

ON ORAL ATTAINMENT IN A FOREIGN LANGUAGE

MAGDALENA POSPIESZYŃSKA-WOJTKOWIAK

Abstract. The article is devoted to oral attainment in a foreign language. It revises the term speaking, and explains its difficult role, yet fundamental both in teaching and in learning. Although oral communication is not easy to success in, it is, inevitably, one of the main determinants of students' achievements in a classroom situation. The article presents what happens during a typical classroom interaction situation, emphasizing the cooperation between the speaker and the listener, quoting Grice's (1975) four maxims governing the cooperation in communication, namely, the maxim of quantity, quality, relation and manner. Then, various features, factors and functions related to speaking are enumerated, underlining the difficult role of speaking in the whole process of foreign language learning. According to Common European Framework (Council of Europe 2001), all the features of spoken language, also mediating and interaction, are important in designing speaking tasks/tests, setting the assessment criteria, establishing and organizing principles and curricula for the teaching programs. The last section of the article is devoted to assessing oral skills, hence different approaches and scales are presented as examples, together with a brief explanation of the constructs of reliability and validity is speaking assessment. Special reliability and validity procedures have to be employed by test designers in order to obtain objectivity of the scores.

Key words: FLA, speaking skill, oral interaction, speaking assessment scales

1. Introduction

Language is a means of communication that is learned or acquired as a system. This system implies perceiving the language via the set of (language) skills, which, as Dakowska (2007: 179) points out, constitute a fundamental part of verbal communication, and thus help categorize and systemize our (verbal) presentation. The following article will be devoted to verbal, namely, oral skill. The author conscious-

ly does not call it *speaking*, as the article not only will present the traditionally understood speaking skill but also oral skills such as mediating and interactive ones – according to Common European Framework (CEF).

2. The speaking skill revisited

Traditionally, all foreign language skills were divided into *receptive* – listening and reading and *productive* – writing and speaking. The speaking skill has been regarded as the one whose development contributes to transfer to the remaining skills (Bygate, 1987). Many a method emphasised its importance in gaining competence in a foreign language. Being, in its nature, so difficult to success in, especially in a classroom situation, it has been treated as the most fundamental skill to work on both by teachers and learners. The Audiolingual Method concentrated on speaking from linguistic principles considering the speaking skill itself as the repetitive and imitative routine tasks in predictable, teacher-initiated situations with emphasis on pronunciation learning. In the times of the Cognitive Method, the view upon speaking became counterbalanced by grammar and activities in all skills supporting the grammatical syllabus. Now, in the ‘communication era’ (Dakowska, 2005: 231), the stress on the speaking skill has become ‘more realistic’ (ibid.). Practising speaking does not imply audiolingual dialogue imitating any more, but it became more communicative, involving not only teacher-student(s) but – most of all – student(s)-student(s) interactions. These interactions may take different forms – “dialogues, role-plays, speech acts and functions of English, solving a puzzle or a dilemma, reaching a consensus, giving directions, spot-the-difference activities, developing a story, simulations, interviews, discussions and debates, drama and games, projects and many other information gap activities” (Dakowska, 2005: 231-32).

2.1. Speaking as interaction

A typical classroom spoken interaction involves the teacher and the learner talking to each other. They, interchangeably, become both speakers and listeners building the spoken event together and sharing the right to affect the results – which can be either mutual or/ and individual (Luoma, 2004: 20). This cooperation between the speaker and the listener is called the cooperative principle (Buck, 2001: 24). Grice (1975) worked out four maxims governing this cooperation in communicative interaction:

- the maxim of quantity (the speaker gives sufficient information but not too much),
- the maxim of quality (the speaker says only what he knows to be true),

- the maxim of relation (the speaker says what is relevant),
- the maxim of manner (the speaker is brief, clear and avoids ambiguity).

When people communicate their goal is meaning – not only the informative meaning (what happened) but also attitudinal meaning (what the speaker thinks about the topic), expertise meaning (the speaker's knowledge about the history of the topic), and judgement meaning (the speaker's view or views about what may happen next).

Similarly to the fact that there are different kinds of meaning(s) involved in the speaker's intention, there are basically two purposes of the speech act. First, we talk to chat and secondly, we talk to share information (Brown et al. 1984). What is even more, both chatting and information giving (and getting) may occur within one spoken event. As Luoma (2004: 22) puts it: "information related talk often comes sandwiched between social chat, and a social chat can easily turn into a serious discussion". Brown et al. (1984) define chatting as the exchange of friendly conversational turns with other people whose basic purpose is to commence and to maintain social contact(s) via creating an amicable atmosphere. The authors (*ibid.*) agree that chatting in the speaker's first language is related to personality/ individuality matters as not everybody is a (socially) skilled 'chatterbox'. Chatting in a foreign language is a slightly different case, that is why teachers should pay attention to particular features which – more or less – natural chatting involves, namely personal matters, social behaviour, cultural events etc. When we talk to inform, we usually transfer information and the most essential thing is to get the message across and make sure that the listener understands it. "Establishing common ground, giving the information in bite-sized chunks, logical progression, questions, repetitions and comprehension checks help speakers reach this aim" (Luoma, 2004: 23).

2.2. Features, factors and functions related to speaking

Many researchers underline the difficult 'role' which the speaking skill plays in Foreign Language Acquisition process. Dakowska (2005: 233) pinpoints that this difficulty results from the following needs:

- the need to perform hierarchical operations, especially at the level of communicative intention, first and foremost, deciding what to say,
- the need to integrate these operations in fractions of seconds to keep pace with the demands of communicative fluency,
- the need to do this primarily in the working memory and relying primarily on one's internal (mental) auditory representations.

When people speak – unlike to the situation when people write – they usually do not do it in sentences. Spoken language tends to consist of utterances built with "idea units" (Buck, 2001: 9), namely short phrases or clauses loosely connected

giving an impression of being totally ungrammatical. Since the speaker's utterance is constructed in real time, it frequently lacks the preparation time to organise and control the flow of speech. Spoken language is rather connected by the coherence of the idea units mentioned above than by any formal grammar. Chaffe (1985) enumerates the following features of speaking resulting as main linguistic differences between spoken and written language:

- in spoken language idea units tend to be shorter, with simpler syntax,
- in spoken language idea units tend to be strung together by coordinating conjunctions (and, or, but etc.) whereas normally in written language we join ideas in more complex ways,
- spoken language usually has hesitations (pauses, fillers, repetitions) that give the speaker more thinking time,
- spoken language is full of repairs (false starts, grammatical and lexical corrections, afterthoughts),
- spoken language tends to have more non-standard features such as dialect, slang, jargon and colloquial expressions,
- spoken language tends to be more personal with more personal involvement (and much less precision),
- speakers tend to indicate their feelings and personal opinions more (*I think, I mean, I wonder* etc.), their language is frequently full of gross approximations, overstatements and exaggerations.

According to the Common European Framework of Reference – CEF (*Council of Europe*, 2001: 125) spoken competence in a foreign language is divided into functional categories – macrofunctions and microfunctions. The former category refers to such chunks/ utterances of language which serve the same functional purpose – description, narration, commentary, explanation or demonstration (cf. Brown and Yule, 1983; and Bygate, 1987). The latter refers to individual actions which are often completed within one spoken turn – e.g. inviting, apologising, thanking etc. The microfunctions are then furtherly classified into six main categories:

- 1) giving and asking for factual information, e.g. describing, reporting, asking,
- 2) expressing and asking about attitudes, e.g. agreement/ disagreement, knowledge/ ignorance, ability, permission,
- 3) suasion, e.g. suggesting, requesting, warning,
- 4) socialising, e.g. attracting attention, addressing, greeting, introducing,
- 5) structuring discourse, e.g. opening, summarising, changing the theme, closing,
- 6) communication repair, e.g. signalling non-understanding, appealing for assistance, paraphrasing.

While language learners improve, they gradually take control of more microfunctions (e.g. due to the fact that their lexicon expands). Thus, beginners may initially know only one way for apologising or asking the way, while intermediate and upper-intermediate students may know many more, and advanced/ proficiency stu-

dents may even make use of tones or other suprasegmental phonetic features to express and emphasise their utterances.

All the features of spoken language are important elements in designing speaking tasks/ tests and setting the assessment criteria. When teachers/ instructors establish and organise principles and curricula for their teaching program they should also bear in mind the other oral skills according to CEF (*Council of Europe, 2001*), namely mediating and interactive skills.

3. Other oral skills according to Common European Framework (CEF)

In the times when Europe began ‘expanding’ and accepting new members the international collaboration also entered schools and classrooms and engaged both governmental and non-governmental institutions in the development of teaching and evaluation methods and in the production and use of (teaching) materials. According to CEF, (communicative) language competence functions within various language activities involving reception and production but also interaction and mediation.

3.1. Mediating skill

Mediating skill comprises interpreting and/ or translating. Mediation appears as both a receptive and a productive activity and it facilitates communication between people who cannot, for different reasons, communicate directly. Mediating language skills understood as (re)processing an existing text (either written or spoken) play a vital role in the normal linguistic functioning of our societies (*Council of Europe, 2001: 14*). Examples of mediating activities are, besides spoken interpretation and written translation, also summarising and paraphrasing in the same language (see the table below).

Table 1. Mediating activities according to CEF (*Council of Europe, 2001: 87*).

<i>MEDIATING ACTIVITIES</i>	
ORAL	<ul style="list-style-type: none"> – simultaneous interpretation (conferences, meetings, formal speeches, etc.); – consecutive interpretation (speeches of welcome, guided tours, etc.); – informal interpretation: <ul style="list-style-type: none"> – of foreign visitors in own country – of native speakers when abroad – in social and transactional situations for friends, family, clients, guests, etc. – of signs, menus, notices, etc.
WRITTEN	<ul style="list-style-type: none"> – exact translation (e.g. of contracts, legal and scientific texts, etc.); – literary translation (novels, drama, poetry, libretti, etc.); – summarising gist (newspaper and magazine articles, etc.) within L2 or between L1 and L2 – paraphrasing (specialised texts for lay person, etc.).

While discussing mediating skills and activities one should mention mediating strategies, namely, ways and techniques of coping in a situation in which one wants to process information and establish meaning. This process may involve the following stages: planning, execution, evaluation and repair (see the table below).

Table 2. Mediation strategies according to CEF (*Council of Europe*, 2001: 88).

MEDIATION STRATEGIES	
PLANNING	Developing background knowledge; Locating supports; Preparing a glossary; Considering interlocutors' needs; Selecting unit of interpretation.
EXECUTION	Previewing: processing input and formulating the last chunk simultaneously in real time; Noting possibilities, equivalences; Bridging gaps.
EVALUATION	Checking congruence of two versions; Checking consistency of usage.
REPAIR	Refining by consulting dictionaries, thesaurus; Consulting experts, sources.

To sum up, (oral) mediation facilitates communication between people who are not able to talk to each other in a direct way. "Translation or interpretation, a paraphrase, summary or record, provides for a third party a (re)formulation of a source text to which this third party does not have direct access" (*Council of Europe*, 2001: 14).

3.2. Interactive Skill

Both reception and production (either oral or written) are primary processes in language learning because they are important for interaction. Productive oral activities play a vital role in many academic (and professional) fields (delivering oral presentations, summaries, descriptions, reports, judgements, etc.).

Due to the fact that the process of oral *interaction* is primarily face to face and provides textual and contextual, linguistic and paralinguistic redundancy, it needs constant monitoring using the following interaction strategies: planning, execution, evaluation and repair – depicted in the table below.

For assessment purposes, speaking is divided into spoken production and spoken interaction where students obtaining the highest level of language competence – C2 (see *Council of Europe* 2001: 27), can effortlessly engage in the following activities depicted in a form of a graph in Figure 1 below.

Table 3. Interaction strategies according to CEF (*Council of Europe*, 2001: 85).

INTERACTION STRATEGIES	
PLANNING	<ul style="list-style-type: none"> – framing (selecting praxeogram); – identifying information/ opinion gap (felicity conditions); – judging what can be presupposed; – planning moves.
EXECUTION	<ul style="list-style-type: none"> – taking the floor; – co-operating (interpersonal); – co-operating (ideational); – dealing with the unexpected; – asking for help.
EVALUATION	<ul style="list-style-type: none"> – monitoring (schema, praxeogram); – monitoring (effect, success).
REPAIR	<ul style="list-style-type: none"> – asking for clarification; – giving clarification; – communication repair.

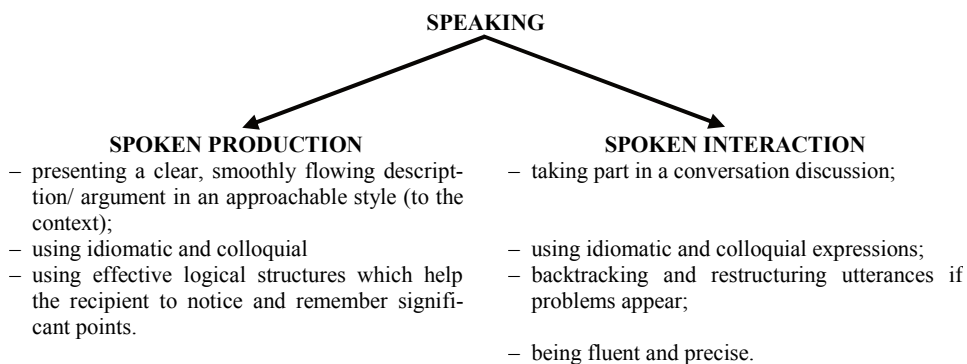


Figure 1. The division of speaking according to CEF.

As far as qualitative aspects of spoken language use are concerned, the student's utterance has been divided into five subcategories – range, accuracy, fluency, coherence and interaction. In order to be classified at 'mastery' level – C2 – the student (*Council of Europe*, 2001: 28):

- 1) shows great flexibility reformulating ideas in differing linguistic forms to convey finer shades of meaning precisely, to give emphasis, to differentiate and to eliminate ambiguity, also has a good command of idiomatic expressions and colloquialisms (RANGE),
- 2) maintains consistent grammatical control of complex language, even while attention is otherwise engaged, e.g. in forward planning, in monitoring other's reactions (ACCURACY),

- 3) can express him/ herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that the interlocutor is hardly aware of it (FLUENCY),
- 4) can create coherent and cohesive discourse making full and appropriate use of variety of organisational patterns and a wide range of connectors and other cohesive devices (COHERENCE),
- 5) can interact with ease and skill, picking up and using non-verbal and intonational cues apparently effortlessly, can interweave his/her contribution into the joint discourse with fully natural turn-taking, referencing, allusion making, etc. (INTERACTION).

Teachers should bear in mind that assessment of speaking is a tough thing. Especially in an exam-situation, students may be so paralysed with anxiety that it becomes difficult to give an objective evaluation of their utterances even if they are subdivided into smaller categories (range, accuracy, fluency, coherence and interaction). As Dakowska (2005: 239) notes, anxiety causes avoidance behaviour which may lead to insufficient participation in speaking tasks making the whole evaluation process futile. The following section is entirely devoted to assessment of oral skills in FLA.

4. Assessing oral skills

Assessing oral skills reliably is a difficult thing. The student's performance is being judged in real time, face-to-face, frequently not only between the learner and the interlocutor but also with one or two assessors accompanying to make the whole process more reliable. The assessor(s) usually must concentrate upon such aspects of the language as pronunciation, accuracy (grammar and vocabulary) and fluency, and they may be equipped with special evaluation scales with more or less detailed lists of criteria to be taken into account while evaluating, or the assessor(s) can grade the whole utterance globally – at a holistic level.

4.1. Assessing pronunciation, accuracy and fluency

Pronunciation is the first thing that strikes the listener's (interlocutor's/ assessor's, teacher's) ears. Luoma (2004: 11) calls pronunciation more broadly – the *sound of speech* because it refers to many items of the speech stream – not only individual segmental elements but also stress, intonation, tones, pitch, volume and speed. Luoma (ibid.) doubts whether all these things should be assessed under one broad criterion – pronunciation.

Another vital part in assessing speaking is accuracy of the utterance with emphasis on grammar and lexicon. If (whether good or bad) pronunciation is the first

thing that strikes the listener's ears, grammar is undoubtedly the first thing to be assessed as an undeniable proof of students' progress or lack of it (Larsen-Freeman and Long, 1991: 38-41). Luoma (2004: 12) concludes in a similar vein: "Learner grammar is handy for judging proficiency because it is easy to detect in speech (...), and because the fully-fledged grammars of most languages are well known and available for use as performance standards". Nonetheless, spoken grammar is much different from what we usually regard as standard forms. As was already pointed, people usually do not speak in sentences, instead they produce utterances called *idea units* (cf. Buck, 2001; Luoma, 2004) joined – or not – by conjunctions or short pauses and hesitation. Spoken grammar is not that complex (or proficient) as written grammar. Idea units are about two seconds or about seven words long, or even shorter (Chafe, 1985). Luoma (2004) notes that generally idea units are clauses which may contain a verb phrase, a noun phrase or a prepositional phrase – but almost hardly ever do they create a sentence, some of them may even not have a verb. What is even more, in real life conversations, an idea unit may be started by one speaker and be completed by another. The structure of an utterance depends also on whether speech is planned or unplanned. In the case of planned speech, when students have time for both preparation and rehearsal (speeches, lectures, conference presentations, oral exams with the lists of topics given before, etc.), the speaker's grammar may be quite complex and their whole utterance may consist of fully-fledged sentences with a high degree of features identified with written language – which, in contrast, are absent in unplanned speech.

The situation looks similarly with vocabulary, in common everyday conversation (unplanned discourse), people do not tend to use very sophisticated lexical items but normal and simple words. However, in planned discourse (especially in oral exam situations) people may use very refined speech which may become one marker of advanced speaking skills (cf. Read, 2000), and there is usually a 'list' of words, phrases and expressions in every interlocutor's or/ and assessor's head which puts the speaker in a good light and rank him/ her high at a fluency level. In spoken interactions people use words which are not very precise (*this/ that one, good, move, put, fine, this/ that thing, there*, etc.) but are present in almost every conversation due to their comprehensibility. These words are called *generic words*. "They talk about people, things, activities that can be seen or (...) are familiar to the speakers" (Luoma, 2004: 17). Apart from generic words, speakers also use the so-called *vague words* (like *thing, whatsit, something*) when they cannot bring back the word they need to say (cf. Channel, 1994), fillers, fixed phrases and hesitation markers – Haselgren (1998) calls them *smallwords* – when speakers want to delay time to speak. These are expressions like *ah, oh, yeah, you see, you know, kind of, sort of, I think, I guess, I mean, let me think, let me see* and repetitions, reformulations, repeating the interlocutor's questions/ phrases, etc. (cf. Nattinger and DeCarrico, 1992). Yet an-

other difficult, and perhaps the most difficult, thing to precise and assess in oral evaluation is fluency. As Luoma (2004: 88-89) observes:

Fluency is a thorny issue in assessing speaking. This is partly because the word ‘fluency’ has [both] a general meaning, as in ‘she is fluent in five languages’, and a technical meaning when applied linguists use it to characterise a learner’s speech. However, even in technical terminology, fluency can be used in a range of senses. The narrowest definitions only include a few features, typically pausing, hesitations and speech rate, whereas the broadest uses are virtually synonymous with speaking proficiency. (...) Definitions of fluency often include references to flow or smoothness, rate of speech, absence of excessive pausing, absence of disturbing hesitation markers, length of utterances, and connectedness. One central part of fluency is related to temporal aspects of speech, such as speaking rate, speech-pause relationships, and frequency of dysfluency markers such as hesitations, repetitions and self-corrections. These can be evaluated [both] by machine and by human impression. (...) [S]tudies indicate that when speakers become more fluent their speech rate increases and the speech flow contains fewer pauses and hesitations. (...) More fluent speakers tend to speak more and their phrases are longer.

Inevitably, fluency is connected closely to progress and proficiency leading us to the issue of assessment scales and rating – which are presented henceforth.

4.2. Assessment scales – examples

There is a whole range of different scales for speaking evaluation – more descriptive scales and numerical ones, rating checklists as well as similar to personality tests evaluating present-absent feature (0-1) of performance scales. We can distinguish *holistic vs. analytic* scales, *general vs. specific purpose* scales. There will be *intuitive, qualitative and quantitative* assessment scales (Luoma, 2004: 94-95).

The development of each scale is a difficult task, because test-instructors must answer a question – what is the ‘applied’ purpose of the assessment, whether it is prepared for giving final grade in a course or for selecting people to train or to employ, or just to provide feedback for future education, etc. Speaking scales developers must also decide on the number of levels and criteria to be taken into account in their scales and they have to bear with their interpretations and choose between norm-referenced and criterion-referenced scoring (Luoma, 2004: 80-81).

While developing speaking tests and their evaluation scales, test developers may adopt already existing scales and try to relate them to their testing situations. The already existing scales are usually prepared after application of intuitive methods without quantitative or qualitative data analyses (*Council of Europe*, 2001: 207); therefore, it is suggested to adapt them before applying. Intuitive methods in (evaluation) scale preparation are frequently based on developers’ experience in teaching

and materials elaboration (cf. Clark and Clifford, 1988). Qualitative methods involve a group of experts analysing the obtained data and relating it to the scale, where the analysis assumes observing an audio or video samples tapes. Quantitative methods of scales preparation require statistical data interpretation. The expertise is usually prepared in large testing by research institutions or/ and by qualified teachers, instructors or testers (cf. Iwashita et al., 2008).

Hereunder, a few examples of speaking scales are presented. Their choice is a subjective selection of the author in order to depict the distinction between the terminology used, the amount of details and other differences (cf. *Council of Europe*, 2001). Firstly, a holistic scale with six scoring levels is presented – The Finnish National Certificate descriptive scale (*National Board of Education*, 2002).

Table 4. The National Certificate descriptive scale (*National Board of Education*, 2002).

THE FINNISH NATIONAL CERTIFICATE SCALE	
6	Speaks fluently with few if any non-native features, such as a foreign accent. Is capable of expressing even subtle nuances of meaning with precision, and also makes varied and appropriate use of idiomatic expressions. Is able to describe even a complicated topic and to include sub-themes in the description, to develop different viewpoints and to bring the presentation to an appropriate conclusion.
5	Speaks fluently without frequent obvious need to search for an expression. Delivery characterised by naturalness, coherence and appropriate length. Is able to present a clear and detailed description of even a complex topic. Can use idiomatic expressions and everyday expressions, and is able to express nuances fairly well.
4	Copes fairly well even in less familiar speech situations. Makes a distinction between formal and informal registers, at least to some extent. Is able to present and justify an opinion comprehensibly. Is able to talk about and describe sights, sounds and experiences. Is obliged only rarely to use circumlocutions in everyday communication because of inadequate language proficiency.
3	Copes with the most familiar speech situations and is able to take the initiative in everyday language-use situations. Speech may be quite slow but there are few unnatural pauses. Is comprehensible despite transferring native or foreign language structures and vocabulary to the target language. Pronunciation may clearly deviate from target language standards.
2	Copes with routine speaking situations that require a simple exchange of information. Nevertheless, the speaker's language proficiency considerably restricts the range of matters that can be dealt with. Successful communication of a message presupposes that the interlocutor is willing to help the speaker in forming the message. Pronunciation may deviate clearly from the target language norm, thus requiring special effort from the interlocutor and impeding successful communication.
1	Is able to ask and reply to simple questions dealing with immediate everyday needs. Can make use of simple polite forms. Copes with the very simplest speaking tasks, but communication is slow and very fragmented often obliged to resort to non-verbal means in order to be understood.

In holistic evaluation, the assessors may only pay attention to general impression. The scale seems practical, quick, and easy-to-use as it only provides one score. Holistic scales are, indeed, practical to establish strengths and drawbacks between

particular levels but not necessarily to diagnose good and bad points of an individual's performance. One of the weaknesses of holistic scales may be the fact that the assessors will depend too much on such quantifiers as: many, a few, few, etc. (cf. North, 2000).

Another scale – The Test of Spoken English (TSE) rating scale is a combination of both holistic and analytic rating (Lazaraton and Wagner, 1996). The analytic scale usually takes form of a grid containing 3-5 criteria and providing detailed guidance for the tester. The example presented below has five levels (labelled: 20, 30, 40, 50, 60), and possessed descriptions of the examinee language at different levels.

Table 5. The TSE rating scale (Lazaraton and Wagner, 1996).

THE TEST OF SPOKEN ENGLISH (TSE) SCALE	
60	<p>Communication almost always effective: task performed very competently. Functions performed clearly and effectively; Appropriate response to audience/ situation; Coherent, with effective use of cohesive devices; Use of linguistic features almost always effective; Communication not affected by minor errors.</p>
50	<p>Communication generally effective: task performed competently. Functions generally performed clearly and effectively; Generally appropriate response to audience/ situation; Coherent, with some effective use of cohesive devices; Use of linguistic features generally effective; Communication generally not affected by errors.</p>
40	<p>Communication somewhat effective: task performed somewhat competently. Functions performed somewhat clearly and effectively; Somewhat appropriate response to audience/ situation; Somewhat coherent, with some use of cohesive devices; Use of linguistic features somewhat effective; Communication sometimes affected by errors.</p>
30	<p>Communication generally not effective: task performed poorly. Functions generally performed unclearly and ineffectively; Generally inappropriate response to audience/ situation; Generally incoherent, with little use of cohesive devices; Use of linguistic features generally poor; Communication often impeded by major errors.</p>
20	<p>No effective communication: no evidence of ability to perform task. No evidence that functions were performed; No evidence of ability to respond to audience/ situation; Incoherent, with no use of cohesive devices; Use of linguistic features poor; Communication ineffective due to major errors.</p>

The students' performances, in which they are asked to give directions/ opinions, support views, hypothesise about future events, etc. are taped and rated centrally. The students' responses are scored one by one (up to 12) and their final grade is

an average of all task scores rounded to the nearest 5 – 20, 25, 30, 35, 40, 45 and so on (Lazararon and Wagner, 1996).

Another scale, the diagnostic speaking test for medical students (undergraduates) at Melbourne University, is an example of a numerical rating scale (Grove and Brown, 2001) – see the table below.

Table 6. A numerical scale for an informal discussion task (Grove and Brown, 2001: 30).

<i>A NUMERICAL RATING SCALE</i>						
Informal Discussion Adequacy of participation						
Maintenance of interaction	6	5	4	3	2	1
Initiative, expansiveness	6	5	4	3	2	1
Quality of ideas						
Maturity and quality of thought	6	5	4	3	2	1
Interpersonal skills						
Engagement, rapport	6	5	4	3	2	1
Nonverbal behaviour	6	5	4	3	2	1
Coherence and expression						
Clarity of ideas	6	5	4	3	2	1
Cohesion and coherence	6	5	4	3	2	1
Register and tone						
Level of formality	6	5	4	3	2	1
Politeness	6	5	4	3	2	1
Directness	6	5	4	3	2	1
Tone of voice	6	5	4	3	2	1
Linguistic Criteria Language						
Range of structure and vocabulary	6	5	4	3	2	1
Breadth and precision of expression	6	5	4	3	2	1
Accuracy	6	5	4	3	2	1
Production						
Pronunciation	6	5	4	3	2	1
Intonation, stress and rhythm	6	5	4	3	2	1
Voice quality	6	5	4	3	2	1

In this scale, the levels are defined by numbers 6 – 1, Luoma (2004: 76) notes that such a solution may be useful if the raters agree about the meaning of each number on the scale, otherwise (*ibid.*):

there is an even number of levels (...) the raters need to decide whether the examinee is on the weak or strong side of the middle point of each scale. The alternative is to have an odd number, but the difficulty with that solution is that the interpretation of the middle

score may(w)hen be particularly variable – very broad for some raters and rather narrow for others.

Hereunder, in the last example, a rating checklist is depicted (Luoma, 2004: 78-79).

Table 7. An example of a rating checklist scale (Luoma, 2004: 79).

<i>A RATING CHECKLIST (WORKPLACE DESCRIPTION)</i>			
Content and structure of description			
	Yes	No	Comment
Mentions what he/ she does for a living			
Says where the workplace is			
Describes location of room or area			
Describes the room/ area itself			
Identifies the key tools and activities			
Describes atmosphere/ feelings			
Language of description			
+ / – Comment			
Flow of speech			
Rhythm and speed			
Pronunciation			
Intonation			
Stress			
Hesitation			
Vocabulary range			
Filler words			
Grammar			
Other, what?			

The checklist presented above was diagnosed to grade an utterance of a student whose task was to give a description of a room/ place where he/ she works.

Two types of checklist are used in this example. The first is a yes/ no format. For each feature on the list, the rater ticks whether it was present in the performance or not. The space for comments can be used to note what was particularly good or bad about each aspect of the description. The second is a +/ – format. The rater marks strengths with a plus sign and weaknesses with a minus sign, and possibly writes down some quick comments. The more concrete these notes are, for example quotes of phrases or sounds that the examinee used, the more useful the notes are going to be when giving feedback to the speakers. For helping examinees learn more, comments on weaknesses are particularly useful, while comments on strengths make them more aware of their skills and help them feel good about their language learning (Luoma 2004: 78).

Other examples of rating scales are those presented in Common European Framework – CEF (*Council of Europe*, 2001). The descriptors shown in CEF were not compiled for a particular (speaking) test but they form a peculiar collection of

basic test-specific criteria which are analytic and were prepared for general purposes (*Council of Europe*, 2001: 28-29; cf. Luoma, 2004: 71).

The last section of the article is devoted to the construct of reliability and validity in speaking assessment.

4.3. Obtaining reliability and validity in oral attainment

Speaking in educational contexts is usually assessed during live presentations between the testee and the tester(s). In such situations the (test) discourse cannot be predicted, even if the topics are known in advance both to the students and the assessors. There will never be two exact conversations even if the learners have the same aims – to pass the exam and to score at their best. All of these contribute to the difficulties concerning the rating criteria so that special reliability and validity procedures have to be employed to obtain objectivity of the scores.

Reliability tells us whether the test scores are consistent, so if we repeated a given testing situation (the same group of people, the same topic, the same, testers) we should obtain the same results. As Luoma (2004: 179) notes, if “the group of examinees are (...) small, (...) this does not warrant the use of advanced statistical methods. (...) Scoring reliability is maintained by using consistent rating procedures, which can be supported by rating forms”. It is relatively easy to avoid subjectivity in written tests as the scoring procedure can be done anonymously and the written works can be coded, however it is very difficult to retain objectivity in spoken assessment (Brown and Hudson, 2002).

The other factor that a good test should possess is *validity*. “A valid test is one which measures what it is supposed to measure and not something else” (Clegg, 1990: 137). In order to establish validity criteria, test creators should clearly define the purpose and aim of the speaking test, as different approaches to language learning will demand different assessment criteria. The linguistic approach, for instance, will put emphasis on language forms – vocabulary, grammar, pronunciation and fluency, whereas the communicative approach will focus more on particular skills and strategies that the learners employ to do particular activities, and the task-based approach will emphasise “the skills assessed in terms of the situations and roles simulated in the test and [will express] scores in terms of the examinees’ ability to deal with the tasks that were included” (Luoma, 2004: 185).

5. Conclusions

This article was devoted to oral attainment in a foreign language. Speaking is the skill which not only is the most difficult to gain competence in by learners but which is also very difficult to assess by teachers/ instructors. Being so ‘troublesome’ both

to master and to evaluate, it is regarded as the best indicator of language proficiency – the determinant of success of what it means to be able to use a foreign language. Being able to communicate orally in a foreign language student must learn many things – they have to know pronunciation, they have to acquire the necessary lexicon, and they have to be able to put their utterances together grammatically correct in order to gain intelligibility and to minimise hesitation phenomena.

References

- Brown, G., Anderson, A., Schilcock, R. and G. Yule. 1984. *Teaching talk: strategies for production and assessment*. Cambridge: Cambridge University Press.
- Brown, G. and G. Yule. 1983. *Teaching the spoken language: an approach based on the analysis of conversational English*. Cambridge: Cambridge University Press.
- Brown, J.D. and T. Hudson. 2002. *Criterion-referenced language testing*. Cambridge: Cambridge University Press.
- Buck, G. 2001. *Assessing listening. Cambridge language assessment series*. Cambridge: Cambridge University Press.
- Bygate, M. 1987. *Speaking*. Oxford: Oxford University Press.
- Chafe, W.L. 1985. "Linguistic differences produced by differences between speaking and writing". In: Olsen, D.R., Torrance, N. and A. Hilyard (eds.). 105-23.
- Chanel, J. 1994. *Vague language*. Cambridge: Cambridge University Press.
- Clark, J.L. and R.T. Clifford. 1988. "The FSI/ILR/ACTFL proficiency scales and testing techniques: Development, current status and needed research". *Studies in Second Language Acquisition* 10. 129-47.
- Clegg, F. 1990. *Simple statistics. A course book for the social sciences*. Cambridge: Cambridge University Press.
- Cole, P. and J. Morgan (eds.). 1975. *Syntax and semantics: vol. 3, Speech Acts*. New York: Academic Press.
- Council of Europe. 2001. *Common European framework of reference for languages: learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Dakowska, M. 2005. *Psycholingwistyczne podstawy dydaktyki języków obcych*. Warszawa: Wydawnictwo Naukowe PWN.
- Dakowska, M. 2007. *Teaching English as a foreign language. A guide for professionals*. Warszawa: Wydawnictwo Naukowe PWN.
- Grice, H.P. 1975. "Logic and conversation". In: Cole, P. and J. Morgan (eds.). 41-58.
- Grove, E. and A. Brown. 2001. "Tasks and criteria in a test of oral communication skills for first-year health science students". *Melbourne Papers in Language Testing* 10 (1). 37-47.
- Hasselgren, A. 1998. *Small words and valid testing*. Unpublished PhD Thesis. University of Bergen, Norway.
- Iwashita, N., Brown, A., McNamara, T. and S. O'Hagan. 2008. "What features of language distinguish levels of learner proficiency? In-depth analysis of task performance in the context of the speaking scale development". *Applied Linguistics* 29.1. 29-49.
- Larsen-Freeman, D. and M.H. Long. 1991. *An introduction to second language acquisition research*. New York: Longman.
- Lazaraton, A. and S. Wagner. 1996. "The revised test of spoken English (TSE): discourse analysis of native and nonnative speaker data". *TOEFL Monograph Series, Ms-7, Dec. 1996*: Educational

-
- Testing Service. 001-18. Obtained on 12 May 2019 from ETS.ORG (<https://www.ets.org/Media/Research/pdf/RM-96-10.pdf>).
- Luoma, S. 2004. *Assessing speaking. Cambridge language assessment series*. Cambridge: Cambridge University Press.
- National Board of Education. 2002. *The framework of the Finnish national certificates*. Helsinki: National Board of Education.
- Nattinger, J. and J. DeCarrico. 1992. *Lexical phrases and language teaching*. Oxford: Oxford University Press.
- Olsen, D.R., Torrance, N. and A. Hilyard (eds.). 1985. *Literacy, language and learning: the nature and consequences of reading and writing*. Cambridge: Cambridge University Press.