

# Awareness-Raising in Training Student Teachers to Rate Written-Performances in line with the CEFR

Jana Bérešová, Trnava University, Slovakia

<https://doi.org/10.37546/JALTSIG.CEFR7-3>

This article is open access and licensed under an Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) licence.

*Language teaching in Slovakia is based on the concepts presented in the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR) and CEFR Companion Volume (CEFR/CV). Despite efforts to align the learning outcomes in primary and secondary education to CEFR proficiency levels in the national curricula and significant changes in approaches to language teaching and learning in previous years, teachers' familiarity with particular level descriptors is unbalanced. Their marking of students' written performances does not explicitly reflect the different proficiency levels. To change this situation in the country, intensive training of student teachers became necessary.*

*Assessing written performances requires systematic training based on marking criteria related to a particular reference level. To prepare future teachers of English to be able to distinguish between particular proficiency levels, several activities were designed to ensure that student teachers acquire a detailed knowledge of a targeted set of descriptors. Among other documents, a written assessment grid (Appendix 4, CEFR/CV) was introduced to sensitise student teachers to the need to familiarise themselves with the descriptors and apply them, specifying what is appropriate to expect at different levels of achievement.*

*A new academic course on assessment and testing of language competence has been introduced in conjunction with a pre-service training course. The aim is to ensure that student teachers receive the necessary training to apply marking criteria when assessing written and oral performances. This study explores the approach adopted using a written assessment grid from the CEFR Companion Volume. Particular activities and data that were collected and analysed during the course of the present study are furthermore presented.*

**Keywords:** written performances, rating, pre-service teacher training, raters' judgements, reference descriptors

## 1 The impact of the CEFR on language education in Slovakia

The first provisional version of the Common European Framework (CEF), as it was initially called, in 1996 and 1998, later published as *A Common European Framework of Reference for Languages: Learning, Teaching Assessment* (COE 2001), significantly influenced language education in Slovakia. This was a time when the educational system itself in Slovakia was in transition due to political changes, and language learning – a subject area that had been underestimated for decades – required a substantial shift in focus from one focused on learning language systems to an approach that focuses on the use of languages for real-life purposes.

The concepts that underpin the CEFR were immediately introduced in the school-leaving examination reform for foreign languages in 1997 when particular descriptors for reception skills were applied in test construction. Such impact is evident in the compilation of new standards (Štátny pedagogický ústav 1999) and new curricula (Bérešová et al. 2002) for foreign languages taught in primary and secondary education, such as English, French, German, Italian, Russian and Spanish, which are currently the

languages of the school-leaving examination. Curriculum writers were inspired by 'can do' statements, which in turn were incorporated into learning outcomes. Alongside officially claimed requirements (Bérešová et al. 2002; Štátny pedagogický ústav 2016), they are now considered the alpha and omega for item writers in the construction of tests for different reference levels.

### **1.1 The impact of the CEFR on the Slovak school-leaving examinations in languages**

The reform of the school-leaving examination was initiated by language teachers who sought to change the assessment of learners' foreign language proficiency. The objective was to establish a valid and reliable measurement. Previously, final examinations conducted in schools lacked objectivity and evidence of reliability and validity in the measurement tools. This led to the implementation of high-stakes examinations, which provide stakeholders with valid and reliable data on secondary school leavers' language competence, transparently displayed on the website of the National Institute for Education and Youth (<https://www2.nucem.sk/sk/merania/narodne-merania/maturita>).

After a three-year piloting process of English tests, the Ministry of Education officially recognised the monitoring process and accepted the necessity of introducing an external part of language school-leaving examinations administered by the ministry-governed testing institution. This process ended in 2004-2005, and since then, an external part of high-stakes language examinations has been administered every year, except for two years of the pandemic situation in the country (2020 and 2021). However, from the very beginning, English has been assessed at two levels, called lower and higher, based on the students' selection. In 2008-2009, following a large number of interventions into the system of testing foreign languages, the Ministry of Education officially recognised three levels of completing language education for secondary school leavers: B1 for students studying at secondary technical schools, B2 for school-leavers from secondary grammar schools and C1 for students studying at bilingual schools or bilingual sections of secondary schools. All the requirements based on the CEFR 'can do' descriptors and officially recognised standards are available in the Catalogues of Requirements for each level (Štátny pedagogický ústav 2016) and school-leaving examination specifications, modified regularly and adapted to specific situations if needed.

Initially, language teachers were enthusiastic about the changes and getting objective data about their students' achievements. The government was, therefore, urged to adopt more objective methods for the assessment of writing and speaking skills as well. However, the costs of hiring teachers to assess papers or oral performances, administrative costs and employing statisticians hindered progress towards consistency between external measures of listening comprehension, reading comprehension, language in use and the measurement of two productive skills.

### **1.2 The current status of the issue**

After twenty years, the situation is entirely different. The enthusiastic teachers who volunteered in the activities carried out as part of the piloting process were replaced by a new generation of teachers facing new challenges, such as earning money to survive in current economic conditions and coping with a lack of EFL teachers in the state sector. Due to the overload of classes (26-32 hours per week), teachers of English rely on published documents or coursebooks printed in the UK and labelled with proficiency levels and do not commonly consult either an English version of the CEFR or its Slovak translations (Spoločný európsky referenčný rámce pre jazyky: učenie sa, vyučovanie a hodnotenie 2006, a revised version in 2017). Teachers are familiar with the common reference levels of the CEFR. However, they may not have a comprehensive understanding of the model of language-related competences and language use. This task demands a thorough comprehension of and specific reference to descriptive examples that are pertinent to the CEFR levels that English teachers work with.

When the provisional document *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR): a Manual* was published in 2009, the

Slovak Republic started the process of aligning language tests with the CEFR, publishing several articles on the processes, such as familiarisation, standardisation, benchmarking, and standard setting. As the setting of cut scores requires the involvement of several parties in a multi-stage, judgemental process, test developers, item writers, policymakers (Berk 1986), teacher trainers and language teachers were invited to participate in several workshops to discuss a linking process. Consultants from the European Centre for Modern Languages were also invited for a workshop to train test developers and item writers to design tests and construct items in line with the CEFR.

Despite efforts to adopt approaches to assessment that are in line with the CEFR, the Ministry of Education remains reluctant to do any research. It supports the previously set cut scores at 33% of test achievement (listening, language in use and reading) and 25% in testing writing. If a test taker achieves less, they fail. If they achieve more than cut scores, test takers are seen as learners of a respective level. Despite criticism in this regard, decisions have so far remained unchanged. As a result of this ministerial approach, test developers and item writers have stopped working on the task of aligning language examinations with the CEFR.

On the other hand, one needs to consider the new generation of teachers who have entered the profession and are not familiar with the process of alignment of examinations with the CEFR initiated in previous years. Although changes in language teaching and testing are significantly influenced by the CEFR (COE 2001) and CEFR/CV (COE 2020), language teachers are the driving force in terms of the practical implementation and application of CEFR philosophies and related concepts. Language teachers need to be thoroughly informed about the aims of language education and to comprehend consistencies and discrepancies between practical findings and theoretical postulations. Only then can the required changes be operationalised.

In-service teacher training used to be organised through a well-developed system that catered for life-long education provided in eight regional centres, situated in each regional capital city. Currently, workshops for language teachers are not commonly organised since language teaching is not the focus of in-service teacher training. It can be concluded from consultancy experience that there is a gap between knowledgeable and more experienced teachers who received regular training some years ago and new teachers graduating from several universities with methodology courses having different aims that are not necessarily aligned with the CEFR perspective.

In addition, not all the methodology courses at Faculties of Arts or Faculties of Education focus specifically on the CEFR. Future language teachers are consequently not aware of concepts and approaches emphasised in the CEFR and CEFR/CV. In workshops, language teachers usually claim that they are familiar with particular reference levels; however, it soon becomes apparent that they have never gone through the process of familiarisation during which participants undergoing training grasp the notions of particular descriptors. Becoming familiar with the wording of particular descriptors is crucial as it enables both students and their teachers to see what students can do at a particular level of proficiency, matching students' performances against relevant descriptors. However, the most challenging aspect is that the matching actual performances against described competences should be justified and enough evidence needs to be provided.

## 2 Methodology

Following an analysis of problems Slovak teachers of English face (Bérešová 2019, Bérešová 2020), Trnava University introduced a new academic course related to assessing and testing learners' language competence in English. In their master-degree programme, student teachers are presented with fundamental considerations in language testing (Bachman 1990; Bachman and Palmer 1996; Council of Europe 2011), test construction (Alderson et al. 1996; Weir 1993), statistical analyses (Bachman 2006), and other test-related topics as the basis for seminars during which student teachers experience particular aspects, applying theory in practice.

During the practicum at primary and secondary schools, student teachers are exposed to much assessment-based input, which is strongly influenced by traditional testing. Their supervisors use formative tests for summative purposes and spoken and written performances are not assessed against any marking criteria in the classroom context. Student teachers thus report their experiences in the ELT seminars and expect to be provided with evidence on how testing and assessment can be qualitatively improved.

## 2.1 Study questions

The academic course on assessment and testing of communicative competence had 18 enrolled Master's degree students. To respond to their expectations for measurement of communicative competence in line with the CEFR, it was essential to plan some activities that would enable them to become aware of the processes necessary for the implementation of CEFR 'can do' statements into rating learners' written performances. The focused question about the relationship between training and being able to rate learners' written performances in line with the CEFR was the following:

RQ: Does one-semester training affect the student teachers' ability to rate learners' written performances in line with the CEFR?

## 2.2 Participants

To obtain proper data, it was essential to design intensive training for 18 student teachers (15 females, 3 males) who were required to participate in all activities related to the procedures necessary for their mutual understanding of specific reference levels and illustrative descriptors used in various scales. They were obliged to participate continuously for three months. In case of their face-to-face absence, they joined the group online.

## 2.3 Research methods

This study employed quantitative research methods, complemented by qualitative analyses of two written performances. To collect data, student teachers' ratings based on the CEFR written grid (CEFR/CV 2020) and those based on the rating scale for written performances (Appendix B) used in the country in line with the Catalogue of Requirements for Level B2 (Štátny pedagogický ústav 2016) were calculated and summarised in the tables. Quantitative methods provided quantified background data to contextualise the presented study. The data provided a basis for a detailed analysis of particular descriptors in both documents. Qualitative methods helped in data interpretation.

Then, student teachers were required to assess two performances written by B2 learners, who, as secondary grammar school leavers, were officially expected to apply for a B2 test. Qualitative analysis of both performances based on the judgement of written performances against qualitative aspects acted as a source of intuiting, which was then tested by quantitative measurement (student teacher's ratings). The ratings were compared with the official scores of the expert teachers who were in charge of assessing learners' performances according to the officially recognised marking criteria (Appendix B).

## 2.4 Materials

The materials used in both student teachers' assignments were the same: the written assessment grid (CEFR/CV 2020), the rating scale (Appendix B), and a B2 task (Appendix A). For a qualitative analysis, three online tools, such as the *English Vocabulary Profile* (<https://www.englishprofile.org/wordlists>), the *English Grammar Profile* (<https://www.englishprofile.org/english-grammar-profile>) and *Text Inspector* (<https://textinspector.com/>), were useful for obtaining detailed data about the quality of two secondary school-leavers' written performances.

The written performances that student teachers were expected to match to the reference levels, and consequently assess, were part of the school-leaving examination that is a high-stakes examination, externally run by the testing centre. In 2022, the topic of this externally-administered written task, internally assessed by designated teachers, was Transport and Travelling (Appendix A).

## 2.5 The period of awareness raising

This study examines particular procedures and results in the academic course initiated in March 2023 and concluded in a session that culminated in a debate when problematic issues were presented and discussed in May 2023. Two hours per 10 weeks enabled students to become aware of fundamentals (4 hours), become familiar with a variety of scales and illustrative descriptors (4 hours), become aware of the coverage of the high-stakes examination, the official specification of writing and the officially recognised marking criteria (2 hours), to be presented with standardisation procedures (2 hours), experiencing the process of rating productive skills against the qualitative aspects (2 hours), judging the first performance (2 hours), judging the second performance (2 hours) and comparison of data and final discussion (2 hours).

### 2.5.1 Fundamentals and introduction to awareness-raising stages

Language test development and examination is a challenge not only for test developers and item writers but also for language teachers. As mentioned above, language testing is a minor topic of methodological courses. Language teachers strive to prepare their students for being tested with limited knowledge related to language test construction and evaluation, which is a complex area and requires theoretical background as well as practical experience.

The CEFR raised many questions concerning the testing of language competence, which was later reflected in the *Manual for Language Test Development and Examining* (COE 2011). These developments have largely contributed to the resources that make up the Council of Europe's toolkit, the users of which need to familiarise themselves to be able to make effective use of the CEFR in their own contexts. The CEFR proposes a general model of language use and language learning. To operationalise this model in language testing, two aspects of authenticity (situational and interactional) must be considered while constructing test items and tasks (COE 2011). Language testing can be viewed from different perspectives, but fundamental considerations that underlie the practical development and use of language tests, proposed by Bachman (1990), significantly influenced the CEFR model of language use. Due to a growing need of the users that decided to follow the CEFR paradigm shift and tended to change the nature of language assessment by aligning their language tests and examinations to the CEFR, *Relating the language examinations and tests to the Common European Framework of Reference: Learning, teaching, assessment – Manual* (COE, 2009) was regarded as encouragement in their endeavours to situate their national language examinations with the CEFR perspective.

Five-interrelated stages described thoroughly in the *Relating the language examinations and tests to the Common European Framework of Reference: Learning, teaching, assessment – Manual* (COE, 2009) emphasised the necessity of being consistent in demonstrating the validity of the claims made about the relationship between language examinations and the levels of the CEFR. The validation of the claim requires both theoretical and empirical evidence. The linking process presupposes standard setting referring to content standards and performance standards that are both defined in the CEFR in the form of level descriptors.

To make students familiar with the interrelated stages of situating tests in relation to the CEFR, student teachers were invited to be actively engaged in the activities and all stages recommended in the *Relating the language examinations and tests to the Common European Framework of Reference: Learning, teaching, assessment – Manual* (COE 2009). Focusing on the project, only the activities concerning written production will be mentioned in this study.

### 2.5.2 The stage of familiarisation

The main aim of the familiarisation stage was to help student teachers become aware of CEFR ‘can do’ descriptors related to written production. The scales selected for writing were the three most relevant ones, namely those for *Overall written production*, *Creative writing* and *Reports and essays* (COE 2020). Each scale contained a randomly changed order of the levels without being indicated. Student teachers were expected to indicate a level, underline key words in the descriptors provided and discuss it in contrast with other levels. During this activity, it was possible to recognise the influence of their previous experience of learning English during their secondary-school studies when their teachers focused on the accuracy of grammar and vocabulary rather than text types, established conventions of the genre, the complexity of discourse at higher proficiency levels, grammar accuracy and vocabulary range. It was essential to emphasise the difference between B1 and B2 in argumentation while writing essays, as B2 writers are expected to give some reasons in support of or against a particular point of view, and B1 can write one-sidedly (COE 2001). To support student teachers’ awareness of linguistic competences, the scales for *Vocabulary range*, *Vocabulary control* and *Grammatical accuracy* (COE 2020) were integrated into training.

### 2.5.3 The stage of specification

The specification stage included student teachers’ familiarisation with the official specifications of three school-leaving examination levels and awareness of what learners are expected to perform in writing. The transparently displayed specification for testing writing at level B2 raised discussion on the number of tasks, and student teachers agreed upon a minimum of two tasks to get more evidence about language learners’ competencies, as proposed by Weigle (2002). However, a school-leaving examination contains only one task due to several reasons, such as the length of an examination, the burden placed on language teachers who are to mark their students’ written performances due to a lack of funding for external raters as well as the status of this type of testing – externally assigned and internally marked. Table 1 displays the official specification obligatory for item writers and assessors for level B2.

**Table 1. Specification for Writing at Level B2**

Aim	To measure learners’ ability to write independently, genre-based, stylistically and grammatically adequately, clearly, comprehensibly and at an appropriate level stated in the Catalogue of Requirements.
Time	60 minutes
Number of tasks	One task, thematically consistent with the topics presented in the Catalogue of Requirements.
Tested skills and subskills	Task achievement, the use of adequate linguistic structures, a range of vocabulary, composition and stylistics, paragraphing and orthography.
Task	A structured task based on the written input. The task is structured into 3-5 points.
Range/scope	200-220 words (minimal number of words: 120)
Rating criteria	Officially recognised marking criteria

Task design is viewed as the most challenging stage in testing writing. Although it is reasonable to state that being given a choice of prompts to write on may be preferred by students, writing on different topics can potentially make the results less reliable (Weigle 2002). Therefore, in Slovakia, students are

given structured tasks on a particular topic, embedded in one of the CEFR domains in which social life is organised (COE 2001) and differentiated on the basis of proficiency levels. Based on Harsch and Rupp (2011), the school-leaving examination can be viewed as a level-specific examination aimed at assessing and reporting school-leavers' proficiency with a focus on one proficiency level. In the project, the focus was only on one reference level (B2), and the intention was to discuss learners' performances in terms of the CEFR descriptors. It is implied that besides the relationship between the targeted proficiency level and the task characteristic, the relationship between the proficiency level and the rating scale level needs to be transparent (Harsch and Rupp 2011).

When being given CEFR descriptors for level B2, student teachers estimated that the task given to students to test their language competence in English enabled them to provide enough evidence on being able to write clear, detailed texts, synthesising and evaluating information and arguments. In line with theoretical approaches, the task needs to provide an opportunity for language learners to show their range (Tardieu et al. 2010; COE 2001). The number of arguments to address also makes the task more challenging and an array of processing and reasoning required to solve the task (Harsch and Rupp 2011). The latter was confirmed by student teachers while assessing the difficulty of the task (Appendix A). The descriptors related to the B2 level describing general linguistic range, vocabulary range, grammatical accuracy and vocabulary control, as well as orthographic control, were consistent with the expected complex language requirements.

Each task should be consistent with marking criteria. At this stage, student teachers were presented with the officially recognised marking criteria in the form of an analytic scale. According to Harsch and Rupp (2011), there are not many studies on the effect of holistic or analytic criteria on the variability of level-specific ratings. However, their study aimed at presenting the data on task difficulty, rating criteria difficulty, and other aspects that influence rating variability, such as raters and learners' abilities (Harsch and Rupp 2011). Conversely, Rea-Dickins and Germaine (1998) imply that analytic-marking schemes have the advantage of providing diagnostic detail of use as learner's performance is described at a range of different areas. Moreover, analytic scoring (sometimes called multi-trait) enables raters to evaluate different aspects of performance separately (Weigle 2012).

The school-leaving exam rating scale (Appendix B) covers six bands (from 0 to 5) focusing on four aspects of written production: task achievement (content), organisation (genre, coherence and cohesion, stylistic quality), grammar (syntactic variability and complex grammatical structures) and vocabulary (range, variability and appropriacy). Regarding the first aspect, student teachers discussed the CEFR descriptors related to thematic development and propositional precision, namely developing a clear argument, clearly signalling the difference between fact and opinion or passing on detailed information reliably (COE 2020) and compared them with the descriptions in the examination rating scale. The same process was followed, using scales related to coherence and cohesion mentioned in the second marking criteria. For the purposes of grammar and vocabulary areas, a variety of scales (general linguistic, vocabulary range, grammatical accuracy, vocabulary control and orthographic control) encompassed in the 'can do' statements describing linguistic competence (COE 2020) were then judged and juxtaposed with the descriptors in the rating scale. The latter were estimated to be consistent with level B2 descriptors taken from the CEFR. However, during the discussion of problematic areas in relation to a partial inconsistency in the judgements of student teachers, the following point emerged: the wording of the rating scale descriptors for each band seemed to be challenging and student teachers agreed upon a necessity of intensive training to ensure consistency of marking.

#### *2.5.4 The stage of standardisation*

Standardisation is seen as a process of consensus building concerning what learners can do at a given level and whether that corresponds to the level claimed by the resource (British Council, UKALTA, EALTA and ALTE 2022). As it is based on arriving at a common understanding of what a language learner can

do at a particular CEFR level, the calibrated performances seemed relevant for training student teachers to become aware of existing exemplary resources.

One of the valuable materials containing illustrative samples of production, already aligned to the CEFR, are the DVDs, which enable us to gain a clear understanding of the relevant CEFR level. Although the DVDs provide performances of oral production, they are a good demonstration of the performance quality at the required level. The calibration was based on the criteria grid that contains similar qualitative aspects, namely range, accuracy and coherence (COE 2001), which were considered as consistent with 'can do' statements referring to any production performance. Student teachers worked individually, matching the selected performances with CEFR descriptors to become aware of what language learners are expected to do at level B2, namely in terms of grammar and vocabulary.

### 2.5.5 *The stage of standard setting*

When describing standard setting, there are many methods for setting cut scores that should be based on a generally accepted methodology and reflect the judgement of qualified people (Zieky and Perie 2006). Standard setting can be viewed as 'a blend of judgements, psychometrics, and practicality' (Hambleton and Pitoniak 2006: 235); however, judgements are commonly considered the cornerstone on which cut scores are based. Regarding test items or examinees, the two main approaches adopted are test-centred methods and examinee-centred methods. Using judges' estimations, it is possible to recognise inconsistencies since their judgements are influenced by their experience (e.g., thinking about students they had taught), and they employ different standards when placing students into performance categories (Van Nijlen and Janssen 2008; Engelhard 2009; Béréšová 2017).

Reckase (2009) summarises the standard-setting process, which is usually required by the 'agency,' and claims that the final numerical score needs to be consistent with test design and content, elaborated description and policy definition of a standard. However, different methods or different implementations of the same method used in standard setting may not provide results that are of equal quality.

As regards testing writing, holistic judgements on work samples seem to be relevant. It is the Body of Work method (COE 2009), which allocates the student's performance to one of the predefined levels for which panellists are to set the standard. This method is commonly set on two rounds; if more are needed, a third round can be added. The scores of the students' performances are not known by the panellists, and their judgements are converted into cut scores, using logistic regression (Noijons et al. 2011).

North (2014) claims that the first method that was proposed to situate results on a test to several levels was the Carroll method, based on the use of "real data from teacher assessments and piloting it against the real test scores of the same group of learners" (North 2014: 216). This standard setting focuses on a correlation between the two sets of results for the same learner.

Standard setting is embedded in the empirical process of gathering quantitative evidence to make appropriate cut-score scales (British Council et al. 2022). This applies to any standard-setting methodology. Once employing test-centred methods, judges estimate at what reference level a test taker can be expected to respond correctly to a set of items. However, when testing writing, the concern is that examinee-centred methods sometimes referred to as empirical-judgemental methods (Berk 1986), where someone who knows test takers provides a holistic assessment of the CEFR levels are applied. However, the analytic judgement method, mentioned in *The CEFR Alignment Handbook* (British Council et al. 2022), is based on reviewing actual performances on the writing test. The performances are expected to be scored by "trained raters using the scoring scales developed for the test in question" (British Council et al. 2022: 57).

This stage of the course was more theoretical than practical. Student teachers were provided with the above-mentioned theoretical approaches to standard setting to allow them to understand the complexity and importance of the standard-setting process. It deserves a great deal of attention and a professional approach. The reason for not going through the standard-setting process was that two

main preconditions were unmet, such as experienced assessors and an insufficient number of samples. However, this stage aimed to make student teachers aware of how important this process is once test takers need proof of their language competence. Standard setting, officially documented, enables stakeholders to judge the quality of the assessment.

### 3 Results

A group of 18 student teachers experienced an intensive awareness-raising training to analyse and assess test tasks and performances in relation to the CEFR levels. As part of this study, two real-life pieces of work were the subject of two different scoring procedures. In one case, the rating scale (Appendix B) was applied first and only then, the performance was linked to the target level; in the other case, the CEFR assessment grid (COE 2020) was applied before the rating scale.

Due to a lack of time, limited by the hours of the academic course, it was possible to provide student teachers only with two written performances. Since the processes of judging those performances differed, both are described separately to clearly recognise particular steps and problems that occurred while working on the assignments.

The ratings of 18 students are presented in the tables to clearly show the student teachers' judgments. Their final ratings of the learners' written performance are compared with those of the officially appointed assessors. The data obtained from the online tools are presented in the tables to analyse inconsistencies in the assessments.

The school that enabled student teachers to assess real performances disclosed only information that could be provided with respect for confidentiality. Other learners' scores, namely those achieved in the external part of the B2 examination and the Speaking test, were added to their written performances to get a complete picture of the learners' abilities. Student teachers were not informed of these achievements or the scores the performances received from the officially appointed teachers during the process of their rating.

#### 3.1 Student teachers' ratings of the first written performance

The student teachers were given the marking criteria (Appendix B) and one school leaver's performance. Based on the marking criteria, each aspect can be given 5 points as the best performance and 0 as the lowest performance. Using the marking criteria and linking the learner's performance with the task formed the first round of judgements (Table 2), revealing that student teachers were most consistent while rating learner's ability to organise their text, meeting a majority of the characteristics of the genre, linking all the ideas mostly logically, and using appropriate connectors. Other aspects were judged in two different bands. The most significant difference emerged while rating the grammatical competence of the learner as one group of raters focused on correctness. In contrast, the rest focused on a range of grammatical patterns and the use of complex language expected at level B2.

**Table 2. Student teachers' ratings in the first round—the first learner's performance**

Points	Task achievement	Organisation	Grammar and spelling	Vocabulary
4	13	18	8	3
3	5	-	10	15

In the second round, student teachers were asked to work in smaller groups of 3 or 4. Consulting all the previously given CEFR scales, they had to present their estimations justifying their judgements. After the second round and a long discussion, student teachers agreed upon the final estimate for all four

qualitative aspects of written performance as 4+4+3+3, converted into 70% of successful performance. During their justification, it was revealed that while discussing the performance, they had not linked the rating scale performance descriptions and CEFR descriptors with the performance but had used their previous learning experience or had been influenced by their practicum. Therefore, a different approach directly linked to the CEFR descriptors was used in their assessment of learners' performances.

Student teachers were asked to use a written assessment criteria grid (COE 2020) and estimate CEFR level of performance. Their judgments were distributed amongst three reference levels, although B1+ is not officially worded in the grid (Table 3). Several recognised that the performance does not fully match any of the officially formulated performances and estimated intuitively that the level between B1 and B2 might be B1+. As the aspect of the overall performance at B2 includes descriptors, such as *can write clear, detailed official and semi-official texts on a variety of subjects* or *can make a distinction between formal and informal language*, the aspect of accuracy entails *showing a relatively high degree of grammatical control and not making errors that cause misunderstandings*, and a learner could write *an essay, which develops an argument* (COE 2020), most student teachers estimated that the learner could perform at level B2 in these three areas.

**Table 3. Student teachers' judgements based on the written assessment criteria grid—the first learner's performance**

Levels	Overall	Range	Coherence	Accuracy	Argument
B2	13	8	4	10	10
B1+	2	10	2	-	-
B1	3	-	12	8	8

Based on the yes/no judgement round, it can be concluded that the overall performance of the first student was estimated at level B2; however, while judging range and coherence, most student teachers claimed that the performance did not match B2 descriptors. Therefore, student teachers were addressed to discuss their judgements precisely and to provide supportive arguments for their choice. In their pre-service teacher training academic course, student teachers became aware of the labelled words both in the Cambridge Learner's Dictionary (McIntosh 2013), gained from the production of test-takers, while the Oxford Learner's Dictionary (<https://www.oxfordlearnersdictionaries.com/>) word labelling is based on the words used in the English coursebooks published in the Oxford University Press. During evidence claims, student teachers consulted words labelled in the English Vocabulary Profile and found that only five words used in the learner's paper are labelled B2, such as *affect, firstly, means, pollute* and *harmful*.

In this case, it was decided to use the Text Inspector system (<https://textinspector.com/>) to measure the quality of the learner's performance (Table 4). At the vocabulary level, the word list types revealed that the learner's performance expected at level B2 was represented by lower-level types of words. The total number of analysed tokens was 245, of which nine were unlisted. Analysing the number of words labelled by CEFR levels, a large number of used words was more relevant for A levels users as B levels were represented only by 12% out of all the words used in the paper. This supported the student teachers' uncertainty when they were asked to judge the criterion concerning the range of language used to express opinions. As far as grammar is concerned, the sentences were in present and past tenses, once the learner used *to be going to* and *will*. There were mistakes when he/she wanted to use more advanced patterns.

**Table 4. Words and their labels taken from the Text Inspector system**

CEFR levels	A1	A2	B1	B2
Words/%	90 (64.75%)	22 (15.83%)	13 (9.35%)	5 (3.60%)
Tokens/%	185 (75.51%)	26 (10.61%)	19 (7.76%)	5 (2.04%)

The second round was less variable as student teachers focused on other descriptors related to general linguistic range and could not match the first student's performance with B2 level descriptors, such as *developing arguments without much conspicuous searching for words, using some complex sentence forms to do so or can vary formulation to avoid frequent repetition, shows a relatively high degree of grammatical control, can use a variety of linking words efficiently to mark clearly the relationships between ideas* (Tardieu et al. 2010). The most useful information appeared on the right side of the scale related to pragmatic competence, in which B2 descriptors give a clear approach to what a B2 user is expected to do in the language, such as *can highlight the most important aspects of a topic, can employ the rules that concern going from the general to details, can deliver all of the contents and components that are expected for the text concerned* (COE 2020). The findings showed that 44% of raters estimated the performance to be B1, reasoning that the use of language is lower than what learners can do at level B2. However, 66% were consistent in estimating the performance as B2, providing a lot of evidence, matching the first learner's performance with exact descriptors of the B2 reference level.

To conclude the rating of the first learner's written performance, it is necessary to disclose the official rating of the paper. As mentioned above, every paper was assessed by two assessors. After synchronising both judgements, the test taker was given 3+3+2+2 = 10 points (50%). The assessors' notes in the paper showed that his/her initial score (9 points = 45%) after the first round of assessment was crossed out and replaced by a new score after the second round. It can be inferred that it may arise from different factors, for example, due to his/her better achievements in an external part of the school-leaving examination, labelled as B2. The officially administered measurements of his/her listening, language in use and reading reached 78.3% (72 percentile), and his/her spoken performance was marked as 1, the best mark in the marking system. However, once CEFR reference levels are implied, his/her written performance slightly contradicts his/her ability to perform receptively. It is significantly in contrast with the achievements in another productive skill (speaking). Since there is no evidence of the learner's spoken performance during the discussion with student teachers, it seemed to be reasonable to disregard the mark for this performance. The achievements in the external part of the school-leaving examination in B2 English and the written performance proved that learner's language competence is at B2.

### **3.2 Student teachers' ratings of the second performance**

The second paper was judged differently. The first process of estimation was based on the use of the written assessment criteria grid (CoE 2020). Having experienced the first paper estimation, student teachers started to read a 269-token long text without focusing on the task, though matching the performance against the criteria described in the grid. In the first round of their initial judgement, their estimation arrived at two levels – C1 and B2. Once the decision is C1, it can be concluded that the learner can achieve B2 (Table 5).

**Table 5. Student teachers' judgements based on the written assessment criteria grid – the second learner's performance**

	Overall	Range	Coherence	Accuracy	Argument
C1	-	5	5	-	-
B2	18	13	13	18	18

Student teachers were positively impressed by a lengthy text containing much information produced by the second learner. The language constructions, such as -ing and -ed participles, the use of different tenses and the use of advanced cohesive devices influenced their estimation. However, when they were given the marking criteria (Appendix B) and the task (Appendix A), their first estimation resulted in the rating presented in Table 6. Using the marking criteria, the performance appeared to be weaker than that of the first student. The weakest aspect was task achievement, as student teachers could recognise memorised parts of the text that did not match the task.

**Table 6. Student teachers' ratings in the first round – the second learner's performance**

Points	Task achievement	Organisation	Grammar and spelling	Vocabulary
4	2	-	-	-
3	6	18	18	8
2	10	-	-	10

During the discussion, student teachers admitted that after the first reading without focusing on the task, the second performance seemed to be written by a good user of English, providing a lot of information in a more advanced language. However, it did not match the bullet points that were clearly stated in the task, enabling markers to be more objective, not being biased by learner's ability to produce a lot of language related to the topic, but not matching the task.

The Text Inspector data proved that the learner used two C1 words; however, one of them (commuting) was used in the rubrics, and the second (sector) has its Slovak form with a letter k. In addition, it can be inferred that the learner used four B2 words, such as *causing*, *secondly*, *traffic jam* and *decade*. In contrast to the previous learner's performance, the number of unlisted words was 17 types representing 19 tokens and, due to their misspelling, such as 'almost,' 'busses,' and 'enourmous,' they were not included in the labelling process. Comparing both text inspector data, it can be concluded that the second performance contained more A2 and B1 words, but the text produced seems to be based on the text from the coursebook the learner memorised while preparing for an oral examination to achieve a good mark in speaking.

Based on the English Grammar Profile, the structures used in the second paper were labelled B1. The same conclusion can be made, using the Slovak Catalogue of Requirements for B2, in which exact structures are mentioned and categorised. When student teachers participated in the second round, their estimations became more consistent, and their assessments were unequivocally 2+3+3+2, which finally meant 10 points. It can be concluded that the performance was given a 50-percent success rate following the transfer of points into percentages.

As mentioned above, a pragmatic competence scale includes descriptors related to the scales of coherence and thematic development. While the second learner could structure the text logically, maintaining a clear development, the text he/she produced was not based on the bullet points of the task, but on the topic, in essence. However, the learner could make links between different parts of the text and construct the text by applying rules that involve moving from the general to the detailed.

The official raters seem to have been biased as well. While in the first round of the judgement, their decision was  $2+2+2+2 = 8$  points (40%), after the second round, the learner achieved  $3+3+3+3 = 12$  (60%). The achievements in the external school-leaving examination were 78.3% (72 percentile), and speaking was marked with the highest mark (1). Based on the use of the CEFR, it can be concluded that the performance of the second learner can be labelled B2, although some doubts arose during the rating process.

### 3.3 Conclusive remarks on the rating process

In training student teachers to rate B2 written performances in English, the analytic judgement method procedures were followed. Although two different approaches were used: the rating scale (Appendix B) as the first and the written assessment criteria grid (CEFR CV 2020) as the second in the first learner's rating and vice versa in the second learner's rating did not significantly influence student teachers' judgements. The failure to adhere to the structured task resulted in student teachers being initially impressed by the second learner's performance. However, they later realised that the learner had not produced appropriate content as it was related to the topic rather than the task. The officially appointed assessors gave the second learner lower scores as the main criterion in the rating scale (Appendix B) is task achievement. The decisions made by designated assessors did not have an impact on student teachers, as they were not informed about the availability of official scores. The precise information was obtained afterwards.

According to the official regulations in the country, language learners can achieve only one score point higher in other criteria than in the first criterion, ensuring that memorised text used inappropriately to complete the task cannot enable learners to pass. If task achievement is scored as 0, all other aspects of the assessment are to be marked as 0.

The descriptions of the summative profile were assessed through binary judgements as to whether the learner's performance demonstrates the required characteristics or not, as suggested by Brindley (2001). Due to qualitative analysis and careful reading of the CEFR, its descriptors and CEFR-related documents, both performances judged holistically matched reference level B2.

## 4 Discussion and conclusions

Each government has its language policy, considering the educational background and history of language education in a respective country. As the CEFR is descriptive, it enables policymakers, curriculum and test developers, teachers, and language users to use it so that each country or group of people affected can benefit from it.

This study aims to highlight the importance of awareness-raising training and the factors that can influence marking and raters' decisions. These factors include the task, marking scale, raters, and their training. The CEFR provides scales with descriptors that enable raters to match learners' or test takers' performances against specific reference levels. A shared understanding of specific levels can reduce inconsistencies between raters. Encouraging raters to justify their ratings and exposing them to other raters' opinions during discussions after each round can also help. It is essential to provide intensive training to ensure a common understanding of the reference scales, their level descriptors, and illustrative samples.

This study presents an overview of the stages required for student teachers to feel confident in demonstrating that their assessments are in line with the respective reference level. The most challenging aspect was assessing learners' written performance and justifying that the scores given were within the reference level. Once the ratings were agreed upon, student teachers could compare them with the scores of officially appointed teachers. This confirmed to the student teachers that the areas they had identified as problematic were similar to those identified by other evaluators. However, the student

teachers' approach to addressing the issues was consistent with CEFR-based materials. Although the number of student teachers included in this study was limited, the findings support further research in both pre-service and in-service teacher training to obtain more data on assessors' ability to combine particular CEFR level descriptors with the marking criteria to achieve consistency in rating language learners' performances at a particular level. Consequently, it is also essential to incorporate a more significant number of samples of language learners' written performances. This would enable gathering sufficient evidence to validate the claim about the relationship between learners' ability to use English in written production and a specific CEFR level.

The idea behind this study was to emphasise the significance of training the users of the CEFR. Achieving a common understanding of reference levels and descriptors is a rigorous task. The training materials used were all aligned with the CEFR, including a task, marking criteria, the written assessment grid, the English Vocabulary Profile, the English Grammar Profile, Text Inspector, and the wording used in discussions. The keywords used throughout were taken from the CEFR descriptors. The student teachers were actively involved in the process as they recognised the opportunity to experience detailed reading of the CEFR descriptors presented in different scales. They were able to discover the many possibilities of applying the CEFR in various contexts. This knowledge should be reflected in the development of classroom materials.

Based on the study results, it is recommended that a re-evaluation of teaching practices should be undertaken. Involving language teachers in the implementation of the CEFR-based marking criteria and in the construction of level-descriptor matching tasks can aid learners in performing better by familiarising them with the requirements for a particular proficiency level. Improving one's understanding of CEFR levels and their descriptors can positively impact the development of classroom materials and, subsequently, enhance the learning process.

## 5. References

- Alderson, J. Charles, Caroline Clapham & Dianne Wall. 1996. *Language test construction and evaluation*. Oxford: OUP.
- Bachman, Lyle F. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, Lyle F. 2005. *Statistical analyses for language assessment*. Cambridge: Cambridge University Press.
- Bachman, Lyle F. & Adrian S. Palmer. 1996. *Language testing in practice*. Oxford: Oxford University Press.
- Bax, Stephen. 2012. <https://textinspector.com/> (accessed May 2023).
- Bérešová, Jana. 2017. *The impact of the CEFR on language examinations in local contexts*. Frankfurt am Main: Peter Lang.
- Bérešová, Jana. 2019. The importance of objectivity in assessing writing skills. *INTED Proceedings*, Valencia: IATED, 4376-4380.
- Bérešová, Jana. 2020. An analysis of B2 tests in English in the Slovak educational context. *EDULEARN Proceedings, 12th International Conference on Education and New Learning Technologies*. Valencia: IATED, 4833-4838.
- Bérešová, Jana, Marta Hosszúová & Marta Macková. 2002. *Učebné osnovy pre štvorročné gymnáziá*. [The national curriculum for four-year secondary grammar schools]. Bratislava: Štátny pedagogický ústav.
- Bérešová, Jana, Marta Hosszúová, Marta Macková & Edita Mareková. 2002. Katalóg cieľových požiadaviek na vedomosti a zručnosti maturantov z anglického jazyka – úroveň B1 a B2. [The catalogue of English language requirements – levels B1 and B2]. Bratislava: Štátny pedagogický ústav.

- Berk, Ronald A. 1986. A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56(1), 137-172. <https://doi.org/10.2307/1170289>.
- Brindley, Geoff. 2001. Outcomes-based assessment in practice: Some examples and emerging insights. *Language Testing* 18(4): 393-407. <https://doi.org/10.1177/026553220101800405>.
- British Council, UKALTA, EALTA & ALTE 2022. *Aligning language education with the CEFR: Handbook*. <https://www.ealta.eu.org/documents/resources/CEFR%20alignment%20handbook.pdf> (accessed 25 April 2023).
- Council of Europe. 2001. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Council of Europe. 2009. *Relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (CEFR): A manual*. Strasbourg: Council of Europe.
- Council of Europe. 2011. *Manual for language test development and examining*. Strasbourg: Council of Europe.
- Council of Europe. 2020. *Common European framework of reference for Languages: Learning, Teaching, Assessment. Companion Volume*. Strasbourg: Council of Europe.
- Engelhard, George, Jr. 2009. Monitoring raters in performance assessments. In Gerald Tindal & Thomas M. Haladyna (eds.), *Large-scale assessment programs for all students: Validity, technical adequacy, and implementation*. 261-287. Mahwah, NJ: Erlbaum. <https://www.taylorfrancis.com/chapters/mono/10.4324/9781410605115-17/monitoring-raters-performance-assessments-george-engelhard-jr-gerald-tindal-thomas-haladyna> (accessed 8 May 2023).
- English Grammar Profile. <https://www.englishprofile.org/wordlists> (accessed May 2023).
- English Vocabulary Profile. <https://www.englishprofile.org/wordlists> (accessed May 2023).
- Hambleton, Ronald K. & Mary J. Pitoniak. 2006. Setting performance standards. In Robert L. Brennan (ed.), *Educational Measurement* (4<sup>th</sup> edn.) 433-470. Westport, CT: American Council on Education/Praeger.
- Harsch, Claudia & Alexander André Rupp. 2011. Designing and scaling level-specific writing tasks in alignment with the CEFR: A test-centered approach. *Language Assessment Quarterly* 8(1). 1-33. <https://doi.org/10.1080/15434303.2010.535575>
- McIntosh, Colin. 2013. *Cambridge Advanced Learner's Dictionary*. Cambridge: Cambridge University Press.
- Národný inštitút vzdelávania a mládeže [The National Institute for Education and Youth]. 2023. Merania. Maturita. [Measurements. The school-leaving examination.] <https://www2.nucem.sk/sk/merania/narodne-merania/maturita> (accessed April-May 2023).
- Noijons, José, Jana Bérešová, Gill Breton & Gábor Szabó (eds.). 2011. *Relating language examinations to the Common European framework of reference for languages: Learning, teaching, assessment (CEFR). Highlights from the Manual*. Graz: European Centre for Modern Languages / Strasbourg: Council of Europe.
- North, Brian. 2014. *The CEFR in Practice*. Cambridge: Cambridge University Press.
- Oxford Learner's Dictionaries. <https://www.oxfordlearnersdictionaries.com/> (accessed April-May 2023).
- Rea-Dickins, Pauline & Kevin Germaine. 1998. *Evaluation*. Oxford: Oxford University Press.
- Reckase, Mark D. 2009. Standard setting theory and practice: Issues and difficulties. In Neus Figueras & José Noijons (eds.). *Linking to the CEFR levels: Research perspectives*. Arnhem: Cito, EALTA. [https://www.ealta.eu.org/documents/resources/Research\\_Colloquium\\_report.pdf](https://www.ealta.eu.org/documents/resources/Research_Colloquium_report.pdf) (accessed 28 April 2023).
- Štátny pedagogický ústav [The National Institute for Pedagogy]. 1999. Vzdelávací štandard pre gymnázium – začiatčníci [Educational standards for secondary grammar schools – beginners]. Bratislava: Štátny pedagogický ústav.

- Štátny pedagogický ústav [The National Institute for Pedagogy]. 2006. Spoločný európsky referenčný rámec pre jazyky: učenie sa, vyučovanie, hodnotenie. [The Common European Framework of Reference for Languages: Learning, teaching, assessment]. Bratislava: Štátny pedagogický ústav.
- Štátny pedagogický ústav [The National Institute for Pedagogy]. 2016. Cielové požiadavky na vedomosti a zručnosti maturantov z anglického jazyka – úrovne B1 a B2. [The catalogue of English language requirements – levels B1 and B2]. [https://www.statpedu.sk/files/articles/nove\\_dokumenty/cielove-poziadavky-pre-mat-skusky/anglicky-jazyk\\_b1b2.pdf](https://www.statpedu.sk/files/articles/nove_dokumenty/cielove-poziadavky-pre-mat-skusky/anglicky-jazyk_b1b2.pdf) (accessed 10 May 2023).
- Štátny pedagogický ústav [The National Institute for Pedagogy]. 2017. Spoločný európsky referenčný rámec pre jazyky: učenie sa, vyučovanie, hodnotenie. [The Common European Framework of Reference for Languages: Learning, teaching, assessment]. Bratislava: Štátny pedagogický ústav.
- Tardieu, Claire, Raili Hildén, Magda Lehmann, & Monique Reichert. 2010. *The CEF-estim grid*. Graz: European Centre for Modern Languages / Strasbourg: Council of Europe.
- Van Nijlen, Daniel & Rianne Janssen. 2008. Modeling judgments in the Angoff and contrasting-groups method of standard setting. *Journal of Educational Measurement* 45(1). 45-63.
- Weigle, Sara Cushing. 2002. *Assessing Writing*. Cambridge: Cambridge University Press.
- Weigle, Sara Cushing. 2012. *Assessing Writing*. In Christine Coombe, Peter Davidson, Barry O'Sullivan & Stephen Stoyhoff (eds.), *The Cambridge Guide to Second Language Assessment*. Cambridge: Cambridge University Press.
- Zieky, Michael, & Marianne Perie. 2006. *A primer on setting cut scores on tests of educational achievement*. Princeton, NJ: Educational Testing Service. [http://www.ets.org/Media/Research/pdf/Cut\\_Scores\\_Primer.pdf](http://www.ets.org/Media/Research/pdf/Cut_Scores_Primer.pdf) (accessed 12 May 2023).

## 6 Biography

Jana Bérešová is a professor at the Faculty of Education of Trnava University, Slovakia, where she is involved in pre-service teacher training. In her academic courses on methodology, she introduces the CEFR and its concepts, which significantly influence language teaching, and shares her expertise with student teachers. She introduced a new academic course on testing communicative competence, which helps prepare student teachers to become aware of basic considerations in language testing and to better understand CEFR levels and descriptive 'can do' statements and their role in assessment procedures. Twenty years ago, she initiated a reform of the country's school-leaving exams, focusing on testing language use in line with the CEFR. Still an item writer, she contributes to improving the effectiveness of the examinations developed and administered in the country. She has written several monographs on language teaching and testing, articles on various aspects of language education and has presented her work at international conferences in Europe and beyond. As a member of the RELANG team initiative offered by the ECML in Graz, Austria, she contributes to changes in language education in several European countries, focusing mainly on aligning curricula, language tests and examinations with the CEFR, and encouraging local examination providers to ensure quality in high-stakes testing.

## Appendix A

In your English class, you have to write an opinion essay entitled 'Transport and Travelling in My Life' (200-220 words). Follow these points:

- the influence of the transport and travelling on the quality of your everyday life,
- your positive contribution to the environment – your choice of travelling/commuting,
- an unforgettable experience from travelling by any means of transport.

## Appendix B

Table 7. *An analytic rating scale to assess writing*

	Relevance and adequacy of the content	Discourse (genre, organisation)	Grammar	Vocabulary
5	The content is totally relevant to the task. All the points of the task are thoroughly and evenly elaborated. Main ideas are consistently developed.	The text meets all the characteristics of the genre. All the ideas are clearly and logically linked. Wording and cohesive devices are effectively used.	Grammar structures are used accurately and appropriately for the task to a large extent. The text is characterised by adequate syntactical variability and complex grammatical structures. Grammatical and syntactical errors occur sporadically.	Vocabulary is rich and relevant to the topic. The text is characterised by rich lexical variability, appropriate collocations and idioms.
4	The content is relevant to the task. All the points are adequately but not evenly elaborated. In general, the main ideas are developed.	The text has a majority of characteristics of the genre. Logical linking of ideas prevails. Wording and cohesive devices are appropriately used.	Language structures are used accurately and appropriately for the task to a considerable extent. The text is characterised by certain syntactic variability and complex grammatical structures. Grammatical and syntactical errors occur to a limited extent.	Vocabulary is rich and prevalently relevant to the topic. The text is characterised by proper lexical variability and correctly used collocations and idioms.
3	The content is almost relevant to the task. One point is not adequately elaborated. The main ideas are sufficiently developed, but not all are relevant.	The text does not have a majority of characteristics of the genre. Ideas are not always sufficiently linked. Wording and cohesive devices are sufficiently used.	To a certain extent, language structures are used accurately and appropriately for the task. The text is characterised by limited syntactical variability and complex grammatical structures to a small extent. Grammatical and syntactical errors occur more frequently.	Vocabulary is appropriately rich and relevant to the topic. The text is characterised by minor flaws in using collocations and idioms.

	<b>Relevance and adequacy of the content</b>	<b>Discourse (genre, organisation)</b>	<b>Grammar</b>	<b>Vocabulary</b>
2	The content is relevant to the task to a limited extent. Two points are not adequately elaborated. The main ideas are not sufficiently comprehensible.	The text has the characteristics of the genre to a limited extent. The flow of the ideas is, for the most part, not linked logically. Wording and cohesive devices are used in a limited way.	Language structures are used accurately and appropriately to the task to a lesser extent. The text is characterised by sporadic syntactical variability, and simple structures prevail. Some grammatical and syntactical errors interfere with comprehension of the text.	Vocabulary is simple and not always relevant to the topic. The repetition of the same words characterises the text. The incorrect use of some words interferes with the comprehension of the text to a certain extent.
1	The content is minimally relevant to the task. The points are not sufficiently elaborated. The main ideas are not comprehensible to a large extent.	The text has the characteristics of the genre to a minimal extent. The flow of the ideas is not logically linked, which causes incomprehension. Wording and cohesive devices are inadequately used.	Language structures are often used inadequately and inappropriately for the text. The text is characterised by minimal syntactic variability and basic grammatical structures. Grammatical and syntactic errors interfere with comprehension of a certain part of the text.	Vocabulary is simple and relevant to the topic to a limited extent. The text is characterised by frequent repetition of the same vocabulary. The incorrect use of the words often causes misunderstanding.
0	The content is not relevant to the task. Points are elaborated by irrelevant ideas. The main ideas are not comprehensible.	The text does not have the characteristics of the genre. The flow of ideas is chaotic and illogical. Wording and cohesive devices are not used, which causes incomprehension.	Language structures are used prevalently inappropriately and inaccurately to the task. The text is not characterised by syntactic variability and contains basic grammatical constructions. Grammatical and syntactical errors prevent understanding of the major part of the text.	Vocabulary is very simple, prevalently irrelevant to the topic. The text is characterised by limited vocabulary. The incorrect use of vocabulary prevents understanding to a large extent.