

## **DEFINING AN ALTERNATIVE ASSESSMENT CONSTRUCT AS A RATING SCALE FOR EVALUATION OF LSP ORAL PRESENTATION\*\***

**Vesna Petrović\***

University of Kragujevac, Faculty of Technical Sciences, Čačak

---

### **ABSTRACT**

The subject of the research is grounded on the theoretical analysis of different models of communicative competence in order to develop rating criteria for assessment of oral presentation in language for specific purposes. The aim of the research is defining validity of an alternative form of assessment, such as oral presentation, by defining the construct of assessment in the manner of the standardised testing procedure. More precisely, starting from the students' performances, i.e. oral presentations in language for specific purposes that need to be evaluated, we analysed theoretical frameworks of communicative competence and specific purpose language ability, as well as speech parameters and public speaking rating scales' characteristics. This approach was adopted in order to reach an optimal solution for an assessment construct of oral presentation performance. The initial hypothesis posits that it is possible to ensure the *a priori* validity of alternative assessment by defining the construct and its operationalisation through rating criteria. Therefore, we designed a unique rating scale for oral presentation assessment based on the results of the theoretical analysis. Thus, the rating scale for oral presentation assessment is composed of the parameters of linguistic, sociolinguistic, discourse and functional competences as well as on performance strategies and background knowledge. The main purpose of such rater-oriented scale is to increase objectivity and *a priori* validity of assessment procedure in the real classroom context. The contribution of this paper is a unique, theoretically-grounded, analytic rating scale which is suitable for application in higher education context.

### **Key words**

*construct, alternative assessment, validity, rating scale.*

---

\*\* This study was supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia, and these results are parts of the Grant No. 451-03-136/2025-03/200132, with University of Kragujevac Faculty of Technical Sciences Čačak.

\* vesna.petrovic@vstss.com, <https://orcid.org/0000-0002-1321-3449>

## Introduction

Assessment plays a pivotal role in foreign language teaching, as it enables a solid foundation for reaching conclusions and making further decisions related to learning and teaching (Safarzadeh & Taghipour Bazargani, 2023). There are a number of assessment approaches such as diagnostic, summative, formative, classroom-based assessment, sustainable assessment, authentic/alternative assessment, performance assessment, etc. In higher education context, summative assessment, or assessment based on numerical grades, is most frequently used as it enables earning an academic degree (Boud & Falchikov, 2005). Vukasović (2006: 102–104) confirms the previous standpoint, stating that there has been substantial progress towards continuous assessment, self-assessment and peer assessment in higher education in Serbia. In order to engage students into active learning, Koh, Tan and Ng (2012) claim that it is necessary to make progress from traditional towards authentic assessment. In the same vein, Lavinia & Alina (2018: 201) state that the researchers support alternative assessment as it creates opportunities for reflective thinking about learning process and achievements, which in turn leads to self-planning of further learning aimed at improving students' performances.

Alternative assessment characteristics vary according to different researchers' standpoints (Norris, Brown, Hudson, Yoshioka, 1998). Herman, Aschbacher and Winters (1992: 2) use alternatives in assessment as the most comprehensive term that includes different forms such as alternative assessment, authentic assessment and performance-based assessment. All the previously mentioned terms are used in research to denote the assessment of a student's creative performance as opposed to, for example, multiple-choice tests. The same authors claim that authentic assessment implies that a student should employ all the skills and knowledge at their disposal in order to solve realistic, i.e. authentic situations. Bachman & Palmer (1996: 23) define authenticity as a degree in which certain test task characteristic respond to the qualities of a similar task that a student could meet in a real life situation. According to them, this is situational authenticity. As they further explain, if a task is relatively authentic, its assessment results could imply that a student would achieve the same results in the target language use situation. Lavinia & Alina (2018: 201) mention portfolios, diaries, exhibitions, oral presentations, interviews, experiments and projects as specific examples of alternative forms of assessment. Certain authors even consider research on authentic assessment as a solid knowledge foundation that may serve as a basis for forming a distinctive framework for authentic assessment (Ashford-Rowe, Herrington, Brown, 2014).

The previously mentioned standpoints show that there is a need for more thorough research on alternative assessment. This implication derives from the research that indicates the movement from traditional towards authentic forms of assessment (Koh, Tan, Ng, 2012). Additionally, alternative assessment offers a wide range of challenging activities that promote students' motivation to use all the

disposable language knowledge to solve a task which is not only linguistic, and therefore requires certain performance.

However, although there is a bulk of research that supports authentic assessment, there are also some studies that argue in favour of ensuring its validity and reliability (Bachman, 2000). Namely, Brown and Hudson (1998) consider validity and reliability to be essential characteristics of the standardised testing which endures rigorous evaluation. Thus, the standardised tests are initially subjected to measurement characteristics such as validity, reliability, objectivity, discrimination, initial setting, practicality and efficiency (Dimitrijević, 1999: 73), which is not undertaken in alternative assessment (Brown & Hudson, 1998). Moreover, a weakness of alternative assessment lies in the fact that its validity is not examined prior to its implementation, which may lead to inconsistencies in assessment (Clapham, 2000). Additionally, it is inevitable to determine the degree of alternative assessment validity concerning the task complexity, as well as the rater reliability (MacLellan, 2004). Janković, Tomović & Aleksić (2020) also advocate that the language proficiency should be evaluated on the basis of clearly defined and uniform rating criteria, thus making the assessment procedure as similar as the standardised assessment.

Taking previously mentioned arguments into account, we created this research to develop a valid and reliable alternative form of assessment based on the measurement characteristics of the standardised tests.

In order to create an alternative form of assessment aimed at *a priori* determining its validity and reliability, prior to its implementation, we started with the oral presentation in LSP since the task represents a key component of alternative assessment cycle. The oral presentation task is a complex activity that encourages students to use a foreign language to complete a purposeful assignment that is not purely linguistic, since an oral presentation involves presenting a specific topic to an audience (Petrović, Radović Firat, Plaurović, 2022: 415). This task is typically assessed at the tertiary level of education, especially in English for specific purposes, as the presentation skills are required in the contemporary business context (Makrieva, 2017). Oral presentation is a performance-based task. It is highly comprehensive and includes several phases: (1) selecting a topic and its accompanying material, (2) careful analysis of the selected texts, which requires reading skill, (3) negotiating meanings about the shortlist material that should be presented (if oral presentation is designed as a group activity), which requires listening and speaking activities, and finally (4) delivery of speech on the selected topic (Morita, 2000). Moreover, it is advisable to adjust the supporting material such as video presentation, slides, images, graphs, etc. Students also rehearse before the final topic delivery. The most contemporary teaching approaches enable online oral presentations (Hill & Storey, 2003). During the development of the assessment cycle, the creator selects appropriate tasks that encourage a student's language performance, which in turn serves as an indicator of their communicative competence (North, 2003: 1).

As we have focused on ensuring the validity of assessment cycle prior to its implementation, we will describe the concept of validity in more detail in the following segments. Apparently, Dimitrijević (1999: 73–74), provided the simplest definition of validity. He claims that a test is valid if “It measures exactly what we intend to measure”. However, over time, both the forms of assessment and the requirements for ensuring validity have been developed and adjusted. Therefore, as opposed to Dimitrijević (1999), Weir (2005) considers it unnecessary to examine the validity of the test itself. It is important to investigate the validity of the inferences based on the test results, and consequently the validity of the interpretations of these conclusions regarding the assessment outcomes. Weir (2005) also explains that until the 1980s, the construct validity was determined statistically, after the test had been done. This allowed us to determine whether the test measured a particular competence. However, Weir, (2005: 17) made a significant progress by introducing the concepts of *a priori* and *a posteriori* validity.

When advocating *a priori* validity, he argues that it is necessary to investigate the degree of validity at the initial stages of test development, define the constructs precisely, and thus facilitate the interpretation of test results obtained through statistical analysis. Weir actually considered that construct definition represents the foundation of ensuring validity (Chan & Latimer, 2020: 105). *A priori* validity encompasses (1) theory-based validity and (2) context validity. *A posteriori* validity comprises: 1) scoring validity, 2) criterion-related validity and 3) consequential validity (Weir, 2005). Additionally, the construct is operationalised through the assessment criteria (North, 2003). *In other words, the rating scale, like the construct itself, should include a description of the learner’s competences* (Fairbairn & Dunlea, 2017: 7). The scale must be grounded on appropriate theoretical principles in order to be valid, while also taking into account the context in which it is applied during its development (North, 2003).

## **Research approach**

This study adopts exploratory research approach (Branković, 2014: 17). Namely, taking into account the previously mentioned arguments, we started from defining the theory-based validity by defining the construct we intend to assess. The following research methods were employed: analysis, classification and synthesis (Raičević, 2009: 28). Specifically, the technique of systematic observation that includes data collection, description, analysis, generalisation and interpretation was applied (Raičević, 2009: 67). We were guided by the idea that establishing the validity of the test prior to its administration would facilitate the interpretation of the results after their analysis, following the approach of the Weir (2005). We established *a priori* validity by grounding the core of the research in the definition of the assessment construct. In other words, we created a rating scale with criteria based on theoretical models of communicative competence and specific purpose language ability in the context of higher education. Thus, our rating scale is based on the selected

components of theoretical models of communicative competence. Namely, the rating criteria defined in this research were based on the parameters we selected by analysing the existing communicative competence models proposed by Bachman & Palmer (1996), specific purpose language ability developed by Douglas (2000), and the components of communicative competence that were suggested by CEFR (Council of Europe, 2001) (see Petrović, Jovanović, Đukić, 2024). Therefore, our rating scale comprises the following communicative competence components: (1) linguistic competence, (2) discourse competence, (3) sociolinguistic competence (4) functional competence, (5) communication strategies and (6) background knowledge (Petrović, 2022: 77). On top of this, we considered the speech parameters proposed by Luoma, (2004), Richards (2008) and Fulcher (2014) when designing a rating scale for LSP oral presentation assessment. They include: (1) use of fixed phrases and formulaic expressions, (2) clarity of presentation, (3) effectiveness of presentation, (4) fluency, (5) grammar accuracy, (6) meeting the needs of the audience and their management, (7) managing eye-contact and (8) structure and organisation of presentation (see Petrović, Jovanović, Đukić, 2024: 69). Through the synthesis of relevant characteristics, we formed an optimal framework for constructing our scale for assessing oral presentations in language for specific purposes. Additionally, we considered the instructions for rating scale criteria development. Therefore, the following sections of the paper are devoted to the process of developing the construct for assessment of oral presentation in language for specific purposes. We also describe the procedure of defining the rating criteria with their descriptors. We will demonstrate that the assessment based on this rating scale actually represents an alternative form of assessment, or performance assessment as oral presentation requires certain performance along with the foreign language knowledge.

### **Rating scales for foreign language proficiency assessment**

Rating scale is an instrument for proficiency assessment (North, 2003: 1). It consists of rating criteria or proficiency levels with different number of credits on whose grounds assessment process is performed (Chen, 2016). Certain rating scales also include the descriptors for measuring students' performances (Fairbairn & Dunlea, 2017: 7).

Rating scales vary according to their purpose, target groups intended to be assessed, terminology used for proficiency descriptors or the degree of systematicity, etc. (Alderson, 1991). As a result, one universal scale cannot satisfy all the assessment needs in different contexts and developing a specific scale with a concrete purpose is recommended. Alderson (1991: 72–74) defines (1) constructor-oriented rating scales, (2) assessor-oriented rating scales or (3) user-oriented rating scales. In this study, we start from Alderson's classification while developing an assessor-oriented scale which focuses on the assessment procedure and the performance intended to be assessed. The scale should be developed to be assessor-oriented if the aim of evaluation is to assess a specific performance and the underlying communicative competence (North,

2003: 3). Regarding the research instrument development, we employed a task-based, i.e., performance-based approach that takes the context into account and may often incorporate task characteristics when creating scale descriptors (North, 2003: 26, 63). Fulcher, Davidson & Kemp (2011: 5) also state that the development of a performance-based scale entails a detailed observation and analysis of the characteristics of the performances being assessed, taking into account the context in which they occur. Thus, the characteristics of performance are integrated into the process of scale development. Nevertheless, consideration of the context in scale development may limit the extent to which the conclusions can be generalized to different contexts (Fulcher et al., 2011; North, 2003). It is also considered that this problem can be overcome if we evaluate a student's competence rather than performance (North, 2003: 2). Moreover, if the theoretical foundation on which the criteria are validly balanced is taken into account during scale development, the assessment and the drawing of conclusions based on such a scale can be highly meaningful and substantive (North, 2003: 78).

Furthermore, we created an analytic rating scale. Most researchers adopt two basic types of rating scales, analytic and holistic (Chen, 2016; Luoma, 2004). An analytic scale is employed to analyse and independently assess individual components of language performance. This enables a highly precise assessment process, in which evaluators focus on the different criteria defined within the scale and thus arrive at a comprehensive conclusion regarding the stronger and weaker aspects of a student's performance (Luoma, 2004: 68). The use of an analytic scale in assessment enables multiple evaluators to focus on the same criteria and aspects of performance. Analytic scales are also suitable for diagnostic assessment aimed at providing feedback to students. However, they may have certain weaknesses. In other words, evaluators may become overloaded by numerous categories requiring attention, which can lead them to assign similar ratings for different qualities (North, 2003: 72). Therefore, the number of levels representing different performance characteristics should be limited to an appropriate extent, enabling evaluators to make accurate judgments (Council of Europe, 2001: 21). Furthermore, the question arises as to whether the criteria within the same scale should have equal weight in the final score. All criteria are evaluated equally in most analytic scales (North, 2003: 73). In contrast, a holistic scale enables the rater to regard and assess the overall language performance. By applying a holistic scale, raters assess the overall performance and, consequently, do not evaluate according to separate criteria. Therefore, the holistic approach is efficient and suitable for real-time application (Fairbairn & Dunlea, 2017; Luoma, 2004). However, it often happens that when using a holistic scale, a rater pays more attention only to certain aspects of a student's performance (North, 2003). Furthermore, there are numerous standardised scales for assessing performance in any foreign language or solely in the English language. Most of these scales are used for assessment of spoken interaction, while the rating scales for assessment of public performance, i.e. oral presentation as an activity of oral production lack (Iberri-Shea, 2017). CEFR is an exception as it offers a rating scale for assessing public performance activities that contains rating criteria with descriptors (Council of Europe, 2001: 60). When analysing the criteria and their descriptors we notice that all in the levels, from the highest to the lowest

one, the affirmative statements are used to show what a candidate can do. Although such formulations represent a challenge, it is advisable to use positive statements which show a candidate's achievement rather than those that highlight what they cannot do (Council of Europe, 2001: 205). Besides, regardless of the type of scale, the descriptors should be specific, clear, concise, and independent (Council of Europe, 2001: 206–207). The expressions that are used to distinguish among the levels should be concrete and precise so that the difference is meaningful and clearly noticeable. Additionally, the expressions that describe proficiency should be clear and simple. Conciseness is recommended to facilitate drawing conclusions about a candidate's achievement; a descriptor should contain two sentences at most, as longer explanations may overload a rater. Independence can be obtained by checking whether each descriptor can be transformed into an independent category of a can/ cannot do list. If a rater can answer yes or no to the descriptor, it can be claimed that it is clear, concise and specific. Descriptors that are used for assessing public announcements (Council of Europe, 2001: 60) represent a highly suitable resource for creating a specific scale. This is due to the fact that the CEFR rating scale refers to a particular context and thus can be used as a solid ground for creating assessment criteria in specific circumstances along with appropriate modifications of descriptors which could satisfy concrete needs for assessment in a specified context (Luoma, 2004).

## Results

Taking the aforementioned theoretical and empirical research into consideration, we formulated the rating criteria with the corresponding descriptors in our study. We operationalised the construct of communicative competence of oral presentation in LSP through the rating criteria. A total number of 19 criteria were developed and divided into 3 groups: (1) language knowledge, (2) presentation content and structure, and (3) presentation strategies. Table 1 presents the criteria with their descriptors that refer to language knowledge.

*Table 1. Criteria for assessment of oral presentation in language for specific purposes – group: language knowledge (Petrović, 2022: 104–105)*

Criterion	Descriptor
Clarity of speech	A student talks loud enough, at appropriate speed, with appropriate intonation and rhythm.
Fluency	A student speaks fluently, spontaneously, and without hesitation.
Vocabulary	A student uses vocabulary that is in line with a specific LSP topic.
Use of appropriate register	A student uses appropriate register that is in line with a specific LSP topic.
Use of appropriate style	A student adapts their delivery style to suit appropriate level of formality.

Criterion	Descriptor
Grammar accuracy	A student mostly uses language constructions correctly (e.g. tenses, sequence of tenses) with an appropriate word order. Occasional errors do not impede communication.

---

It is obvious that Table 1 presents the overview of the language knowledge criteria. This group consists of 6 parameters: clarity of speech, fluency, vocabulary, use of appropriate register, use of appropriate style and grammar accuracy. A survey of the rating criteria that refer to content and structure of presentation is given in Table 2. This group also encompasses 6 criteria: appropriacy of topic, background knowledge, structure and organisation of presentation, audio-visual aids, alignment of thematic content with supporting visual material and alignment of visual materials' delivery with the context of entire presentation.

*Table 2. Criteria for assessment of oral presentation in language for specific purposes – group: content and structure of presentation (Petrović, 2022: 104–105)*

Criterion	Descriptor
Appropriacy of topic	The chosen topic is relevant for specific LSP area.
Background knowledge	A student shows appropriate subject knowledge.
Structure and organisation of presentation	A student's presentation is clear and well-structured and it includes appropriate presentation elements, such as an introduction of main points and a relevant conclusion, a suitable range of expressions and fixed phrases to introduce the topic in a logical manner, as well as to introduce their team members or a new presentation topic. The student's interaction with other team members is appropriate, including the proper ending of the presentation and inviting the audience members to ask questions and participate.
Audio-visual aids	A student uses appropriate software for the visual presentation of the text, such as images, animation or clips.
Alignment of thematic content with supporting visual material	Oral presentation is in line with visual aids used in the slides. Material is effortlessly delivered and is straightforward to follow.

---

Criterion	Descriptor
Alignment of visual materials' delivery with the context of entire presentation	Delivery of visual materials is in line with the oral presentation material and pace.

Finally, the third group comprises 7 criteria that describe presentation strategies: topic complexity appropriate for student's level of LSP proficiency, topic appropriacy relevant to the interest of audience, planning and preparation, effectiveness of presentation, managing eye-contact, handling follow-up questions and time management. The overview of the rating criteria is presented in Table 3.

*Table 3. Criteria for assessment of oral presentation in language for specific purposes – group: presentation strategies (Petrović, 2022: 104–105)*

Criterion	Descriptor
Topic complexity appropriate for student's level of LSP proficiency	A student chooses a topic that is compatible with their level of LSP proficiency.
Topic appropriacy relevant to the interest of audience	A student has delivered a presentation relevant to audience interest.
Planning and preparation	A student has prepared a straightforward presentation and has shown the ability to work as a team member.
Effectiveness of presentation	A student presents effectively, persuasively and enthusiastically using appropriate non-verbal presentation skills (e.g. gestures or miming)
Managing eye-contact	a student manages to keep eye-contact with the audience in order to check whether they are being understood, whether the topic is of interest, or whether they have lost the audiences' attention.
Handling follow-up questions	A student mostly gives appropriate answers or comments in response to the audience.
Time management	A student talks within an appropriate time frame.

Additionally, Table 4 presents the overview of the criteria based on the communicative competence characteristics in order to provide necessary theoretical background.

*Table 4. Construct based on communicative competence components according to Bachman and Palmer (1996), Douglas (2000) and Council of Europe (2001) (Petrović, 2022: 77)*

Components of Communicative Competence	Components of construct
Grammatical knowledge/ Linguistic competence	Clarity of presentation Fluency Vocabulary Grammar accuracy
Textual knowledge/ Discourse competence	Structure and organisation of presentation
Sociolinguistic knowledge/ Sociolinguistic competence	Use of appropriate register  Use of appropriate style
Functional knowledge/ Functional competence Strategies	Handling follow-up questions  Topic complexity appropriate for student's level of LSP proficiency Planning and preparation Audio-visual aids Alignment of thematic content with supporting visual material Alignment of visual materials' delivery with the context of entire presentation Topic appropriacy relevant to the interest of audience Effectiveness of presentation Managing eye-contact Time management
Contextual knowledge	Appropriacy of topic Background knowledge

Table 4 contains of clarity of presentation, fluency, vocabulary and grammar accuracy, which represent communicative competence parameters based on grammar knowledge and linguistic competence. Structure and organisation of presentation show communicative competence component based on textual knowledge and discourse competence. In order to assess the component of communicative competence founded on sociolinguistic knowledge and sociolinguistic competence, we developed the rating criteria referring to usage of appropriate register and style. Handling follow-up questions is part of the construct based on functional knowledge and functional competence. Communication strategies encompass: topic complexity appropriate for student's level of LSP proficiency, planning and preparation, audio-visual aids, alignment of thematic content with supporting visual material,

alignment of visual materials' delivery with the context of entire presentation, topic appropriacy relevant to the interest of audience, effectiveness of presentation, managing eye-contact and time management. Contextual knowledge consists of appropriacy of topic and background knowledge. However, it is important to highlight that beside appropriacy of topic and background knowledge which directly refer to LSP knowledge, both vocabulary and register which comprise grammar knowledge, also relate to LSP knowledge. This is due to the fact that that students need to use appropriate vocabulary and register in line with the selected LSP topic and therefore represent a key component of LSP knowledge.

## **Discussion**

The previously mentioned arguments lead us to the conclusion that the rating scale developed in this research is a unique, specific and intended for assessment in particular context. In other words, the rating scale is created for the assessment of students' communicative competence when performing oral presentation in LSP or in Business English. It is systematic and contains the rating criteria with detailed descriptions of construct components being assessed. Furthermore, it is assessor-oriented as it is based on assessment procedure and students' performance evaluation. While developing performance descriptors, we aimed to ensure that the formulations would be comprehensible to teachers as raters, as well as to students as recipients of the feedback based on the achievement descriptions within the criteria.

Our scale is performance-based rating scale as it was created on the grounds of an analysis of performance characteristics being assessed. It takes into account the specific LSP context and oral presentation characteristics as a task being evaluated. In order to address the problem inherent in a performance-based approach, which refers to the limited possibility of generalising conclusions drawn from the assessment process, we theoretically grounded the criteria in this research following North's (2003: 2) proposal. This approach allowed us to assess a student's communicative competence in performing oral presentation task by using the rating criteria we formulated. Each criterion within the scale can be evaluated with an appropriate grade or a specified number of points. In conclusion, a final grade can be entitled on the basis of previously evaluated rating criteria or a total number of the calculated points. This approach creates the possibilities for both summative and diagnostic assessment (North, 2003: 77).

Additionally, we followed CEFR (Council of Europe, 2001: 205–207) recommendations for specificity, clarity, conciseness and independence of performance description within the rating criteria. Namely, the descriptors contain affirmative sentences with simple formulations and expressions which show what a student can do/cannot do. There are no descriptors which contain a description that is longer than two sentences. Longer descriptors could cause confusion with raters as they would have to recall the information at the beginning of the description. The independence of the criteria was also ensured by formulation that shows what a

student can do/cannot do independently. Additionally, the scale for public announcements by CEFR (Savet Evrope, 2002: 62–63) and the scale for assessment of public speaking suggested by Iberri-Shea (2017: 11–14) represented a fruitful source of words and phrases selection for creating our rating scale. It is notable that each of the three groups of criteria contains the same or approximately the same number of parameters. As it is apparent that the rating scale contains a total number of 19 criteria, we considered that it would overload a rater. Petrović, Jovanović & Đukić (2024: 73, 74) propose 5 different models of the scales, each containing 9 criteria, whose validity and reliability were established and proved by the mathematical multicriteria decision-making analysis.

## Conclusion

This research presents the development of the rating scale for assessment of oral presentation in LSP. The aim of the study is to establish *a priori* validity at the initial stage of test creation. We formed our validity concept following the theoretical model of Weir's (2005) validity concept. We established validity prior to creating and administrating an assessment task, i.e. oral presentation. We also defined the construct we intend to assess and it is operationalised through LSP communicative competence within the activity of oral production. Therefore, we can expect that we created the possibility for establishing *a posteriori* validity (scoring validity, criterion-related validity and consequential validity) as well. Namely, we consider that we ensured the conditions for establishing scoring validity which further leads to drawing valid inferences about a candidate's performance. Additionally, we based our rating criteria following appropriate theoretical grounds and we believe that such approach enables establishing the consequential validity. According to Weir (2005) we actually allowed for generalisation of the judgements about a student's performance in the target language use situation. Following the aforementioned procedure, we also ensured that this scale, with its criteria and descriptors, should be comprehensible and applicable for teachers, as it will help them correctly interpret the assessment construct (Council of Europe, 2001: 21, North 2003: 17). Taking the previous arguments into account, we expect that the rating scale for oral presentation in LSP will help teachers in the assessment procedure. To put it differently, we believe that teachers will be able to pay attention to the same qualities of a student's performance and that we will prevent the possibility of evaluating irrelevant criteria that are not part of construct (e.g. effort). We also consider that we will possibly decrease the occurrence of certain assessment mistakes such as excessive severity or leniency, mean score tendencies, forming a grade on the basis of the whole group's performances, mistakes occurring due to a rater's exhaustion, etc.

The weakness of this research lies in the fact that the specific rating scale has not yet been used in an actual assessment context, in which *a posteriori* validity would be investigated. However, this creates the potential for future research which may test practical usage of different combinations of the criteria that would additionally examine the reliability of the rating scale itself.

## References

- Alderson, J. C. (1991). Bands and Scores. In J. C. Alderson & B. North (Eds.): *Language Testing in the 1990s* (pp. 71–86). London: Macmillan.
- Ashford-Rowe, K., Herrington, J. & Brown, C. (2014). Establishing the critical elements that determine authentic assessment. *Assessment & Evaluation in Higher Education*, 39(2), 205–222. <http://dx.doi.org/10.1080/02602938.2013.819566>
- Bachman, L. F. & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*, vol. 1. Oxford: Oxford University Press.
- Bachman, L. F. (2000). Modern language testing at the turn of the century: assuring that what we count counts. *Language testing*, 17(1), 1–42.
- Boud, D. & Falchikov, N. (2005). Redesigning assessment for learning beyond higher education. *Research and development in higher education*, 28(special issue), 34–41. [https://conference.hersda.org.au/2005/pdf/refereed/paper\\_398.pdf](https://conference.hersda.org.au/2005/pdf/refereed/paper_398.pdf)
- Branković, S. (2014). *Metodologija društvenog istraživanja [Methodology of Social Research]*. Beograd: Zavod za udžbenike.
- Brown, J. D. & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly*, 32(4), 653–675.
- Chan, S. & Latimer, N. (2020). Placing construct definition at the heart of assessment: research, design and a priori validation. In L. Taylor & N. Saville (Eds.): *Lessons and Legacy: A Tribute to Professor Cyril J. Weir (1950–2018)* (pp. 105–132). Cambridge: Cambridge University Press. <https://www.cambridgeenglish.org/Images/582822-silt-volume-50.pdf#page=129>
- Chen, G. (2016). Developing a Model of Analytic Rating Scales to Assess College Students' L2 Chinese Oral Performance. *International Journal of Language Testing*, 6(2), 50–71.
- Clapham, C. (2000). Assessment and Testing. *Annual Review of Applied Linguistics*, 20, 147–161.
- Council of Europe (2001). *Common European framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Dimitrijević, N. (1999). *Testiranje u nastavi stranih jezika [Testing in Foreign Language Teaching]*. Beograd: Zavod za udžbenike i nastavna sredstva.
- Douglas, D. (2000). *Assessing Language for Specific Purposes*. Cambridge: Cambridge University Press.
- Fairbairn, J. & Dunlea, J. (2017). Speaking and Writing Rating Scales Revision. *Aptis Technical Report*, TR/2017/001, 1–36.
- Fulcher, G. (2014). *Testing Second Language Speaking*. London and New York: Routledge Taylor and Francis Group.
- Fulcher, G., Davidson, F. & Kemp, J. (2011). Effective rating scale development for speaking tests: Performance decision trees. *Language Testing*, 28(1), 5–29.

- Herman, J. L., Aschbacher, P. R. & Winters, L. (1992). *A practical guide to alternative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Hill, M. & Storey, A. (2003). Speak Easy: online support for oral presentation skills. *ELT Journal*, 57(4), 370–376. <https://maaz.ihmc.us/rid=1K7F73WTZ-1LYSH40-V0B/Hill%20-%20SpeakEasy.pdf>
- Iberri-Shea, G. (2017). Adaptation and assessment of a public speaking rating scale. *Cogent Education*, 4(1), 1–16. <http://dx.doi.org/10.1080/2331186X.2017.1287390>
- Janković, N., Tomović, N. & Aleksić, M. (2020). Principi i kriterijumi ocenjivanja kao osnove vrednovanja postignuća u obrazovanju [Principles of assessment and assessment criteria as essentials in educational evaluation]. U V. Plovina i B. Kovačević (ur.): *Primenjena lingvistika danas – jezik, književnost i interdisciplinarnost*, Vol. 6, 123–139. Beograd: Filološki fakultet, Univerzitet u Beogradu. [https://doi.org/10.18485/dpls\\_pld.2020.6.ch7](https://doi.org/10.18485/dpls_pld.2020.6.ch7)
- Koh, K. H., Tan, C. & Ng, P. T. (2012). Creating thinking schools through authentic assessment: the case in Singapore. *Educational Assessment Evaluation and Accountability*, 24, 135–149. <https://link.springer.com/article/10.1007/s11092-011-9138-y>
- Lavinia, N. & Alina, B. T. G. (2018). Innovative Methods for Assessing Students' Performance in the ESP Classroom. *Ovidius University Annals, Economic Sciences Series*, 18(1), 200–205. <https://stec.univ-ovidius.ro/html/anale/RO/wp-content/uploads/2018/08/15.pdf>
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Maclellan, E. (2004). How convincing is alternative assessment for use in higher education?. *Assessment & Evaluation in Higher Education*, 29(3), 311–321.
- Makrieva, I. (2017). The changing attitude towards oral presentations in the ESP classroom. In E. Kurteva, M. Belova & I. Georgieva (eds.): *Foreign Languages and the contemporary higher education*, Conference Proceedings, VIII International Scientific Conference, 23–25 June 2017, Varna (pp. 398–404). Varna: Medical University.
- Morita, N. (2000). Discourse Socialization through Oral Classroom Activities in a TESL Graduate Program. *TESOL Quarterly*, 34(2), 279–310. <http://www.jstor.org/stable/3587953>
- Norris, J. M., Brown, J. D., Hudson, T. & Yoshioka, J. (1998). *Designing second language performance assessments*. Second Language Teaching & Curriculum Center, Mānoa: University of Hawai'i.
- North, B. (2003). Scales for Rating Language Performance: descriptive models, formulation styles, and presentation formats. *TOEFL Monograph*, 24, 1–96.
- Petrović, V. (2022). *Evaluacija i unapređivanje kriterijuma za ocenjivanje usmene prezentacije na engleskom jeziku struke* [Evaluation and Improvement of Criteria for Assessment of Oral Presentation in English for Specific Purposes] (Neobjavljena doktorska disertacija). Univerzitet u Beogradu.
- Petrović, V., Radović Firat, A. & Palurović L. (2022). Oral presentation as authentic material for providing ESP instruction. In I. Milićević (Eds.): *Proceedings TIE 2022*,

- 9th International scientific conference Technics and Informatics in Education, 16–18 September 2022, Čačak (pp. 414–419). Čačak: Faculty of Technical Sciences. <https://doi.org/10.46793/te22-414p/>
- Petrović, V. M., Jovanović, A. S. & Đukić, M. R. (2024). Developing Multicriteria Rating Scale for Assessment of Communicative Competence in LSP Oral Presentation. *Inovacije u Nastavi*, 37(4), 64–84. <https://doi.org/10.5937/inovacije2404064P>
- Raičević, V. (2009). *Metodologija istraživanja u metodici nastave stranih jezika [Research Methods in Foreign Language Teaching Methodology]*. Beograd: Čigoja.
- Richards, J. C. (2008). *Teaching Listening and Speaking: From Theory to Practice*. Cambridge: Cambridge University Press.
- Safarzadeh, M. & Taghipour Bazargani, D. (2023). Exploring Iranian ESP Teachers' Assessment Practices in Online Specialized English Courses. *International Journal of Research in English Education*, 8(4), 51–60. [https://ijreeonline.com/files/site1/user\\_files\\_68bcd6/bazargani74-A-10-938-3-25b4871.pdf](https://ijreeonline.com/files/site1/user_files_68bcd6/bazargani74-A-10-938-3-25b4871.pdf)
- Savet Evrope (2002). *Zajednički evropski okvir za učenje, nastavu i ocjenjivanje jezika [Common European framework of Reference for Languages: Learning, Teaching, Assessment]*. Podgorica: Ministarstvo prosvjete i nauke.
- Vukasović, M. (2006). *Razvoj kurikuluma u visokom obrazovanju [Curriculum Development in Higher Education]*. Beograd: Alternativna akademska obrazovna mreža.
- Weir, C. J. (2005). *Language Testing and Validation: An Evidence-based Approach*. Houndmills, Basingstoke, Hampshire. New York: Palgrave Macmillan.