, J. (1995). A case study of student self-assessment in upper secondary education. In J. Lasonen & M.-L. Stenstrom (Eds.), *Contemporary issues of occupational education in Finland* (pp. 199–215). University of Jyvaskyla, Institute for Educational Research.

Léger, de S. (2009). Self-assessment of speaking skills and participation in a foreign language class. *Foreign Language Annals*, *42*(1), 158-177.

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.

Mayring, P. (2010). *Qualitative Inhaltsanalyse: Grundlagen und Techniken.* Beltz Verlag.

Mayring, P. (2014). *Qualitative content analysis: theoretical foundation, basic procedures and software solution.* SSOAR Open Access.

McDonald, B., & Boud, D. (2003). The impact of self-assessment on achievement: The effects of self-assessment training on performance in external examinations. *Assessment in Education: Principles, Policy & Practice*, *10*(2), 209–220.

McMillan, J. H. (2013). Why we need research on classroom assessment. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 3–16). Sage.

McMillan, J. H., & Hearn, J. (2008). Student self-assessment: The key to stronger student motivation and higher achievement. *Educational Horizons*, *87*(1), 40-49.

Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, *82*(2), 213–225.

MŠMT (2004). *Školský zákon 561/2004 Sb.* MŠMT. http://www.msmt.cz/uploads/skolsky_zakon.pdf.

MŠMT (2017). *Rámcový vzdělávací program pro základní vzdělávání.* MŠMT. https://www.msmt.cz/file/41216_1_1/.

Nicol, D. J., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review Perspective. *Assessment & Evaluation in Higher Education*, *39*(1), 102–122.

Nurov, A. (2000). *Self-Assessment of foreign language achievement: The relationship between students' self-assessment, teachers' estimates and achievement test.* Master's Thesis. Bilkent University.

Oscarson, A.-D. (2009). *Self-assessment of writing in learning English as a foreign language.* ACTA UNIVERSITATIS GOTHOBURGENSIS. http://files.eric.ed.gov/fulltext/ED505960.pdf.

Panadero, E., Brown, G. T. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*, 803–830.

Pereira, M. J. C., Bermúdez, J. A. L., & Medina, L. A. V. (2018). Improving L2 oral accuracy and grammatical range through self-assessment of video speech drafts. *Profile: Issues in Teachers' Professional Development*, *20*(2), 127–142.

Poehner, M. E. (2012). The zone of proximal development and the genesis of self-assessment. *The Modern Language Journal*, *96*(4), 610–622.

Raya, M. J. (2014). Lehrerautonomie: Auf dem Weg zu Freiheit und Eigenverantwortung. *Jahrbuch Deutsch als Fremdsprache*, *40*, 147–162.

Rheinberg, F. (2008). Bezugsnormen und die Beurteilung von Lernleistung. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch der Pädagogischen Psychologie* (pp. 178–186). Hogrefe.

Rolheiser, C., & J. A., Ross (2001). *Student self-evaluation: What research says and what practice shows.* http://eloewen.pbworks.com/w/file/fetch/118309278/Student%20Self%20Evaluation%20What%20Research%20Says%20and%20What%20Practice%20Shows.pdf

Ross, S. (1998). Self-assessment in second language testing: A meta-analysis and analysis of experimental factors. *Language Testing*, *15*(1), 1–20.

Ross, J. A. (2006). The reliability, validity, and utility of self-assessment. *Practical Assessment Research & Evaluation*, *11*(10), 1–12.

Ross, J. A., Rolheiser, C., & Hogaboam-Gray, A. (1998). Skills training versus action research in-service: Impact on student attitudes to self evaluation. *Teaching and Teacher Education*, *14*(5), 463–477.

Schneider, G. (1996). Selbstevaluation lernen lassen. *Fremdsprache Deutsch. Zeitschrift für die Praxis des Deutschsprachenunterrichts Sondernummer 1996*, 16-23.

Seliger, H. W., & Shohamy, E. (1989). *Second language research methods.* Oxford University Press.

Stiensmeier-Pelster, J., & Schöne, C. (2008). Fähigkeitsselbstkonzept. Self-copcept of Ability. In W. Schneider & M. Hasselhorn (Eds.), *Handbuch der Pädagogischen Psychologie* (pp. 63–73). Hogrefe.

Taras, M. (2010). Student self-assessment: Processes and consequences. *Teaching in Higher Education*, *15*(2), 199–209.

Tassinari, M. G. (2010). *Autonomes Fremdsprachenlernen: Komponenten, Kompetenzen, Strategien.* Lang.

Vašátková, D. (1995). *Atribuční procesy: Studie.* Gaudeamus.

Volante, L., & Beckett, D. (2011). Formative assessment and the contemporary classroom: Synergies and tensions between research and practice. *Canadian Journal of Education*, *34*(2), 239–255.

Weiner B., & Kukla A. (1970). An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology*, *15*(1), 1–20.

Weskamp, R. (2007). Self-assessment/Selbstkontrolle, Selbsteinschätzung und -einstufung. In K. R. Bausch, H. Christ, & H. J. Krumm (Hrsg.), *Handbuch Fremdsprachenunterricht.* (pp. 382-384). Francke Verlag.

Wilkening, M. (2013). *Selbst- und Partnerevaluation unter Schülern.* Beltz Verlag.

William, D. (2013). Feedback and Instructional Correctives. In J. H. McMillan (Ed.), *Sage handbook of research on classroom assessment* (pp. 197–214). Sage.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, *41*(2), 64–70.

Author

Mgr. Jana Veličková, Ph.D., Masaryk university, Faculty of Education, Department of German language and literature, Poříčí 9, 603 00 Brno, e-mail: velickova@ped.muni.cz

Rozvíjení sebehodnocení zaměřeného na ústní prezentace u budoucích učitelů němčiny jako cizího jazyka: Analýza sebehodnoticích výroků

Abstrakt: Sebehodnocení učícího se se jeví jako významný prediktor učebního úspěchu (Hattie, 2018), přesto je v českých základních školách realizováno nedostatečně (ČŠI, 2021). Jednou z příčin může být chybějící zkušenost učitelů s rozvíjením vlastního sebehodnocení. Předložený text představuje studii, jejímž rámcem byl 12týdenní intervenční program zaměřený na rozvoj sebehodnocení v oblasti prezentování v němčině u budoucích učitelů němčiny jako cizího jazyka (n = 15). Cílem studie bylo zjistit, jaké obsahové charakteristiky vykazují sebehodnoticí výroky participantů před a po absolvování intervenčního programu. Za použití metody *Lautes Erinnern* (vzpomínání nahlas) bylo získáno 25 sebehodnotících výpovědí (13 před intervencí, 12 po intervenci), které byly analyzovány pomocí induktivní tvorby kategorií kvalitativní obsahové analýzy (Mayring, 2014). Analýza ukázala tři hlavní charakteristiky rozvoje sebehodnocení: nárůst evidence sebehodnocení, přesun od zaměření primárně na jazykově nespecifické aspekty výkonu i k jazykově specifickým a vývoj od převažujícího negativního hodnocení k zastoupení také pozitivního hodnocení. V závěru studie jsou diskutovány implikace pro kvalitnější vzdělávání učitelů v oblasti adekvátního rozvíjení jejich sebehodnocení.

Klíčová slova: sebehodnocení učícího se, rozvoj sebehodnoticích dovedností, obsahová analýza, němčina jako cizí jazyk, vzdělávání učitelů

The role and perception of peer and teacher feedback in multiple-draft feedback provision on foreign language learners' writing

Blanka Pojslová

Masaryk University, Language Centre

Received 28th December 2022/ final version received 26th April 2023 /
/ accepted 10th May 2023

Abstract: This paper presents the findings of an experimental study which examines how effective peer feedback is as a substitute for teacher feedback in computer-mediated multiple-draft feedback provision on undergraduate EFL learners' writing. Sixty-five university students were assigned to two comparison groups to receive different feedback treatments. The first group (N = 33) was given multiple-draft feedback on three subsequent drafts of the same text only by the teacher, while the second group (N = 32) was given feedback by three peers on the first draft, and by the teacher on the second and third drafts. The study adopted a quasi-experimental pre-test/post-test design, with two comparison groups which differed in the source of feedback they received on their writing. The data analysis was conducted by employing the Wilcoxon rank test to evaluate changes in writing quality scores after the treatments. Moreover, the paper discusses how learners in the comparison groups perceived teacher-only and combined peer-teacher feedback, specifically focusing on giving and receiving peer feedback. The findings of the study indicate that both peer-teacher and teacher-only feedback contributed to significant improvement in writing quality in both comparison groups regarding all three perspectives from which the writing quality was assessed – overall quality, genre, and register. The findings confirm learners' strong preference for teacher feedback, but also show that peer feedback helps develop learners' writing ability and performance, and aids learners with their own learning process.

Keywords: computer-mediated feedback, peer feedback, teacher feedback, writing quality, feedback perceptions, English as a Foreign Language

Assessment is an essential part of teaching, curriculum development, and student learning, and can be seen from two different perspectives – as assessment *of* learning, and as assessment *for* learning. While the former focuses on how well the skills, subskills, and content have been learned, the latter aims to determine the learner's incremental improvements (Newton et al., 2018, p. 66). It is the assessment *for* learning which provides learners

https://doi.org/10.5817/PedOr2022-4-444

with opportunities not only to reflect on their learning but also to receive feedback on their learning.

Feedback is one of the most critical factors contributing to learning, and underpins the other factors influencing learning (Hattie, 2009, p. 253). Hattie and Timperley (2007, p. 81) conceptualise feedback as "information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding." However, feedback is more than information about what is wrong or what can be improved. Feedback is an interactive process between a feedback giver and a feedback receiver in which learners, as active agents, seek and use information from different sources and decide which feedback to use and how to use it (Boud, 2015, p. 4).

The current study was conducted to help foreign language teachers decide how and when to implement computer-mediated peer feedback in their classes. It aims to determine whether making peer feedback a part of multiple-draft feedback provision while adhering to best practice principles of feedback provision can contribute to improvements in the quality of learners' writing after they were instructed on conventions of academic writing and genre requirements. The study also shows how students perceived the computer-mediated feedback they received, focusing on the perceived amount of feedback and attention they paid to different categories of feedback. Finally, perceptions related to giving and receiving peer feedback were investigated.

Specifically, this study seeks to expand on existing research into computer-mediated written feedback and answer the following research questions:

RQ1: How do two feedback treatments with different sources of feedback (teacher-only and peer-teacher) compare regarding changes in writing quality?

RQ2: How do students' perceptions of teacher-only and peer-teacher feedback treatments compare?

RQ3: How do students perceive giving and receiving peer feedback?

The following text presents a theoretical background to the study, a brief review of related research, and a description of the methodology, including research design, instruments, treatment, and data analysis. This is followed

by a report of the findings and a discussion. The article ends with some suggestions for future pedagogical practice.

1 Theoretical background

Teachers, the most common feedback givers, should see feedback as a loop. This loop involves not only giving feedback but also detecting that the feedback was understood and, most notably, that feedback led to a change in learning (Boud, 2015). Therefore, teachers need to ensure that feedback has been effective, and that the information provided has been apprehended and transformed into learning by feedback recipients. To achieve this effectiveness, feedback as information about the gap between the current and desired level of understanding needs to be specifically related to the task to fill this gap (Sadler, 1989). Moreover, feedback must be situated in a learning context to which the feedback is related, and it must happen after the learner's response to the teacher's instruction. Feedback is also most effective for learners if it is based on their faulty interpretations rather than on a complete lack of understanding (Hattie & Timperley, 2007).

There has been extensive research into best practices of feedback giving for effective language learning and development of learners' writing skills. Specifically, feedback on writing should be balanced and timely. Besides corrective feedback on linguistic errors, feedback should include comments on the structure, organisation, content, and style of the learner's writing (Zamel, 1985; Ferris, 2003; Hyland, 2007). Furthermore, feedback should be provided multiple times on the same text and related to the teacher's instruction (Ferris, 1997; Hyland & Hyland, 2006). As for the forms of feedback, Ferris (2010) suggested combining direct and indirect feedback methods, as they may deliver different but complementary results. Indirect feedback methods might be preferred to direct feedback methods in the case of advanced writers, since indirect feedback leads to problem-solving and reflection on existing knowledge, which is more likely to contribute to long-term acquisition and promotes responsibility for learners' own writing (Bitchener & Knoch, 2010; Ferris et al., 2013). Low-proficiency learners will appreciate direct feedback, as their linguistic resources are relatively limited. Finally, feedback should be specific and selective rather than covering all instances of problematic language to prevent feedback from being frustrating for writers and exhausting for feedback givers (Mantello, 1997; Ferris, 2002).

Feedback on learners' writing might be conveyed in oral or written mode. The oral mode takes the form of teacher-student in-person conferences in which the teacher and student interactively negotiate the meaning of a text through dialogue (McCarthy, 1991). Written feedback can be defined as comments written on students' texts to provide a reader response to students' efforts while helping them improve and learn as writers (Hyland, 2007). Written feedback provides fewer opportunities for clarification and is less immediate than oral feedback; however, students can return to it and take time to consider it.

With the development of information technologies, these feedback modes are increasingly mediated by computers, mainly at the tertiary education level (Elola & Oskoz, 2017). In classes where a process-genre approach to writing is adopted, computer-mediated feedback refers to human feedback given by exchanging texts and comments through computer networks, either synchronously, in real-time, or asynchronously (Ware & Warschauer, 2006). Advocates for feedback as a critical element of the process-genre approach to teaching writing recommend that students receive feedback from a range of sources given on multiple drafts (Badger & White, 2000). Thus, teacher feedback, the traditionally dominant form of feedback (Paulus, 1999; Montgomery & Baker, 2007), should be complemented by other sources of feedback, one of which can be peer feedback.

Liu and Hansen (2005) define peer feedback as "the use of learners as sources of information and interactants for each other in such a way that learners assume roles and responsibilities normally taken on by a formally trained teacher, tutor, or editor in commenting on and critiquing each other's drafts in both written and oral formats in the process of writing" (p. 1). To reflect the learner's dual role as a writer and reviewer in this process, Wakabayashi (2013, p. 181) considers this dual role and redefines peer feedback as "a collaborative learning task by which learners acquire revision procedures while taking on the dual role of writer and reviewer."

Benefits of peer feedback for its recipients include positive effects of peer feedback on writing quality (Hedgcock & Lefkowitz, 1992; Villamil & Guerrero, 1998; Berg, 1999), an enhanced sense of audience (Mangelsdorf, 1992; Carson & Nelson, 1994; Ho & Savignon, 2007) and ownership of the text (Tsui & Ng, 2000). When describing the benefits of peer feedback, Mendonça and Johnson (1994, p. 746) emphasise the possibility for students

to "reconceptualise their ideas in light of their peers", and Mittan (1989) stresses the importance of receiving reactions and responses from authentic readers and a clearer understanding of reader expectations. Furthermore, peer feedback is often easier to understand and more adequate to the developmental level of the learners (Chaudron, 1984; Allison & Ng, 1992). Most importantly, it develops critical evaluation and self-revision skills, and it supports learner autonomy (Villamil & DeGuerrero, 1998; Tsui & Ng, 2000; Rollinson, 2005).

The benefits of peer feedback for its givers were examined by Tsui and Ng (2000), who found that learners learned more about writing by reviewing peer texts than by receiving peer comments. Lundstorm and Baker (2009) showed that the group in which students only gave peer feedback but received none significantly outperformed the group which only received peer feedback.

However, peer feedback has its limitations, as peers tend to give comments on a surface level and neglect global issues (Leki, 1990). Furthermore, peer comments can be vague, unhelpful, and even counterproductive as students may have inappropriate expectations about the content and structure of peers' text (Ferris & Hedgcock, 2005, p. 227).

Despite the benefits of peer feedback, teachers, especially in the EFL context, might remain sceptical about implementing peer feedback in their classes because they find it time-consuming, unreliable, and hard to monitor (Meletiadou & Tsagari, 2022). This is particularly relevant for peer feedback given asynchronously in a computer-mediated mode where teachers have little control over peer interactions.

2 Literature review

There is an extensive body of research exploring written feedback from numerous perspectives; for the purpose of this study, which investigates the role of computer-mediated peer and teacher feedback in improving the quality of EFL learners' writing, the following literature review focuses only on the studies which measure the impact of computer-mediated feedback given by teacher and/or peers on learners' writing production, and on how learners perceive feedback they received. To identify the relevant studies, ScienceDirect, Sage Journal, ERIC, Scopus, and Elsevier databases were

searched for the following key words: peer feedback, writing quality, and feedback perceptions while covering the period of 1995-2020.

2.1 Impact of computer-mediated feedback on writing quality

With ICT developments, computer-mediated feedback has become more visible in writing classes, mainly in tertiary education (Hyland & Hyland, 2006; Elola & Oskoz, 2017). However, studies examining the effect of computer-mediated written feedback on the writing quality of EFL writers are relatively scarce. AbuSeileek and Abualsha'r (2014) compared one control and three experimental groups that received different computer-mediated feedback treatments on their writing using track changes, recast feedback, and metalinguistic feedback. All three experimental groups outperformed the control group that did not receive any feedback, and the group that received feedback in the form of track changes significantly outperformed the other two experimental groups on writing quality.

Pham et al. (2020) explored the effect of peer feedback on global and local aspects of EFL academic writing production. They found that post-test writing production improved significantly from global (organisation, idea development, flow) and local (accuracy, punctuation, syntax, lexical choice) perspectives. Motallebzadeh et al. (2011) compared the effect of traditional pen-and-pencil teacher feedback (control group) with computer-mediated teacher and peer feedback (experimental groups) on writing quality. The results showed that both experimental groups outperformed the control group, and the peer feedback group outperformed the experimental group, which received computer-meditated feedback from the teacher.

Al-Olimat and AbuSeillek (2015) compared three computer-mediated feedback treatments: teacher-only, peer-only, and combined peer-teacher feedback. The findings revealed that all three experimental groups, which received one of the computer-mediated feedback treatments, significantly outperformed the control group, which neither received nor provided feedback. The group that received combined peer-teacher feedback significantly outperformed the other experimental groups in writing quality.

2.2 Students' perceptions of feedback

Learners' perceptions of feedback should be taken into consideration, as learners' beliefs and attitudes are "a significant contributory factor in the

language learning process and success" (Breen, 2001). Studies on students' perceptions of traditional pen-and-paper written feedback suggest that students appreciate teacher feedback and prefer it to other feedback forms, such as peer and self-evaluation (Saito, 1994; Zhang, 1995). Students overwhelmingly (94%) prefer teacher feedback to non-teacher feedback, but the majority (61%) preferred peer feedback over self-feedback (Zhang, 1995). Nevertheless, students recognise the importance of peer feedback. Yang et al. (2006) claim that reading peers' writing and giving peer feedback was perceived as useful by 70% of the peer feedback class students because they can learn from each other's strong points, which compensate for their own weaknesses. Moreover, mutual communication contributes to understanding and finding better solutions to writing problems. Research on perceptions of computer-mediated feedback suggests that students perceive computer-mediated feedback as useful and relevant (Liu & Sadler, 2003; Lu & Bol, 2007; Elola & Oskoz, 2016; Ene & Upton, 2018) but usually prefer face-to-face feedback on their writing to computer-mediated feedback, even though the latter leads to deeper revisions (Schultz, 2000; Liu & Sadler, 2003; Tuzi, 2004; Guardado & Shi, 2007).

3 Method

3.1 Ethical considerations

All participants agreed to take part in this study, and a consent form was obtained from each of them. Also, there was no control group that did not receive any treatment, since not giving feedback on participants' writing might have impeded their successful completion of the course.

3.2 Context of the study and participants

The study was conducted in the last semester of the four-semester English for Specific Purposes (ESP) course at the Faculty of Economics, Masaryk University, Brno, Czech Republic. This ESP course aims to develop students' communicative competence in Business English with a target CEFR level of C1. Each semester of the course focuses on a different aspect of foreign language communicative competence. The semester in which the study was conducted aims to familiarise the students with selected conventions of academic writing relevant to their needs, and with the genre requirements of an expository essay.

The study participants consisted of sixty-five undergraduate EFL students from four intact classes of a total of fourteen classes. The intact classes were utilised to avoid interfering with normal university schedules and activities. However, the intact classes were randomly assigned to comparison groups. Two classes each were randomly selected as Group 1 (N = 33) to receive teacher-only feedback and Group 2 (N = 32) to receive combined peer-teacher feedback. The participants, aged 21–24, were homogenous regarding their language proficiency, as they had to undergo three prerequisite courses that were completed by standardised end-of-course pro-achievement tests. The detailed description of participants' profiles can be found in Appendix A.

3.3 Research Design

The current study mostly adopted a quantitative research design, with some qualitative features in the form of open-ended questions in the student survey on feedback perceptions. The quantitative research took the form of a Comparison Group Pretest Postest design (Mackey & Gass, 2005, pp. 146–147) with two comparison groups each receiving a different treatment, which was complemented by a survey on feedback perceptions. The research adopted a quasi-experimental design, since it was not feasible to randomly assign students to comparison groups due to institutional constraints. The classes, which constituted the two comparison groups, were taught by the same teacher, who was also the researcher and feedback giver. The student survey on feedback perceptions was designed as Likert-scale questionnaires with open-ended items that prompted students to elaborate on some Likert-scale items.

The study was conducted over 13 weeks. In the first six weeks, the participants were introduced to selected conventions of academic writing and genre requirements of an expository essay, specifically a problem-solution essay (PSE). Having been given this input, they were assigned to write the first draft of the problem-solution essay, on which they received three-draft computer-mediated feedback. The first drafts were collected in two pre-test learner corpora.

The feedback treatment that each comparison group received differed in the source of feedback. Comparison Group 1 received computer-mediated teacher-only feedback on all three drafts of the problem-solution essay. In contrast, Group 2 received computer-mediated peer feedback on the

first draft, and teacher feedback on the second and third drafts. After the treatments, participants in both comparison groups were assigned to write post-test essays that were collected in two post-test learner corpora. Finally, the questionnaires were administered to examine how students in both comparison groups perceived the feedback treatment they were given.

3.4 Data collection

The pre-test and post-test essays were elicited using two different prompts, and the results of each prompt were compiled in separate corpora – thus resulting in two pre-test corpora and two post-test corpora. The prompts, piloted on a similar population before, offered two topics, and participants could choose either one, depending on their preferences and content knowledge. This decision was based on the findings of Laufer & Nation (1995) that when students are able to choose their topic, it increases their interest in the writing task. The prompts did not explicitly state genre, stylistic, or formal requirements, as the participants had already been familiarised with these in the contact classes. The prompts used for eliciting the learner corpora can be found in Appendix B.

Two raters independently rated the essays in pre-test and post-test corpora. Both raters hold an MA degree in English language and literature and have had ten years of experience teaching and assessing students in English for Specific Purposes courses at the tertiary level. The raters gave scores to anonymised students' essays using three different rating scales to measure three different aspects of writing quality. Their two scores on each essay were averaged to compose a final score for each rating scale. If the raters disagreed by more than one point in any of the assessment criteria of a given essay, that essay was rated by a third rater to grade its disputed criterion. The scores given by the third rater were then averaged with whichever of the two scores was closest to it (Paulus, 1999).

The questionnaires were administered electronically at the end of the semester, a week after the submission of the post-test essays. Although they were administered in English, participants could respond in their L1 (Czech or Slovak) in the open-ended items.

3.5 Instruments

Rating scales

Both pre-test and post-test essays were scored using three different assessment scales, which evaluated the writing quality from different perspectives: overall writing quality, genre, and register. The overall writing quality was assessed using the *Certificate in Advanced English* (CAE) assessment scale for Overall writing quality, which consists of four subscales: content, communicative achievement, organisation, and language. The responses were marked on each subscale from 0 to 5 (Appendix C). To evaluate writing quality from the perspective of genre and register, assessment scales were developed by the researcher following Bachman and Palmer (2010, pp. 229–254) and responses were marked on each subscale by the raters from 0 to 4. The scale evaluating writing quality from the register perspective consists of nine criteria that relate to selected conventions of academic writing as they reflect lexico-grammatical features of academic discourse (Appendix D). Similarly, the scale evaluating writing quality from the genre perspective consists of six criteria that relate to the genre requirements of a problem-solution essay (Appendix E).

Questionnaires

To examine students' perceptions of feedback they received on their essays, students completed questionnaires (Appendix F and Appendix G) based on Ferris's (2003) questionnaire *Student survey on teacher feedback*. Ferris's survey was adopted for the needs of the current study by using three original items (3, 4, 9), which were rephrased and renumbered to follow the research design. Two items on giving peer feedback in the questionnaire for Group 2 were added together with open-ended items and the item on feedback usefulness. Both questionnaires had been piloted by administering them to a similar population a year earlier, and administered electronically with a setting that ensured that all respondents had to fill out all items including the open-ended ones.

On this questionnaire, students in both comparison groups shared how much feedback in the category of *Genre*, *Organisation*, *Grammar*, *Vocabulary*, *Academic writing*, and *Mechanics* they think they received on the first and second drafts, and how much attention they think they paid to feedback in the same categories on the first and second draft. Students were further asked

to share how they perceived the usefulness and effectiveness of feedback they received regarding improvement in their writing skills. Students in Group 2 who received combined peer-teacher feedback were asked whether reading their peers' texts and giving peer feedback improved their writing skills. Students ranked their answers on a Likert scale with the choices "A lot", "Mostly/Some", "A little", and "Not at all/None".

3.6 Treatment

The treatment under investigation consisted of two computer-mediated feedback strategies in the form of multiple-draft feedback provision on the same text with a different source of feedback. Group 1 received teacher-only feedback on all three drafts, while Group 2 received peer feedback on the first draft and teacher feedback on the second and third drafts. Before giving peer feedback, students in Group 2 were given a 45-minute training session to familiarise themselves with the rationale and techniques of giving computer-mediated peer feedback. Such training has been shown to significantly improve students' peer reviewing skills (Berg, 1999; Min, 2005). Students were trained to give peer feedback in a similar manner to the teacher's way of giving feedback.

The logistics of the computer-mediated peer feedback were handled by an online application called *Peer Review*, which randomly and anonymously assigned each essay to three peers. The number of peer feedback givers was set to three to compensate for a lower number of peer comments as compared to the number of teacher's comments (Hublová, 2016, p. 141).

Because of the high language proficiency of the participants, indirect forms of feedback were preferred to direct forms. To make the indirect feedback as specific as possible while meeting the student's needs, the indirect feedback combined colour-coded feedback with MS Word comments. The coded feedback covered five broad categories: *Organisation*, *Academic writing*, *Vocabulary*, *Grammar*, and *Mechanics* and a feedback giver used different colour codes to highlight problematic language in the text in relation to these categories. The coded feedback was complemented by MS Word comments mostly on genre-relevant problems and links to external sources that offered more detailed explanations or metalinguistic information.

The feedback giver also completed a feedback checklist with a 4-point scale to specify the extent to which the writer met the expectations regarding the genre requirements, conventions of academic writing, and organisation.

3.7 Data Analysis

The data were analysed to examine how writing quality changed between the pre-test and post-test in the comparison groups. Since the sample size was small (N = 33, resp. N = 32), Shapiro-Wilk tests were performed. The tests did not show evidence of normal distribution (p-values < 0.05) for variables in Group 2, but in Group 1, they showed evidence of normality for some variables (p-value>0.05). Based on this outcome, and after visual examination of the histograms, nonparametric Wilcoxon signed-rank tests were used to make comparisons possible. To measure the magnitude of the experimental effect, the effect size was calculated as *Pearson r* and interpreted as small for r of 0.1-0.29, as medium for r of 0.3-0.49, and as large for r greater than 0.5 (Cohen, 1988, p. 25).

In order to carry out a statistical comparison between questionnaires administered in the comparison groups, numerical values were assigned to the four quantity options given on each question: “A lot” was coded as 4, “Mostly/Some” as 3, “A little” as 2 and “Not at all/None” as 1. After the numerical values were assigned, the students’ responses were averaged for each response item and each feedback category. Open-ended responses were coded using thematic analysis (Suter, 2012).

4 Findings

RQ1: How do two feedback treatments with different sources of feedback (teacher-only and peer-teacher) compare regarding writing quality?

The data in Table 1 show that the means of students’ scores for all three aspects of writing quality increased between the pre-test and post-test in both comparison groups. The coefficient of variation for all three aspects of writing quality decreased in both comparison groups, which means that both feedback treatments contributed to more homogeneous post-test writing production. The reductions in the variation were higher in Group 2 with peer-teacher feedback, with a decrease of 7.18 percentage points (pp) for overall quality as compared to a decrease of 5.91 pp in Group 1, a decrease of 11.52 pp for the genre as compared to a decrease of 7.02 pp in Group 1, and a decrease of 5.07 pp for register as compared to 2.19 pp in Group 1. The results suggest that combined peer-teacher feedback contributes to levelling students’ writing production more than teacher-only feedback.

Table 1

Descriptive statistics for three aspects of writing quality: overall quality, genre, and register

Writing quality	Group 1 Pre-test		Post-test		Group 2 Pre-test		Post-test	
	Mean/SD	V(%)	Mean/SD	V(%)	Mean/SD	V(%)	Mean/SD	V(%)
Overall quality	13.38/2.78	20.77	15.53/2.31	14.86	14.44/3.36	23.27	16.23/2.61	16.09
Genre	16.26/4.01	24.69	18.50/3.27	17.67	18.02/4.31	23.91	20.56/2.55	12.39
Register	28.65/3.52	12.30	30.55/3.09	10.11	29.27/4.50	15.39	31.47/3.25	10.32

Table 2 shows results of the Wilcoxon test that revealed a statistically significant increase in writing quality between the pre-test and post-test in both comparison groups regarding all three aspects of writing quality. In Group 1 with teacher-only feedback, the effect size was large (r = 0.6) for the increase in overall quality and register, and medium (r = 0.4) for genre. In Group 2 with peer-teacher feedback, the effect size was large (r = 0.5) for genre, and medium (r = 0.4) for overall quality and register. The results suggest that teacher-only feedback was more effective regarding improvements in the students' production from the perspective of overall quality and register. In contrast, peer-teacher feedback was more effective regarding improvements from the perspective of genre.

Table 2

Results of the Wilcoxon signed-ranked test for the changes in writing quality

Writing quality	Group 1 Z	**p**	**r**	Group 2 Z	**p**	**r**
Overall quality	-3.360	**0.001**	**0.6**	-2.490	**0.013**	**0.4**
Genre aspect	-2.534	**0.011**	**0.4**	-2.970	**0.003**	**0.5**
Register aspect	-3.360	**0.001**	**0.6**	-2.485	**0.013**	**0.4**

RQ2: How do students' perceptions of teacher-only and peer-teacher feedback treatments compare?

The second research question compared and explored how the participants perceived the feedback treatments they received. First, students in the comparison groups were asked how much feedback they thought they had received on the first and second drafts in various feedback categories. Table 3 shows that the perceived amount of feedback in Group 1 decreased between the first and second draft in all feedback categories. In Group 2, with peer-teacher feedback, the perceived amount of feedback increased in the categories of Genre, Organisation, and Academic writing and decreased in the categories of Grammar, Vocabulary, and Mechanics between the first and second draft.

Table 3

The perceived amount of feedback in feedback categories

Feedback category	Group 1				Group 2			
	1st draft		2nd draft		1st draft		2nd draft	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Genre PSE	2.76	0.92	2.24	0.95	1.97	0.81	2.58	1.23
Organisation	3.03	0.58	2.41	0.99	2.21	0.70	3.00	0.83
Grammar	2.68	0.73	2.32	0.81	2.64	0.74	2.48	0.83
Vocabulary	2.59	0.89	2.21	0.81	2.58	0.66	2.58	0.83
Academic writing	3.38	0.60	2.47	0.83	2.82	0.73	3.09	0.72
Mechanics	2.21	0.73	1.79	0.77	2.36	0.86	2.15	0.83

The results of the Wilcoxon test in Table 4 revealed that the reductions in the perceived amount of teacher feedback in Group 1 in all feedback categories were statistically significant, with a large effect size for the categories of *Organisation* ($r = 0.5$) and *Academic writing* ($r = 0.7$), and with a medium effect size for the categories of *Genre*, *Grammar*, *Vocabulary*, and *Mechanics* ($r = 0.4$). In Group 2, the results revealed that the increase in the perceived amount of feedback was statistically significant in the categories of Genre and *Organisation*, with a large effect size of $r = 0.5$ for *Genre* and $r = 0.7$ for *Organisation*. These results suggest that students in Group 1 perceived that they had received significantly more teacher feedback on the first draft than on the second draft. In contrast, students in Group 2 perceived that they had received significantly more feedback from the teacher on the second draft than from the peers on the first draft in the categories Genre and Organisation.

Table 4

The results of the Wilcoxon test for change in perceptions of feedback amount between drafts

	Group 1			Group 2		
Feedback category	Z	p	r	Z	p	r
Genre PSE	-2.643	**0.008**	**0.4**	-2.877	**0.004**	**0.5**
Organisation	-3.207	**0.001**	**0.5**	-3.912	**0.000**	**0.7**
Grammar	-2.676	**0.007**	**0.4**	-1.076	0.282	0.2
Vocabulary	-2.457	**0.014**	**0.4**	0.000	1.000	0.0
Academic writing	-4.337	**0.000**	**0.7**	-1.889	0.059	0.3
Mechanics	-2.501	**0.012**	**0.4**	-1.377	0.169	0.2

The students in the comparison groups were then asked how much attention they thought they had paid to feedback in various feedback categories on the first and second drafts. Table 5 shows that the perceived amount of attention in Group 1 decreased in all feedback categories between the first and second drafts. In contrast, in Group 2, the perceived amount of attention increased, except for in the category of *Mechanics*. These results suggest that students paid more attention to their first round of teacher feedback, which in Group 1 was the feedback on the first draft, and in Group 2 the feedback on the second draft (except for *Mechanics*).

Table 5

The perceived amount of attention paid to feedback in feedback categories

	Group 1				Group 2			
Feedback category	1st draft		2nd draft		1st draft		2nd draft	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Genre PSE	3.24	1.21	3.03	1.19	1.79	1.52	2.7	1.69
Organisation	3.38	0.70	3.09	1.14	2.58	1.20	3.3	1.16
Grammar	3.41	0.82	2.88	1.10	2.76	1.12	3.03	1.31
Vocabulary	3.35	0.77	2.79	1.12	2.88	1.05	3.03	1.40
Academic writing	3.47	0.71	3.26	0.99	3.09	0.98	3.36	1.20
Mechanics	2.88	1.27	2.24	1.39	2.73	1.32	2.15	1.77

Table 6 shows the results of the Wilcoxon test that revealed that the reductions in the amount of attention paid to feedback in Group 1 were statistically significant in the categories of *Grammar*, *Vocabulary*, and *Mechanics*, with a large effect size for *Grammar* (r = 0.5) and *Vocabulary* (r = 0.5), and a medium effect size for *Mechanics* (r = 0.4). In Group 2, the perceived amount of attention paid to feedback increased significantly in the categories of Genre and *Organisation*, with a large effect size (r = 0.5) for both categories.

Table 6

The results of the Wilcoxon test for change in the amount of attention paid to feedback

	Group 1			Group 2		
Feedback category	z	p	r	Z	p	r
Genre PSE	-0.701	0.484	0.1	-2.637	**0.008**	**0.5**
Organisation	-1.248	0.212	0.2	-2.687	**0.007**	**0.5**
Grammar	-2.887	**0.004**	**0.5**	-1.402	0.161	0.2
Vocabulary	-2.883	**0.004**	**0.5**	-0.739	0.46	0.1
Academic style	-1.064	0.287	0.2	-1.933	0.053	0.3
Mechanics	-2.371	**0.018**	**0.4**	-1.613	0.107	0.3

Finally, students were asked how useful they found the feedback they received and how effective in improving their composition writing skills the feedback was. The data in Table 7 show that 73% of the students in Group 1 thought that teacher-only feedback was useful “a lot”, 21% of the students found it “mostly” useful, and 6% thought it was useful “a little”. In Group 2 16% of the students thought that peer feedback on the first draft was useful “a lot”, 49% of the students found it “mostly” useful, 32% thought it was useful “a little”, and 6% of the students thought it was not useful at all. The mean values show that students in Group 2 with peer-teacher feedback found teacher feedback (mean = 3.61) more useful than peer feedback (mean = 2.61) and more useful than students in Group 1 (mean = 3.56) who received teacher-only feedback.

Table 7

Students' perceptions of feedback usefulness

	Group 1		Group 2			
	1st and 2nd draft		1st draft		2nd draft	
	n	%	n	%	n	%
A lot	24	73	5	16	26	84
Mostly	7	21	15	49	5	16
A little	2	6	10	32	0	0
Not at all	0	0	1	3	0	0
Mean	3.56		2.61		3.61	

Table 8 shows that 94% of the students in Group 1 with teacher-only feedback thought that feedback was effective in improving their writing skills either "a lot" (52%) or "mostly" effective (42%), while 6% of these students found teacher-only feedback effective in improving their writing skills "a little". In Group 2, 68% of the students thought that peer feedback on the first draft was either "a lot" (16%) or "mostly" (52%) effective in improving their writing skills, while 33% of these students thought that peer feedback was either "a little" effective (22%) or not effective at all (10%). However, no student thought that teacher feedback on the second draft was effective "a little" or "not all." Students in Group 2 found teacher feedback on the second draft either "a lot" (68%) or "mostly" (32%) effective in improving their writing skills. The mean values show that students in Group 2 with peer-teacher feedback found teacher feedback (mean = 3.45) more effective in improving their writing skills than peer feedback (mean = 2.58) and more effective than students in Group 1 (mean = 3.35) who received teacher-only feedback.

Table 8

Students' perceptions of feedback effectiveness

	Group 1		Group 2			
	1st and 2nd draft		1st draft		2nd draft	
	n	%	n	%	n	%
A lot	17	52	5	16	21	68
Mostly	14	42	16	52	10	32
A little	2	6	7	22	0	0
Not at all	0	0	3	10	0	0
Mean	3.35		2.58		3.45	

When students in Group 1 with teacher-only feedback were asked to elaborate on how useful and effective the teacher feedback was, they stated that teacher-only feedback contributed to improving their texts and writing ability (e.g., "Owing to the comments and recommendations I received I think there is a *huge improvement*[1] between the first and the last draft. They were really *useful* for me."/R17). They valued the specificity of teacher feedback and appreciated the links to external sources and metalinguistic information (e.g., "It is helpful to see the *comments being linked to the problems* in the text. Then I know what I need to change and how it should be done."/R23; "The corrections and comments were *very factual*."/R9). However, some of the students remained sceptical about the teacher-only feedback (e.g., "Some of the advice I may remember, but most of it will *be forgotten* for sure." / R28; "I had to write it according to teacher's feedback, which is harder than writing on my own."/R22).

RQ3: How do students perceive giving and receiving peer feedback?

The third research question investigated how the students in Group 2 with peer-teacher feedback perceived receiving peer feedback as compared to receiving teacher feedback, as well as their perceptions of giving peer feedback. Table 9 shows the results of the Wilcoxon test that revealed that changes in the perceptions of peer and teacher feedback between the first and second drafts regarding feedback usefulness and effectiveness were statistically significant. Students in Group 2 found teacher feedback on the second draft statistically more useful than peer feedback on the first draft

[1] Keywords in excerpts from qualitative data are italicised.

with a large effect size (r = 0.8) and statistically more effective in improving their writing skills than peer feedback with a large effect size (r = 0.7).

Table 9

The results of the Wilcoxon test for change in feedback perceptions in Group 2

	z	p	r
Perceived usefulness	-4,443	**0.000**	**0.8**
Perceived effectiveness	-4,058	**0.000**	**0.7**

In open-ended questions, students elaborated on the perceived usefulness of peer feedback on the first draft. Some wrote that peer feedback gave them other views on the topic of the essay (e.g., "Thanks to the (peer) feedback I *added* my own *views* to my essay."/R35). Some said it drew their attention to mistakes they would not have otherwise noticed (e.g., "Their feedbacks point to *mistakes* I haven't noticed before."/R52). Some said they realised the importance of the comprehensibility of the text for the reader (e.g., "Moreover, they show me that not every idea which is understandable for me must be *clear for the others*."/R52).

Nevertheless, about one-third of the students in Group 2 did not find peer feedback useful (35%) or effective (33%) in improving their writing skills. These students, in open-ended questions, wrote that they received very little or no feedback from their peers (e.g., "I don't think so… two of three peers just filled in the form where I can see almost *nothing* and added *no comments.*"/R62). Furthermore, they did not consider peer feedback as valuable or knowledgeable as teacher feedback (e.g., "I don't feel I or my colleagues are *eligible* to assess someone's else English."/R36). Some stated that peer feedback comments did not cover the aspects of genre or text organisation (e.g., "Peer's feedback is not very oriented on *composition and structure*."/R47).

When commenting on teacher feedback on the second draft, students from Group 2 expressed more trust in and preference for teacher feedback. They appreciated that teacher feedback was specific and knowledgeable (e.g., "In the teacher's feedback I feel there was *more helpful advice* for improving my writing."/R60; "I can be sure that the teacher only corrects what is *relevant* and I can then use this feedback *without worrying* about it being wrong."/R65) and provided them with comments on genre and organisation (e.g., "Teacher's feedback does not lack comments on *structure and composition*."/R55).

As for perceptions of giving feedback, students were asked whether they found reading peers' texts and giving peer feedback effective in improving their writing skills. Table 10 shows that 58% of the students from Group 2 thought that reading peers' text was "a lot" (10%) or "mostly" (48%) effective in improving their writing skills, as opposed to 23% of students who found reading peers' effective in improving their writing skills "a little" (11%) or "not at all" (2%).

Table 10

Students' perceptions of peer feedback for improvement in writing skills

	Reading peers' texts		Providing peer feedback	
	n	%	n	%
A lot	3	10	7	23
Mostly	15	48	13	44
A little	11	36	9	30
Not at all	2	6	1	3

When asked to elaborate on these questions, students wrote that reading their peers' text helped them realise their own mistakes, compare their level of writing with their peers' level of writing (e.g., "When you see the mistakes of the others you can *become aware of your own mistakes*."/R39), find inspiration, and reflect on their own writing (e.g., "I might *inspire*, learn from *mistakes* and compare my *level of writing* with others."/R34; "I could get some inspiration from essay, which I consider good."/R51).

Regarding the effectiveness of giving peer feedback for improving peer feedback givers' writing skills, 67% of the students found giving peer feedback either "a lot" (23%) or "mostly" (44%) effective in improving their writing skills, as opposed to 33% of the students who found it either "a little" (30%) or "not at all" (3%) effective. In an open-ended question, the students wrote that by seeing peers' *mistakes* they realised their own *mistakes* which they want to avoid next time and saw the mistakes as an opportunity to learn (e.g., "I find beneficial to think about mistakes in others' PSEs so I can avoid make them in my writing."/R57; "When I find the mistakes of my classmates, it is a sign that I realise these mistakes and then I know I should avoid them."/R61).

Furthermore, they stated that giving peer feedback helped them with understanding genre requirements and their application (e.g., "Yes, as I try to look for the composition and structure and so *strengthen my own automation of applying it in my essays*."/R58; "It helps me grasp *the concept of the essay*."/R45).

In contrast, the students who did not find giving peer feedback effective in improving their own writing skills doubted their peer's expertise to give feedback or questioned the effort the peer had put into feedback provision (e.g., "It depends if the colleague has *all necessary skills* and as well *how much work* does the colleague put in the review."/R44). Some students did not find peer feedback specific enough (e.g., "*Inappropriate* color use together with *minimum of comments* made me mainly confused."/R37).

5 Discussion

The findings of this study revealed that both treatments significantly contributed to improving writing quality regarding all three aspects of writing quality. Teacher-only feedback was more effective in terms of overall quality and register, while peer-teacher feedback was more effective in terms of genre. The larger effect of teacher feedback on register might be attributed to the novelty of this aspect of writing for students where the teacher's expertise plays a crucial role in offering support and drawing students' attention to this aspect of writing. This might seem contradictory, as genre was an equally new aspect of writing for students, but here there was a larger effect of peer-teacher feedback on improving writing quality. However, in this case, it might be assumed that peer training and giving peer feedback contributed to the students internalising the genre requirements more effectively than internalising conventions of academic writing. Furthermore, the findings suggest that both treatments might have contributed to more homogenous writing performance of the students in both groups regarding all three aspects of writing quality. Peer-teacher feedback seems to level individual differences in writing performance more than teacher-only feedback in all three aspects, in genre aspects most prominently. This might be attributed to multiple-draft feedback provision which clarified the expectations by indicating where the desired level of performance is and showing how to achieve this desired level.

In Group 1, the perceived amount of feedback decreased significantly between the first and second draft in all feedback categories, while in Group 2 the perceived amount increased in all categories except for the category of *Mechanics*. The increase was statistically significant in the categories of Genre and Organisation, which were at the centre of the feedback treatment along with the category of *Academic writing*. This might be attributed to previous findings (Leki, 1990) that peers tend to give comments on a surface level (*Grammar*, *Vocabulary*, *Mechanics*), and also to the novelty of genre requirements and conventions of academic writing.

Students in both groups paid more attention to the first round of teacher feedback on their writing, which was on the first draft in Group 1 and the second draft in Group 2. In Group 1, the reductions in the perceived amount of attention between the first and second draft were significant in the categories that were not the focus of the feedback treatment (*Grammar*, *Vocabulary*, *Mechanics*). In contrast, in Group 2, the perceived amount of attention increased significantly in the categories of *Genre* and *Organisation*, which were at the heart of the feedback treatment. These results suggest that students realised the gap between their current level of understanding and the desired one and focused more on feedback related to these gaps. Group 2, with combined peer-teacher feedback, then managed to transform this focus into significantly better writing performance regarding the genre aspect of writing, while Group 1 was significantly more successful regarding the register aspect of writing.

As for the perceived usefulness and effectiveness of feedback treatments, the findings revealed that students appreciate and value teacher feedback and found it both useful and effective in improving their writing skills. Students in Group 2 with peer-teacher feedback valued teacher feedback as more useful and effective than students in Group 1, and significantly more useful and effective than peer feedback. This result might be attributed to the varying quality of peer feedback they received on the first draft. Nevertheless, students in Group 2 realised the importance and value of peer feedback, as they found reading peers' text (58%) and giving peer feedback (67%) effective in improving their writing skills. This is in line with previous research (Yang et al., 2006).

When asked about how they perceived feedback, the students mentioned that the most important benefits of peer feedback were self-reflection,

the importance of comprehensibility of the text for the reader, and seeing mistakes as an opportunity to learn. However, students also mentioned the lack of expertise, specificity, and trust as drawbacks of peer feedback. Considering this, the benefits of peer feedback are primarily associated with giving peer feedback whereas the drawbacks are associated with receiving it. This result supports previous studies that also found that giving peer feedback contributes more to improving the quality of students' writing production than receiving (Lundstorm & Baker, 2009) and that receiving feedback is where students can benefit most from peer feedback as independent writers (Tsui & Ng, 2000). When giving peer feedback, students take an active role in their learning and are forced to exercise their thinking rather than passively receiving information, which gradually leads to developing the strategies necessary for generating ideas, editing, and revising their own writing.

The findings of this study have some limitations. Firstly, the quasi-experimental design of the study together with the size of the sample might lower its internal validity and generalizability. Secondly, the study excluded a control group for ethical reasons. Thirdly, the time constraints did not allow for a more sophisticated method of data-collection for the qualitative part of the research in the form of structured or semi-structured interviews. Finally, the different levels of interpretation of the Likert scale by the respondents should be considered.

Despite these limitations, this study has value as one of the few attempts so far to explore the phenomenon of computer-mediated multiple-draft feedback in the context of higher education in the Czech Republic, where the issue is underresearched. Moreover, the focus on the change in genre and register aspects of writing quality after feedback treatments makes this research original and highly relevant for developing academic writing in the EFL context at the tertiary level.

6 Conclusion and recommendations

This paper presents the findings of an empirical study which examined whether peer feedback can be an effective substitute for teacher feedback in multiple-draft computer-mediated feedback provision on foreign language students' writing, and how students perceived the feedback they received. The main aim was to evaluate and compare how feedback treatments, which took the forms of multiple-draft feedback given by the teacher on three drafts

and multiple-draft feedback given by peers on the first draft and teacher on the second and third draft, contributed to improving the writing quality of ESP undergraduate students and how the students' perceptions of these treatments compare.

The quality of writing performance was assessed from three perspectives: overall writing, genre, and register. The perspective of genre covered the genre requirements of an expository essay, and the perspective of register covered the conventions of academic writing as they reflect in linguistic features of writing production. Both aspects were linked to class input which preceded the feedback treatments and were the primary focus of feedback.

Despite approximately 33% of the study participants being sceptical about the usefulness and effectiveness of peer feedback, and despite teachers' negative assumptions about implementing peer feedback in their classes (Meletiadou & Tsagari, 2022), this study shows that making peer feedback part of multiple-feedback provision might benefit both students and teachers. For students, such feedback might help improve the quality of their writing in certain aspects, such as genre, especially if there is sufficient training and a direct relation to class instructions. And for teachers, not being the only ones who provide feedback on all students' drafts could save them time and energy. Peer feedback could also help teachers by levelling their students' writing performance.

Furthermore, peer feedback should be perceived as complementary to teacher feedback rather than as a replacement for it. To make peer feedback complementary and beneficial, students need to be given training on practical aspects of giving feedback via a variety of activities (Liu & Hansen, 2005) using authentic students' written production with examples of both teacher and peer feedback. This training might include explaining how they might benefit from peer feedback, not only as feedback receivers but also as feedback givers. Finally, the entire process of multiple-draft feedback should be supported by careful scheduling so that the activities do not come all at once for the students.

Teachers should also consider the order in which types of feedback are given. Giving teacher feedback before peer feedback might assist peers in giving more specific feedback on the second draft, but some peers might fear that after teacher feedback their feedback will not be trusted by the peers, or that

there might be little to comment on (Yang, 2006). Combining peer written computer-mediated feedback with oral peer feedback, possibly as a part of peer feedback training, might be also considered. Nevertheless, for teachers to make informed decisions about peer feedback implementation in their classes, teacher training in this area is of the upmost importance, especially in the EFL context.

References

AbuSeileek, A., & Abualsha'r, A. (2014). Using peer computer-mediated corrective feedback to support EFL learners' writing. *Language Learning & Technology*, *18*(1), 76–95. http://dx.doi.org/10125/44355

Al-Olimat, S. I., & AbuSeileek, A. F. (2015). Using computer-mediated corrective feedback modes in developing students' writing performance. *Teaching English with Technology*, *15*(3), 3–30. https://files.eric.ed.gov/fulltext/EJ1138425.pdf

Allison, D., & Ng, P. (1992). Developing text revision abilities. In M. Lau & M.J. Murphy (Eds.), *Developing writing: Purposes and practice* (pp. 106–130). Institute of Language Education.

Bachman, L. F., & Palmer, A. S. (2010). *Language assessment in practice: Developing language assessments and justifying their using in the real world.* Oxford University Press.

Badger, R. G., & White, G. (2000). A process genre approach to teaching writing. *ELT Journal*, *54*(2), 153–160. https://doi.org/10.1093/elt/54.2.153

Berg, E. C. (1999). The effects of trained peer response on ESL students' revision types and writing quality. *Journal of Second Language Writing*, *8*(3), 215–241. https://doi.org/10.1016/S1060-3743(99)80115-5

Bitchener, J., & Knoch, U. (2010). Raising linguistic accuracy level of advanced L2 writers with written corrective feedback. *Journal of Second Language Writing*, *19*, 207–217.

Boud, D. (2015). Feedback: Ensuring that it leads to enhanced learning. *Clinical teacher*, *12*(1), 3–7. https://doi.org/10.1111/tct.12345

Breen, M. P. (Ed.). (2001). *Learner contributions to language learning: New directions.* Routledge.

Carson, J. G., & Nelson, G. L. (1994). Writing groups: Cross-cultural issues. *Journal of Second Language Writing*, *3*(1), 17–30. https://doi.org/10.1016/1060-3743(94)90003-5

Chaudron, C. (1984). The effects of feedback on students' composition revisions. *RELC Journal*, *15*(2), 1–14. https://doi.org/10.1177/003368828401500201

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Elola, I., & Oskoz, A. (2016). Supporting second language writing using multimodal feedback. *Foreign Language Annals*, *49*, 58–74. https://doi.org/10.1111/flan.12183

Elola, I., & Oskoz, A. (2017). Writing with 21*st* century social tools in the L2 classroom: New literacies, genres, and writing practices. *Journal of Second Language Writing*, *36*, 52–60. https://doi.org/10.1016/j.jslw.2017.04.002

Ene, E., & Upton, T. A. (2018). Synchronous and asynchronous teacher electronic feedback and learner uptake in ESL composition. *Journal of Second Language Writing*, *41*, 1–13. https://doi.org/10.1016/j.jslw.2018.05.005

Ferris, D. R. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, *31*, 315–339. https://doi.org/10.2307/3588049

Ferris, D. R. (2002). *Treatment of error in second language student writing.* University of Michigan Press.

Ferris, D. R. (2003). *Response to student writing: Implications for second language students.* Routledge, Taylor & Francis Group.

Ferris, D. R. (2010). Second language writing research and written corrective feedback in SLA: Intersections and practical applications. *Studies in Second Language Acquisition*, *32*(2), 181–201. https://doi.org/10.1017/S0272263109990490

Ferris, D. R., & Hedgcock, J. (2005). *Teaching ESL composition: Purpose, process, and practice*. Lawrence Erlbaum.

Ferris, D. R., Liu, H., Sinha, A., & Senna, M. (2013). Written corrective feedback for individual L2 writers. *Journal of Second Language Writing*, *22*, 307–329. https://doi.org/10.1016/j.jslw.2012.09.009

Guardado, M., & Shi, L. (2007). ESL students' experiences of online peer feedback. *Computers and Composition*, *24*(4), 443–461. https://doi.org/10.1016/j.compcom.2007.03.002

Hattie, J. (2009). The black box of tertiary assessment: An impending revolution. In L. H. Meyer, S. Davidson, H. Anderson, R. Fletcher, P. M. Johnston, & M. Rees (Eds.), *Tertiary assessment & higher education student outcomes: Policy, practice & research* (pp. 259–275). Ako Aotearoa.

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

Hedgcock, J., & Lefkowitz, N. (1992). Collaborative oral/aural revision in foreign language writing instruction. *Journal of Second Language Writing*, *1*(3), 255–276. https://doi.org/10.1016/1060-3743(92)90006-B

Ho, M. C., & Savignon, S. L. (2007). Face-to-face and computer-mediated peer review in EFL writing. *CALICO Journal*, *24*(2), 269–290. https://doi.org/10.1558/cj.v24i2.269-290

Hublová, G. (2016). *Rozvoj diskurzivní competence v cizojazyčném akademickém psaní v anglickém jazyce v důsledku účasti v e-learningovém kurzu* [Doctoral thesis]. Masarykova univerzita. https://is.muni.cz/auth/th/63926/pedf_d/DISERTACNI_PRACE.pdf

Hyland, K. (2007). *Second language writing.* Cambridge University Press.

Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing. *Language Teaching*, *39*(2), 83–101. https://doi.org/10.1017/S0261444806003399

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in l2 written production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Leki, I. (1990). Potential problems with peer responding in ESL writing classes. *CATESOL Journal 3*, 5–17. www.journals.aiac.org.au/index.php/alls/artic

Liu, J., & Hansen, J. G. (2005). *Peer response in language writing classrooms.* The University of Michigan Press.

Liu, J., & Sadler, R. W. (2003). The effect and affect of peer review in electronic versus traditional modes on L2 writing. *Journal of English for Academic Purposes*, *2*(3), 193–227. https://doi.org/10.1016/S1475-1585(03)00025-0

Lu, R., & Bol, L. (2007). A comparison of anonymous versus identifiable e-peer review on college student writing performance and the extent of critical feedback. *Journal of Interactive Online Learning*, *6*(2), 100–115.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of peer review to the reviewer's own writing. *Journal of Second Language Writing*, *18*(1), 30–43. https://doi.org/10.1016/j.jslw.2008.06.002

Mackey, A., & Gass, S. M. (2005). *Second language research: Methodology and design.* Lawrence Erlbaum Associates.

Mangelsdorf, K. (1992). Peer reviews in the ESL composition classroom: What do the students think? *ELT Journal*, *46*(3), 274–284. https://doi.org/10.1093/elt/46.3.274

Mantello, M. (1997). Error correction in the L2 classroom. *Canadian Modern Language Review*, *54*, 127–131.

McCarthy, M. (1991). *Discourse analysis for language teachers.* Cambridge University Press.

Meletiadou, E., & Tsagari, D. (2022). Exploring EFL teachers' perceptions of the SSE of peer assessment in external exam-dominated writing classes. *Languages*, *7*(16), 1–18. https://doi.org/10.3390/languages7010016

Mendonça, C., & Johnson, K. (1994). Peer review negotiations: Revision activities in ESL writing instruction. *TESOL Quarterly*, *28*, 745–769. https://doi.org/10.2307/3587558

Min, H. T. (2005). Training students to be successful peer reviewers. *System*, *33*(2), 293–308. https://doi.org/10.1016/j.system.2004.11.003

Mittan, R. (1989). The peer review process: Harnessing students' communicative power. In D. M. Johnson, & D. H. Roen (Eds.), *Richness in writing: Empowering ESL students* (pp. 207–219). Longman.

Montgomery, J. L., & Baker, W. (2007). Teacher-written feedback: Student perceptions, teacher self-assessment, and actual teacher performance. *Journal of Second Language Writing*, *16*, 82–99. https://doi.org/10.1016/j.jslw.2007.04.002

Motallebzadeh, K., & Amirabadi, S. (2011). Online interactional feedback in second language writing: Through peer or tutor? *Theory and Practice in Language Studies*, *1*(5), 534–540. https://doi.org/10.4304/tpls.1.5.534-540

Newton, J., Ferris, D., Goh, C. C., Grabe, W., Stoller, F. L., & Vandergrift, L. (2018). *Teaching English to second language learners in academic contexts: Reading, writing, listening, and speaking.* Routledge, Taylor & Francis Group.

Paulus, T. M. (1999). The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, *8*(3), 265–289. https://doi.org/10.1016/S1060-3743(99)80117-9

Pham, T. G., Lin, M., Trinh, V. Q., & Bui, L. T. P. (2020). Electronic peer feedback. EFL academic writing and reflective thinking: Evidence from a Confucian context. *SAGE Open*, *10*(1), 1–20. https://doi.org/10.1177/2158244020914554

Rollinson, P. (2005). Using peer and teacher feedback on student writing. *ELT Journal*, *59*(1), 23–30. https://doi.org./10.1093/elt/cci003

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, *18*, 119–144. https://doi.org/10.1007/BF00117714

Saito, H. (1994). Teachers' practices and students' preferences for feedback on second language writing: A case study of adult ESL learners. *TESL Canada Journal*, *11*, 46–70.

Schultz, J. (2000). Computers and collaborative writing in the foreign language curriculum. In M. Warschauer, & R. Kern (Eds.). *Network-based language teaching: Concepts and practice* (pp. 121–150). Cambridge University Press. https://doi.org/10.1017/CBO9781139524735

Suter, W. N. (2012). *Qualitative data, analysis, and design.* SAGE Publications, Inc., doi.org/10.4135/9781483384443

Tsui, A. B. M., & Ng, M. (2000). Do secondary L2 writers benefit from peer comments? *Journal of Second Language Writing*, *9*, 147–170. https://doi.org/10.1016/S1060-3743(00)00022-9

Tuzi, F. (2004). The impact of e-feedback on the revisions of L2 writers in an academic writing course. *Computers and Composition*, *21*(2), 217–235. https://doi.org/10.1016/j.compcom.2004.02.003

Villamil, O. S., & De Guerrero, M. (1998). Assessing the impact of peer revision on L2 writing. *Applied Linguistics*, *19*(4), 491–514. https://doi.org/10.1093/applin/19.4.491

Wakabayashi, R. (2013). The effects of the peer feedback process on reviewers' own writing. *English Language Teaching*, *6*(9), 177–192. http://www.ccsenet.org/journal/index.php/elt/article/viewFile/29656/17594

Ware, P.D., & Warschauer, M. (2006). Electronic feedback and second language writing. In K. Hyland & F. Hyland (Eds.), *Feedback in second language writing* (pp. 105–122). Cambridge University Press.

Yang, M., Badger, R., & Yu, Z. (2006). A comparative study of peer and teacher feedback in Chinese EFL writing class. *Journal of Second Language Writing*, *15*, 179–200.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, *21*, 697–715. https://doi.org/10.2307/3586773

Zhang, S. (1995). Reexamining the affective advantage of peer feedback in the ESL writing class. *Journal of Second Language Writing*, *4*, 209–222.

Author

Ing. Mgr. Blanka Pojslová, Ph.D., Masaryk University, Language Centre, Komenského náměstí 2, 602 00 Brno, Czech Republic, e-mail: Blanka.Pojslova@econ.muni.cz

Role a vnímání vícenásobné vrstevnické a učitelské zpětné vazby při rozvoji cizojazyčného psaní

Abtrakt: Příspěvek představuje výsledky experimentální studie, která zkoumala, jak se nahrazení učitelské zpětné vazby vrstevnickou zpětnou vazbou projeví v kvalitě cizojazyčné písemné produkce pregraduálních studentů angličtiny jako cizího jazyka. Ve studii byl použit kvazi-experimentální design s využitím pretestu a posttestu se dvěma porovnávanými skupinami, které se lišily zdrojem zpětné vazby k písemné produkci. Pro posouzení změny v kvalitě písemné produkce vlivem intervence v podobě dvou typů zpětné vazby byl využit Wilcoxonův test. Účastníci studie (N = 65) byli rozděleni do dvou porovnávaných skupin, které se lišily zdrojem vícenásobné, počítačem zprostředkované zpětné vazby, kterou obdrželi ke své písemné produkci.

První skupina (N = 33) obdržela ke třem průběžným verzím textu výhradně učitelskou zpětnou vazbu, zatímco každý student ve druhé skupině (N = 32) obdržel k první verzi textu zpětnou vazbu od tří vrstevníků a ke druhé a třetí verzi textu od učitele. Příspěvek se dále zabývá tím, jak účastníci studie vnímali zpětnou vazbu, kterou ke svým textům obdrželi. Výzkumná zjištění ukazují, že oba typy zpětné vazby významně přispěly ke zlepšení kvality písemné produkce účastníků studie, a to z hlediska všech tří zkoumaných aspektů kvality písemné produkce – její celkové kvality, žánru a registru. Výzkumná zjištění také potvrzují výrazné preference studentů pro učitelskou zpětnou vazbu, ale současně ukazují, že vrstevnická zpětná vazba napomáhá procesu učení a přispívá k rozvoji schopnosti psát v cizím jazyce.

Klíčová slova: počítačem zprostředkovaná zpětná vazba, vrstevnická zpětná vazba, učitelská zpětná vazba, kvalita psaní, vnímání zpětné vazby, angličtina jako cizí jazyk

Appendix A: Participants' profiles

		Group 1	*Group 2*
Gender	*Male*	17	11
	Female	16	21
Age	*Mean*	21.4	21.4
	Range	21 – 24	21 – 23
L1 background	*Czech*	21	15
	Slovak	11	17
	Other	1	0
English proficiency test (***CEFR based***)	*Mean Score*	59.3	61.8
	SD	11.3	14.3
Course test 1 results	*Mean Score*	54.7	56.3
	SD	6.5	7.4
Course test 2 results	*Mean Score*	52.7	53.8
	SD	6.4	6.8
Course test 3 results	*Mean Score*	44.5	46.8
	SD	5.9	5.6

B1: 42-63; B2: 64-86; C1: 87-95

Course test 1+2: Max.: 75pts. / Min. to pass: 45pts.

Course test 3: Max.: 65pts. / Min. to pass: 39pts.

Appendix B: Prompt for eliciting pre-test and post-test corpora

Write the first draft of a problem-solution essay of 350-450 words on ONE of the following topics that will include:

- *introducing the situation*
- *stating the problem and its solutions*
- *concluding by summarising and evaluating*

1. **A domestic appliance company is facing decreasing sales.**
2. **A country's economy is suffering from rising unemployment.**

Prompt for eliciting pre-test learner corpora.

Write the first draft of a problem-solution essay of 350-450 words on ONE of the following topics that will include:

- *introducing the situation*
- *stating the problem and its solutions*
- *concluding by summarising and evaluating*

1. **A small Czech brewery has recently been acquired by an American multinational.**
2. **A corporate customer has started defaulting on payments to its supplier.**

Prompt for eliciting post-test learner corpora.

Appendix C: CAE Assessment scale for overall writing quality

C1	Content	Communicative Achievement	Organisation	Language
5	All content is relevant to the task. Target reader is fully informed.	Uses the conventions of the communicative task with sufficient flexibility to communicate complex ideas in an effective way, holding the target reader's attention with ease, fulfilling all communicative purposes.	Text is a well-organised, coherent whole, using a variety of cohesive devices and organisational patterns with flexibility.	Uses a range of Vocabulary, including less common lexis, effectively and precisely. Uses a wide range of simple and complex grammatical forms with full control, flexibility, and sophistication. Errors, if present, are related to less common words and structures, or occur as slips.
4	*Performance shares features of Bands 3 and 5.*			
3	Minor irrelevances and/or omissions may be present. Target reader is on the whole informed.	Uses the conventions of the communicative task effectively to hold the target reader's attention and communicate straightforward and complex ideas, as appropriate.	Text is well organised and coherent, using a variety of cohesive devices and organisational patterns to generally good effect.	Uses a range of Vocabulary, including less common lexis, appropriately. Uses a range of simple and complex grammatical forms with control and flexibility. Occasional errors may be present but do not impede communication.
2	*Performance shares features of Bands 1 and 3.*			
1	Irrelevances and misinterpretation of task may be present. Target reader is minimally informed.	Uses the conventions of the communicative task to hold the target reader's attention and communicate straightforward ideas.	Text is generally well organised and coherent, using a variety of linking words and cohesive devices.	Uses a range of everyday Vocabulary appropriately, with occasional inappropriate use of less common lexis. Uses a range of simple and some complex grammatical forms with a good degree of control. Errors do not impede communication.
0	Content is totally irrelevant. Target reader is not informed.	*Performance below Band 1.*		

Source: https://www.cambridgeenglish.org/exams-and-tests/advanced/

Appendix D: Assessment scale for writing quality from the perspective of register

	4 *Complete*	3 *Extensively*	2 *Moderate*	1 *Limited*	0 *Zero*	Descriptor
The language of the essay is **formal**						The essay is **formal**. It demonstrates an appropriate choice of academic and specific Vocabulary; strong and one-word verbs are preferred. Also, informal phrasal verbs (e.g., *put off*, *get better*, *go on*), contractions (e.g., *don't*, *shouldn't*), run-on expressions (e.g., etc., *so on*), rhetorical questions, idiomatic and colloquial words (e.g., *stuff*, *a lot of*, *sort of*), slang, abbreviations and sub-headings, numbering, bullet-points are avoided.
The language of the essay is **objective**.						The essay is **objective** which might be achieved via using impersonal language, i.e., *there is …*, *it is …*, passive voice, specific nouns, or adverbials (e.g., *arguably*) and via avoiding personal pronouns (*I*, *we*, *you*, *they*) and words showing emotions (e.g., *luckily*, *remarkably*, *amazingly*).
The language of the essay is **explicit**.						The essay is **explicit** about the relationships in the text, i.e., the ideas are effectively connected, which is achieved via the link between the thesis statement, topic sentences and the conclusion, between paragraphs, sentences, and words. These connections are expressed via using cohesive devices appropriately and naturally.
The language of the essay is **accurate**.						The use of Vocabulary is **accurate**, e.g., *economics* is not confused with *economy*, and also grammatical structures are used accurately.

	4 *Complete*	3 *Extensively*	2 *Moderate*	1 *Limited*	0 *Zero*	Descriptor
The language of the essay is **concise** and **precise**.						The essay is **concise** and **precise**. The facts and figures are given precisely; the redundancies are avoided, e.g., repeating the same words and ideas, including irrelevant points and using empty, vague words (e.g., *people* instead of *employees*, *thing* instead of *factor*, *issue*, *topic*, *budget*) or expletive constructions (e.g., *due to the fact that instead of because*).
The language of the essay is **tentative**.						The language of the essay is **tentative**, i.e., categorical statements are avoided, and hedged sentences are used instead.
The language of the essay is **complex**.						The language of the essay is relatively **complex**, i.e., longer words are used, and is lexically denser with varied Vocabulary. It prefers noun-based phrases to verb-based phrases. The language is grammatically complex, i.e., more subordinate clauses, more *that/to* complement clauses, longer sequences of prepositional phrases, participles, more attributive adjectives, and more passives than spoken language.
The mechanics of the essay are **correct**.						Punctuation, spelling, and capitalisation is correct.
The use of the above conventions of academic writing is **consistent**.						The conventions of academic writing are followed **consistently**.

Appendix E: Assessment scale for writing quality from the perspective of genre

	4 *Complete*	3 *Extensive*	2 *Moderate*	1 *Limited*	0 *Zero*	*Descriptor*
The **conventions** of the problem-solution essay genre are observed.						The text considers the **problems** of a particular situation and gives **solutions** to these problems. The description of the situation is included in the essay prompt. The structure of the essay follows **the general essay structure** of an academic essay, consisting of an introduction, the main body, and a conclusion. The main body of the problem solution essay may have a block or a chain structure. For the block structure, all of the problems are listed first (2nd essay paragraph), and all of the solutions are listed afterwards ($3^{rd}/4^{th}$ essay paragraph). For the chain structure, each problem ($2^{nd}/4^{th}$ paragraph) is followed immediately by the solution to that problem ($3^{rd}/5^{th}$ paragraph).
The **introduction** gives background information.						The **introduction** of essay introduces its topic by giving some general background information and restating the situation.
The essay has a clear **thesis statemen**t.						The introduction has a clear **thesis statement** which states the specific topic of the essay (problem(s)) and lists controlling ideas (solutions) that will be discussed in the main body.
Each paragraph has a clear **topic sentence**.						Each essay paragraph has a clear topic sentence that identifies the main ideas of the paragraph, and that is further developed or explained in the paragraph. The supporting sentences in the paragraph develop the **topic sentence**.

	4 *Complete*	3 *Extensive*	2 *Moderate*	1 *Limited*	0 *Zero*	*Descriptor*
The essay has strong **support**.						The essay has **strong support** in the form of facts, reasons, and examples which are relevant, specific, and well-developed. The essay contains enough material to satisfy the reader on the topic that is being addressed.
The **conclusion** summarises the main points.						The **conclusion** of the essay restates the situation, the thesis statement and problem(s) and gives a final evaluation, recommendation, suggestion or comment on the subject or the solutions while adding no new information.

Appendix F: Student survey on teacher-only feedback in Group 1

1. How useful do you find your teacher's feedback on your drafts?

   a) A lot b) Mostly c) A little d) Not at all

2a. Do you feel that your teacher's comments and corrections were effective in improving your composition writing skills?

   a) A lot b) Mostly c) A little d) Not at all

2b. Please, specify

3. How much of the comments and corrections on the 1st draft involve

	A lot	Some	A little	None
Genre				
Organisation of ideas				
Grammar				
Vocabulary				
Academic style				
Mechanics (punctuation, spelling)				

4. How much of the comments and corrections on the 2nd draft involve

Genre	A lot	Some	A little	None
Organisation of ideas				
Grammar				
Vocabulary				
Academic style				
Mechanics (punctuation, spelling)				

5. How much attention do you pay to the comments and corrections on the 1st draft involving

	A lot	Some	A little	None	Not applicable
Genre					
Organisation of ideas					
Grammar					
Vocabulary					
Academic style					
Mechanics (punctuation, spelling)					

6. How much attention do you pay to the comments and corrections on the 2nd draft involving

	A lot	Some	A little	None	Not applicable
Genre					
Organisation of ideas					
Grammar					
Vocabulary					
Academic style					
Mechanics (punctuation, spelling)					

This copy of the survey includes only those parts of the survey analysed in this study.

Appendix G: Student survey on combined peer-teacher feedback in Group 2

1a. How useful do you find your peers' feedback on your 1st draft?

a) A lot b) Mostly c) A little d) Not at all

1b. How much useful do you find your teacher's feedback on your 2nd draft?

a) A lot b) Mostly c) A little d) Not at all

1c. Please, specify

2a. Do you feel that your peers' comments and corrections were effective in improving your composition writing skills?

a) A lot b) Mostly c) A little d) Not at all

2b. Do you feel that your teacher's comments and corrections were effective in improving your composition writing skills?

a) A lot b) Mostly c) A little d) Not at all

2c. Please, specify

3. How much of the comments and corrections on the 1st draft involve

	A lot	Some	A little	None
Genre				
Organisation of ideas				
Grammar				
Vocabulary				
Academic style				
Mechanics (punctuation, spelling)				

4. How much of the comments and corrections on the 2nd draft involve

Genre	A lot	Some	A little	None
Organisation of ideas				
Grammar				
Vocabulary				
Academic style				
Mechanics (punctuation, spelling)				

5. How much attention do you pay to the comments and corrections on the 1st draft involving

	A lot	Some	A little	None	Not applicable
Genre					
Organisation of ideas					
Grammar					
Vocabulary					
Academic style					
Mechanics (punctuation, spelling)					

6. How much attention do you pay to the comments and corrections on the 2nd draft involving

	A lot	Some	A little	None	Not applicable
Genre					
Organisation of ideas					
Grammar					
Vocabulary					
Academic style					
Mechanics (punctuation, spelling)					

7a. Was reading your peers' texts effective in improving your own composition writing skills?

a) A lot b) Mostly c) A little d) Not at all

7b. Please, specify

8a. Was providing peer feedback effective in improving your own composition writing skills?

a) A lot b) Mostly c) A little d) Not at all

8b. Please, specify.

This copy of the survey includes only those parts of the survey analysed in this study.

The more the merrier? Analysing self-assessment tool set use

Martina Šindelářová Skupeňová

Masaryk university, Faculty of Education, Department of English language and literature

Self-assessment is seen as a crucial component for successful language learning in autonomous settings. Since self-assessment is a metacognitive competence of a language learner, it can be practised and developed by students themselves, and, more importantly, it can also be fostered by language teachers or language advisors. This text describes self-assessment tools and practices that are suggested to students of the English Autonomously course (EA) taught at Masaryk University. The goal of the text is to investigate whether the range of self-assessment options and the tool set offered to students in the course help them to evaluate their language learning in a personalised and effective way.

The term *self-assessment* is defined with regard to this particular course and the course context is shortly described in the second chapter of this text. The third chapter explains the individual self-assessment tools and how they are presented in the course. The fourth chapter compares and analyses how individual students approach the tool set and leads into a conclusion.

1 Background

1.1 Definition of self-assessment

The term *self-assessment* has been defined in multiple ways and in various ranges. This text is based on the definition by Panadero et al. (2016) who sees *self-assessment* as a “wide variety of mechanisms and techniques through which students describe (i.e., assess) and possibly assign merit or worth to (i.e., evaluate) the qualities of their own learning processes and products” (p. 804). Therefore, this article describes steps students take to self-assess their language skills as well as to evaluate their previous language learning experience. Both the terms *self-assessment* and *self-evaluation* are going to be used to portray the various tools and their complementary functions in the English Autonomously course. Furthermore, the text follows Andrade’s

https://doi.org/10.5817/PedOr2022-4-484

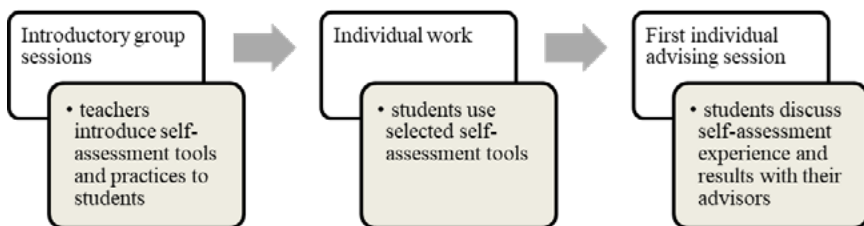
broad understanding of this concept which does "include *self-assessment* of one's abilities, processes, and products" (2019, p. 2). As a result, the term self-assessment can be applied to the complete set of tools and activities that are catered to students in the investigated course to help them to "inform adjustments to processes and products that deepen learning and enhance performance" (Andrade, 2019, p. 2).

1.2 Context: English Autonomously course (EA)

English Autonomously is an elective course at Masaryk University which is open to students of all faculties, and it aims at developing their language skills as well as their metacognition. The course gives students a chance to identify their individual language learning needs, to set their own goals, to create their study plans and to include activities and materials of their choice. To be able to make all those self-regulating steps, they are supported by a team of Language Centre teachers and the course provides them with a clear framework for learning in an autonomous way. The framework consists of two introductory group sessions, a series of individual advising sessions and a set of tools; all the framework components are provided in the target language as explained later. This article focuses on the initial stage of the course when self-assessment plays a crucial role. The scheme below shows the sequence of self-assessment activities at the beginning of the course.

Figure 1

English Autonomously course – initial stage

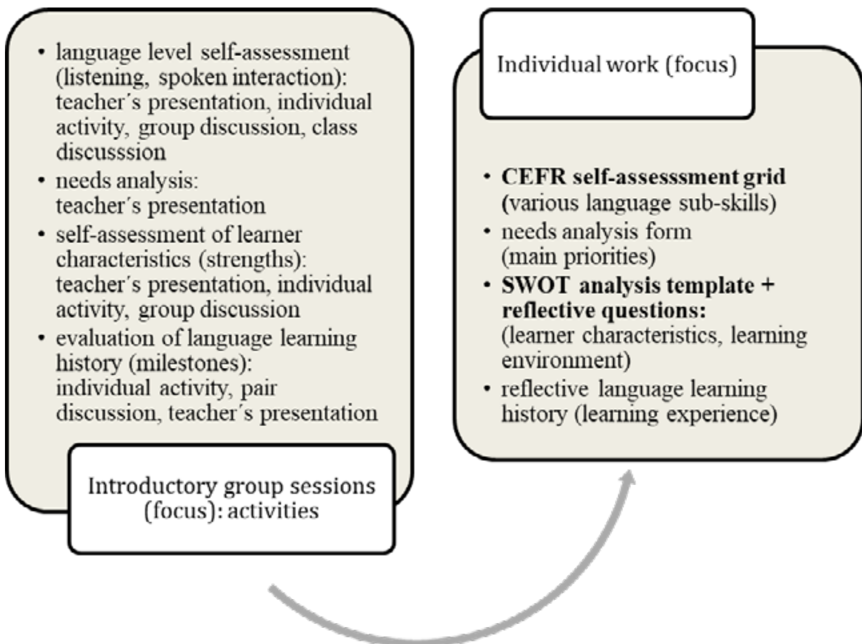


2 Self-assessment tools and practices in EA

This chapter describes the set of self-assessment tools that is recommended to all English Autonomously students, it explains their origin and purpose, and shows how they are launched through (group) activities during the introductory sessions. In chapter 4, it will be investigated how two selected students approach these (optional) tools when working with them on individual basis and whether the tool set provides them with diversified support. Figure 2 shows which self-assessment activities are included in the group sessions and which options students have to self-assess their learning while working individually. The recommended tools for individual work are highlighted.

Figure 2

Self-assessment activities and tools in EA



It needs to be pointed out that self-assessment, as well as other meta-cognitive and self-regulating activities are only conducted in English in the course. The fact that target language use in tightly interconnected with self-regulation corresponds to the double aim of the English Autonomously course. It also reflects the underlying principle of language learner autonomy development. As explained by Little (2022), when learners plan, implement, monitor, and evaluate their own learning in English, i.e., "by exercising agency in the target language they gradually develop a proficiency that is reflective as well as communicative" (p. 64). It will be shown in this text that self-assessment in the EA course entails using the target language in introspective, reflective activities and in interactive, communicative situations too.

2.1 Self-assessment of language skills

The first tool that is offered to EA students is the CEFR self-assessment grid[1] because most of them are familiar with the system of A1 – C2 language levels. At the introductory session, a paper version of the grid is handed out and the teachers explain the concept of "can do statements." The aim is to make students realize that their current language level can be described in a more precise and detailed way and that this description will typically cover more than one CEFR level. As authors of the N*ew CEFR Companion Volume* suggest, the teachers "encourage users to develop differentiated profiles" (p. 38).

During the introductory session, the students are asked to focus on descriptors for two selected subskills, usually spoken interaction and listening, and to identify those descriptors that apply to their current competences. Then, the students decide which level best describes their respective skills and they share the results of this mini self-assessment activity anonymously using an online tool e.g. Menti or Google forms. The overall results are shown to the students, and they are invited to comment on them in small groups. In group discussions, students discuss the level differences between the two skills or between their individual and group profiles. In the follow-up class discussion, the idea of differentiated profiles is emphasized again by the teacher. After the introductory session, the students are recommended to continue working with the self-assessment grid or with more specific descriptors, and to finalize their profiles. In chapter 4, examples of these CEFR based individual language level profiles will be discussed.

1 Assessment grid – English (https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=090000168045bb52)

2.2 Needs analysis

The second aspect of their language learning situation that the students are asked to self-evaluate is their needs. When designing the English Autonomously course, the team was lucky as they could build on the expertise of the colleagues from Language Centre at University of Helsinki. They shared their best practices gained from running Autonomous Learning Modules (ALMS) and the EA team was allowed to use ALMS tools too including the Needs Analysis form. This form is listing typical students' needs both in a short- and long-term perspective and was slightly adapted by the EA team (see Appendix 1).

When the tool is distributed to the students during the first intro session, it is introduced only shortly, but the idea of prioritising their needs is emphasized. Furthermore, the students are reminded that they can only address a certain number of goals in the short-term perspective. Working with this form should support self-assessment of current and future needs which according to Murray, "provides insight towards learning goals to get closer to their ideal selves" (2011). However, the example listed in chapter 4 shows that even if the tool is clearly structured and easy to use, some students need additional support in bridging their most immediate needs with their ideal selves.

2.3 Self-assessment of strengths, weaknesses, opportunities and threats

The third perspective that the students are invited to evaluate is their approach to language learning in general, they are encouraged to evaluate their internal (language) learning characteristics, to assess themselves as learners and also to consider external factors that affect their learning. An existing tool, the SWOT analysis template, was further developed by the EA team, so that it is more supportive for the students. For each of the four template sections (strengths, weaknesses, opportunities, threats), a set of reflective questions was added to lead students into thinking about their learning preferences and limitations (see Appendix 2).

During the introductory session, the SWOT tool is introduced through a short activity. The students first identify their personal strengths as learners and then exchange information on these "special powers" in small groups, presenting them as resources they can draw on both individually and as a group. This group sharing should initiate deeper reflection and introspection as well as promote the interactive aspect of self-regulated learning. As

explained by Ushioda (2006) it is important to create “a social environment that supports learners’ sense of autonomy and intrinsic motivation to pursue optimal challenges through the zone of proximal development” (p. 15). When they later meet in group activities, the students should be aware of which learning strengths they can offer so that the whole group could benefit. After the intro session, the students are expected to go further and deeper in self-evaluating their weaknesses and assessing their learning opportunities and threats. As the example in the following chapter illustrates, if they choose to use the SWOT tool, the additional questions help them to better self-asses both internal and external factors that impact their learning.

2.4 Language learning history

In contrast to the previously mentioned form-based, clearly structured self-assessment tools, evaluation of students’ learning history was first included into the course as an open writing activity. The EA team soon noticed that many students have problems approaching this open task. When asked to write their language learning histories, the students would often only produce a chronological overview of individual stages in their language learning history. However, the aim of this task is to initiate a process of self-reflection and it should result in a text which evaluates their previous learning experience.

To support the move into evaluation and reflection, a preparational activity was designed for the introductory session. The students are asked to look back at their language learning experience, to perceive it as a journey and to produce an image representing their journey. The teachers make sure that this activity is introduced in an open and safe way, the students are encouraged to be creative and to express themselves in a visual form that is most appropriate for them, it is suggested that they can produce simple pictures, use symbols or metaphors, graphs, or schemes, etc. The students then comment and reflect on their various images in small groups, some examples are discussed in an open class forum. The discussion leads into teacher’s recommendations on how they should transform their images into reflective texts. The teacher aims to point out that they need to go beyond the facts when writing their language learning histories. They are advised to comment on their images, to reflect on the depicted events, to explain the related emotions, and to interpret the meaning of factors forming their language learning experience. The following chapter compares a sketch and

a text produced by a selected student and explains how they complement each other.

3 Analysis: Self-assessment tool set use

This chapter discusses how two students (A and B) with different language learning profiles approached the above-described self-assessment tool set. Their use of the same tools is compared to find out whether the choice of tools supports them in an individualized way when self-assessing their language learning. Students A and B were selected to represent diverse types of English Autonomously population and to investigate whether the tool set provides them with diversified support. Student A is enrolled into a bachelor programme at Faculty of Social Sciences, student B is a Master student at Faculty of Natural Sciences; their entrance language levels varied and as it will be shown below, they approached the tools with various level of learner autonomy too.

When working with the CEFR grid, both students highlighted the most advanced statements that they found corresponding to their level. Student A was able to assess his skills by using a mix of B1, B1+ and B2 descriptors, he even excluded specific descriptors' details that he had not considered appropriate. As a result, he created an individualised language profile. On the other hand, student B completely relied on descriptors for B2 which correspond to the officially declared level of her secondary school leaving exam. Since she accepted them all without any adaptation, the profile very probably did not reflect neither her individualized skill set nor her language learning experience since the exam. By comparing the two profiles, it can be suggested that student B's ability to self-assess her language competences was lower than student A's. The two examples illustrate that CEFR self-assessment grid can lead students in creating personalized profiles, but those who are less advanced or less experienced in self-assessment need to be supported in going beyond the basic level scale and in using more specific can-do statements. In the EA course, such additional support can be offered in the individual advising session.

Interestingly, student B made an additional note on her self-assessment grid, she added a small arrow pointing to C1 level in spoken interaction and marked it as her "next step". Thus, she managed to relate her language level self-assessment closely to needs analysis and goal setting. When she was later

working with the needs analysis form, she incorporated this observation, and emphasized the item *holding social conversations* among the four short-term needs that she selected. This indicates that she is able to identify and prioritise her needs and the EA tools supported her in doing so. In contrast, student A experienced some difficulties in identifying his immediate needs. Using the needs analysis form, he identified all the listed skills as useful and selected twenty-two of the items as skills he needed immediately (now). Such a wide selection would not help in setting goals or designing a study plan for a semester; thus, it can be concluded that student A's ability to prioritise his needs was rather low. It was during the individual advising sessions that the student was able to consult his needs and to gain additional support in prioritising from his advisor. Eventually, the student selected four academic sub-skills which are inter-related (*listening to lectures and talks*, *taking notes*, *taking part in group discussions*, *giving mini presentations*) and highlighted them as "acute needs". His example shows that some students need more explanation and guidance to fully benefit from using the needs analysis tool.

Both students worked with the SWOT tool during the introduction session, but only student A decided to employ the template for his individual self-assessment. He filled all four sections with short statements, besides the language skills e.g., "speaking only about basic topics", his points reflect upon his personal characteristics "ambitious character", "bad time management" and his motivation for learning "need of the English language" too. In each section, he provided answers to two or three leading questions which were relevant for him, and there are logical connections to his previously conducted needs analysis as well as to the CEFR based language level self-assessment. Thus, student A demonstrated that he is able to use the tool to evaluate his learning process and to self-assess him as a learner. Student B opted not to use the SWOT template for her self-assessment, but as it will be shown below, she carefully conducted an analysis of her previous language learning experience using a different tool.

As for her language learning history (LLH), student B created a simple, yet informative image of a plant representing her growth as a learner during the introductory session, and she discussed it with other students. Student A decided not to get involved in this activity at the introductory session and he wrote his language learning history without this preparational stage. The text he submitted was quite long (378 words), but it referred not only

to English (182 words), but to German, Spanish and French languages too. Thus, his account of English language learning was rather superficial and descriptive. The following sample demonstrates that there are only minimal reflective or evaluative comments in student A's description of his secondary school experience[2].

> I continued with English language at secondary school of course, there was the same problem … unfortunately – a new teacher every year. Last teacher we had was without personal approach and we spent lot of lessons watching tv series friends or reading English texts together, top of this were long and useless homeworks during distance-learning.

The reflective moments in student A's text focus mostly on teachers and learning content, they do not evaluate his learning experience.

Student B's account of her language learning history is longer (498 words), and it only concerns the English language. It can be considered not only more detailed, but also more reflective and evaluative which is demonstrated by the following sample also commenting on a secondary school teacher:

> When I was in secondary school, I had to rely on myself more as our teacher didn't show much enthusiasm about teaching. I started to watch videos on YouTube more and I came across some books providing simplified reading. This helped me improve my understanding as well as grammar. Later I chose English as a voluntary subject. At these lessons I met other more experienced students. Once again, I felt like our teacher was passionate about his subject. He also helped me with preparation for an English Olympiad. Although I was very nervous and felt like I knew nothing, I performer quite well in the written test, which boosted my confidence. At that time, I also started to watch British shows like Doctor Who and Sherlock. While I really enjoyed following the plot and characters, I also improved my listening skills and learnt some new words.

The sample shows that Student B is able to reflect on her learning experience in more depth. When comparing her LLH text to the original image, an extension of insight can be observed. The image suggests that there were two "branches", two ways of learning English at secondary school. The text explains in more detail that while the school experience was not satisfactory, the voluntary learning was more beneficial. When contrasted to student A's text, her reflective comments also concern the student herself, her emotions are more often mentioned, and her learning outcomes are more clearly

[2] All samples of students' texts are presented in their original, unrevised form in this article.

evaluated. The comparison shows that a deeper self-reflection of learning processes is a difficult task that needs certain scaffolding. Students A and B each decided to use a different tool to help them reflect on their language learning experience, furthermore, both of them had an opportunity to attend an advising session and to focus on learning history reflection then. Their various approaches to evaluating their learning histories suggest that a choice of optional support elements in the course provides students with additional scaffolding.

4 Conclusion

The aim of the text was to describe the practices and tools that students of English Autonomously are encouraged to use for self-assessment and to analyse whether the tool set provide them with appropriate support to evaluate their learning situations in a diversified way. It was shown how the four basic tools (CEFR grid, Needs Analysis form, SWOT template and LLH) are presented during the introductory sessions of the course and that individual students choose to use them differently afterwards. The students compared in this text demonstrated different levels of various metacognitive sub-skills, e.g., the ability to prioritise needs or ability to create an individualised language profile when working with the tool set. Based on their comparison, it was observed that providing a choice of self-assessment tools and practices is meaningful as it helps to compensate for students' lack of specific self-regulating skills. It can be recommended that the choice and options available to students are well explained. Furthermore, it was confirmed that it is beneficial if self-assessment tools are employed in communicative activities in classroom or in advising sessions. This approach to promoting self-regulation is explained by Tassinari (2016): "Since learners may not be used to this reflection, it is the duty of the adviser and/or teacher to choose settings and pedagogic practices which enhance reflection, and which always take into account the needs and attitudes of the learners." (p. 130) This text was analysing whether, by combining a variety of activities and providing a choice of tools, the self-assessment stage of the English Autonomously course is efficient. Following the examples described in this text, it can be stated that "The more the merrier", because it was the range of self-assessment practices that helped the students successfully self-evaluate their language learning experience, create their language profiles, identify their needs, and become more aware about themselves as language learners.

References

Andrade, H. L. (2019). A critical review of research on student self-assessment. *Frontiers in Education*, *4*(87), 1–13.

Little, D. (2022). First person singular – language learner autonomy: Rethinking language teaching. *Language Teaching*, *55*, 64–73.

Panadero, E., Brown, G. L., & Strijbos, J.-W. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review*, *28*, 803–830.

Tassinari, M. G. (2016). Assessment for learning; assessment for autonomy. In C. Gitsaki & C. Coombes (Eds.), *Current issues in language evaluation, assessment and testing. Research and practice* (pp. 118–136). Cambridge Scholars Publishing.

Ushioda, E. (2006). Motivation, autonomy, and sociocultural theory. In P. Benson (Ed.), *Learner autonomy 8: Insider perspectives on autonomy in language teaching and learning* (pp. 5–24). Authentik.

Author

Mgr. Martina Šindelářová Skupeňová, Masaryk university, Faculty of Education,
Department of English language and literature, Poříčí 623/7, 603 00 Brno, Czech Republic,
e-mail: martina.sindelarova@mail.muni.cz

Appendix A: Needs analysis checklist

PRIORITISING YOUR NEEDS

In the table below, tick the skills you see yourself needing most at present or in the future.

LANGUAGE FOCUS	NOW	FUTURE
READING Reading academic articles or texts		
Reading literature		
Reading texts in Internet		
Reading newspapers or magazines		
Reading advertisements and public announcements		
WRITING Writing essays, report		
Writing academic articles or texts		
Creative writing		
Writing CV		
Writing a diary		
Writing formal letters		
Filling in forms		
Writing informal letters		
Writing texts on the Internet		
Writing memos and messages		
Writing newspaper articles		
Writing scholarship or grant proposals		
Writing for talks and presentations		
LISTENING Listening to lectures, talks and presentations		
Listening to conversations or discussions		
Listening for entertainment (e.g., TV, films, videos)		
Listening to interviews		

LANGUAGE FOCUS	NOW	FUTURE
Listening to news on TV or radio		
Listening to songs or music		
Listening on the Internet		
Listening on the telephone		
SPEAKING Holding social conversations		
Taking part in group discussions (e.g., tutorials)		
Having interviews		
Giving talks and presentations		
Holding telephone conversations		
Attending meetings		
Communicating when travelling (e.g., airports, hotels)		
OTHER NEEDS		

Appendix B: SWOT analysis template

strengths	*weaknesses*
opportunities	***threats***

SWOT ANALYSIS QUESTIONS

STRENGTHS

- What are you good at when you are learning a language?
- What do you like about languages (even your native one)?
- What is your greatest achievement in terms of learning languages? How did you achieve it? Did you enjoy the process?
- What personal strengths (characteristics) could you use for learning English?
- What general learning strategies you are good at could you use for English?

WEAKNESSES

- What problems do you typically encounter when learning languages?
- What do you hate when learning a language?
- What do you find boring?
- What personal weakness could stop you from speaking better English? Can you turn it into a strength?

OPPORTUNITIES

- What is your inspiration/ motivation for learning English?
- Are there any people you can "use"?
- Where can you plunge more into English?
- What are your passions that you could follow in English?
- How can you increase the percentage of speaking English in your life?

THREATS

- What are the biggest obstacles on your way to better English?
- Are you sure you cannot do anything about them?
- When you were learning a language last time, what was most difficult?
- Do you lack anything in order to be successful in learning English?
- Look deep – who says you are not good at learning languages, speaking etc.? How do they know? Can you find one example when this was not true?