

# Testing speaking in English for medical purposes

## Testování mluvení v lékařské angličtině

Petra Zrníková

**Abstract:** The paper deals with testing speaking in OET, sTANDEM, IELTS, and UNICert III. The aim is to describe similarities and differences in examination, connection to CEFR. Theoretical background on task-based testing is developed briefly. We share our experience with UNICert III – “English for medical sciences” course and developing speaking examination.

**Key words:** testing speaking, English for medical purposes, CEFR, certification systems, task-based approach

### Introduction

Medical environment is a very specific in terms of speaking skills. When designing speaking test, tasks should cover the most important communication situations typical for the target group. Test-developers should bear in their mind that candidates take the test because they need it, and the needs depend on employers' requirements. Therefore, an outstanding question lies in the ratio of speaking functions, such as talk as interaction, transaction, and performance. In other words, whether they should be weighted equally or which one is dominant for some reasons. In relation to that, the question of number of tasks testing medical knowledge in comparison with tasks testing use of general vocabulary needs to be answered.

The first section of the paper deals with origin of Common European Framework for languages, functions of speaking, and guidelines, which can be followed when testing speaking. The second part of the paper deals with speaking functions, such as talk for interaction, transaction, and performance. Our primary concern is to the talk as performance, and especially, giving a presentation. It is considered a typical communication situation for doctors, both in academic and professional setting.

Nowadays, the task-based approach in test design is widely discussed. As it is a very complex topic, only a few key terms are outlined, for example, advantages of task-based testing, assessment approach (holistic vs. analytic), authenticity of tasks, kinds of tasks (guided vs. controlled), or kinds of validity.

These aspects of speaking tests are studied in the following certification systems: a certification system focused on General English (International English Language Testing System – IELTS), two medical-oriented systems (Occupational English Test – OET, Standardized Language Certificate for Medical Purposes – sTANDEM), and

one academic-oriented system (University Certification System for Languages – UNICert).

We conducted a survey of tasks in speaking examination, criteria for fluency, accuracy, range of vocabulary, and grammar. Rating and scoring procedures are being presented, too. Moreover, the connection of systems to Common European Framework of Reference for languages (CEFR), especially, correspondence between CEFR descriptors for proficiency level C1 and descriptors of the systems is shown.

The level C1 was chosen because it is the highest level of proficiency in English, which can be achieved in Slovakia. The other reason is our UNICert course and certification examination. Since there have been several cases when overseas doctors misunderstood the patient resulting in misdiagnosis or even patient's death, demand for proving language proficiency has become more important than ever before. To avoid similar affairs we want to provide our medical students and junior doctors with language education of the highest quality. The target group is undergraduate students and junior doctors who have already mastered general English and medical-oriented English course, both at the level B2.

In the last section, the results of speaking examination in UNICert III (C1) provided by Jessenius Faculty of Medicine in Martin are being presented. Giving a presentation for academic purposes on the topic of candidate's choice comprises 50% of their final grade in speaking examination. Presentation skills of undergraduate students and postgraduates are being compared.

## **1 Common European Framework of Reference for Languages**

The attempt to create CEFR for languages dates back to 70's and 80's of the 20th century. CEFR for languages is a basic document for other standards in language teaching and testing in Europe. CEFR descriptors can be used as guidelines transferable to any language learning and testing. They can also serve as a starting point for interpreting and comparing various certification systems, their standards and criteria.

Two subskills of speaking skills are distinguished in CEFR. "Spoken interaction: I can express myself fluently and spontaneously without much obvious searching for expressions. I can use language flexibly and effectively for social and professional purposes. I can formulate ideas and opinions with precision and relate my contribution skilfully to those of other speakers. Spoken production: I can present clear, detailed descriptions of complex subjects integrating subthemes, developing particular points and rounding off with an appropriate conclusion." (Self-assessment grid for speaking). In more detail, descriptors of speaking skills of English user who has achieved the level C1 are defined below (CEFR, pp. 74 to 82).

Overall spoken interaction: Can express him/herself fluently and spontaneously, almost effortlessly. Has a good command of a broad lexical repertoire allowing gaps to be readily overcome with circumlocutions. There is little obvious searching for expressions or avoidance strategies; only a conceptually difficult subject can hinder a natural, smooth flow of language.

Communication with a native-speaker he/she: Can understand in detail speech on abstract and complex topics of a specialist nature beyond his/her own field, though he/she may need to confirm occasional details, especially if the accent is unfamiliar.

Conversation: Can use language flexibly and effectively for social purposes, including emotional, allusive and joking usage.

Informal discussion with friends: Can easily follow and contribute to complex interactions between third parties in group discussion even on abstract, complex unfamiliar topics.

Formal discussion: Can easily keep up with the debate, even on abstract, complex unfamiliar topics. Can argue a formal position convincingly, responding to questions and comments and answering complex lines of counter argument fluently, spontaneously and appropriately.

Interview: Can participate fully in an interview, as either interviewer or interviewee, expanding and developing the point being discussed fluently without any support, and handling interjections well.

For example, spoken interaction can be tested on the basis of holding a conversation between the assessor and the candidate, or between the two candidates. The goal is to assess candidates' ability to express their opinions and ideas fluently and spontaneously. In case of spoken production, giving a lecture on a complex topic might be required in order to see candidates' ability to convey facts, explain theories, describe graphs, etc., and to respond to topic-related issues.

When developing a speaking test, the CEFR Grid for Speaking (ALTE CEFR SIG, 2014) can be used. It is intended to stimulate critical reflection amongst those involved in the preparation and evaluation of speaking tests, and to facilitate precise reporting by testing bodies participating in audits carried out by organizations such as Association of Language Testers in Europe (ALTE). The document is divided into two main parts. The first part is composed of 34 questions about different aspects of the speaking test as a whole and its individual speaking tasks. In the second part, the explanatory notes are provided. Next, notes explaining some items in the document are mentioned below (The CEFR Grid for Speaking, 2014, pp. 8–11).

The choice of channel (means) used in the speaking test is determined by the construct, context and purpose of a test. The channel influences how the test is administered and marked. It may also influence test-taker's performance and motivation. The guide also says that each test is different, so it is impossible to offer one-size-fits-all solution when it comes to choosing a testing channel.

When it comes to the term construct, it refers to the theory underlying the design of a test – that is, the way the test developers to be explicit about their test construct, as the choice of construct will affect the decisions they make about the content of the test, the marking criteria, and the boundaries between different levels of ability. There are different ways of defining language ability. Some experts see language in abstract terms, describing the competences that test-takers need in order to produce the right kind of language: linguistic competence, sociolinguistic competence, pragmatic competence etc. Others see language in terms of the skills that test-takers need to display. These experts might look at the skill of speaking and break it down into different subskills. Another way of viewing language is in terms of “can do” statements – e.g. the test-takers can express simple opinions or requirements in a familiar context.

There are several rating methods which need to be taken into consideration when preparing a speaking test. In a holistic approach, the test-taker's performance is judged as a whole. The rater does not give separate scores for different features of the performance (grammar, vocabulary, etc.). In an analytic approach the rater gives separate scores for several different language features. This approach recognises that a test-taker's grammar may be very good, but his/her vocabulary may be weaker. It has been claimed that the holistic approach more closely resembles how language production is judged in real life, and can be quicker than using an analytic approach. However, analytic marking can offer rich diagnostic information for L2 learners. There are mixed results from research into the reliability of using the two approaches.

Another method of rating is using the task rubrics. In rigidly controlled tasks the task determines the structure of the test-taker performance, leaving no room for spontaneous interaction. Partially controlled tasks may present a scenario in which the main conversational path is outlined, leaving some room for spontaneous interaction. Tasks with an open format may depend entirely on the interaction between the examiner and the test-taker or may require the test-taker to produce a monologue. Rigidly controlled tasks may seem inauthentic at times, but they make it easier to compare test-taker performances. Open tasks may seem more authentic, but it can be more difficult to assess the resulting interaction.

Candidate's speaking skills assessment can be based on integration of skills. Test-takers' speaking scores may depend not only on their speaking skills but also on their other skills, e.g. skimming a text to comment on it (reading), taking notes

while conducting a telephone call (writing), or understanding an audio prompt (listening). Test developers may consciously choose to integrate other skills with speaking or they may choose to assess speaking alone. The choice depends on the construct underlying the test. If the speaking required in the target language use situation involves other skills, then it may make sense to design test tasks that involve these skills. The test developer should be aware of the problems of “construct-irrelevant variance”, however, where the test-taker’s ability in the other skills may affect their speaking performance unintentionally.

Specifying the communicative purpose of a task is important, both for the test developer and for the test-taker. It helps to control a task difficulty and allows for criteria that focus on the most valid aspects of a task, as well as to be aware of the fact that different communicative purposes require very different skills. A task with a referential communicative purpose might require a test-taker to summarise a lecture by rephrasing the main and supporting ideas in a structured way. Alternatively, the test-taker could be asked to agree or disagree (emotive), add a convincing personal assessment of the input material’s content (conative), or engage in meaningful conversation about the lecture (combination of referential, conative, emotive and phatic).

## 2 Functions of speaking

According to the function, Brown and Yule (1983, in Richards, 2008) differentiate talk as interaction, transaction, and performance.

Talk as interaction, or conversation, is described as “interaction that serves a primarily social function. When people meet, they exchange greetings, engage in small talk, recount recent experience.” (Richards, 2008, p. 22). The purpose is to be friendly, to develop relationships. The focus is more on the speakers than on the message. The main features include reflection of speaker’s identity, politeness; use of conversational register; use of formal or casual language, etc. The speaker should be able to open and end conversation, choose topics, make jokes, react to others, recount personal experiences, etc.

Talk as transaction is aimed at what is said or done. “The message and making oneself understood clearly and accurately is the central focus, rather than the participants and how they interact socially with each other”. (Richards, 2008, p. 26). Participants involved in this kind of communication need to be able to negotiate the meaning, and so apply skills, such as describing something, asking questions, asking for clarification, asking for information, justifying an opinion, making suggestions, clarifying understanding, making comparisons, agreeing and disagreeing.

Burns (1998, in Richards, 2008, p. 26) distinguishes two types of talk as transaction. The first type involves situations where the focus is on giving and receiving

information (classroom group discussion and problem solving, making telephone call to obtain information, tour guide, asking for directions on the street). The second type is transaction, which focus is on obtaining goods and services (checking into hotel, airport, buying ticket, ordering food, buying something a shop).

Talk as performance refers to “public talk, that is, talk that transmits information before an audience, such as classroom presentation, public announcements, and speeches... Talk as performance tends to be in the form of monolog rather than dialog, often follows a recognizable format (e.g., a speech of welcome), and is closer to written language than conversational language. Similarly, it is often evaluated according to its effectiveness or impact on the listener”. (Richards, 2008, p. 27). The most common communication situations are giving a presentation, giving a lecture, and conducting a debate. The focus is both on message and audience, organization and sequencing of the speech. In terms of language, form, range of vocabulary and accuracy (grammar and correct pronunciation) are important; it is more like written language, often like a monologue. It is necessary to attract the audience, maintain their engagement.

According to Ellis and O’Driscoll (1992), there are three main parts of the presentation: the introduction, the body of the presentation and the conclusion. A good presentation starts with a greeting followed by a friendly smile, then the speaker and the topic, an explicitly identified goal of the presentation are introduced, and outline moving on to the brief theoretical background necessary for understanding the key terms and concepts. One may begin with the hot news, an interactive task or a funny real-life story to get the audience motivated to listen to. The goal of the informative presentation is to inform the audience about the facts or research results, whereas the persuasive presentation should strengthen or change the listener’s opinion providing him/her qualitatively and quantitatively sufficient amount of arguments. Based on the speaker’s goal, the body of the presentation might be built up according to several models. Starting with the brief outline of the problem, the theoretical background and key terms definitions, one can slightly move on to pros and cons arguments, their consequences supporting the claims by statistic data, examples and results from the research or case studies. A good conclusion is as important as a good beginning. Similarly, as the speaker should not forget to introduce him/her-self and the topic, definitely he/she should not forget to mention a brief summary of what has been said. After finishing the presentation (saying thank you for your attention) there is open space for sharing ideas and opinions in discussion).

### **3 Task-based testing and assessment**

Task-based language teaching (TBLT) is based on learning by experience or a form of experiment (trial – error). This approach enhances motivation of learners, fos-

ters real reason for communication in the target language, allows to see learner's development, to see "a linkage between functions, the language forms that realized them and the meaning-bearing uses to which they were put" (Norris, 2009, p. 580).

Task-based language assessment (TBLA) measures what a candidate is able to do with the language. "Key is the idea that holistic activity structures, such as tasks, offer an ideal frame within which knowledge use can be experienced and understood, and from which learning opportunities should be developed" (Norris, 2009, p. 579). TBLA has strong theoretical background in communication language teaching which is characterized with the "form-function-meaning relationship". Therefore, this approach can be referred to as a direct approach.

Proper summative TBLA of tasks should have the following qualities: "dependence on representative tasks that can be trusted to reflect language use in actual targeted domains (general or specific); replication of authentic task performance conditions and criteria; consistency in administration and reliability in rating, scoring, or otherwise judging task performances". (Norris, 2009, p. 586).

According to Hughes (1989, p. 15), direct testing approach has three advantages: "First, provided that we are clear about just what abilities we want to assess, it is relatively straightforward to create the conditions which will elicit the behaviour on which to base our judgements. Secondly, at least in the case of the productive skills, the assessment and interpretation of students' performance is also quite straightforward. Thirdly, since practice for the test involves practice of the skills that we wish to foster, there is likely to be a helpful backwash effect." In other words, the biggest challenge in TBLA is to use as much authentic material as possible. However, candidates who take speaking examination are aware that they are tested, and so it decreases authenticity of the situation. Nevertheless, the developers should focus on creating the content as realistic as possible.

Some examples can be found in Madsen (1983, p. 158) who distinguishes guided and controlled tasks. For example, a guided task is paraphrasing which can be based either on listening or reading, or describing graphs and diagrams, and using role-playing. He distinguishes two types of role-plays: the open-ended one, which can suit, especially, more talkative people, and the guided role-play characterised by fixed roles and brief description of scenarios.

The interview can also be used in speaking test. But a rapport needs to be developed to be as much objective as possible. It is recommended to use the guided interview. "Parts of the interview – especially the initial warm-up can appear so relaxed that the student may not even be aware that he is being evaluated at that moment. If you know something about the student, you can tailor and personalize the questions." (Madsen, 1983, p. 163). Examples of tasks are wh-questions, re-

sponding to statement, qualifying information, revising information, or correcting information. The aim in the guided interview is to get the student talking on his own, so the interview should not be limited to asking/answering questions.

When speaking about assessment, descriptors for fluency (speech rate, amount of information conveyed per minute or general impression of fluency) and accuracy are discussed frequently. Other factors, such as listening comprehension, correct tone, reasoning ability, and initiative in asking for clarification play a crucial role as well.

Weir (2005) characterizes several kinds of validity related to standardized tests and objective assessment. Context validity is concerned with the social dimensions of a task, such as task setting (purpose, format, time constraints) and linguistics demands (channel, discourse mode, length, topic, lexical, structural, functional). Theory-based validity is oriented to candidate's ability to cope with the contextual variables at the particular level of proficiency by CEFR. There are many cognitive and metacognitive processing activities which need to be studied and detected more precisely. The assumption is that much more complex and comprehensive processing is demanded at C1 level than at A1 level. He emphasizes that defining such progression will not be easy at all. Theory-based validity includes executive processes (goal setting, topic and genre modifying, idea generation, organization of ideas, translation), context knowledge (internal, external), and language knowledge (grammatical, discursal, functional, sociolinguistic). Last but not least, scoring validity is the issue of quality dealing with how well the participant is expected to carry out a task at a particular level in terms of clear and explicitly specified criteria of assessment that are symbiotically linked to the context-based and theory-based parameters of the construct being measured. Scoring validity includes the following criteria: rating scale, raters, rating procedures (rating training, standardization, rating conditions, rating, moderation, statistical analysis), and grading and awarding.

## **4 Testing and Assessing Speaking in Various Certificates**

In this section, speaking examination within particular certification systems will be described.

IELTS certificate is world-wide recognized. Depending on test-taker's needs, there are two kinds of IELTS examinations: academic English and general English. Regardless test-taker's subject focus, IELTS Academic test measures English proficiency to approve that the candidate can study at university or high school abroad. IELTS General Training is aimed at proficiency in common workplace and social situations. The difference is only in writing and reading part. In academic test, more specific genre or discourse is used, more difficult vocabulary is included, and complexity of style is more considered than in the test for general English.



IELTS speaking examination is divided into three parts. Firstly, test-takers answer general questions about themselves (their home, family, work, study, hobbies). Secondly, they are given a particular topic, about which they will speak for two minutes. They might be asked additional questions afterwards. In the last part of the exam, which takes up to five minutes, the topic from the second task is discussed in more detail. The whole speaking exam takes 11–14 minutes.

In IELTS assessment of speaking skills is of an analytic kind of assessment because it is based on criteria such as fluency and coherence (25%), lexical resource (25%), grammatical range and accuracy (25%), and pronunciation (25%). Then the candidate is given a score within the scale from 1 to 9. The overall band score is calculated by taking mean of the total of the four skills scores.

#### *4.1 UNiCert framework*

More focused on academic language is “UNiCert©” system. UNiCert is recognized in the Central Europe. The framework is an important step on the way to a common university-specific training and certification system for modern languages. UNiCert aims to equip students with communicative skills typical professional and academic situations at home or abroad where the target language is spoken.

At UNiCertLevel III (C1 in CEFR) students should be fully capable of meeting the linguistic requirements of a work placement or period of study in the country of the target language without the need for additional explicit language training. This level is the recommended level of mobility for an academic stay abroad.

Cumulative assessment may be achieved by the accumulation of final grades in a number of sections of the training programme or through the accumulation of assessments carried out in the final section of each level, i.e. all four skills must be assessed during the final section of each level and each must be passed in order for an overall pass to be awarded for that level.

The holder of this certificate [at the level C1] is proficient in general, academic and professional language which enables him/her to communicate easily with others, demonstrating flexibility and variety in his or her modes of expression. He/she can understand the vocabulary and structures of demanding, extensive, original written and spoken materials on general topics as well as on those related to his or her intended profession, can comprehend both explicit and implicit information and understand the content of presentations and lectures in detail. He/she can express him/herself fluently and effectively both in speech and writing on a range of complex topics relevant to his or her field of study within the context of his or her work and study abroad, making use of sophisticated structures and an extensive range of general and specialised vocabulary. He/she can express his/her own

views coherently, cohesively, logically and in a stylistically appropriate manner. (UNlcert Framework, 2012, p. 14).

#### 4.2 *Medical-oriented certification systems*

Next, three occupation-oriented tests for medical professionals will be described. Neither OET (Occupational English test), nor sTANDEM (Standardized Language Examination for Medical Purposes) measure the range of candidate's medical knowledge, but aim at communication skills, such as communication with an employer and colleagues, undertaking courses of further medical training, history taking, answering patient's questions, dealing with different age groups and challenging patients, or explaining medical conditions and treatment. In both of them, the examination tasks have a form of a role-play, which is based on real-life workplace situations specific to the particular medical profession.

In OET speaking test consisting of two role-plays, in total it takes 20 minutes, the candidate always takes the role of the health professional and the interviewer plays the patient or patient's relative, so there is no time for any preparation. The candidate's oral performance is recorded and assessed by two assessors independently. The assessor scores the candidate for each role-play respectively.

The score from 1 (the lowest) to 6 (the highest) is assigned for each of the criteria. The total score comprises the overall performance across the two role-plays, and all criteria are weighted equally. Then, the score for speaking skills is expressed in the form of grade: A (very high level of performance), B (high level: able to use English fluently and accurately for professional needs), C (good level: however it is not acceptable to a range of health and medical councils), D (moderate level which requires improvement), and E (low level requiring considerable improvement).

In contrast, the test-taker in sTANDEM (Charpy & Carnet, 2014) is given cards with instructions and prompts, which can he/she use in the role plays. The second task is focused on speaking independently on the topics, such as choosing specialty, daily routine, medical education, health care system, alternative medicine, etc. In the third task, a test-taker is required to give a presentation.

Assessment is holistic, following the "can do statements" (the level C1 in CEFR) below:

- can provide detailed descriptions of complex topics and link the individual sub-topics,
- can relate individual topics and use appropriate closing,
- can make announcements fluently,
- can express subtle shades of meaning using appropriate intonation and accent,
- can give clear, well-structured presentations on complex topics,

- can easily handle instructions.

UNICert III (C1 in CEFR) – “English for medical sciences”, provided by Jessenius Faculty of Medicine (JFM) in Martin, Slovakia, is aimed at the undergraduate students who passed the compulsory course of “English for General Medicine” in the 1st year of their study as well as at the postgraduates who graduated at the faculty.

The speaking test is a bit different in examination tasks from the other certification tests. We have decided to test candidate’s knowledge, their ability to paraphrase facts in logical sequences, to explain medical conditions and procedures both medical and non-medical language. The assessors may ask questions to test deeper knowledge or to provoke discussion on the current medical issues. Next, the candidate is required to give a presentation for academic purposes. They have to prepare a Powerpoint presentation on the topic of their choice within the range of topics included in the preparatory course. Candidate’s ability to convey facts effectively and clearly as well as to respond to questions with confidence and flexibility is being assessed. The emphasis is on fluency as well as accuracy.

Holistic approach, including analytic assessment with rubrics, is applied. In assessment of overall speaking performance, the candidate is scored at the scale 100–91% (A – excellent; numerical value 1), 90–81% (B – very good; numerical value 1,5), 80–73% (C – good; numerical value 2), 72–66% (D – satisfactory; numerical value 2,5), 65–60% (E – sufficient; numerical value 3), and below 59% (Fx – insufficient; numerical value 4). For setting final grade, all parts of examination are taken into consideration equally. This is called cumulative approach – the candidate must pass the test in each skill in order to pass the certification examination as a whole.

## **5 Results of the presentation examination in UNICert JLF 2014/2015**

In total 8 students took part in the oral examination. There were 5 undergraduate students (2nd to 5th grade) and 3 post-graduate students (1st grade). The presentation takes 20 minutes for a candidate. The candidate can achieve 100 points (percent) in total. There are five rubrics (each for 20 points in total), and within the rubric the candidate can be given 5 points at maximum for each criterion (Figure 1).

In both groups, the highest score was achieved for the criterion “Content”. The topics were original, showing creativity in processing the facts and implementing demonstrative videos, pictures and figures. Information was presented from the new viewpoint. In terms of validity and appropriateness, a senior medical doctor

**EVALUATION FORM FOR ORAL PRESENTATION**

Student's name: \_\_\_\_\_ Date: \_\_\_\_\_

Topic: \_\_\_\_\_  
 Indicate the appropriate score from A (5 points) = excellent, B (4 points) = good, C (3 points) = average, D (2 points) = below average, E (1) = poor, FX (0) = missing or failed; 100 points = maximum

<b>Presentation Structure</b>	<b>Presentation Content</b>
Introduction of topic and its content	Appropriateness of topic for audience
Organization into logical sections	Validity of information
Summary and conclusions	Handling the questions
References/Literature/Other sources	Length of talk (timing)
Closing the presentation	
<b>Visual Aids</b>	<b>Performance Impression</b>
Balance between pictures and text	Engagement, enthusiasm
Clear, catchy slides, not overcrowded	Smooth transition from issue to issue
Contribution of colours and sound effects to understanding, not distracting	Body language (posture), gestures, interaction with audience, eye contact
Font (size, style, quality), legible figures conveying results effectively	Voice and pacing, articulation, word crutches (um, okay, you know, etc.)
	Independent speech (no reading of text)
<b>Language</b>	<b>Notes/Committed Errors</b>
<b>Fluency</b>	
<b>Grammar Accuracy</b>	
<b>Pronunciation</b>	
<b>Vocabulary</b>	

Additional comments: \_\_\_\_\_  
 Total score: \_\_\_\_\_ Evaluated by: \_\_\_\_\_

**ASSESSMENT CRITERIA FOR ORAL PRODUCTION**

<b>Grade</b>	<b>Description of skills</b>
<b>A</b>	<ul style="list-style-type: none"> <li>- introduction of the topic, giving brief outline, the content points are expanded appropriately</li> <li>- clearly organized ideas with use of connectors and signalling phrases</li> <li>- use of wide range of vocabulary (synonyms) and sentence structures</li> <li>- language errors are minimal and do not disturb global listening comprehension</li> <li>- summary of the key information is included; conclusion with own viewpoint</li> <li>- style, format and graphics requirements are appropriate, used sources are listed</li> <li>- pronunciation is accurate throughout, with good rhythm and intonation for this level</li> <li>- excellent use of the time and multimedia</li> <li>- high level of presentation skills is shown, audience is involved from the very beginning up to the end, interactive and confident performance</li> </ul>
<b>B</b>	<ul style="list-style-type: none"> <li>- introduction of the topic without outline, the content points are expanded appropriately</li> <li>- clearly organized ideas with suitable use of connectors and signalling phrases</li> <li>- use of good range of vocabulary and sentence structures</li> <li>- language errors occur on the sentence level, but do not disturb the global comprehension</li> <li>- summary of key information is included; conclusion with own viewpoint is very short</li> <li>- style format and graphics requirements are appropriate, used sources are listed</li> <li>- pronunciation is generally comprehensible, with few errors</li> <li>- good use of the time and multimedia</li> <li>- good level of presentation skills, audience is involved at the beginning or at the end</li> <li>- more interactivity and self-confidence in public performance is needed</li> </ul>
<b>C</b>	<ul style="list-style-type: none"> <li>- introduction of the topic without outline, the content points are expanded appropriately</li> <li>- ideas are organized adequately, use of simple connectors and signalling phrases to the context</li> <li>- number of language mistakes, but does not impede global understanding</li> <li>- pronunciation with more "foreign" accent, but errors are rare</li> <li>- summary is present, but own viewpoint is missing</li> <li>- style, format and graphics requirements are appropriate, used sources are listed</li> <li>- use of the time and multimedia is acceptable</li> <li>- average level of presentation skills is shown, performance is learnt by heart or read</li> </ul>
<b>D</b>	<ul style="list-style-type: none"> <li>- no introduction of the topic and outline</li> <li>- incomplete ideas, weak text organization, rare use of connectors and signalling phrases</li> <li>- use of simple vocabulary</li> <li>- frequent errors in pronunciation of verb or noun endings partially impede global understanding</li> <li>- inappropriate use of style, format and graphics elements, missing list of sources</li> <li>- missing summary and own viewpoint</li> <li>- use of the time is not always appropriate</li> <li>- low level of presentation, rare eye contact with audience</li> <li>- performance is read fluently for whole duration of presentation</li> </ul>
<b>E</b>	<ul style="list-style-type: none"> <li>- no introduction of the topic and outline</li> <li>- partially misunderstood instructions, incomplete ideas, weak text organization, no use of connectors and signalling phrases</li> <li>- use of simple vocabulary and high number of mistakes make performance slow and impede global understanding</li> <li>- mispronunciations force the speaker to interpret</li> <li>- inappropriate use of style, format and graphics elements, missing list of sources, summary and own viewpoint</li> <li>- insufficient level of presentation skills, performance is read for whole duration of presentation</li> <li>- no communication with audience; no eye contact</li> </ul>
<b>FX</b>	<ul style="list-style-type: none"> <li>- no performance or little language for assessment</li> <li>- performance is irrelevant, total misunderstanding of instructions</li> <li>- deadline for the task is not met</li> </ul>

Fig. 1: The assessment rubrics and criteria for giving a presentation, UNICert JFM

(an associate professor) who is a member of the committee, confirmed that the content corresponded with knowledge and skills required from a junior doctor.

When considering “Structure”, the data showed that “Logical organisation” was managed very well. All candidates achieved 20 points. Next, 4 undergraduates and 2 post-graduates achieved 5 points for “Introduction” of their topic. Then, three undergraduates and three post-graduates were assigned 5 points for “References”, and 4 undergraduates and 3 post-graduates for the rubric “Closing”. The criterion “Summary and conclusion” was covered by 4 undergraduates and all post-graduates.

Despite of widespread assumption that the young generation possesses high level of skills in computer and information literacy, it can be concluded that the undergraduates undervalued graphic design of their presentations. Only 2 of them were given 5 points for “Non-distracting background” and 3 candidates for “Legible font and figures” and “Clear slides”. The criterion “Text-picture balance” was managed the best. When talking about the post-graduates, all of them achieved maximum score in all criteria. The reason for that might result from their experience of giving a talk at conferences or giving a lecture for undergraduate students when leading courses. It can be said that the postgraduates are more aware of the fact how important is to catch and hold attention of the audience.

In addition, the overcrowded slides often occurred in presentations of the undergraduates. They put down complex clauses, or even whole sentences instead of keywords. Consequently, it took the speaker away from their natural speech rate to reading. We believe that it may be due to their nervousness or low self-confidence in using English. Three undergraduates achieved 5 points for “Enthusiasm” and “Smooth transition”, and two for “Body language” and “Independent speech”. In contrast, all post-graduates were assigned maximum in all criteria.

To sum up, on the scale of the grades (A-FX), four undergraduates were awarded grade “C” (72%, 75%, 75%, 80%) and one “A” (100%). In the group of postgraduates, two candidates achieved “A” (100%) and one got “C” (75%).

Candidates’ feedback revealed that learning terminology and its pronunciation have made them to feel more confident when teaching foreign students. They appreciated acquisition of wide range of vocabulary from various branches of medicine. However, discussion on topics, such as abortion or euthanasia, was found difficult because they have a lack of professional experience; the topics are too abstract to them. In general, they were satisfied with the course content and approach to testing. They would strongly recommend taking part in the course to their colleagues.

## Conclusion

Nowadays, call for developing more specified and unified descriptors is the primary concern of many language experts and test-developers dealing with English for specific (not only medical) purposes (e.g. Fulcher). The outstanding question is what should be tested, in what kinds of tasks, and for what reason.

The aim is to find the way how to make various certification systems equivalent in terms of candidate's profile. In other words, how to ensure that descriptors are understood by test assessors in the same way and tested skills meet the same criteria. If this happens, the certificates would be recognized for academic and professional purposes without any doubts about candidates' proficiency level achieved across various certification systems.

Despite the fact that certificates recognized in Europe and worldwide are closely connected with CEFR, some scholars and stakeholders criticize on several weaknesses of CEFR. Each of the four certification systems understands the role of CEFR as general guidelines, which can be further adapted to meet the certification aims and needs of candidates.

Description of the assessment policy between the certification systems provided in this paper showed that rubrics are named more or less in the same way. The difference is in the degree of their vagueness. While IELTS and OET use rubrics followed by the brief description of candidate's ability, the guides of sTANDEM and UNiCert framework provide "can do statements" in their guides.

Another difference was found in the types of tasks, instructions given, or preparation time. IELTS speaking test consists of three tasks oriented on use of General English, even though the candidate enrolled for the academic-oriented certificate. In sTANDEM, the candidate is provided with all the information at the day of the examination, so individual's medical knowledge is not tested, and time for preparation is very short. In OET the candidate has to succeed in medical-oriented role-plays. In UNiCert at JFM in Martin, the candidates show their medical knowledge, give a presentation on a specific topic of their choice. The specific topic has to correspond to topics of the course syllabus (e.g. the course topic Neurology – the presentation topic Alzheimer disease). The course topics are drawn by candidates at the last course lesson, so the preparation is a part of their self-study to speaking examination.

Development of assessment criteria and descriptors along to appropriate tasks is a long-term process. Therefore, further study of various sources, their analysis and comparison has to be done in order to achieve agreement between experts in order to create a new common framework for language testing.

## Reference

- AKS (2012). *The UNICert Framework*. Retrieved May 5, 2016, from [http://www.unicert-online.org/sites/unicert-online.org/files/beispiel\\_ro\\_unicert\\_framework\\_english.pdf](http://www.unicert-online.org/sites/unicert-online.org/files/beispiel_ro_unicert_framework_english.pdf)
- ALTE, Language Policy Division (2011). *Manual for language test development and examining: For use with CEFR*. Retrieved May 10, 2016, from [http://www.coe.int/t/dg4/linguistic/ManualLanguageTest-Alte2011\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/ManualLanguageTest-Alte2011_EN.pdf)
- British Council (2013). *IELTS Guide for teachers*.
- British Council (2016). *How should the CEFR be used by recognising institutions wishing to set language ability requirements?* Retrieved May 15, 2016, from [http://www.ielts.org/researchers/common\\_european\\_framework.aspx](http://www.ielts.org/researchers/common_european_framework.aspx)
- Cambridge English Language Assessment (2015). *OET Preparation Support Pack*. Australia: Melbourne: Box Hill Institute.
- CHARPY, J. P., & CARNET, D. (2014). The European sTANDEM Project for Certification in Medical English: Standards, Acceptability and Transgression(s). In *ILCEA* (19, pp. 1–14). Místo vydání: vydavatelství.
- Council of Europe (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Retrieved April 10, 2016, from [http://www.coe.int/t/dg4/linguistic/source/framework\\_en.pdf](http://www.coe.int/t/dg4/linguistic/source/framework_en.pdf)
- Council of Europe (2001). *Common European Framework of Reference: European language levels: self-assessment grid*. Retrieved April 10, 2016, from [http://www.britishcouncil.mk/sites/default/files/self\\_assessment\\_grid.pdf](http://www.britishcouncil.mk/sites/default/files/self_assessment_grid.pdf)
- ELLIS, M., & O'DRISCOLL, N. (1992). *Giving presentation*. Essex: Longman Group UK, Ltd.
- FULCHER, G. *Language Assessment in Medical Contexts*. Retrieved May 16, 2016, from <http://languagetesting.info/whatis/scenarios/medical.php>
- HUGHES, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- MADSEN, H. S. (1983). *Techniques in testing*. New York, NY: Oxford University Press.
- NORRIS, J. (2009). *Task-based teaching and testing*. Retrieved April 15, 2016, from <https://larc.sdsu.edu/testassesswebinar/jnorris/Norris2009-Handbook-of-Language-Teaching.pdf>
- RICHARDS, J. C. (2008) *Teaching listening and speaking: From theory to practice*. New York: Cambridge University Press
- sTANDEM. *Descriptors recommended for sTANDEM*. Retrieved May 18, 2016 from [http://www.standem.eu/wp-content/uploads/2012/07/sTANDEMDescriptors\\_03.08.2012.pdf](http://www.standem.eu/wp-content/uploads/2012/07/sTANDEMDescriptors_03.08.2012.pdf)
- sTANDEM (2013). *Why sTANDEM is superior to general academic language proficiency testing* Retrieved May 18, 2016 from <http://www.standem.eu/wp-content/uploads/2013/01/Why-sTANDEM-is-superior-to-general-academic-language-proficiency-testing.pdf>
- UNICert III, English for medical sciences. (2016, May) Jessenius Faculty of Medicine in Martin. <http://www.jfmed.uniba.sk/pracoviska/vedecko-pedagogicke-pracoviska/teoreticke-ustavy/ustav-cudzich-jazykov/unicertiii/>
- WEIR, C. J. (2005). *Limitations of the Common European Framework for developing comparable examinations and tests*. Retrieved April 20, 2016, from <http://ltj.sagepub.com/content/22/3/281.full.pdf>

## Author

**Mgr. Petra Zrníková, PhD.**, e-mail: [zrnikova@jfmed.uniba.sk](mailto:zrnikova@jfmed.uniba.sk), Department of Foreign Languages, Jessenius Faculty of Medicine in Martin, Comenius University Bratislava, Slovakia.

Author is a lecturer at the Department of Foreign Languages. She teaches English for General Medicine, Nursing, Midwifery, and Public Health. She teaches voluntary courses, such as English for specific purposes (students of 4th grade in General Medicine) and UNICert III – “English for medical sciences”. Her major interest is need analysis, developing course materials and tests used in teaching English for medical purposes.