

TESTING SPEAKING ABILITY AT ORAL EXAMINATIONS IN DEPARTMENTS OF ENGLISH STUDIES

MAŁGORZATA JEDYNAK
University of Wrocław, Poland

Abstract. The paper discusses the problem of assessing spoken English production starting from the theoretical deliberations on testing the communicative ability. In the empirical part, the author reports on the two approaches to testing the students' oral performance at the final oral examinations in the Departments of English Studies at the University of Wrocław and the Higher Vocational School in Nysa, both located in the south-west of Poland. The impact of preparation time, the examination format, and testing techniques on the students' oral performance were investigated in a qualitative study. Being an examiner the author examined the oral material collected during the two final examinations in the above mentioned Departments. The findings show that the two approaches to oral performance testing have an influence on the students' productions. The paper concludes with some recommendations for the examiners related to the introduction of a more complex testing system in which a versatile student-student and student-teacher format and at least two different testing techniques are used. Additionally, the allocation of a relatively short time for students to prepare at an oral examination is advocated for enhancing natural everyday communication which is a priority in the currently dominant communicative approach to foreign language learning and teaching.

Key words: speaking, testing, oral examinations, communicative ability

INTRODUCTION

As Broadfoot (2005: 123) notices 'a world has been developed into a testing society' where 'standardized testing constitutes the main part of the assessment regime'. According to the author, testing starts in some countries as early as at the age of six or seven and lasts throughout the individual's university career or even professional life. In contemporary Europe, where mobility and language learning are stressed, testing speaking skills in a foreign language is of paramount importance. Furthermore, the intercultural component in oral communication is nowadays recognized as a priority since it allows people in Europe to function adequately and knowledgeably in intercultural contact situations (cf. Sercu, Paran 2010). The very title of the book by Sercu and Paran (2010) *Testing the Untestable in Language Education* suggests that everything we teach in a classroom, though may seem untestable at first glance, can in fact be tested, speaking being not an exception. A number of models of language competence such as Bachman's (1990) model or the Common European Framework of Reference for Languages:

Learning, Teaching, Assessment (2002) have an important effect in the field of testing and the assessment of speaking skill.

Testing speaking skill is necessary for both learners and teachers. In this way, they both acquire more profound knowledge on what has been assimilated in a practical English use course. According to the author of the article, administering an oral proficiency test regularly has a positive backwash effect, as more time can be spent in a course to develop students' speaking skill.

Teachers accept testing and evaluating students as a part of foreign language curriculum at all stages of teaching. Systematic evaluation of learners' progress and final achievement allows teachers to identify the errors students commit, diagnose their weak points and implement some forms of corrective feedback. Additionally, testing students enables teachers to evaluate to what extent the employed teaching materials, techniques or syllabi were effective. Another advantage of oral tests is that they may be motivational to students who speak well but have problems with reading or writing.

The problem is that it is predominantly reading or writing skill which is tested. Testing and evaluating the ability to speak is clearly neglected by many foreign language teachers. This happens for a number of reasons, one of the most significant being apparently the difficulty of its administering as well as evaluating its scores. Teachers testing oral production easily get discouraged by the fact that they are bound to appeal to their subjective judgments while assessing students' oral performance. In Poland, testing and evaluation are regulated by some external criteria established only for middle school or secondary school final examinations. At higher education level, testing and evaluation are more problematic since there are no formal principles for teachers to follow. Departments of English Studies in Poland decide individually on a structure of oral examinations and components of oral performance which are evaluated. For all these reasons, there is a need to have an insight into the nature of testing and evaluation applied at the university level. Consequently, the objective of this article is (1) to present the theoretical framework for testing communicative ability; (2) to sensitize university foreign language teachers and academics to the impact of various factors such as timing, the examination format and testing technique on the final attainment of the students in speaking on the basis of the empirical data collected at the oral examinations in two types of Departments of English Studies; (3) to recommend some changes in the current examination systems in these two departments.

THEORETICAL FRAMEWORK FOR TESTING COMMUNICATIVE ABILITY

Prior to testing the ability to speak, the authors of tests and examiners should fully understand its nature. There is some confusion around the notion *ability to speak* since it is identified with oral proficiency and communicative competence.

One of the definitions of oral proficiency, probably the most traditional one, is provided by Lado (1961) who describes it as 'ability to use in essentially normal communication situations the signalling systems of pronunciation, stress, intonation, grammatical structures and vocabulary of the foreign language at a normal rate of delivery for native speakers of the language' (1961: 241). It seems, however, that viewing oral proficiency in this way is superficial. Nowadays the native speaker norm is a target only for some students, for example the ones studying at Departments of English in Poland who are trained to be translators, teachers or linguists. However, many learners and their teachers depart from RP accent and the British variety of English to the use of English as a lingua franca in Europe with its numerous varieties as Euro Englishes. The use of non-native pronunciation and grammar may give rise to a more effective communication with non-native speakers of English than sticking to the norms imposed by prescriptive grammars. For these reasons, it seems justified to separate oral proficiency from communicative language ability which nowadays should constitute a basis for the construction of tests, especially the oral ones. Researchers attempting to describe the concept of the communicative language ability agree upon the point that it refers to the use of language communicatively and involves the knowledge of or competence in the language as well as the capacity for implementing or using this competence (see for details Hymes, 1972: 269-29, 1973; Canale and Swain, 1980; Savignon, 1983; Widdowson, 1983; Candlin, 1986: 38-57). It is worth mentioning the contribution of Professor of Applied Linguistics Lyle Bachman (1990) to the previously outlined ideas of competence since he characterised the processes by which its various components interact with one another and also with the context in which language use takes place. Before Bachman's works Lado's (1961) and Carroll's (1961) framework for describing the measurement of language proficiency was used. It viewed proficiency as being incorporated in four skills (listening, speaking, reading, and writing) and the components of knowledge related to these skills. Unlike Bachman's framework, Lado's and Carroll's framework did not indicate in any way how these components are related. It was not clear whether the skills are just the manifestations of the knowledge components in different modalities, or whether they were different in some ways.

A much more profound conception of language proficiency has been provided in the works of Savignon (1983) and Kramersch (1986: 366-72) who reformulated communicative competence. The researchers recognised the importance of context beyond the sentence, which includes both the discourse of which separate utterances are part of, and the sociolinguistic situation which to a large extent governs the nature of discourse in form and function. Consequently, in the definition of communication by Savignon (1983: 8) the context is perceived as dynamic rather than static, dependent on 'negotiation of meaning between two or more persons', and 'context specific'. He notices that communication takes place in an infinite variety of situations, and its success in a particular role depends on one's understanding of the context and on prior experience of a similar kind' (Savignon, 1983: 9). The same viewpoint is held by Kramersch (1986: 367).

All these deliberations about the nature of communication bring us to the framework of communicative language ability (CLA) proposed by Bachman (1990). It is CLA that should be nowadays tested at various oral examinations. It basically comprises three components. The first one defined as *language competence* includes a set of specific knowledge components used in communication via language. The second one called *strategic competence* refers to 'the mental capacity for implementing the components of language competence in contextualised communicative language use' (Bachman, 1990: 105). The last CLA components are *psychophysiological mechanisms* related to 'the neurological and psychological processes involved in the actual execution of language such as sound perception and production' (Bachman, 1990: 106-107). These components interact with each other with the language use context and language user's knowledge.

In a discussion on testing the communicative ability, it is worth mentioning Canale and Swain's (1980) theoretical framework in which the fundamental distinction is made between *communicative competence* and *communicative performance*. The authors of the model suggest that 'communicative testing should be devoted not only to what the learner knows about the second language and about how to use it (competence) but also to what extent the learner is able to actually demonstrate this knowledge in a meaningful communicative situation (performance)' (Canale and Swain, 1980: 34).

As it has been already mentioned above, one of the characteristic features of recent frameworks of communicative competence is the recognition of language use as a dynamic process, which involves the evaluation of relevant information in the context and the negotiation of meaning on the part of the language user. This view of communication corresponds to another element of communicative ability called *strategic competence* which is defined by Tarone (1981: 288) as 'the mutual attempt by two interlocutors to agree on a meaning in situations where the requisite meaning structures do not seem to be shared'. In this interactional definition of strategic competence, she seems to include both linguistic and sociolinguistic rule structures in her notion of meaning structure. For Faerch and Kasper (1984: 45), however, her definition is too narrow since it only applies to a situation when the negotiation of meaning is a joint effort between two interlocutors. It is much more frequent to negotiate the meaning without a participation of another interlocutor which is visible in the communicative language use in writing or reading. Tarone does admit that there is also a place for production strategies which are distinct from the language user's language competence, yet, unlike communication strategies, they lack the interactional focus on the negotiation of meaning.

Undoubtedly, strategic competence plays an important role in oral performance testing since it enables us to make the most effective use of the available abilities in performing a given task and consequently achieving better results on an oral test. The students with the same or similar level of language competence may be perceived differently by examiners, namely one of them may be considered a more effective language user due to his or her ability to use

various strategies. For example, in an oral examination students are expected to describe a picture using some specific vocabulary. They may complete the task successfully without employing this vocabulary thanks to adopting various strategies such as ignoring all the elements of a stimulus and switching to a topic related to some aspect of a stimulus.

Apart from language competence and strategic competence one also needs to mention physiological mechanisms which are used in the execution phase of language use. As Faerch and Kasper (1983) claim they refer to the neurological and physiological processes. In receptive language use, 'auditory and visual skills are used' whereas in productive use 'the neuromuscular skills are employed' (Faerch and Kasper, 1983: 11). For example, a student in an oral test makes use of his/her linguistic competence to formulate correct utterances in a foreign language. She/he uses the visual skills to obtain the non-linguistic information from the picture stimulus, auditory skills to comprehend the examiner's instructions, and articulatory skills to provide correct stress, intonation and pronunciation of words (cf. Bachman, 1990).

MAIN PROBLEMS WITH ORAL TESTING

There are many advantages of administering an oral test which have been outlined in the introduction to the article. Still, many teachers withdraw from testing the speaking ability substituting it with written tests. It is due to the fact that testing speaking skill is difficult in many respects. The first problem appears at the stage of *designing an oral test* that should be *valid, reliable, scorable, practical* (economical) and *administrable* (Lado, 1961). The three qualities that constitute the so called *sine qua non*, without which a test is not worth spending any time over it, are *validity, reliability, and practicality*. Let us have some insight into these three concepts which seem to be crucial for testing any skills, including speaking skills.

The validity of a test is the extent to which it measures what it is supposed to measure and nothing else (Heaton, 1988). In order to be able to regard the test as valid for our purposes, two questions have to be answered, namely: (1) *What* precisely does the test measure? and (2) *How well* does it measure? If there is enough evidence of a high correlation between the test scores and the student's actual ability in the skill tested, one may assume that the given test is valid.

There may be a problem with validity in an oral test if it measures not only the speaking ability but also external knowledge or other skills. For example, a student has limited vocabulary and extensive knowledge of a topic but still his/her overall oral performance is highly evaluated.

With regard to reliability, it refers to the stability of test scores (Harris, 1969). An oral test will be reliable if it produces essentially the same results consistently on different occasions when the conditions of the test remain the same and when there has been no instruction intervening. As Lado (1961) asserts, a reliable test will yield dependable scores in the sense that they will not fluctuate much if we

were to give the same test to the same student in another time, or if it was to be assessed by another competent scorer.

A highly reliable and valid instrument may not be at the same time practical or usable. Its practicality is related to economical factors (the time spent on its preparation), the ease of administration and scoring, and the ease of interpretation. Whatever scoring format a tester uses, he or she should remember to develop a scoring key precisely in order to avoid the approach of having a general idea what he/ she is looking for (Krypsin and Feldhusen, 1974). One way to derive the best scoring key is to take the test oneself, or have another tester to take it.

There are various types of scoring which may be implemented in an oral examination. Drawing up a precise banding system or rating scale seems advisable. This scoring method is commonly known as *analytic*. Clear and accurate description of the various characteristics of performance at each level makes it possible for a teacher to identify precisely what he or she expects for each band and assign the most appropriate grade to a student. In the table below, there is a sample of analytic scoring system.

Table 1 Example of analytic rating scale related to pronunciation

11-12 points	excellent	<ul style="list-style-type: none"> • native-like English pronunciation • no problems with understanding • no L1 interference of prosodic features and individual sounds • word-stress/rhythm always appropriate • individual sounds are unambiguous and well articulated
10-8 points	good	<ul style="list-style-type: none"> • almost native-like English pronunciation • at some points problems with understanding • noticeable L1 interference of prosodic features and individual sounds • word-stress/rhythm occasionally inappropriate • individual sounds may be sometimes ambiguous and not well articulated
7-5 points	pass	<ul style="list-style-type: none"> • far from native-like English pronunciation • frequent problems with understanding • many instances of L1 interference of prosodic features and individual sounds • word-stress/rhythm frequently inappropriate • individual sounds are frequently ambiguous and not well articulated
4-0 points	poor	<ul style="list-style-type: none"> • very far from native-like English pronunciation • gross errors and a very heavy L1 pronunciation/rhythm make • understanding impossible • abundant instances of L1 interference of prosodic features and individual sounds • word-stress/rhythm always inappropriate • individual sounds are always ambiguous and not well articulated

As one may notice, the analytic method ensures consistency of marking and avoidance of subjective considerations. Designing a rating scale may be time consuming, especially when it must be restructured in order to match students' proficiency level. The analytic scoring system may also entail a separation of various categories related to students' performance such as *Pronunciation, Fluency, Use of Vocabulary, Use of Grammar, Intelligibility, Repair Skills, or Task Completion*. Such a profile is suggested for example by Harmer (2001). Apart from the elements mentioned above, one may also wish to test other items which are tested by the English Speaking Boards such as *Voice* (its range, flexibility, tone accent), *Clarity, Vocabulary and Composition, Audience Control and Relationship, Imagination, Sincerity, and Spontaneity*. All of them may be assessed on a five point scale.

It seems that one of the greatest problems related to evaluating oral tests is the difficulty in defining the nature of the speaking skill itself and a lack of consensus among teachers as to what constitutes speaking ability. Examiners agree upon a division into such components as grammar, vocabulary and pronunciation. However, such aspects as fluency or appropriateness of expressions are equally important, although not frequently applied in oral production evaluation. Madsen (1983) also mentions some other essential elements of speaking ability which should be analyzed, namely *listening comprehension, correct tone* (e.g. sadness or anger), *reasoning ability, or initiative in asking for clarification*. As one may see, there are numerous ingredients of speaking and many of them are very difficult to identify. Even if one identifies them, it is still difficult to test them objectively. For example, fluency may be understood differently by different testers. For one rater, it is associated with quickness of response, whereas for another rater it is more related to the amount of information conveyed within one minute. Similarly, some doubts may appear while assessing the pronunciation component. A question arises whether a native-speaker standard should be looked for in students' performance or rather non-native standard which is more frequent as an effect of teachers' use of non-native pronunciation. For these reasons, it seems justified to standardize the descriptors referring to individual markers.

As one can see, there are many problematic issues related to evaluation of the analytic method. However, this type of scoring various features of the performance seems to be worth implementing because it may constitute an effective feedback for a student who is able to see how his or her particular skills have been graded (Heaton, 1988).

The *impression method* as opposed to the analytic method or guided judgment according to the CEFR, entails one or more markers awarding a single grade based on the total impression of students' performance as a whole. Since the method is highly subjective its application is disputable. However, even with the analytic method it is hard to say that it ensures complete objectivity. It is possible for two testers assign a different number of points for the same oral performance. It is generally acknowledged that the speaking ability is hardly possible to be tested in a systematic and objective way. This point of view is also shared by Lado (1961) and Komorowska (2002). One of the major problems is

inter-rater reliability which means that we can hardly count on reliable scoring unless we list specifically what the examiner is to listen for in a student's response. This is mainly because different examiners pay attention to different aspects of a student's oral performance. For example, one examiner attaches great importance to pronunciation while another to grammar. Furthermore, some testers are stricter in applying certain evaluating criteria while others are more lenient. For this reason, standardization of raters seems necessary.

Once a tester decides on an evaluation method and components to be tested he/ she needs to design a test bearing in mind the three basic features abovementioned, namely, validity, reliability, and practicality. Successful oral tests should provide a student with sufficient clues to generate certain responses containing language areas we intend to test. However, it is not as easy as it seems. The main objective of an oral test is to elicit a quick response from a student without actual saying what we wish to check. An interaction in such a test should be natural and resemble real life communication. Yet, a student may produce utterances which we did not intend to elicit. Thus, a tester needs to develop tests using such techniques and such test components which will check, in a relatively short time, what he or she wishes to test. There are numerous testing techniques which may be applied in an oral examination. One of them is the use of pictures by means of which a tester may elicit various utterances. However, a visual stimulus may evoke various associations in students and consequently it may happen that only brighter ones will provide a tester with utterances intended to be elicited. Thus, the selection of an appropriate picture seems crucial if the technique is to be fairly objective. Some testers prefer a language stimulus finding it more reliable than a visual one. However, a tester faces another problem here, namely, the language of a stimulus. For this purpose some examiners use exclusively the target language during an oral examination while others, in order to avoid students' anxiety, introduce the language stimulus in the students' native language.

Finally, there is a problem with oral test administration. Timing seems a crucial aspect here. Whether a student is allowed to have some time for preparation or not depends on examination procedures set by individual departments. If a tester decides on an examination format in which one student prepares his speech outline while another one is already involved in oral production, the latter may disturb the former. It seems that an ideal situation is when a test is administered individually for each student. There are various oral examination interaction patterns such as student-student, student-teacher, and finally students-teacher, all of which having obvious advantages and drawbacks.

The tester as a person is as important for the interaction pattern. Being an examiner (tester) and an interlocutor involved in a conversation with a student is a difficult task. Undoubtedly, such a situation negatively affects the evaluation process. Thus, it seems advisable to separate the interlocutor's function from the *assessor's* function. One of the solutions to the problem is having at least two testers on an examination board. Another one is the introduction of a recording system, which allows for post-examination evaluation of students' productions.

As it can be seen, there are numerous problems with designing oral tests, the evaluation of students' oral performance, and the administration of such tests. The most comprehensive work which provides extensive information on foreign language testing and assessment is the Common European Framework of Reference (CEF) (2002).

It provides a comprehensive description of what learners should learn to use a language in communicative situations and what knowledge and skills learners should develop to be able to act effectively. The description also takes into account the cultural context in which the language is set. The CEF also defines common reference levels (A1-C2) which allow learners to measure their language progress at subsequent stages of learning.

RESEARCH QUESTIONS AND DESIGN OF THE STUDY

The students of Departments of English Studies are a very specific group of learners. Since they are trained to become specialists in the English language, they are expected to represent almost native-like competence and performance. A graduate of the English philology programme should be at an advanced level of the target language.

As it has been already mentioned in the introduction, in Poland there are no formal procedures imposed by any supervising institutions for testing university, college or higher vocational school students aiming at becoming an English teacher, a translator of English, or an expert in various domains of the English language. The departments of English need to establish their own guidelines regulating the length of an examination, the examination format, and testing techniques. For the last decade, the author of the article has been involved as a tester in numerous oral examinations with university students. Hence, the primary aim of the study was to make a comparative analysis of two ways of testing the speaking ability practised at the Department of English Studies at the University of Wrocław and the Department of English at the Higher Vocational School in Nysa, both of which train foreign language teachers and translators. The study is qualitative in nature. The researcher did not entertain any hypotheses at the outset of a study, but she was just taking notes and making recordings of the students' productions during the observation. Since there was no a priori hypothesis which could be tested, the researcher felt free to discover any potential factors influencing the students' oral performance. It was during the study when the researcher realized that various factors contribute to different oral performances. Hence, the study may be referred as hypothesis-generating. The main hypothesis was that there is some relationship between a set of variables such as timing, the examination format, testing technique and the final attainment of the students in speaking. The concept of the final attainment is difficult to operationalise since the researcher relied more on her subjective non-participant observation of the oral performance and the notes taken during the examinations than on the final mark assigned by the examination board to a particular student.

The following variables were under investigation: preparation time, the examination format, and testing techniques, all of which were different in the two Departments of English Studies. The oral examinations gave rise to the following research questions formulated by the author:

- 1) What is the relationship between the preparation time and the students' final attainment?
- 2) What is the relationship between an examination format and the students' final attainment?
- 3) What is the relationship between a testing technique and the students' final attainment?

Although the study, unlike typical non-participation research, was a short-term one it allowed for collecting much data on the students' speaking ability. The subjects of the study were 26 undergraduate students of English: 12 from Higher Vocational School in Nysa (the HVS students) and 14 from the University of Wrocław students (the WU students). They were all enrolled in the third year of studies. The former represented B2/C1 level, while the latter C1/C2 level in accordance with the Common European Framework of Reference for Languages. Their proficiency levels were established on the basis of the annual final proficiency tests. All the students attended regular practical classes on communication and language use with a division into the following components: practical English grammar, English phonology/phonetics, and finally conversations in English. There were, however, some differences in the intensity of the provided courses (the HVS students had by 50 percent less phonology/phonetics) and in the teacher's first language (the WU students had a native speaker teacher). In both departments, the students were provided with clear and detailed evaluation criteria so they could know what aspects of their oral production would be assessed. The students were also informed in advance on a testing technique used at the oral examination.

There were some similarities between the oral examinations practised at the two departments. In the Higher Vocational School and at the University of Wrocław there were always three testers at the examination boards. In both cases, only one tester was involved in the interaction with the student, whereas the others followed the student's speech and evaluated his/ her oral performance by means of assessment grids. At the same time the testers took notes of all the errors they noticed in oral productions to discuss them after the examination with other testers and to assign relevant ratings from the scale.

There were also some differences in the way the oral examinations were conducted. Although the analytic method was applied by the testers in the two departments, various components were considered while evaluating the oral performance. For the HVS students, the testers prepared just three categories for assessment such as pronunciation, grammar, and vocabulary, while the WU students were rated for seven components such as grammar, vocabulary, fluency, cohesion, pronunciation, comprehension, and interaction.

The examiners at both universities relied on a five point Likert's scale. The final number of points was converted into the grades (2 being the lowest and 5 being the highest mark to obtain).

Different time was also allotted for the performance. The HVS students had no time for preparation but almost unlimited time for speaking, while in the case of the WU students timing was controlled and restricted to 20 minutes (10 minutes for preparation and 10 minutes for oral performance).

The examination format also differed. In the Higher Vocational School, one-to-one teacher-student format was practiced, whereas at the University of Wrocław a versatile student-student and students-teacher format was introduced. As far as testing techniques were concerned, the HVS examiners applied mainly conversation on the basis of the stimuli and role-plays, whereas the WU testers relied only on role-plays and a presentation of a topic. In a typical role-play, the examinee was requested to take on a certain role and act out a given real-life situation with the interlocutor. The role-play was usually initiated by the examiners. The role-play instructions may involve the technique of asking questions (e.g. asking the interlocutor about his/her plans), providing explanations or descriptions (e.g. telling what he/she is going to see in London), or making negotiations (e.g. encouraging the interlocutor to do something, suggesting some ways of doing something). Some role-plays resemble more monologues since the examinee needs to present extensively his/her point of view. The language of instruction was English; however in the Higher Vocational School the examiners occasionally switched to the students' native language. Table 2 below displays the similarities and differences between the two ways of conducting oral examinations at the departments.

Table 2 Comparative analysis of oral examinations in the two English Departments

	HVS English department	WU English department
Number of testers	three	three
Components under evaluation	pronunciation, grammar, and vocabulary	grammar, vocabulary, fluency, cohesion, pronunciation, comprehension, and interactions
Timing	no time for preparation; almost unlimited time for speaking	20 minutes (at least 10 minutes for preparation and 10 minutes for oral performance)
Exam format	one-to-one teacher-student format	a versatile student-student and students-teacher format
Oral performance testing technique	conversation on the basis of the stimuli and role-plays	role-plays presentation of a topic
Language of instruction	the target language	the target language and native language

FINDINGS

During the study, the researcher noticed that the way the oral tests were administered in the two settings had both positive and negative effects on the students' final attainment in the oral performance.

With regard to timing, it seems that the time allotted for preparation, as it was the case with the WU students, gave rise to more accurate and well-structured productions. The WU students committed fewer grammar mistakes. Their average score for grammar part was 4.5 points, while the HVS students score was only 3.5 points. The WU students also used richer vocabulary surpassing the HVS counterparts by 0.5 point. The former also seemed to organize more logically their speech than their HVS counterparts. This observation is not, however, supported with any reliable data since a component related to the organization of speech was not under evaluation. The researcher also noticed that the WU students' performance lacked spontaneity and naturalness. Once again these two elements were not evaluated by means of any scale. Yet, the recordings allow tracing spontaneity and naturalness of students' performance. Some students followed the outline of speech that they had prepared or even read out the whole sentences they had managed to write at the preparatory stage as exemplified in the recording extract below:

Student: It is generally acknowledged that (a pause) German or French are less widely taught languages. Therefore, it seems justified to pass a new law that would progressively introduce these languages in the foreign language curricula in a primary school replacing the dominant English. Undoubtedly, this scenario has a chance of implementation provided the lobby for this project ...

Examiner: But do you think it is possible?

The HVS students in turn, without prior preparation, used more informal language with idiomatic expressions characteristic of natural spontaneous everyday communication. They also reacted more naturally in various situations when the flow of discourse could be hardly predictable as exemplified in the speech sample below:

Student: This is a fact, upon my word.

Examiner (with disbelief): There's not the least doubt of it.

Student: I told you the naked truth... it was like...you know a kind of ... shock. I guess not just for me. But for my driving instructor as well! He was like 'Failed! Failed!' screaming at me. Can you believe I failed at the same instructor six times?

What may seem to be negative at first glance is the allocation of unlimited time for oral performance, yet all the HVS students found 10 minutes as

being sufficient for the oral presentation, which they communicated after the examination. From the psychological viewpoint, however, this awareness of having unlimited time may act for students as both a facilitator and an inhibitor.

It seems to the researcher that timing of the examination preparation should be short enough to allow a student just to familiarize with the instruction, or contents of the examination set, and to prepare a draft with bullet points reflecting the main issues to be discussed. From the researcher's experience as an examiner and the students' accounts it seems that an optimal time for efficient oral examination preparation is five minutes. It was supported by the results obtained during the two examinations described in the study. The students who prepared for their oral performance within five minutes committed more grammar mistakes than the students who were allotted more time for preparation (the average score for grammar was 3.5 and 4.0 points respectively). However, the former produced the speech which was more natural and resembled everyday life situations. The students who opted for a longer preparation time produced speech which was artificial and more formal.

The issue of time preparation should be considered by examiners. When an examination task does not entirely resemble situations encountered in everyday life, it seems advisable to allocate more time for preparation, for example, when a student needs to discuss the future of the English language or prevention of environmental problems. Without a prior preparation the examiner may expect to hear from the student 'I don't know' answer. A remedy to the problem may be discussing all the topics with the students prior to the oral examination so they know what to say and they only decide on the form of what they intend to convey.

As to the examination format, it was noticed by the researcher that the teacher-student format resulted in worse oral performance than the student-student and students-teacher format implemented at the University of Wrocław. The student-examiner format did not seem to meet entirely a condition of a good test, namely its reliability. An interaction with the examiner was an obvious emotional burden for all the students who reacted either with inhibition or nervousness. From the researcher's position, these were the poor students who were prone to experience high debilitating anxiety. Frequently, in the situations when being exposed to a more complex lexical item or grammar structure produced by the examiner, these students easily gave up without making any attempts to find out the meaning, to ask for clarification or repetition as exemplified in this speech sample recording:

Examiner: Do you think he commutes every day?

Student: I don't know this word. Sorry.

Examiner: Oh, we had this word so many times in our integral skills classes. Do you remember?

Students: No.

Examiner: Never mind. Please continue.

Student: Well, that's it. I have nothing more to say.

Examiner: Are you sure?

Student: Yes, I have a completely blank mind.

Since the student-examiner format leaves the undergraduate at an obvious linguistic disadvantage, it seems that a versatile format introduced at the University of Wrocław is more advantageous. The WU students displayed higher motivation and self-assurance at the oral performance stage. Comparing the mean ratings obtained by the two groups of students for their oral performance, it is evident that the versatile format results in fewer grammar errors (M for WU students=4.5 and M for HVS students= 3.7) and fewer pronunciation errors (M for WU students=4.2 and M for HVS students= 3.5). The versatile examination format also seemed to influence speech fluency; however, it cannot be supported with any objectively measured data since this component was not considered in the assessment grid at the Wrocław University.

Undoubtedly, the student's speaking ability depends to a great extent on a student with whom he or she is paired. From the researcher's observations and the analysis of the speech sample recordings, it is clear that the best pairing is when two students represent approximately the same level of language proficiency. It is disputable whether the personality factor should also be taken into consideration while matching students for oral examinations. It seems that if the objective of an oral examination is to allow a student to experience a real life communication in which speakers with various personality traits are involved, pairing students of a similar psychological profile is not recommended. In the study, the WU students were paired at random by a computer, which had an impact on their oral performance. Two students both of whom represented leadership and extrovert skills found the task too overwhelming. Despite having some time to prepare a draft of a dialogue in which they were supposed to persuade an interlocutor of one's arguments, they were only able to produce a very heated and unproductive discussion as exemplified in this interaction recording transcript:

Student A: So you are for the capital punishment and I am against.

Student B: What?! Adam, we agreed on something different. Miss (to the examiner) we drew lots and he was to be for it not me [...]

Student A: But I will not say something that is contrary to my personal beliefs!

Student B: Neither will I.

Examiner: Adam you are 'for'.

Students A: (in a native language) Ok, but Ladies first, let her start.

Nevertheless, sometimes pairing students both of whom are leaders may give rise to a very interesting and fruitful discussion. Currently, the WU students are allowed to choose an interlocutor for the oral examination. What can be observed

now is a tendency for the students to pair with the students of a similar level of proficiency and a similar personality profile.

It is difficult to evaluate which of the testing techniques are most productive since their effectiveness depends on a number of factors such as students' personality, preferences, prior experience with the presentation technique. From the researcher's observation, the greatest objection may be directed at a topic presentation technique which did not involve any interaction with another interlocutor. While presenting a topic, the WU students utilized conventional ideas and memorized chunks of vocabulary. Their speech did not resemble communication one may encounter in everyday life situations. However, the productions of the WU students contained fewer grammar mistakes in comparison to the HVS students.

A role play was a technique applied with both target groups. The author believes that it is effective for assessing oral performance. Its great advantage is high realism and real communication provocation. The WU and HVS students seemed to have positive attitudes towards being involved in the fictitious roles and situations which were highly realistic and encouraged real-life communication. This observation was supported by the very students who expressed their preferences for such tasks after the examination. The students had total freedom in the choice of vocabulary and grammar structures; however, they made more grammar errors than in the case of presenting a topic (the average score for grammar component was 3.5 and 4.5 respectively). They felt emotionally comfortable, especially if the oral examination was conducted in the student-student format. Furthermore, the technique is appreciated by the examiners since it allows eliciting functional language such as agreeing, persuading, complaining etc. Conversation on the basis of a stimulus is also a very effective technique that was introduced at an oral examination only with the HVS students after the role play testing technique. A discussion on the basis of a visual stimulus was apparently more effective than a role-play. First of all, it resulted in more ideas since the stimulus tapped some associations related to the topic. It was reflected in the vocabulary used by the students (the average rating for vocabulary was 4.75 in comparison to 4.0 in the presentation of a topic technique). Consequently, it initiated a more interesting discussion with the interlocutor. For example, a stimulus in a form of a flashcard with a spider's web was associated by one student with Katie Melua's song *Spider's Web* and gave rise to his discussion of the issue of tolerance in the world. However, similarly to a role-play technique, a discussion on the basis of a visual stimulus resulted in more grammar errors (the average rating was 3.5).

It is generally acknowledged that testing is inseparably connected with the concepts of reliability and validity, both of which have already been discussed. As concerns the oral performance testing in the present study an attempt was made to maximize testing reliability by minimizing the effects of a few factors which can be the source of measurement error. For example, both the WU and the HVS students could opt for a time of testing most convenient to them and had been provided with the precise testing procedures. As to the procedure

of assigning points for various components under evaluation, it might be questioned for its subjectivity. It is manifested by the fact that one student's oral performance may be evaluated in different ways by various testers. However, in order to ensure the reliability of testing the examiners at the two universities had detailed descriptions of students' oral performance assigned to particular points on a rating scale. For example, assigning 5 points for a pronunciation component meant that a student's pronunciation was always at a native-like level, both at the level of segment and suprasegment production, while assigning 4 points meant that a student's pronunciation was occasionally native-like. Still, divergent ratings might have appeared due to the examiners' various interpretations of the words 'occasionally', 'always' or 'never'. Another problematic situation observed in the oral examinations referred to the cases when a student had a limited vocabulary but very extensive knowledge of a topic. Again, the examiners had to stick to the precise descriptions assigned to the ratings and evaluated only the component under evaluation. For example, such a student would be assigned a rating of 3 points as his/her vocabulary was not sophisticated and did not correspond to the C1 proficiency level.

CONCLUSIONS, IMPLICATIONS AND DIRECTIONS FOR FUTURE RESEARCH

In the study, the researcher attempted to show that there is an impact of preparation time, the examination format and testing techniques on the students' final attainment in oral performance. Allocation of preparation time resulted in fewer grammar errors, richer vocabulary and logical organization of speech. Yet, it also gave rise to unnatural and artificial speech. As to the examination format, the versatile format proved to be more advantageous resulting in fewer grammar and pronunciation errors. Pairing students may also be a factor influencing oral performance. The issue of testing anxiety in the context of various examination formats was not formally investigated in the study, but it seems justified to give more insight into this problem in the future. The relationship between the testing techniques and attainment in oral performance was most difficult to show. Undoubtedly, there is a need to investigate this relationship more thoroughly in the future. Presentation of a topic, though not embedded in everyday life communication, resulted in fewer grammar errors. Role-play, in turn, enhanced students' freedom in the choice of vocabulary and grammar structures but resulted in numerous grammar errors. A discussion on the basis of a visual stimulus resulted in the richness of vocabulary but worse grammar.

As it can be noticed, testing the speaking skill is difficult in many respects. Taking into consideration the complexity and vagueness of the nature of speaking ability, it is hard to establish what expectations of language performance we should have when testing. For different examiners, various aspects of oral performance are important. Additionally, there are very many factors influencing the speaking ability such as proficiency level, speaking apprehension, or

extroversion/introversion dichotomy. In the study, the author as an examiner had an insight into just three variables, namely timing, the examination format, and testing technique.

On the basis of the study, the author is able to say that the oral examinations administered at the two Departments of English Studies have both advantages and drawbacks. The WU students would undoubtedly benefit if the technique of conversation on the basis of a stimulus was introduced. The HVS students, in turn, would benefit if a more versatile examination format was implemented.

Taking into consideration the study findings, the author feels tempted to propose an improved system of oral performance testing at a university level. First of all, it is advised to introduce a complex testing system in which a student is tested by means of at least two different techniques such as a role-play enriched with a free topic discussion on the basis of a stimulus such as a picture or a newspaper extract. The former would test a student's communicative ability in informal situations, while the latter would check it in formal situations. As to timing, the author advocates abandoning the idea of allocating some time for preparing a role play since natural and spontaneous everyday life communication is nowadays a priority. However, a free discussion on a given topic should always be preceded by maximum five minute preparation time to allow a student to draft his/her talk. Allocating more time for preparation is not advisable because as it is apparent at the oral examinations at the Wroclaw University, students may be tempted to write whole sentences and to read them out later for examiners. Finally, the versatile examination format (student-student and student-examiner) should be implemented to ensure that a student will fully use his/her communicative potential. As the study findings showed, the student-student format may lead to unproductive and meaningless oral performance which is difficult to evaluate and the help of an examiner is then necessary. In the student-examiner format, in turn, inhibition and reluctance to speak might appear that can be avoided if an examiner allows a student to talk to another student.

Finally, it should be considered whether too great a concentration in oral testing on phonology, vocabulary or grammar does not have a detrimental effect on the communicative teaching of the language. There is also insufficient knowledge about the weighting which should be given to specific language elements at various stages of learning a language.

Nowadays not only communicative ability but also intercultural communicative ability should be tested (Sercu and Paran, 2010). In the case of the WU and the HVS students, this element of communication was not taken into consideration. It seems advisable to introduce the intercultural component into assessment in the future.

With regard to future research, it is tempting to focus on other aspects related to the speaking ability as for example an impact of the native speaker versus non-native speaker examiner on students' performance, or examining attitudes of examiners to various evaluation grid components.

REFERENCES

- Bachman, L. F. (1990) *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. and Palmer, A. S. (1982) The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16 (4): 449-65.
- Bachman, L. F. and Cohen, A. D.. (1998) *Interfaces Between Second Language Acquisition and Language Testing Research*. Cambridge: Cambridge University Press.
- Broadfoot, P. (2005). Dark alleys and blind bends. Testing the language of learning. *Language Testing*, 22 (2): 123-141.
- Canale, M. and M. Swain.(1980) Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1: 1-47.
- Canale, M. (1983) On some dimensions of language proficiency. In J. W. Oller, Jr (ed.) *Issues in Language Testing Research* (pp. 333-42). Rowley, Mass.: Newbury House.
- Candlin, C. N. (1986). Explaining communicative competence limits of testability? In C. W. Stansfield (ed.) *Toward Communicative Competence Testing: Proceedings of the Second TOEFL Invitational Conference* (pp. 38-57). Princeton, NJ: Educational Testing Service.
- Carroll, J. B. (1961) Fundamental considerations in testing for English language proficiency of foreign students. In *Testing the English Proficiency of Foreign Students* (pp. 30-40). Washington, DC. Center for Applied Linguistics.
- Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. (2002) Council of Europe: Cambridge University Press.
- Faerch, C. and Kasper, G. (1984) Two ways of defining communication strategies. *Language Learning* 34, (1): 45-63.
- Harmer, J. (2001) *Writing English Language Tests*. Essex: Pearson Education Limited.
- Harris, D. P. (1969) *Testing English as a Second Language*. New York: Mc Graw-Hill.
- Heaton, J. B. (1988) *Writing English Language Tests*. Harlow, Essex: Longman.
- Hymes, D. H. (1972) On communicative competence. In J.B Pride and J. Holmes (ed.) *Sociolinguistics* (pp. 269-29). Harmondsworth: Penguin.
- Hymes, D. H. (1973) Toward linguistic competence. *Texas Working Papers in Sociolinguistics*, Working Paper, 16. Austin, Texas: Center for Intercultural Studies in Communication, and Department of Anthropology, University of Texas. Available from <http://azargoshnasp.net/languages/talyshi/TALYSHSUMGAYIT.pdf> [Accessed January 3, 2013].
- Hymes, D. H. (1982) *Toward Linguistic Competence*. Philadelphia, Pa. Graduate School of Education, University of Pennsylvania.
- Komorowska, H. (2002) *Sprawdzanie umiejętności w nauce języka obcego – kontrola – ocena – testowanie*. Warszawa: Fraszka Edukacyjna.
- Kramsch, C. (1986) From language proficiency to interactional competence. *The Modern Language Journal*, 70 (4): 366-372.
- Krypsin, W. J. and Feldhusen, J. F. (1974) *Developing Classroom Tests: a Guide for Writing and Evaluating Test Items*. Mineapolis, Minn: Burgess.
- Lado, R. (1961) *Language Testing*. New York: McGraw-Hill.
- Madsen, H. S. (1983) *Techniques in Testing*. Oxford: Oxford University Press.
- Munby, J. (1978) *Communicative Syllabus Design*. Cambridge: Cambridge University Press.
- Savignon, S. J. (1983). *Communicative Competence: Theory and Classroom Practice*. Reading, Mass. Addison-Wesley.

- Sercu, L. and Paran, A. (2010). *Testing the Untestable in Language Education*. Clevedon: Multilingual Matters.
- Tarone, E. E. (1981) Some thoughts on the notion of communicative strategy. *TESOL Quarterly* 15: 285-295.
- Widdowson, H. G. (1983) *Learning Purpose and Language Use*. Oxford: Oxford University Press.

Malgorzata Jedynak works as an academic in English Studies Department, University of Wrocław in Poland. In 2004 she obtained her PhD degree in applied linguistics with the thesis entitled *Critical Period Revisited: the Impact of Age on the Final Attainment in L2 Pronunciation*. Her research papers are related to the acquisition of L2 aspects and skills by foreign language learners. The recent research she is involved in concerns different aspects related to the acquisition of L1 and L2 by the visually impaired learners: the choice and use of learning strategies, the role of autonomy in the learning process, teaching techniques implemented in a L2 classroom, the issue of giftedness in a foreign language. Email: gosiajedynak@poczta.onet.pl