

**UNIVERSITY OF PARDUBICE
FACULTY OF ARTS AND PHILOSOPHY
DEPARTMENT OF ENGLISH AND AMERICAN STUDIES**

TESTING SPEAKING

THESIS

**AUTHOR: Alena Kombercová
SUPERVISOR: Mgr. Pavel Brebera**

2006

**UNIVERZITA PARDUBICE
FAKULTA FILOZOFICKÁ
KATEDRA ANGLISTIKY A AMERIKANISTIKY**

**OVĚŘOVÁNÍ DOVEDNOSTI
MLUVENÍ**

DIPLOMOVÁ PRÁCE

**AUTOR PRÁCE: Alena Kombercová
VEDOUCÍ PRÁCE: Mgr. Pavel Brebera**

2006

Acknowledgment:

I am grateful to Mgr. Pavel Brebera for his guidance and consultations. Also, I would like to thank to my friends and to both my families for their patience and support.

Abstract:

My diploma paper is focused on testing the speaking skill and the aim of the paper is to evaluate the selected testing techniques from the point of the teacher and the learner.

From the theoretical part we acquire the background of the communicative competence and the speaking skill. We concentrate on how to test, and what criteria we consider when testing speaking. As known, testing speaking is a broad topic and that is why I devoted mainly to describe, analyse and evaluate testing techniques. Testing techniques are selected according to the age and to the language competence of learners. In the final, we learn about the evaluation of learners' performances and moreover, about the evaluative instruments that were used.

In the research part, I tried to apply the acquired theory into the practice and evaluate testing techniques in accordance to the communicative competence that is related to the results of the rating scale. Then, in regard to the evaluative instruments, such as: preparation project and test specification, I mention the process of preparations and structures of each testing technique. Afterwards, testing techniques are analyzed by questionnaires from the point of a learner. The numerical results and data are specified and evaluated.

In conclusion, we find out an overall view of the evaluation of testing techniques which further enhances to testing speaking. Testing speaking is one of the main skills in the foreign language.

Abstrakt:

Moje diplomová práce se zaměřuje na ověřování dovednosti mluvení a jejím cílem je vyhodnotit vybrané testovací techniky jak z pohledu učitele, tak i žáka.

Z teoretické části získáme přehled komunikativní kompetence a dovednosti mluvení. Soustředíme se, jak testovat a jaká kritéria zvážit, když budeme dovednost mluvení ověřovat. Ověřovat dovednost mluvení je široké téma, a proto se hlavně věnuji popisu, analýze a ověřování testovacích technik. Testovací techniky se zřetelem na věk a jazykovou způsobilost žáků. V závěru se dozvíme, jak ověřovat výkony žáků a jak vybírat a používat hodnotící nástroje.

Ve výzkumné části se snažím aplikovat získanou teorii v praxi a zhodnotit testovací techniky vzhledem ke komunikativní kompetenci, která se vztahuje k výsledkům hodnotící škály. Na základě použitých hodnotících nástrojů, jako je např. přípravný projekt a specifikace testu, uvádím průběh přípravy a strukturu každé testovací techniky. Nakonec jsou testovací techniky rozebrány z pohledu žáků pomocí dotazníků. Numerické výsledky a data jsou specifikovány a ohodnoceny.

V závěru získáme celkový pohled na hodnocení testovacích technik, což by mělo napomoci samotnému ověřování dovednosti mluvení, která je jednou z hlavních dovedností při zvládnutí cizího jazyka.

Content:

Theory

1. Introduction	1
2. Speaking skill.....	2
2.1. The definitions of speaking skill.....	2
2.2. Communicative competence.....	2
2.3. Littlewood's classification of communicative activities.....	5
2.4. Conclusion.....	6
3. Testing.....	7
3.1. The definition of a test.....	7
3.2. Types of test.....	8
3.3. Methods of testing.....	10
3.3.1. Direct x Indirect testing.....	10
3.3.2. Norm-referenced x Criterion referenced testing.....	10
3.3.3. Discrete-point x Integrative testing.....	10
3.3.4. Objective x Subjective testing.....	11
3.4. Criteria to the evaluation of tests.....	11
3.4.1. Validity.....	11
3.4.2. Reliability.....	13
3.4.3. Practicality.....	14
3.4.4. Authenticity.....	14
3.5. Preparation.....	15
3.6. Test specification.....	16
4. Testing speaking.....	17
4.1. Why test speaking?.....	17
4.2. Objectives of testing speaking.....	17
4.2.1. Criteria to choose testing techniques.....	18
4.2.2. Testing in pairs x individually x group work.....	19
4.3. Testing techniques.....	20
4.3.1. Information gap techniques.....	21
4.3.2. Oral interview.....	23
4.4. Evaluation of testing techniques.....	24
5. Evaluation of learner's performance testing.....	25
5.1. Introduction.....	25
5.2. Evaluative instruments.....	25
5.2.1. Questionnaire.....	25
5.2.2. Audio-tape recording.....	26
5.2.3. Rating scale.....	26
5.3. Criteria to choose evaluative instruments.....	28
5.4. Evaluation and scores/marks.....	29
Conclusion of the theory.....	30

Research:

1. Introduction.....	31
2. Background of the research.....	32
3. Analyses of five testing techniques.....	38

3.1. Map-direction.....	38
3.2. Picture comparison.....	42
3.3. Discussion about: “Future”.....	46
3.4. Role-play.....	51
3.5. Re-tell a story.....	57
Conclusion of the research.....	63
Resumé.....	66
Bibliography.....	70
Appendices.....	72
Appendix 1.....	73
Appendix 2.....	75
Appendix 3.....	76
Appendix 4.....	77
Appendix 5.....	78
Appendix 6.....	79
Appendix 7.....	80
Appendix 7.1.....	80
Appendix 7.1.1.....	83
Appendix 7.2.....	84
Appendix 7.3.....	87
Appendix 7.4.....	89
Appendix 7.4.1.....	91
Appendix 7.5.....	92
Appendix 8.....	97

Theory:

1. Introduction

To introduce “Testing Speaking” first, I have to mention why I chose the theme. With regard to the previous teaching experience I had problems testing student speaking ability. In order to the fact, I chose a topic related to this area. Testing speaking is considered to be one of the crucial points in teaching English. However, due to teachers’ lack time and some teachers not viewing it as important, teachers tend to drop the testing of this skill. From my point of view, speaking is one of the most important skills in learning the target language. Therefore, testing of the speaking skill is required to make the teaching consistent and at the same time motivate learners for success. The aim of this thesis is to evaluate testing techniques.

Consequently, the theoretical syllabus of the thesis is focused on the speaking skill itself. When testing the skill, we have to be aware what the skill includes. I concentrated more on the communicative competence, which stresses the speaking skill, and focused on delivering the message through different components; in this case, Canale and Swain components of communicative competence. The practical portion of the thesis is further focused on evaluating testing techniques from both the point of view of a teacher and of a learner.

In the next chapter of the theoretical part, I generally concentrated on testing. It gives a view of what we must consider when preparing a test and also it provides hints for what to be careful of when testing any skill. Further, the theoretical part specifies on how to test speaking and what criteria to choose for testing techniques. Finally, I am going to describe certain testing techniques which will be tested and evaluated in the practical part. Thus, the theoretical input to the evaluation of testing techniques and the evaluative instruments are provided in the fifth part of the theory.

2. Speaking skill

2.1. The definitions of speaking skill

Speaking is a skill just as writing, reading and listening are skills. Speaking relates to listening mostly, because people react in speaking due to the previous context they participated in. While, “The teaching of the speaking skill has become increasingly important” (Celce-Murcia 1991, 127), in the classroom, speaking is the most difficult skill to master for the learner and conduct for a teacher. Because, learners approach the speaking skill differently, and because every learner has his/her different abilities, learners can become overwhelmed. Usually, learners fear saying something wrong or, in the worst cases, they do not communicate at all because of this fear. Communication takes place everywhere, either natives or foreigners, and that is why learners are supposed to communicate as much as possible. As Halliwell states: “Children need to talk. Without talking they cannot become good at talking. They can learn about the language, but the only way to learn to use it is to use it” (Halliwell 1992, 8). So, it is important for students to speak, to continually practice speaking, otherwise there is no chance of learning this ability. In other words, Bygate uses the definition of speaking as a comparison to driving a car. He writes that:

What knowledge does a car driver need? Clearly he or she needs to know the names of the controls; where they are; what they do and how they are operated.... However, the driver also needs the skill to be able to use the controls to guide the car along a road without hitting the various objects that tend to get in the way; you have to be able to do this at normal speed; you have to drive smoothly and without getting too close to any dangerous obstacles. And it is not enough to drive in a straight line: the driver also has to be able to manage the variations in road conditions safely.... In a way, the job we do when we speak is similar. (Bygate, 1987, 3)

Bygate’s example focuses on the way students learn the skill of speaking. To learn speaking in a foreign language is difficult; however, it is manageable. Speaking should be practiced among varied situations and students should communicate and interact among themselves to learn it.

2.2. Communicative competence

Speaking itself has a close relationship with communication. People communicate among themselves in order to be understood. Communication itself is covered under the term “communicative competence.” Briefly, communicative

competence means the ability to communicate, and for learners it is important because when using this communicative competence, they get the message or information across.

Furthermore, the process of communicative competence itself developed over years, and through this development, we can now acknowledge communicative competence now. To review this development, I am briefly going to outline the history of communicative competence.

At the beginning, we can name Noam Chomsky who defined the terms: the competence and performance of language. Later, Hymes and his theory became known. He suggested that communicative competence is based on the theory of language use in social life and he invented four levels of communication behavior. The first is "... whether something is possible, the second whether something is feasible, the third appropriate and last one whether something is in fact done" (Luoma 2004, 97; Savignon 1983, 12).

Then, we continue towards Bachman and his communicative-language competence. He further developed the communicative competence. (To study more, see Bachmann, L. F. *Fundamental Considerations in Language Testing*. Oxford: OUP, 1990.)

Many researchers approach the definition of communicative competence in varied ways. According to Savignon: "Communicative competence may be defined as the ability to function in a truly communicative setting – that is, in a dynamic exchange in which linguistic competence must adapt itself to the total informative input, both linguistic and paralinguistic, of one or more interlocutors" (Savignon 1983, 22). All of these methods have one thing in common: how to teach speaking in the target language effectively.

In contrast to the other researchers, Canale and Swain narrowed down the communicative competence into four components and, in my opinion, made it clearer for language teachers. These components are grammatical competence, sociolinguistic competence, discourse competence, and strategic competence. "Together they suggest a model of communicative competence as a basis for curriculum design and classroom practice" (Savignon 1983, 35). Out of these mentioned concepts of communicative competence, I have chosen Canale and Swain's communicative competence for further

research to determine if all of the components of their communicative competence have been included. So, to learn more about the competences, I will give the descriptions of each component.

First, *grammatical competence* has the same meaning as linguistic competence which deals with lexical, morphological, syntactic and phonological features of language; and then, way of putting these features together to make sentences (Savignon 1983, 36).

Second, *sociolinguistic competence* defines the social context and rules of a language for what and how to say something and when it is appropriate to say it (Savignon 1983, 37).

Third, *discourse competence* includes the ability to connect sentences and words, to use them coherently and cohesively. There is a need for shared knowledge between the speaker and listener (Savignon 1983, 37).

And last, *strategic competence* is the ability to use strategies to communicate. “It is an essential component in a descriptive framework for communicative competence” (Savignon 1983, 43). If someone does not understand what a learner says, he/she asks for repetition or through facial expressions he points it out. This skill makes learners aware of being fluent in communication, and they learn from the context which strategies they can make use of to approach the communicative problem. Bygate classifies strategies according to ‘achievement strategies’ and ‘reduction strategies’. “Both these types of strategy aim to compensate for a problem of expression” (Bygate 1987, 42). Whenever test takers hesitate with how they could express themselves, they can use these strategies. In achievement strategies, learners can use guessing strategies. There are the possibilities of borrowing a word from their mother tongue, foreignizing a word or inventing a word (Bygate 1987, 44). Also, to achieve fluent communication, learners can use paraphrasing, describing the word using synonyms or, they can expect help from a listener. These are co-operative strategies (Bygate 1987, 45). In contrast to the achievement strategies, there are reductions strategies, which can for example avoid a word or a phrase. The reductions strategies can be compensatory strategies which repeat a word or they can use the interlocutor’s language as well (Bygate 1987, 48).

Through the process of the speaking skill, Bygate also distinguishes facilitation production and compensation. “Firstly, speakers use devices in order to [facilitate]

production, and secondly they often have to [compensate] for the difficulties” (Bygate 1987, 14). Learners can be advised to facilitate their speech by using simplified structures, ellipses, formulaic expressions or fillers and hesitation markers. Or else, test takers can correct themselves, rephrase or repeat their speaking ability in order to compensate the problems which come out of the time pressure (Bygate 1987, 18 - 19).

“The inclusion of strategic competence as a component of communicative competence at all levels is important because it demonstrates that regardless of experience and level of proficiency one never knows [all] a language” (Savignon 1983, 46). All these components interrelate and increase the importance of each other. To read more about the communicative competence, see *Communicative Competence: Theory and Classroom Practice* by Sandra J. Savignon.

The specification of components of communicative competence helps to create speaking activities, and it documents the purpose of the communicative tasks. “Many classroom teachers have concentrated on promoting communicative competence in language learners by using ‘communicative activities’- those which rely more on the student’s ability to understand and communicate real information...”(Riggenbach and Lazaraton in Celce-Mulcia 1991, 125).

2.3. Littlewood’s classification of communicative activities

I am investigating testing techniques through testing speaking tasks in my research paper; therefore, I am going to analyse the communicative activities in relation to Littlewood’s view. This view is well defined and well structured even for beginner teachers. Littlewood recognizes two distinct communicative activities: pre-communicative activities and communicative activities (in Bygate 1987, 61). Then, the division of pre-communicative activities is divided into *structural activities* and *quasi-communicative activities*. In structural activities, there is mainly a practice of grammatical structures and in quasi-communicative activities a practice of a simple dialogue without any sufficient meaning (in Bygate 1987, 61). Whereas in communicative activities, Littlewood separates the activities into *functional communicative activities* and into *social interaction activities*. In these activities it is important that knowledge of a target language and a skill are intertwined and learners

practice dialogues through information gap activities. Furthermore, in the social interaction activities, there is communication in social context (in Bygate 1987, 61).

2.4. Conclusion

In summary, speaking is varied in the way one approaches the skill. There are a lot of theories on communicative competence, which is still being examined and developed even today. In my opinion, whenever and wherever people live on Earth, there will still be a need for communication and speaking. Therefore, people must learn to communicate in different languages and master the speaking skill. Testing the speaking ability can motivate learners and help them improve their communication. It is called a 'backwash effect' which is "the effect of testing on teaching and learning" (Hughes 2003, 1).

3. Testing

3.1. The definition of a test

“A test is a sample behavior” (Savignon 1983, 232). Savignon uses the definition to emphasize that when test takers are tested, they give a performance which is evaluated according to their speaking competence. However, the most important principles in the test are validity and reliability. These two principles will be discussed later on.

Savignon gives several points of what tests accomplish:

- measure student progress
- serve as a powerful motivating factor
- measure communicative language use
- tests of communicative competence offer the best assurance that we are preparing learners for the real world (Savignon 1983, 247–248)

The author justifies that tests are used for many purposes. Tests tell us what learners can do with the language they are learning, and they can also motivate learners depending on how successful they are in the tests.

Davies defines a test in a different way, he writes that ‘the good test is an obedient servant since it follows and apes the teaching’ (in Hughes 2003, 2). In my belief, a test is partially a servant to teaching but it should not copy the teaching itself. Hughes argues: “We cannot expect testing only to follow teaching...; however ...it should be supportive of good teaching ...and exert a corrective influence on bad teaching” (Hughes 2003, 2).

Similarly, as Bygate compares speaking to driving a car, Fulcher, as well, compares testing speaking to driving a car under different conditions:

The learner driver cannot be tested on all roads, in all possible weather conditions, or in all the potentially dangerous situations that may be encountered in the future. Similarly, the language learner cannot be tested in all speech contexts or on every task that could simulate potential situations in which he or she may have to speak in the future. From a sample performance, we need to make inference about the likely success or failure of the learner’s future performance in non-text contexts. (Fulcher 2003, 47)

Therefore, we design tests to get as much information as possible. We cannot test speaking in all situations using all techniques that exist, but we can pay attention to the purpose of a test and its construction. Furthermore, Heaton reports that “The most useful tests for use in the classroom are those tests which you write yourself” (Heaton

1990, 6). Since a teacher knows his/her students, he/she also can design the test appropriate to the student's level and to their strengths and weaknesses (Heaton 1990, 6). For example, I would not ask the eighth grader to pretend to be in office and I would not put the fourth grader student in a situation of a teenager, etc.

Moreover, the tests should reflect teaching and learning. "The challenge to teachers is to create classroom tests that serve as learning devices through which washback is achieved" (Brown 2004, 29). There are many ways to increase washback, the well-known are marks for learners and teachers, but Brown recommends:

- give praise for strengths and also constructive criticism of weaknesses
- give strategic hints on how a student can improve
- give him scores on subsections of the test (Brown 2004: 29).

Thus, we distinguish formative tests and summative tests. The former gives information to the learner how he/she can reach the goals and the latter does not give any washback to a learner, which is classified as a wrong approach. (Brown 2004, 29-30). Further, we divide tests according to their purpose which is described in the next chapter.

3.2. Types of tests

In general, we need to know what the purpose for testing is, so that we can choose the appropriate type of a test. Heaton in the book *Classroom Testing* defines only four types: progress tests, diagnostic tests, achievement tests, and placement tests. On the other hand, Douglas Brown in *Language Assessment* distinguishes five different types of test. He accepted the original four types, but added language aptitude tests which "measure general ability to learn a foreign language and ultimate success in that undertaking" (Brown 2004, 43). Nowadays, it is seldom used because as studies showed, to measure success is quite impossible.

Here are the descriptions of the four main types of tests. I will start by defining *achievement tests* since I mentioned this type for my research. The achievement tests measure student's learned knowledge and skills. (Heaton 1990, 14). Achievement tests are given mostly at the end of a course or at the end of a year. "In some ways it is like a progress test but it covers a longer period of learning" (Heaton 1990, 13).

The *progress test* is more useful for teachers in order to follow the student progress. Its purpose is to "find out how well the students have mastered the language

areas and skills which have just been taught” (Heaton 1990, 9). Mostly, teachers try to give progress tests regularly in their lessons, for they serve as “safeguards against hurrying on to complete a syllabus or a textbook” (Heaton 1990, 9). For example, after finishing some kind of grammar task, learners would have a test to check how well they have achieved the goal. Additionally, Bachman included proficiency tests as a test type, which are closely related to progress tests and he also included language aptitude tests into his selection of test types (see Chapter 3 in Bachman’s *Fundamental Considerations in Language Testing*. 1996, 70).

Another type of a test is a *diagnostic test*. Its function is to identify the student’s difficulties or weaknesses and then, to help to prevent these problems. “Teachers must diagnose problems in order to teach effectively” (Heaton 1990, 11). According to Douglas Brown, they should “elicit information on what students need to work on in the future” (Brown 2004, 47). In order to elicit this information, teachers test learner’s knowledge, in this paper – the speaking skill and determine their lack of knowledge.

Sometimes, the diagnostic test is a part of some other test type such as a *placement test*. This test, which guides teachers on where to place students into groups with different levels of language ability, the teacher may be able to prevent weaknesses.

In regard to all types of tests we can easily decide what purpose has each test. From my point of view, testing speaking through the achievement tests is manageable, since speaking is a process generally learned on the basis of learning knowledge and skills over the learning period. As Savignon claims, “...achievement tests should reflect the nature of the proficiency or competence toward which learners are supposed to be advancing” (Savignon 1983, 246).

As Bachmann views the type of test, he includes that “the amount and type of testing, if any, that is done depends upon the decision that are to be made and type of information that is needed to make the correct decisions” (Bachman 1996, 56). To sort out different test types, he classifies them according to the “purpose, the use for which they serve, the content, the frame of reference in which we learn the results, the way to score the tests and the specific technique” (Bachmann 1996, 70).

Generally, tests give examiners the information they need to measure student’s ability and compare it to the other performances, depending on the purpose of a test. Although, Hughes discovers that “The information that we hope to obtain will of course

vary from situation to situation” (Hughes 2003, 9), we still can use similar tests in similar situations. If a test will be suitable for one purpose, it can be inappropriate for another one. “In fact there is no best test or best technique” (Hughes 2003, 6). The teacher must decide, for example, what purpose the speaking test will be for teacher’s class. Also, it is teacher’s responsibility to choose the appropriate testing technique to test speaking.

3.3. Methods of testing

Not only do we distinguish test types but also, the methods of testing. These methods of testing can be direct or indirect, norm-referenced or criterion-referenced, discrete-point or integrative, objective or subjective. To make it clear what type of testing I am going to choose, there will follow a description of all methods.

3.3.1. Direct x indirect testing

First, the direct testing of speaking is an interactive case where an interlocutor deals directly with the examinee. In contrast, for indirect testing the examinee may listen to tasks on a tape and try to answer questions or complete tasks on the tape. The examiner, in this case, is indirectly involved. I will use direct testing for my research.

3.3.2. Norm-referenced x criterion-referenced testing

Next, we can divide the tests according to the results of the student’s performances which can be viewed in two ways: either norm-referenced or criterion-referenced. The norm-referenced testing is used if the performance of an individual is to be compared to the performance of another test taker in the group. “We are interested in how much better than the other candidates are” (Heaton 1990, 16). This method is mostly used to select students and compare them to the other students. Opposite to the norm-referenced type is the criterion-reference testing which is concerned with the success of the performance. We are interested in how well the test taker manages the task. “We want to find out only the degree of success someone may have in doing something” (Heaton 1990, 18). When considering both types of testing, the criterion-referenced testing will be the most suitable for my research.

3.3.3. Discrete-point x integrative testing

“Discrete-point testing refers to the testing of one element at a time, item by item” (Hughes 2003, 16). Contrary to this method, integrative testing tests a

combination of items. Hughes reports that “Integrative testing...requires the candidate to combine many language elements in the completion of a task” (Hughes 2003, 16). For example, in a test of speaking, learners must combine subskills (grammar, vocabulary, pronunciation) and skills (listening to the tester, speaking, or reading the instructions).

3.3.4. Objective x subjective testing

Further, we distinguish objective or subjective testing. Objective testing has only one answer, therefore test takers are scored according to the “correctness ...determined by predetermined criteria” (Bachman 1996, 76). Whereas, subjective testing which is mostly used in speaking tests can have several possible answers because “... subjective questions allow for much greater freedom and flexibility in the answers they require” (Heaton 1991, 32). On the other hand, Heaton claims that the answers do not have to be 100 per cent right or 100 per cent wrong (Heaton 1991, 32). Testers evaluate the performance according to their judgment when grading.

3.4. Criteria to the evaluation of tests

Brown suggested five principles to the evaluation of classroom tests: practicality, reliability, validity, authenticity, and washback (Brown 2004, 30). Brown had in mind that all these principles “...go a long way toward providing useful guidelines for both evaluating and existing assessment procedure and designing one on your own” (Brown 2004, 30). Furthermore, he suggests that other principles can be found as well; however, these five he considers as a “good foundation to evaluate existing instruments and to build your own” (Brown 2004, 30). In my research, I thoroughly concentrated on two main instruments of the evaluation of testing methods: authenticity, and practicality. However, the description of validity and reliability will be provided to get an idea of what to consider when designing tests.

3.4.1. Validity

Lado claims both validity and reliability are criteria for the evaluation of language tests (Lado 1961, 30). He uses a question to determine validity: “Does the test measure what it is intended to measure?” (Lado 1961, 30).

First, we have to be assured about the purpose of the test and then, to determine what we want to learn from the test. Savignon states: “The validity of a test is the extent

to which a test measures what it is supposed to measure and nothing else” (Savignon 1983, 236). As we look at speaking tests, they should test speaking ability and not reading. Brown further states: “There is no final, absolute measure of validity, but several different kinds of evidence may be invoked in support” (Brown 2004, 22). He categorizes five types of validity evidence: content-related validity, criterion-related evidence, construct-related evidence, consequential validity, and face validity.

If the content-related validity of the test is focused on student speaking ability then the test cannot be measured by student writing ability and vice versa (Brown 2004, 22-23). Bachman states that the limitation of content validity is that it focuses on tests rather than on test scores (Bachman 2004, 247).

Whereas, criterion-related evidence measures “...to which the ‘criterion’ of the test has actually been reached” (Bachman 2004, 24). And Bachmann views it as “... a relationship between test scores and some criterion which we believe is also an indicator of the ability tested” (Bachman 2004, 248).

In construct-validity, we ask a question: “Does this test actually tap into the theoretical construct as it has been defined?” (Brown 2004, 25). A teacher should be aware of the construct validity when testing in an oral interview, where he/she takes into account criteria like pronunciation, vocabulary, grammar, fluency, and sociolinguistic appropriateness and, if two of these are not present, then the construct validity fails (Brown 2004, 25). Bachmann criticizes construct validity because it emphasizes the test scores and does not provide a “...means for investigating the processes of test taking themselves” (Bachman 1996, 269).

In the case of consequential validity, the consequences of a test are examined. “Consequential validity encompasses all the consequences of a test...” (Brown, 2004, 26).

Last mentioned, face validity “...exists only with respect to the impressions of observers” (Savignon 1983, 236). Brown explains that:

Face validity refers to the degree to which a test looks right, and appears to measure the knowledge or abilities it claims to measure, based on the subjective judgment of the examinees who take it, the administrative personnel who decide on its use, and other psychometrically unsophisticated observers.” (Mousavi in Brown 2004, 26)

A teacher or observer can not objectively test face validity; it is determined by the feeling of the learners whether the test is testing what is intended. As Brown reports, the content validity is an important part in face validity (Brown 2004, 27). According to this fact, learners must, after reviewing a test, determine that the test is acceptable and useful for them (Bachmann 1996, 288-289). Also, "... test appearance is a very important consideration in test use" (Bachmann 1996, 289). Therefore, in constructing a test, it is important that the learners feel the test is manageable, and therefore, acceptable.

Validity, as a complete criterion, becomes a very difficult criterion in the testing of speaking. When designing a test, the validity of the test should be taken into account and searched critically. At the suggestion of Bachman, Test of English as a Foreign Language (TOEFL) is one of "the extensive programs of language testing validation research in the world" (Bachman 1996, 294).

3.4.2. Reliability

A reliable test is said to be consistent and dependable (Brown 2004, 20). Meaning that, a reliable test, given to the same student a second time, should have similar results (Brown 2004, 20). Reliability can be focused on student-related reliability, rater reliability, test administration reliability and test reliability.

Student-related reliability can be influenced by the feeling of the test taker. If he feels sick, he can have different results than those he would receive if feeling well. On the subject of rater reliability, the test results can be influenced by an error of a rater or some bias which influences scoring (Brown 2004, 21). "Rater reliability is concerned with the extent to which two or more raters are capable of agreeing with each other on the score they award to the same individual(s)" (Fulcher 2003, 139). Test administration reliability concern physical or environmental factors. Example of this could deal with unreliable materials given to test takers or the fact that they do not feel comfortable (Brown 2004, 21). In the case of test reliability, if the test is too long, it can be frustrating for examinees (Brown 2004, 22).

"While reliability is a quality of test scores themselves, validity is a quality of test interpretation and use" (Bachman 1996, 25). As stated, the test writers need validity and reliability to achieve a valid and reliable test. Besides, there are many approaches how to measure reliability and validity of the tests, but since the paper is focusing on

different evidence how to evaluate testing techniques, I can only recommend to see *Fundamental Considerations in Language Testing*. OUP, 1996 by Bachman, L. F.

3.4.3. Practicality

The practicality of testing speaking is also a very valuable feature, as well as reliability and validity. It deals with matters like time, money, administration, or scoring procedures of the test. Furthermore, practicality belongs to one of the principles of how to evaluate testing techniques. According to Brown, we can search practicality according to the 'practicality checklist' which asks questions, for instance:

- 1) Are administer details clearly established before the test?
- 2) Can students complete the test reasonably within the set time frame?
- 3) Can the test be administered smoothly, without procedural "glitches"?
- 4) Are all materials and equipment ready?
- 5) Is the cost of the test within budgeted limits?
- 6) Is the scoring/ evaluation system feasible in the teacher's time frame?
- 7) Are methods for reporting results determined in advance? (Brown 2004, 31)

The practical checklist is a valuable inspiration for my research, where I am going to evaluate the speaking test from the viewpoint of a teacher, and it should show how testing techniques are practical, which is one of the evaluation criteria. In other words, the checklist should display whether test takers had sufficient time to answer the question or finish the task, or if scoring procedures were time consuming. "...time often emerges as the most important factor, one that overrides other considerations in evaluating an assessment" (Brown 2004, 31). As Brown also mentions that time is valued in giving feedback on the students' tests (Brown 2004, 31).

3.4.4. Authenticity

"The test used to be boring, not connecting parts without authenticity" (Brown 2004, 28). Mostly, tests do not focus on the content but rather on the importance of the structure, such as: grammar structure. This is a crucial problem because learners feel that tests or tasks do not correspond to real-life situations and are bored completing them. "Yet current teaching practice delays experience in authentic communication until the student has acquired a basic set of grammatically correct utterances" (Savignon 1983, 67). Savignon actually stresses the fact that learners are taught grammar first and then, they communicate in real-life situations. The teaching itself should be based on the authentic materials and tasks, so that learners will perceive them faster. If we say that the test is authentic, we claim that it is like real-life situations (Brown 2004, 28).

Bachman and Palmer define authenticity as “the degree of correspondence of the characteristics of a given language test task to the feature of a target language task” (in Brown 2004, 28). And Brown gives a list how authenticity can be presented in the test tasks:

- the language in the test is as NATURAL as possible
- items are contextualized rather than isolated
- topics are meaningful (relevant, interesting) for the learner
- some thematic organization to items is provided, such as through a story line or episode
- tasks represent, or closely approximate, real world tasks (Brown 2004, 28)

The principle of authenticity can have an impact on test taker’s scores and on their face validity as well. “Authenticity in oral language assessment relates to both the type of language used and the task to which that language is applied” (O’Malley and Valdez Pierce 1996, 59).

3.5. Preparation

It is important to prepare a test by searching relevant materials, designing instructions, and defining the problems the tester can deal with.

It is a good idea to list in writing all the material that you want your test to cover: you can then refer back to the list during and after the test writing to see if you have included all you intended. (Ur 1996, 42)

Heaton recommends that “You should always try to select a representative sample of skills and language areas from your syllabus or from what you have actually taught” (Heaton 1990, 25). He then develops a question to determine whether learners have mastered it (Heaton 1990, 25). Moreover, the instructions should be prepared before the test. “Clear instructions are crucial; otherwise you will test familiarity with the test procedure and not language ability” (Underhill 1987, 40).

The relevant criterion for preparation is also time, either: 1) time to prepare for the testing from the point of the teacher or 2) time needed from the point of a learner. “Tests are announced at least a week in advance in order to give them¹ plenty of time to prepare” (Ur 1996, 43). “Although reporting on what we are doing means spending more time still, it is also helpful because it forces us to think about the activities more carefully” (Luoma 2003, 191). In the same way, Common European Framework

¹ ‘them’ in Ur, Penny means: learners

remarks: “The less time available for task preparation and performance the more demanding the task is likely to be” (CEF 2001, 163).

3.6. Test specification

The test specification helps testers to be careful about the objectives and strategies for testing. “A test’s specifications provide the official statement about what the test tests and how it tests it” (Alderson and Clapham and Wall 1995, 9). For this research paper, I used the test specification framework of Hughes: *Testing for Language Learners*. 2003: 59 – 62) and adjusted it to my needs. In other words, the test specification “...forces teachers to look far more closely at both their short-term and long-term goals, especially the former” (Heaton 1990, 26).

Similarly, Luoma describes that “The specifications record the rationale for why the assessment focuses on certain construct, and how the tasks and criteria operationalize them” (Luoma 2004, 113). For Luoma, one of the advantages in writing test constructions is “that they will help the developers create a coherent system whose parts fit together” (Luoma 2004, 115).

However, teachers, who write their own tests, should be familiar with what the purpose of a test is, whether the test is valid, reliable, practical, and authentic, and also, if the test on speaking ability contains all components of communicative competence. It is said that:

The classroom teacher needs to institute a progression from artificial exercises to real language use, from discrete linguistic objectives to communicative objectives, and from discrete-point tests to tests of communicative competence. (Schulz and Bartz in Savignon 1983, 29)

There will always be a tendency to improve tests, testing techniques or scoring procedures. Also, general conditions for testing should be considered. According to Brown and Yule, there should be brief instructions, clear requirements, and students should be familiar with the task (Brown and Yule, 122).

4. Testing Speaking

4.1. Why test speaking?

Testing the speaking skill is still a myth for teachers, writers, researchers, and examinees. First, Madsen states that the nature of the speaking skill is not well defined. Second, he said that there is a disagreement about how to test all criteria and finally, he says there is difficulty in evaluating the criteria all at once (Madsen 1983, 147). If a teacher takes this all into account, he/she really must fear testing speaking. To narrow the statement down, Brown defined the purpose of a speaking skill as follows: “....a taxonomy of skills from which you’ll select one or several that will become the objective(s) of an assessment task” (Brown 2004, 142).

As Bachman claims: “...the performance we observe and measure in a language tests is a [sample] of an individual’s total performance in that language” (Bachman 1996, 33). It is evident that test measurements cannot be defined as ideal; however, it shows a teacher and a learner the way where to continue in studies.

4.2. Objectives of testing techniques

Mostly, testing techniques of the speaking ability are aimed at the communicative activities because, on the basis of communicative activities, learners can communicate. As an example shows:

Many classroom teachers have concentrated on promoting communicative competence in language learners by using “communicative activities”- those which rely more on the student’s ability to understand and communicate real information... (Riggenbach and Lazaroton in Celce-Murcia 1991, 125)

However, “The broad aim of all these techniques is to encourage learners to speak by giving them something to speak about...” (Underhill 1987, 40). Similarly, as for Underhill who defines the aim of techniques, the same applies to Fulcher who says that “a goal to design a test task is to elicit enough speech to allow a rating to take place” (Fulcher 2003, 50). Comparably, Hughes suggests that techniques:

- will elicit behaviour which is a reliable and valid indicator of the ability in which we are interested;
- will elicit behaviour which can be reliably scored;
- are as economical of time and effort as possible;
- will have a beneficial backwash effect, where this is relevant (Hughes 2003, 75)

Researchers vary in stating what testing techniques are and what objectives they should complete. However, to select specific testing techniques for the learners, we have to consider the criteria to choose the testing techniques. These are described below.

4.2.1. Criteria to choose testing techniques

“Wanting to communicate means having a good reason for doing so” (Halliwell 1992, 14). With regards to this definition, I must agree with the fact that every test must have a purpose. The purpose of a test is one of the main criteria in selecting the appropriate testing techniques. “If the student has a purpose in speaking, he immediately finds himself in a situation in which what he says and how he says it have significance” (Brown and Yule 1983, 117-118). This means that, if a speaker delivers information to a listener, who is unaware of this information, then the speaker is more motivated to deliver the message (Brown and Yule 1983, 117). Despite this fact, we should know “...what it is you want to test” (Brown 2004, 49) when considering the purpose of a test.

Another consideration to choose the testing technique is the context that should be chosen, so that learners can speak based on that context (Luoma 2003, 30). It is easier for students who do not have to think about the input when tested in spoken interaction.

Also, the difficulty of the task may cause many problems for the test taker and tester. “Difficulty is not a direct characteristic of tasks; rather, it is the sum of task characteristics and the conditions under which someone performs the task in relation to the person’s ability in the skills that it requires” (Luoma 2003, 46). For beginners, the testing tasks should be easy and for advanced speakers, the tasks should be more difficult. Every researcher has his own explanation regarding difficulty. For Brown and Yule, the difficulty lays in the complexity of task materials and for Foster and Skehan, it is the planning time, cognitive complexity and the difficulty of task familiarity in designing tasks (in Luoma 2003, 46). However, the difficulty of testing techniques can be avoided in this paper since the test takers are familiar with the testing techniques. What can be new for them is the context or their choice of speaking partners.

The level of learners as related to their age was another criterion for the selection of tasks. Learners in a certain age group are interested in certain topics and the context for testing tasks should be relevant to their age. Above all, the level of learner’s

language competence should be specific to the tasks as well. Learners are to be tested on the basis of their acknowledged level of language competence. Therefore, the testing techniques are adjusted to learner's language competence.

The list of criteria, namely – the purpose of a test, the context, the level of the learners, the level of learner's language competence and the difficulty of a test, is relevant for my further research.

4.2.2. Testing in pairs x individually x group work

While testing speaking, a tester must consider and prepare ahead what kind of interactive patterns will be used. There are well-known interactive patterns used, such as: to test learners individually, in pairs, or in groups.

The individual testing is one of the well-known techniques and most often used when testing speaking. Due to the familiarity, this interaction pattern is more preferred. Savignon classifies this method as 'flexible' for the opportunity that prepared questions can be shaped or switched according to the developed situation, for example in an interview. However, this method also carries a negative side by the fact that the tester gains power over the test taker (Savignon 1983, 35). Still, in my mind, this method cannot be erased and substituted by pair-work right away. Practice and experience are needed for such a shift.

Second, the testing in pairs is connected with a partner relationship. The main focus of testing speaking in pairs is "...to test the examinees to interact with each other" (Savignon 1983, 36). Also, Weir claims that "the concern is that all test takers may not get an equal opportunity to show their speaking skills at their best" (Weir in Savignon 1983, 37). Whereas, the phenomenon of pair-work testing can be crucial for other examinees where it is "...inevitably influenced by the other participants" (Weir in Savignon 1983, 37). This effect can cause as some examinees have a different basis for learning English, or they can have prejudices against each other. Or else, some test takers can be shy among other people, or the question of sexes can play its role as well.

"Like pair work, group interactions tasks are also generally well received by learners" (Savignon 1983, 39). However, learners need to be experienced with group-work interaction and as a result, I will not include group-work pattern in my research. The learners were not used to working or being tested in groups and 'weak' learners tended to be shy among the 'strong' learners. In practice, I focused only on the

individual and pair-work patterns that are further discussed and analysed for use in testing speaking.

4.3. Testing techniques

When interaction patterns are set, we can target at the testing techniques, sometimes called: 'elicitation techniques'. They are mostly divided according to their purpose of testing. Nevertheless, Underhill states: "There is no natural classification of test techniques..." (Underhill 1987, 44). Then, he explains that the testing techniques are sorted into the least controlled and the most controlled in his book. For detailed classification of testing techniques, see Underhill, Nic. *Testing spoken language*.

In my opinion, there is not really important into which section the testing technique belongs to but what focus it has on learners; in other words, how talkative and successful learners are in their speaking skill when they are tested through testing techniques.

To test the speaking techniques, Brown claims that:

First, new and innovative testing formats take a lot of effort to design and a long time to refine through trial and error. Second, traditional testing techniques can, with a little creativity, conform to the spirit of an interactive, communicative language curriculum. Your best task as a new teacher is to work within the guidelines of accepted, known, traditional testing techniques. (Brown 2004, 49)

Therefore, teachers can focus on the test procedure and in addition, they do not have to worry about the innovations of testing techniques. First, learners are mostly familiar with known techniques as well as teachers and teachers can further concentrate on the learner's performance in a more controlled way than on the new technique which has not been experienced.

There were many books written on the types of testing techniques. Nevertheless, the authors designing testing techniques vary in the sub-categories of testing techniques. I am not going to describe these different classifications, but I will focus on the testing techniques selected for my research paper. The main division is devoted to tasks that carry information gap and focus on an oral interview. The oral interview type is one of the ways how to make a student to say something (Madsen 1983, 162). In the interview type, it is important that an interviewee uses "a sincere, open, supportive manner..." (Madsen 1983, 163).

4.3.1. Information gap techniques

In the information gap tasks, “there should be a listener who does not have the information which the speaker has and who needs the information” (Brown and Yule 1983, 111). According to the age and to the level of learners, the testing techniques selected follows:

Picture tasks

According to Madsen, picture tasks belong to the limited activities which contain guided activities with direct responses (Madsen 1983, 148). With regard to Madsen, this type of limited task is suitable for learners with limited skills in English (Madsen 1983, 151). Therefore, the picture tasks are adequate for my learners; furthermore, students enjoy looking at pictures and talking about them. “A good picture can be the stimulus that generates the confidence to speak and a flow of conversation” (Underhill 1987, 67). Heaton claims that “...pictures are very useful for testing the speaking skill (Heaton 1990, 61). There are many possible ways how to test picture tasks, for example: through action pictures, sequence pictures, the use of a map, charts or graphs, or use simple line drawings...(Madsen 1983, 151-153). Besides, the picture tasks contain good face validity (Madsen 1983, 148). However, there are also disadvantages of these picture tasks. It is important to “...choose the picture(s) very carefully as a picture often influences the language forms which the students use” (Heaton 1990, 62).

Re-telling a picture story

This technique can be categorized under the “interactive speaking” tasks classified by Brown where the interaction between a test taker and a tester is evoked. According to Brown, the classification of interactive speaking tasks is proposed to the relationship of the interaction: “Interpersonal interactive tasks where we find role-plays and interviews and transactional tasks looking for individuality communication, for example: speeches, or re-telling long stories” (Brown 2004, 167). I think that learners find these tasks as more motivating and authentic “They show how well the examinees can recount a sequence of event, usually in one time frame, either present or past” (Luoma 2004, 144). In the paper, the task is based on the picture sequence “where the content of the pictures guides what will be said” (Luoma 2004, 144). “They should generate enough talk and provide opportunities for the examinees to show what they

know” (Luoma 2004, 144). Learners show the ability to re-tell a story through: “... features of narratives: setting the scene, identifying the characters and referring to them consistently, identifying the main events, and telling them in a coherent sequence” (Luoma 2004, 144).

In addition to the re-telling of the story, the task contains the description of pictures as well. Brown suggests that the difficulty of description tasks can be narrowed or enlarged according to the students level (Brown 2004, 123- 124). The structure of a descriptive technique relates to the next technique as well.

Discussion-presentation

Underhill states that discussion or conversation is “the most natural thing - two people having a conversation on a topic of common interest“(Underhill1987, 45). On the contrary, it is the hardest task to make in a test. There is “a danger that this technique will reward extrovert and talkative personalities” (Underhill 1987, 46). The problematic point regarding the grading of the discussions is that “Since it is the discussion...there is usually no single correct answer...” (Underhill 1987, 49). What Fulcher recommends is that “The discussion and interaction also makes it an excellent task type for assessing strategic competence...” (Fulcher 2004, 78).

Guided role-play

Also, the guided role-play tasks could be structured according to the freedom of the responses. The distinction between guided role-play and open-ended role-play is seen in the nature of a learner. Usually, the active and extroverted students have better results in the open-ended role-play, and those shy, introverted students pass through this task. That is why; the guided technique is set on the same level for either talkative or shy students (Madsen 1983, 160-161). Also, this technique has some disadvantages. One of them is the fact that other factors like memory, personality of a student, or reading ability can have an effect on scoring (Madsen 1983, 162).

The choice of discussion topics should not be offensive or personal. “Personal stories often reveal embarrassing details that speakers would be shy to discuss in a test or, if not, they may be so uneventful that the speakers would consider them unworthy to tell” (Luoma 2004, 145). This fact is important to realize when a teacher prepares a test. “Topics for role-plays should be taken from student’s current interests and anticipated experiences” (Donahue and Parsons in O’Malley and Pierce, 1996, 86).

4.3.2. Oral interview

Map-direction

Generally, the map-direction technique belongs to the explanation tasks, which Madsen defined. The explanation tasks should be familiar with what they are explaining or describing (Madsen 1983, 160). In order to understand the structure of this technique, it can be also titled as the oral interview, because the tester and the test taker communicate together in order to get the information across. “The oral interview is not really an elicitation device, but rather a kind of framework for using various elicitation techniques” (Madsen 1983, 162). The most important part of an interview is that communication is absorbed between the interviewer and the student. Moreover, the main aim is to get students to talk on their own (Madsen 1983, 165). It is a well-known technique for learners and testers and the scoring is simpler than in guided techniques; however, it is time-consuming if it takes place in a class where there are about twenty learners (Madsen 1983, 162-166).

Picture comparison

In my research paper, the picture comparison was selected according to the criteria above. “The purpose of this activity is to find out in what ways the two pictures are different...” (Heaton 1990, 62). Similarly, Fulcher suggests that “It is possible to develop such tasks into open activities when the pictures are used as a springboard to wider discussions of topical issues” (Fulcher 2003, 73). There are a lot of possibilities what techniques to combine with pictures. In my belief, the selection of picture comparison technique relies on language competence of learners.

Even though the description and purpose of each testing technique used in the research part was expressed, it is only a part of many variations to testing techniques. Halliwell claims that: “Flowers are weeds when there are too many of them or they are in the wrong place!”(Halliwell 1992, 15). In meaning, the testing speaking with the usage of different testing techniques can cause misunderstandings and failures of learners’ performance. In order to this fact, “...we should make sure that the speaker can see a reasonable purpose for performing the task at hand” (Brown and Yule 1983, 111). To choose the testing technique for a particular classroom, we have to take into account the criteria, which should help us to select the techniques. Next chapter is going to focus on how a teacher can evaluate the testing technique.

4.4. Evaluation of testing techniques

As mentioned earlier, there is nothing like the best technique or the best test. When evaluating or assessing² the learner's performance we take into account different principles. O'Malley and Pierce conclude that:

Oral language assessment can take various forms depending on your purpose for assessment, student's level of language proficiency and the purposes for which students use oral language in the classroom. (O'Malley and Pierce 1996, 69).

The purpose of the test is described in the theoretical part of this paper in chapters 3.2. and 3.3. Without the purpose of a test, we cannot evaluate a test. To find out whether the testing technique was prepared ahead, there can be a construction of the preparation projects used in the research. Also, this can be evaluated with regard to the testing techniques.

Next step in testing the speaking ability is to design a test specification where we can include skills to be assessed or scoring procedures used. In this case, it is also helpful to evaluate whether test specifications were completely understood by the tester.

In addition, we can include the analyses of the components of communicative competence (CC) to see if the testing techniques focus on all the components described by Canale and Swain. This evaluation of testing techniques will increase the idea of the testing techniques content. There can be a testing technique with stronger components of strategic competence and sociolinguistic competence and with weak components of discourse and grammatical competence. It depends on many factors that have been already mentioned.

From the point of view of a learner, the testing techniques will be evaluated through the tests of the speaking ability where learners get marks after each testing technique. Testing is evaluated according to the rating scale which contains five criteria. Furthermore, learners evaluate the testing technique on the basis of the questionnaire written in their mother tongue. Then, the output of the questionnaires will be brought in the target language for the research purposes.

5. Evaluation of learner's performance testing

5.1. Introduction

² in this context and the paper, assessing closely relates to evaluation

In order to evaluate learners' performances, we must be experienced in testing the speaking skill and further, we should be acquainted with methods related to the evaluation of learner's performance testing, and the evaluation of testing techniques. It is also advisable to explain the term: evaluation. According to Weiss: "Evaluation can be defined as the systematic gathering of information for the purpose of making decisions" (Weiss in Bachmann 1996, 22).

In addition, the methods used for evaluation are described as qualitative and quantitative methods. (Fulcher 2003, 200-201). Also, Fulcher states that "All evidence is valuable, whether it comes from the application of quantitative and/or qualitative methods" (Fulcher 2003, 200). Consequently, I will introduce qualitative methods since the evaluative instruments ascribed below correspond to this section except the rating scales. "...qualitative studies that are commonly used to study speaking tests: expert judgment, questionnaires and interviews ..." (Fulcher 2003, 216). In order to start with the qualitative methods, I will describe each evaluative instrument below.

5.2. Evaluative instruments

For evaluative instruments, there were chosen questionnaires, audio-tape recordings, rating scales, preparation projects (discussed before), test specifications and tests (both discussed previously). There is a wide range of instruments we can use. Nevertheless, we must be specific and consider what instruments will be useful. In this chapter, the focus is on the evaluative instruments, which were selected for the research paper. To select the relative evaluative instruments we have to consider the criteria for the selection as well.

5.2.1. Questionnaire

Questionnaires contribute to one of the types of qualitative methods. "The use of qualitative methods in studying tests of second language speaking is a much more recent phenomenon, but one that has contributed a great deal of understanding to validity" (Fulcher 2003, 216). Questionnaires are easier for administrators to master, because there are the same questions for every candidate. "Questionnaires are [printed] forms for data collection, which include questions or statements to which the subject is expected to respond, often anonymously" (Seliger and Shohamy 1989, 172). Different

types of questions can be asked, and since they are anonymous, and candidates can respond in a free way (Seliger and Shohamy 1989, 172).

We also differentiate structured questionnaires from unstructured questionnaires. “Structured questionnaires are considered to be more efficient than open ones” (Seliger and Shohamy 1989, 173). For my research purposes, I am going to use unstructured questionnaires to find out how learners view the testing technique. “Unstructured questionnaires vary in their degree of explicitness, will include open questions to which the subject will be expected to respond in a descriptive manner” (Seliger and Shohamy 1989, 172).

5.2.2. Audio-tape recording

The audio-tape recording should be a helpful tool to record learners’ performances, so that a teacher can re-listen to the tape later on and evaluate the performances. Even though it is a time-consuming method, it is the most popular teacher research method (Hopkins 2002, 105). “If the responses are recorded, it is possible to score several problems in each item” (Lado 1961, 246).

However, “Pupils often find the presence of a tape recorder in the class disturbing and have to be introduced to the technique over time” (Hopkins 2002, 105). In order to prevent the learner’s stress with the appearance of a tape-recorder, learners used tape-recorders in their mini-dialogues for speaking practice in their classes.

5.2.3. Rating scale

Rating scales belong to one of the main instruments in testing the speaking skill. Fulcher clarifies that:

Rating scales are important in tests of speaking because they are operationalisations of the construct that the test is supposed to measure, whether the description of the construct is [thick], or [thin] as in the scaling of descriptors. (Fulcher 2003, 113)

There are various types of rating scales distinguished; however, the most common are holistic and analytic scales. “Holistic scales express an overall impression of an examinee’s ability in one score” (Luoma 2004, 61). Analytic scales express the test taker’s strengths and weaknesses and, “...contain a number of criteria” (Luoma 2004, 68).

Besides, the rating scales must be aimed at the learners and therefore, “You should assign the rating that most closely fits the student’s actual performance” (O

'Malley and Pierce 1996, 65). Underhill adds that the rating scale should be kept "as simple as possible" (Underhill 1987, 100). For this reason, testers can more concentrate on the evaluation of performance when knowing the description of rating scales by memory. They do not have to worry about the complex description and can evaluate it clearly. In addition to evaluate learners' performance, I designed a rating scale to the learners' needs for my research paper. It consisted of five criteria: grammar accuracy, vocabulary, pronunciation, interactive communication, and fluency. In this case, the weighting system, which an extra point or percentage can be given to some of the criterion, was not used because each criterion is taken equally. In order to specify the criteria, the description will be introduced.

First, grammar accuracy is specific to the mastery of word order, tense and pronouns, since learners have been acquiring only the basic of English grammar. "When testing speaking it is similarly important to decide what types of errors be penalized in the scoring, and what errors should be ignored or treated with leniency" (Fulcher 2003, 27 – 28).

In pronunciation, we evaluate whether learners understand foreign sounds and are able to pronounce words which do not occur in their native language. As Fulcher reports: "The speaker must first decide what to say, be able to articulate and create the physical sounds that carry meaning" (Fulcher 2003, 25).

The criterion of vocabulary basically tests if learners have achieved the vocabulary which they have been learning and practicing throughout their learning process. The evaluation of this subskill is limited regards to the learner's knowledge; however, it can be further developed by the strategies learned and used during the learner's performance.

Next criterion – fluency was tested whether learners can connect their speech and express what they wanted to say. Fulcher described fluency as:

If speech is going to be fluent, the process of planning what to say, retrieving the necessary grammar and vocabulary, and speaking, needs to be automatic. It is when speech becomes more noticeably automatic that we describe a speaker as being 'fluent'. (Fulcher 2003, 30)

In order to test the fluency, we must concern whether the fluency does not disturb the learner's performance or whether the pauses are not too long.

In the interactive communication, I mostly concentrated to evaluate whether learners understand each other or the interlocutor. Further, the criteria concerned whether learners can experience the ability of taking turns and the ability to grasp the control over the conversation or simply listen to the conversation. “Firstly, a second language learner must be a good listener if he or she also wishes to be a good speaker” (Buck in Fulcher 2003, 35).

When concentrating on the evaluation of the performance according to the rating scales, we should be careful not to misjudge the overall performance (Johnson 2001, 307). Hughes advises that “...great care must be taken to ignore personal qualities of the candidates that are irrelevant to an assessment of their language ability” (Hughes 2003, 133 – 134). This seems to be humorous but a similar case can also happen during the evaluation of the speaking ability:

I remember well the occasion when raters quite seriously underestimated the ability of one young woman who had dyed her hair blonde. In an oral test it can be difficult to separate such features as pleasantness, prettiness, or the cut of someone’s dress, from their language ability – but one must try! (Hughes 2003, 134)

Therefore I try to avoid the test taker’s appearance and to focus only on his/her language ability. It is difficult not to look at the qualities of the learners we teach, but we must at least try.

In summary, relating to the evaluative instruments, I wanted to give a brief description why I chose certain instruments for the research part. Every technique has its advantages and disadvantages and researchers must evaluate the possible techniques according to the criteria, they set for themselves.

5.3. Criteria to choose evaluative instruments

In this research paper, the criteria considered in choosing the evaluative instruments were the learners’ age and time specification. As there is a possibility to fill questionnaires, give interviews or write journal reports I find that questionnaires for eighth graders will be a good way how to evaluate the testing technique. Certainly, the questionnaires will be constructed in their mother tongue because their language competence cannot still comprehend their feelings or attitudes. Therefore, the choice of their native language was included.

The next criterion is time because, for example: doing interviews would be a time-demanding for me as a teacher, tester, and evaluator. For learners, it would be a stressful technique as well, because they would go through a testing process of their oral skill and if possible through the oral interview.

5. 4. Evaluation and scores/marks

The varieties of evaluation differ according to the purpose of our tests. For example, to test speaking ability, the performance evaluation plays a significant role. The performance is tested under certain circumstances and Luoma explains that:

It is often tested in live interaction, where the test discourse is not entirely predictable, just as no two conversations are ever exactly the same even if they are about the same topic and the speakers have the same roles and aims. (Luoma 2004, 170)

As a matter of fact, there are different types of evaluation: a teacher's evaluation, self-evaluation, or peer-evaluation. For my purposes, I selected the teacher's evaluation according to the learner's experiences. In order to use self-evaluation method, "we should be teaching the learner to know how he is getting on independently of the teacher" (Underhill 1987, 23). And similarly in the peer-evaluation, learners must be taught to evaluate their peers. Furthermore, "some techniques will be easier to self-correct than others" (Underhill 1987, 23).

Mostly, the evaluation is measured through scores, similarly marks or grades. And Brown claims that "Grades must be the most-talked-about topic in anyone's school years" (Brown 2004, 281). In addition, grades still play a leading role among teachers and learners. "It's ironic – when our study life – is inserted into [marks] only... (Brown 2004, 282).

In particular, scores must show criteria as well. "If we are to interpret the score on a given test as an indicator of an individual's ability, that score must be both reliable and valid" (Bachmann 1996, 24). On the other hand, test takers should not see that we are writing notes or marks while they are being tested. Heaton reports: "Never mark in front of a student" (Heaton 1990, 67). Due to the fact, test takers can be stressed out or their performance can be reduced.

Conclusion of the theory

Providing the theory to the thesis of the paper, it should increase the understanding of testing speaking in the terminology area. It should give a general overview of the thesis; however, it should provide the basis for the practical part.

At the beginning of the paper, I was discussing communicative competence principles and described speaking skill which is a part of a communicative competence. Through the content of the theory, we could recognize types of tests and testing, and defined the testing of speaking. Generally, it is important to know that: "Testing is sometimes the art of the possible" (Johnson 2001, 306). The explanation is not needed here, since the statement summarizes it clearly. Besides, the main attention was given to the testing speaking techniques that are further going to be evaluated in the practical part.

To be able to evaluate testing techniques, I attempted to give a detailed look at the evaluation generally, and I stated the criteria according to which I could focus on the evaluation of learners' performances and on the testing techniques from the point of the teacher and a learner. Therefore, the various evaluative instruments were also provided in the theory, so that the advice or relevant steps in the construction of instruments could be provided for teacher's needs.

In my opinion, we must learn throughout our lives' experiences and in this case, to evaluate the testing techniques we should devote main time to practice and to the critical evaluation of the prescribed aim. There always will be things we will not know. Then, we will appreciate if someone else can explain them to us what to improve or what to learn. Similarly, "...learning about other speaking tester's experiences can help us learn more" (Luoma 2004, 191). As Savignon claims: "Teaching always has been and always will be as much art as it is science" (Savignon 1983, 9).

Research:

1. Introduction

With regards to the theory of the paper, I am now going to introduce a practical part. The aim of the research was to evaluate testing techniques aimed at the speaking ability from the teacher's and learner's point of view.

In the following part, I mention the background information of the research, such as: when and where the research was taken, etc. After the introduction, I explain the usage of evaluative instruments and criteria, namely: preparation projects, authenticity, overall practicality of testing techniques, test specifications and analysing the communicative competence. All of these should display the usefulness of each testing technique and its characteristics in testing speaking.

The evaluative criteria from the point of a learner are the evaluations of the learner's performances through the rating scale criteria and the analyses of anonymous questionnaires. In other words, learners are evaluated by marks which serve as connectors among the analyses of communicative competence (CC). Each criterion in the rating scale is evaluated towards the components of CC. Nevertheless, learners have the possibility to express why the testing techniques are interesting or why they are difficult through the questionnaires constructed towards the testing techniques. In consequence of introducing part, I focused on individual testing techniques that were selected and examined for the research. The evaluation and analyses of testing techniques were provided according to the evaluative instruments and criteria.

Through reading the research, a reader can get an idea how to proceed through testing the speaking ability and what he should take into account when preparing the testing technique. By the way, the analyses of testing techniques in relation to communicative competence show what each testing technique includes. The research also gives an idea that we can evaluate the testing technique from various angles. And I think that the evaluation of testing techniques should be considered as one of the positive purposes to test the speaking ability increasingly.

2. Background of the research

The research was performed in the elementary school where I was doing my teaching practice during 2005. The teaching practice was designed as a clinical year at the University of Pardubice. The aim of my research was to evaluate the testing techniques according to the criteria mentioned in chapter 4.2.1. Moreover, the techniques are evaluated from the point of view of a teacher and a learner.

I was teaching learners for a whole year at the village school and for this reason I could monitor learner's testing, the evaluation; and thus, I could concentrate on the research paper. So, I knew learner's strengths and weaknesses and that is why I could choose the relevant materials for them, from mostly real-life situations according to their syllabus.

The learners at the age of 14 – 15 (attended the eighth grade) allowed me to do the research in their class. They attended English class three times a week, and they have been studying English for four years. The class consisted of eight learners – six boys and two girls. Testing the speaking ability was thoroughly practiced first and sufficient time was devoted to explaining the tasks, instructions, methods and strategies. Learners got familiar with the way of testing speaking ahead of time; and furthermore, they met the criteria of the rating scale that was set to be equal and in relations to this, they knew the requirements for the testing speaking. In addition to that, if learners received the instructions, they were given sufficient time to prepare. Further, it is recommended to “give each student a few minutes to examine the picture before trying to elicit language” (O'Malley and Pierce 1996, 80).

Testing in pairs x individually

As mentioned earlier, the testing spoken techniques were designed through the individual and pair-work methods. Both were chosen on further studies of my pupils and their experiences with these interaction patterns.

I tried to test speaking in groups because I thought that eight pupils would have been easy to manage. In practice it showed that learners were passive in groups, either afraid or shy to speak. The reason of this can be seen in a short period of practice time. So, for testing speaking it would be difficult to score the group-work method and evaluate which testing technique was appropriate for the learners; therefore, I considered pair-work and individual testing, as appropriate.

The process of testing from the point of a teacher

Despite the fact that I contributed to testing as a teacher of this class, I also had to be a tester at the same time. The tests were designed as direct, criterion-referenced, subjective, discrete-point, and achievable tests. Later on, I evaluated recorded learners' performances after each testing technique, so that I could have the basis and feedback for the evaluation which was done later. To evaluate the performances, I constructed the rating scale for the learners' needs in my classroom. It consisted of five criteria: pronunciation, vocabulary, grammatical accuracy, fluency, and interactive communication. Each criterion had five evaluative level descriptions. To mark the learners, I evaluated the test taker's performance with regards to each criterion and then calculated the average out of all five marks. The usage of marks from testing speaking served as a feedback for learners and further, served to the overall assessment of the English subject. The teacher learned about what strengths and weaknesses learners had. Also, the marks served as a helpful device for a teacher to analyse if the testing techniques were appropriate for learners according to the purpose of a test, to the learner's level, and to the designed concepts of communicative competence.

Scoring procedure

Moreover, the evaluation was not perceived at once to test takers, even though one of the pedagogical requirements of evaluation is to deliver the results to students immediately, so they can have a quick feedback. However, I was the only tester and evaluator at the same time. For this purpose, I recorded all the students' performances on the tape-recorder. Next lesson, learners received the results, but first they listened to their speaking performance, to see whether they met the task goal. Then, they could compare the results to their self-evaluation attitudes and I could give some ideas for their improvement. Sometimes, they found out that they repeat words to be fluent or that others have a wrong pronunciation, etc.

Evaluative instruments

Based on the theory, the description of constructing the particular evaluative instruments will be contained and generally analysed.

- Preparation project

One of the first steps in testing was to prepare an outline which further controlled testing itself. As a result, I constructed the preparation project to thoroughly prepare for each testing technique and to be prepared for the further steps in testing (see Appendix

4). In addition, the project focuses on the preparation of materials and instructions to the specific testing technique. The instructions should be designed ahead, so that it would enhance the learners' performances, and serve as a teacher's note.

- Practical checklist

This evaluative instrument serves only for the general evaluation of testing techniques' practicality. I will provide a detailed analyses in the following paragraph and not in the evaluation of each testing technique as the following instruments were arranged. The usage of practical checklists was completed, after the testing speaking was done (see Appendix 2). I was inspired by the practical checklist by Brown. (*Language Assessment*, 2004, 31). To analyse the practicality generally we can say that learners were familiar with the test instructions before the tests and also, they met the requirements for the evaluation out of the rating scale. Further, the sufficient time before and for testing was also fulfilled. Some test takers answered quickly and only in some cases, they looked for words or for help.

Since, the rating scale was prepared before the test and the evaluation was done after the re-recording the testing of all individuals after the test, we can conclude that the administration was planned and fulfilled. Furthermore, tests were all recorded, so I, as an evaluator, could re-listen to the testing again and mark the learners' performance according to the rating scale.

Related to the practicality of the scoring system, I learned that the evaluations of testing the speaking ability done by a teacher were time-consuming. It took generally three hours to master each evaluation of the testing technique: to re-record the student's performance, to evaluate the performance according to criteria, and to sum up the mark.

In conclusion, from my point of view, I found the selected techniques practical, even though the procedure of the scoring system took a long time due to the fact that I was the sole test administrator, tester, and evaluator. As suggested, when designing a test, there should be considered not only the practical issues of the test but also the structure of the test.

- Test specification

I designed the test specification instrument in order to follow whether tests aimed at testing techniques' characteristics were structured carefully (see Appendix 5). Each testing technique contained this test specification for purposes mentioned earlier.

- Questionnaire

To gather the evidence regarding the testing techniques' evaluations from the student's perspective, students were asked to fill in questionnaires after the testing was done. The questionnaires were designed to the different types of testing techniques in their mother tongue (see Appendix 7.4.1.). In addition, I translated the learners' answers for the purpose of writing the paper in English. For my purposes, I chose unstructured questions because I wanted the learners to express their own feelings about testing techniques. Besides, the questionnaires were anonymous which enabled learners to respond freely. The aim of questionnaires was to find out why testing techniques are interesting or difficult for learners from their point of view, and how testing techniques generally influenced their performance. The evaluation of questionnaires was processed through the qualitative research.

Consequently, questionnaires had been pre-tested before the actual usage. I found out that some answers were too general and had to be adjusted. In the question: whether you liked the testing technique, I had to change the question to: "Why did you like the activity?" Mostly, learners answered "yes" and "no" and on this basis, I could not really determine what they liked or did not like.

In order to gather quantitative data, I used other evaluative instrument:

- Rating scale

To use this instrument in the research, I had to construct the rating scale for the needs of my learners (see Appendix 3). The design of the rating scale is analytic because the analytic process enables the tester to see the strengths and weaknesses of learners' performances. As described earlier, the analytic scale is divided into five criteria: grammar accuracy, vocabulary, pronunciation, fluency, and interactive communication. The criteria were discussed and selected with the learners.

Further, the criteria were also selected according to learner's language competence. Each criterion contains five levels with specific descriptions where it helps a tester to evaluate the performances through its analysis. The description of each criterion is suggested in the theory part of the paper. Since, I consider each criterion equal, I did not use the weighting process.

The rating scale has been piloted and practiced before actual testing. The descriptions of some criteria were changed to be more specific, because the evaluation

was subjective and difficult to master. I asked another teacher, from the elementary school where I was doing my 'clinical' year while studying in University of Pardubice, for a help. After the instructions, she evaluated the role-play and picture comparison technique according to my design of the rating scale. Her results were not different from my results; therefore, I continued in using the rating scale for other performances.

Analysis of testing techniques regarding the components of CC

The collection of testing techniques for testing speaking is also evaluated on the basis of the components of CC (see Appendix 1). In the evaluation of testing techniques, the components of communicative competence are all included as in the Canale and Swain theory. It is hard to distinguish what components of CC are included in each chosen testing technique. Every testing technique in this research includes more or less the components of CC. The difficulty in stating certain components of CC rely on the instructions as well. The instructions can either be focused on mainly grammar testing or on discourse competence. This is one of the main reasons why the testing techniques to be evaluated are challenging according to the components of CC. On the other hand, it is obvious that the testing technique is composed of the four basic components of CC.

As a result, I decided to apply the components of CC to the relationship of the rating scale. The rating scale consists of five criteria which should correspond to the components of CC. Through the studies, I proposed to include pronunciation, vocabulary, and grammar accuracy under the grammatical competence. Next criterion – fluency, would be included in the discourse competence, where we can analyse the production of fluent speech. Another criterion, interactive communication, would be a part of sociolinguistic competence. Though, it was demanding to ascribe the particular criterion to the relevant components of CC, the recommendations were done only for the purposes of the paper to be able to compare the learners' results. To see the results, I calculated averages of all criteria in each testing technique (see Appendix 8, Table 1) and further, the average was calculated out of three criteria of the grammar competence in order to analyze and compare all four components of CC.

In conclusion to the background of the testing techniques, I am going to analyse five testing techniques which were chosen according to the learners' competence and to their experiences. Through the further analyses of recorded performances I used my feedback notes to comment on the complete evaluation of each testing technique.

3. Analyses of five testing techniques

3.1. Map-direction

Description

The first technique I will mention is the interactive 'map directions'. As pupils have already used the direction expressions in their lessons, they were familiar with the way of describing the map. Each test taker received a map of a town with all sorts of shops or buildings. At the beginning, they were supposed to choose a shop in the picture they liked, and also to explain why they chose the particular shop. This task should meet the requirements to get familiar with the map design. In the second task learners should give a direction to the tester for a certain place. Finally, after getting to a certain place, in this case – the movie theater, the test takers could talk about their favorite movie – to say the title and if manageable for them, the content of the film as well. The transcription of learners' performances is provided in Appendix 7.1.

Skills to be assessed

The main skill to be assessed in the 'map-direction' technique was to give directions of a town to a teacher, where test takers would be ready for real-life experience. Since learners were supposed to master prepositions, I designed a test for the technique: map-direction. Moreover, the syllabus for the eighth graders contained the types of films which fit into the test. The topics were following each other and there were no interruption elements. Other skills, learners were assessed the ability to answer simple questions, ability to be fluent, and ability to react in interaction.

Authenticity

Concerning the authenticity of this task, I proposed that topic was chosen appropriately from the position of a teacher. Learners had already experienced searching through a map, and when traveling abroad, they would be able to use their communicative skills. I tried to design the tasks to be followed and to be connected to each other. For example, getting to the movie theatre through giving instructions, we could discuss their favorite films or whether they like going to the cinema. Concerning the nature of the language, it was relevant to the learners' needs. In conclusion, the authenticity of the language and the content was conducted from my perspectives.

Preparation project

The preparation of materials was not time-consuming because the students and I were accustomed to the structure of this technique. Since the technique closely relates to the oral interviews, I did not have to prepare written instructions. Mostly, the time was spent reading and consulting the specific literature. In the literature part of the

preparation project, I took notes of tips to follow, for example: if learners get lost, I would help them and give them tasks they would achieve according to their level. The total time devoted to the preparation project was seven hours.

Reactions to the test specification

In the test specification, the map-direction technique was constructed to be focused at the interaction: student – teacher (S – T), which was related to the interview type. After testing, I realized that some test takers had problems answering the tester's questions. Even though this test was an achievement type, learners mostly looked for the words and hesitated to answer correctly. While scoring learners' performances, I could say that map-direction technique was time-consuming because the learners' interactions were short and difficult to evaluate.

The test was constructed of three tasks which were related to the main topic. The planned time for the interview was planned up to five min., but through analysing the testing part, the test was finished in three min. Learners gave short answers and mostly, used avoidance strategies described in the next section.

Analysis regarding the components of CC

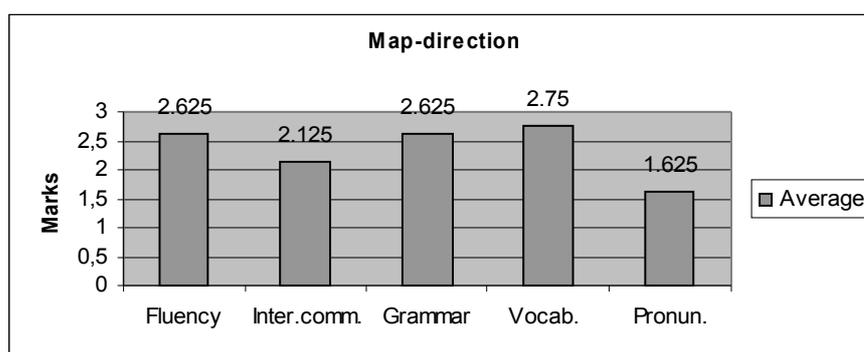
From the standpoint of the components of CC, I tried to analyse the testing technique (see Appendix 1). In my opinion, there was strong strategic competence because learners could choose their own way of how to get to the destination. The learners used avoidance, co-operative and compensatory strategies, for example: "A place...a pet shop because ...are... ehm.. ... animals" (see Appendix 7.1.). Test takers used self-corrections and repetitions as well. For example, "They're.... they weren't... they were having some child...", "Go....go ..turn left."

On the other hand, the sociolinguistic competence, from my point of view, was quite weak in this technique because the interaction was between the tester and the test taker. The social classes were not equal; the teacher had a dominant role. Thus, the interactive communication which is ascribed as a part of sociolinguistic competence reached the average of 2.125, which is measured as the second highest criterion in this technique.

In the discourse competence, there was an impulse for a test taker to find directions from a tester as if he was lost. To this concept corresponds fluency out of the rating scale, where the learners' total average measured was 2.625.

The grammatical competence average was 2.333 which included grammatical accuracy, vocabulary, and pronunciation. The learners' strength among these criteria was pronunciation (1.625) and on the other hand, the learners' weaknesses were seen in the vocabulary criterion (2.75). To illustrate the calculations better, there is a survey of each criterion averages in Table1:

Table 1: The average of criteria in the map-direction technique



Analyses of the learners' questionnaires

Learners had an opportunity to evaluate the testing technique as well according to the questionnaires and to the results of the rating scale. In the map-direction technique, the first question for learners was

→ Why learners liked or did not like the activity?

Half of the respondents did like the activity. They liked the activity because: “It wasn’t boring.”, “Because it was interesting.”, “Because we practice speaking.”, “Sometimes you (a teacher) help me, and when I describe a picture, I try to read it out of it.” There were a few negatives to this question: “I was nervous and I did not know some words.”, “I did not know how to say it in English.” To summarize the answers, the students liked the map-direction and mostly, they complained about the lack of vocabulary.

The next question followed,

→ Why the task was interesting?

The respondents reacted successively: “It was, because I learned new vocabulary.”, “Because I haven’t tried it so often”, “Something new.”, “The situation can happen in your life.” On the contrary, there were negative answers as well: “It wasn’t interesting

because I don't like maps", "The map was small and confusing." Regards to these answers, I think that learners were aware of the advantage of knowing the map-direction technique from their further language experience. It is important for learners to see the other motivations why they are tested in speaking ability.

Further, learners could react to the question:

→ Whether it was difficult to give map instructions or not and why?

Mostly, they did not find the instructions difficult since they answered: "You only added prepositions and directions to the names of the streets.", "It was simple.", "Simple sentences", "We have done it several times."

To the conclusion of questionnaires, I view the learners' answers to this testing technique positively. Mostly, they did not find the activity boring, or difficult, but interesting. Ur states that "...students need a reason to speak more than they need something to speak about; once they have such a reason; however, the fact that the topic is stimulating will make the whole discussion more interesting" (Ur 1981, 6).

Analyses of the rating scale

According to the rating scale, learners' average of the marks was 2.625. Mostly, learners had problems through the criteria like vocabulary, grammar accuracy, and fluency. In the questionnaires, some of the learners admitted the fact that they did not know the vocabulary or appropriate words to use. As an example of grading individual learners on the bases of criteria in the rating scale, see Appendix 7.1.1.

Conclusion

Generally, the map-direction technique can be assessed as practical according to the practicality checklist, authentic, and interesting for learners. Learners have been familiar with the technique and I think they found the technique important for their real-life language experiences. From the point of the components of CC, the technique has positive results in the sociolinguistic competence. In my opinion, this was caused also by the relieved tension of test takers at the beginning, where they were given easy questions and also, I as a tester played a role of a prompter during the testing. With regards to the knowledge about the testing technique evaluation, I would certainly use it for further testing speaking ability because it fulfilled the learners' needs.

3.2. Picture comparison

Description

To introduce the 'picture comparison' technique, I will first describe the technique itself. A learner was given two pictures where he/she had to first compare them at first. The pictures included summer photos from two different places. Further, test takers were asked to choose holidays where they would prefer to go. It was suggested that learners could choose the holidays as a result of the photos or they could share their own ideas. Later on, their next task was to explain why they would choose the place or why not. The technique was selected according to learners' positive attitudes to the pictures. As a matter of fact, the topic was relevant at that time when summer holidays were approaching and the topic corresponded to the syllabus practice as well (see Appendix 7.2.).

Skills to be assessed

The objective of this technique was that children should be tested to compare two different pictures and express where they would have preferred to go on holidays. By this time, learners had been practicing vocabulary related to summer holidays. Other skills, which were assessed, for example: to express likes and dislikes, or to express their ideas, correspond to the overall skills of picture comparison technique.

Authenticity

In my opinion, the topic was interesting for test takers because six out of eight test takers stated that the activity was interesting. The language was natural and comprehensive for the learners. The technique in the first part can rarely be found in the real-life situations; and therefore, it is not authentic. For a test taker it loses the point of completing the task when the tester knows the picture, as there is no information gap. However, in the second task, the learner had an opportunity to express what he/she likes or where he/she prefers to go on holidays. This technique was appealing for a test taker and the tester as well.

Preparation project

In the preparation project, the most time-consuming part was to study the literature. It is recommended to give learners time to look at the pictures and also, the content of the pictures should not be too comprehensive or on the contrary, it should not be too simple either. Additionally, the literature list contained in this project was helpful for finding or revising the points in the testing technique.

In the second part, the materials were all self-made; nevertheless, learners were not guided by any written instructions. The technique was mainly designed to the oral-interview technique and the use of written instructions was avoided. To summarize the preparations, the total time reached to seven and half hours. It was mainly to the study of the literature which provided the theoretical way of testing the technique.

Reactions to the test specification

According to the test specification, the main assessed skill was to compare two different pictures. At the beginning, learners were assigned to compare the pictures by describing them. Therefore, the comparative task has been changed into the descriptive task. The learners easily accepted the task as a descriptive one, as noted because even though the time limit was set to five min., in many cases the test was finished early. It was also structured that learners had one min. to look at the pictures before testing.

By this time, the scoring procedures (as discussed in the test specification) and the testing had already been done for a month and it was clear that there was a high level of comprehension of the rating scale and as a result, evaluation was easier.

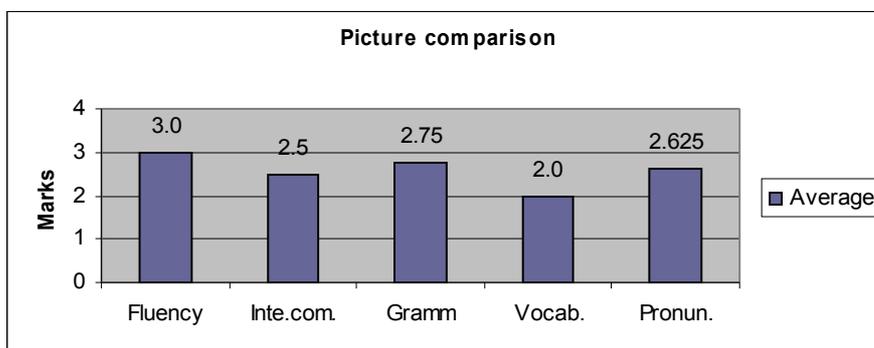
Analysis regarding the components of CC

Looking at the table of the communicative competence, the weakest component in this technique from my standpoint was the sociolinguistic competence. There is an evident dominant role of a teacher interviewing a test taker. The average of interactive communication was 2.5 and it is not the strongest component of the CC.

On the other hand, the discourse component was very strong because the tasks interrelated among each other and there was strong lexical cohesion. To evaluate these two components, it is important to analyse the criterion of the rating scale, in this case – fluency, even though, the average of fluency which belongs under this category was very low (3.0).

To consider the average of all criteria, the grammatical competence of the technique was seen as the relevant part of the testing. The average of grammatical competence became a leader among the other competences. It was 2.46 which specifically: the test takers were mostly successful with pronunciation (2.0) and then, followed by vocabulary success (2.63) and the final position of the success belonged to the grammatical accuracy (2.75). To compare the average of all criteria, see Table 2:

Table 2: The average of criteria in the picture-comparison technique:



In the 'picture comparison' technique the strategic competence was not fully used because test takers used mostly the avoidance strategies to reduce the structures with the aim of transferring the message, for example: "The picture...sea and beach and ...hotels.", "Because....American beach very nice." Even though, different strategies mentioned were used among test takers, they still did not increase better results to the overall testing of the picture comparison technique. On the other hand, test takers used repetitions and self-corrections as well: "Go... we go in swimming pool.", "In Carribic is ...is big big city and eh big sea and beaches and..."

However, when not looking at the averages of the criteria or at the test takers marks, I perceived the learners' strengths to be in the grammatical competence. Consequently, the results out of the criteria to the grammatical competence were analysed as the strongest as well. Besides, learners were talking about the topics they knew in relation to the pictures. Then, the strategic competence was a strong element in this area.

Analyses of the learners' questionnaires

To evaluate the picture comparison task from the point of view of a learner, we have to analyse the answers from the learners' questionnaires. During the testing half of the respondents claimed to be nervous, and we can define the technique as not easy-manageable for learners.

The following question helps us to understand:

→ Why learners liked or did not like the activity?

They liked the activity because: “It is better than to write a grammar test.”, “I like when the lesson is changing, I don’t mind this type of testing.”, “It is fun to talk to the tape-recorder and then, to hear myself.” Two learners did not like the activity because: “I prefer testing with my friend.”, “I did not like the activity because it is boring to describe picture.” Not only do learners appreciate the various interactions but they also like the type of activities, however, some prefer to talk rather than to write. Other analyses related to the characteristics of the task were approached through the question:

→ Why the task was interesting?

This question seemed to be positive because six learners agreed with the statement. Learners found the technique interesting because: “On each picture there was something else.”, “There could be a lot of comparisons found – whether sea or village.” “It was interesting to talk about summer holidays.” Nevertheless, there was also an answer that: “The activity didn’t seem interesting to me.” For sure, there will not be a task which will suit everyone; instead, teachers are at least trying to attract the biggest audience of learners.

Next question is focused:

→ Whether it was difficult to give map- instructions or not and why?

Six learners out of eight did not find the test task difficult because: “From the beginning it was clear, so I didn’t have to compare anything.”, “It was easy because the pictures were different”, “Because I liked the pictures”.

I believe learners only viewed the task positively when colorful pictures were used. Besides, the pictures were taken from real-life situations and learners mostly recognized the village from their area and further, it surprised them at the beginning. Both pictures were received positively, so learners could easily talk about them. Moreover, the testing was one of the shortest interviews and learners did not mind speaking.

Analyses of the rating scale

Considering the learners’ ratings, this technique became the weakest of all testing techniques. The average of marks was 2.687. After reading the questionnaires, learners seemed to be positive about testing; however, the evaluation showed the ignorance of grammar structures, and the lack of vocabulary. But I appreciated when

learners repeated their speech fluently, used simple sentences, used fixed phrases or succeeded to transmit the information gap.

Conclusion

The picture comparison technique was not found to be difficult from the standpoint of a teacher and a learner even though the ratings did not correspond to it. Also, some learners said that the technique was simple and interesting. From the point of view of a teacher, the technique is practical, but in a large class an extra interviewer would be recommended.

3.3. Discussion about “Future”

Description

This testing technique was realized as a mixture of a presentation, comparison and discussion techniques. Learners knew the aim of the test, and had been familiar with the criteria that were to be tested. At first, learners in pairs compared life at the present time and in the future, and they were discussing how life would look like in the future. Further, they discussed what changes might happen in their lives in the future (see Appendix 7.3.).

Skills to be assessed

The technique was aimed at testing the ability of learners to share the ideas among themselves. Learners should be able to make comparisons and express the ideas about their own future. Likewise, they should be able to react to their partner’s answers or questions. By this time, learners have been practicing the present tense expressing future and were familiar with expressing their attitudes. In their eighth grade syllabus, there is a topic devoted to the future. Therefore, the test was designed as an achievement test to find out the learners’ strengths and weaknesses in their speaking ability.

Authenticity

In my opinion, the authenticity of this testing technique was strong, for learners could generally compare the life people live now and life which people will live in the future. Also, they could plan or imagine what their future will look like. And, for the 14-15 years old students, it is an exciting experience. As a result, the future discussion technique can be defined as authentic since the topic is appropriate for learners of this

age and the general topic is combined into those two tasks, which means that the task is contextualized.

Preparation project

The design of the preparation project corresponded fully to the syllabus of learners. The choice of a topic was relevant and learners had been practicing discussion in pairs several times before they were tested in this testing technique. It was a part of the preparation to testing and also, after consulting literature, it was well recommended. Consulting the literature sources was time consuming because it lasted six hours. Due to the literature study, I suggested written instructions for learners. They were written in the target language including the tips of topic that were related to the main topic. Learners were then given choices to talk about things they preferred (see Appendix 4.). The preparation project was not as long as the previous testing technique, only eight hours. It was caused by the fact that the topic was directly proposed on the syllabus basis, and it did not have to be searched or selected for further needs. From the position of a teacher, this technique was thoroughly prepared and planned.

Reactions to the test specification

In the test specification list, the test structure was complex since the skills that were supposed to be assessed were complex as well. It consisted of two main tasks; however, to reach the tasks, there had to be structured pre-tasks, for example: the description and the comparison of life nowadays and life in future. Learners' syllabus included the mentioned topic; therefore, I wanted to test where learners' strengths and weaknesses were.

In my belief, learners liked to express their imaginative future and what was interesting, test takers tried to make up things according to their language competence. The time limit of the test takers' performances was set up to ten minutes, and it was achieved. On the other hand, there were performances in which, it was difficult to stop the test takers' discussion because they got very excited about the topic.

Concerning the scoring procedures, it was structured and completed after the re-listening the tape-recorded performances. It was a long process of transcribing the performances and the evaluation of all individual test takers according to the rating scale criteria, which lasted three hours.

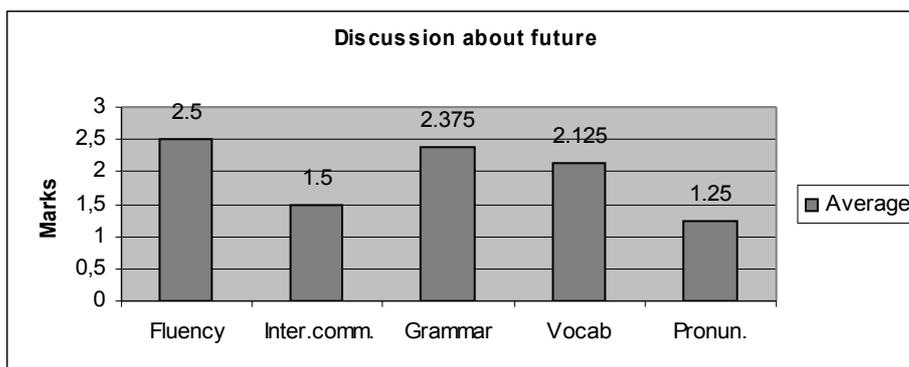
Analysis regarding the components of CC

First, the grammatical competence was focused on the ability to express ideas in the future. There were structures containing negative and positive sentences, furthermore the questions were given in present tense and in future tense. According to the rating scale average and to the criteria belonging to the grammatical competence, the component was considered as strong one since the average was estimated at 1.916. Even though, the grammatical accuracy criterion reached only 2.375. Other criteria namely: vocabulary (2.125) and pronunciation (1.25) enhanced learners to attain better overall marks.

Sociolinguistic competence became very valuable in this specific technique. Learners were tested in pairs based on an interesting topic which enabled learners to talk about the topic in a free and informal manner. However, the strong learners were paired up with weak learners, so that it could have increased the performance of the weak learners. Therefore, the sociolinguistic competence approached the learners' strengths and reached an average 1.5 which was the best result besides the pronunciation measurement.

On the contrary, the discourse competence was evaluated as the weakest component of CC, since the fluency reached only 2.5 marks average. However, the opportunity of pairing up learners between the weak and strong, enabled learners to use similar grammatical or vocabulary structures. As a matter of fact, learners cooperated among each other, and this was the main goal of this technique. To see the average of all criteria, look at Table 3:

Table 3: The average of criteria in the discussion about future



Hence, in the strategic competence, we could find a valuable device by using mostly the co-operative strategies. For example: George saying: “And in nature ... nature will be destroyed.” – And Jessica adds: “Nature ..I believe...I believe that that there nature will be destroyed.” Moreover, the reduction and avoidance strategies were used as well, for example: “I will five dobermans.” or “Everywhere...everywhere will factory and doesn’t will tree.” Then, test takers used repetitions (“I think that people will travel ..will travel to space and and.”), self-corrections (“I have ..I will have got a family.”) or hesitation devices. Learners attempted to communicate among each other, and tried to use different types of strategies as seen in Appendix 1.

Analyses of the learners’ questionnaires

Learners have already been used to fill in the questionnaires. Moreover, they were more self-confident in writing their answers. First, test takers were not nervous during the testing task, which was positive. Second, it is suggested that pair-work techniques have a great impact on the good feelings of test takers.

In the following questions we learn:

→ Why learners liked or did not like the activity?

It was found out that six out of eight respondents claimed that they liked the activity and also, they answered why: “I like to talk about future.”, “I liked the pair-work because I could work with my friend.”, “I liked the task but I would prefer a different topic – what about the holidays.” The rest of respondents did not like the activity because as they said: “I’d rather work alone”, or “I don’t like testing but I like to talk to my friends.” Through the answers, we can get an idea that the topic and even the selected interaction matters to some learners. It depends on many factors, such as: on

the learner's individuality, and his preferences and attitudes to learning. It is essential that a tester or a teacher at least tried to satisfy learners' needs as much as possible.

The following question introduces regarding strengths of the testing technique:

→ Why the task was interesting?

Half of the class found the activity interesting and they explained why: "Future interests me.", "It was interesting to say what kind of future...", "Better than looking for vocabulary.", "It was funny to learn something about my friend's future." The other half of test takers did not find the activity interesting because: "It would have been interesting if I could have talked about something else like sports, etc.", "I don't know." or "I don't like anything today." The responses were mostly oriented at the topic of the task. We can conclude that learners were interested in discussing the future, and the others would have rather enjoyed talking about a different topic.

Next question asked, was concerning:

→ Whether it was difficult to discuss the future or not, and why?

For only three pupils, the task was not difficult. Others claimed that "I was unable to talk about this topic.", "It was difficult a little bit because I have a lack of vocabulary.", or "I don't know vocabulary." Again, learners thought that the discussion task was difficult with regards to their lack of vocabulary.

In conclusion to the analyses of questionnaires, we can say that the topic for discussion was well received among learners. Learners preferred to speak and be tested with their friends. Nevertheless, there were negative responses about the difficulty of the task. Test takers explained that the difficulty was due to the lack of their language knowledge. Even though they liked to talk about the topic, the problem was with lexis. In this case, I could note that learners relied on the feeling and fact that they must have learned all the words. Mariani mentions that "We are still very much concerned with exact communication – something which perhaps does not even exist" (Mariani 1994). Learners try hard to speak correctly, and forget or are not accustomed to use paraphrasing or guessing strategies which are still difficult for them to manage.

Analyses of the rating scale

On account of the evaluation of questionnaires, it is important to consider the marks from the rating scales criteria. The average of learners' total marks was 2.25. There was no test taker who got a mark worse than 3. To explain this, one should

remember the divisions of pairs for the strong and weak test takers. This is probably shown by the successful marks of weak learners that were enhanced by the strong learners. Incidentally, the marks corresponded to the test takers answers in the questionnaires. Learners were positive about the technique and so were their marks.

Conclusion

To conclude this technique, it has positive face validity. Learners were satisfied with the technique; the topic was interesting for them. Moreover, they discussed and presented the topic in pairs. From the viewpoint of a teacher, the technique was difficult in regard to the planning, preparation, and evaluation of the scoring procedures. The most challenging question dealt with the fact of whether or not to pair up strong and weak students, but this proved to be a good decision.

On the other hand, the testing technique from the point of a learner was satisfactory. Learners only complained at the lack of subskills that they saw as a problem in this technique.

3.4. Role-play

Description

The role-play technique as a test was enjoyable for test takers. Before the task started, learners were given instructions, which they were supposed to read before the test. Every test taker received different instructions, either A or B list. Then, I explained that their task was to persuade each other about their goal and come to a final decision together. Also, there were hints in the instruction paper which helped learners to choose what topic they would like to raise. The first impression of the test takers who were supposed to play a parent was not motivating at all; however, learners easily adapted to their role-plays (see Appendix 7.4.).

The technique was chosen with regards to the learners' age and to the language competence they had. Role-play is a well-known technique and the learners practiced it many times during their English lessons. The purpose of the technique was targeted at the achievement test.

Skills to be assessed

Certainly, learners should be able to persuade their classmates and give good reasons to do so. Besides, the expressions of agreement and disagreement should be also

tested, since learners have been practicing them in their class. Moreover, according to their syllabus they are obliged to express their opinions and ideas. Test takers were aware of this, since the main skills, which had been described, were included in the role-play technique.

Authenticity

The language test takers used for their role-play technique was natural because pairs of test takers were combined at the similar level of language competence. The topic was chosen to attract the test takers attention. In fact, we can classify the content as authentic; however, this technique cannot be prescribed to real-life situations. The way of playing the parent and the child amused learners but did not seem to be real in this case. They were happy to try these roles, especially role-plays of children who could have gone to the birthday party very late if parents allowed them. This situation would be more possible if I was a parent instead of a tester. Learners also needed the visual connections to get into the situation better; therefore the characteristics of the parent role were difficult to comprehend for learners. To summarize the task, I would classify the content of the technique authentic, but the technique itself as not authentic.

Preparation project

To prepare for the technique, I also searched sources to get some ideas or suggestions how to test the role-play most successfully. Consulting the literature was not as demanding as in other testing techniques preparations as it lasted only four hours. Through the literature, I concentrated more on the role-play topics and situations in which there were some conflicts to solve or where a part of the information was missing.

In addition, the process of preparing the materials and instructions was considered, so that test takers had clear ideas about the testing. They were delivered written instructions of the task. The instructions given to the test takers should encourage them in speaking and help learners to choose a topic which they had been familiar with.

It was difficult to select the content of the situation so that it would motivate learners. As Heaton advises: “Remember that stimulating material can reinforce learning, whether used as part of your teaching or your testing” (Heaton 1990, 24-25).

The total preparation time took six hours but the test construction had to be designed as well.

Reactions to the test specification

The test specification list focuses on the large amount of skills contained in the task. Test takers are thought to be prepared for the skills ahead of time by practice through the syllabus in this grade. With regards to the syllabus, learners must be able to react in the prescribed role and persuade someone that he/she is or is not wrong. Therefore, the achievement test was aimed at these skills and the content applied to the way of teenagers' behavior.

Also, the way of scoring procedures was planned and seen as time-consuming. The interval of three hours was sufficient to evaluate the test takers' recorded performances. This technique was later discussed with learners and their performances were re-recorded. The feedback was beneficial for learners because they could figure out the mistakes they made or found and learned new strategies for their next dialogues.

Analysis regarding the components of CC

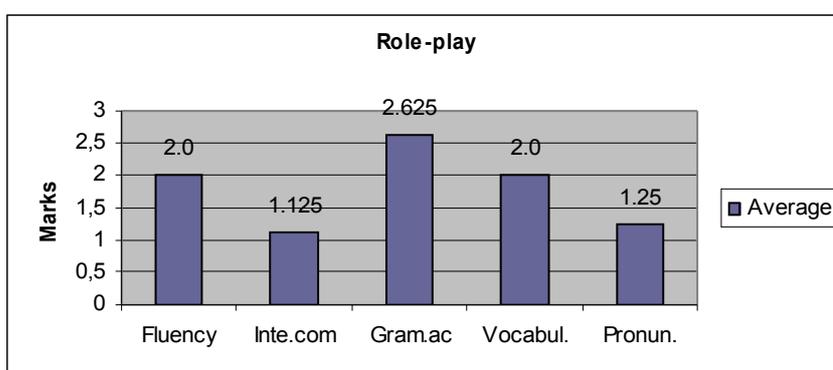
Concerning the components of CC, I can state that the strongest components of CC are the sociolinguistic, discourse, and strategic competence. First, the task was perceived in a pair-work pattern: learners used an informal register and the appropriate level to each other. Mostly, learners chose their partners according to the same language competence. There was a strong sociolinguistic competence; however, the role of a parent can be considered as socially higher than the role of a child in the testing the speaking ability. Through the rating scale criterion, the interactive communication was high at 1.125 which corresponds to the strengths of sociolinguistic competence.

Similarly, in the discourse competence, test takers talked about the topics they were assigned and there was not one who would have talked out of the topic or would not have been cohesive and coherent in a speech. The lexical cohesion mainly concerned the family discussion which was, in fact, the shared topic of both partners. Learners tried to get the message across in order to complete the task and it was considered as a great start, for example: in speaking and in communication with a foreigner. The rating scale criterion – fluency was estimated at 2.0, which belongs to the successful results of the role-play technique.

At the same time, the strategic competence was considered as a strong component during the learner’s performance from the teacher’s view point. Learners at approximately the same level could give ideas to each other and they used compensation and avoidance strategies (“But we ...holidays.”, “I homework – it’s come. I will do homework, then B-day party.”) Further, learners repeated words or phrases, for example: Bill saying: “You don’t go party.” - Mike answers: “Please?”- Bill: “You won’t go.” – Mike: “Please?” Moreover, test takers used formulaic expressions (“You are kidding.”) or self-corrections (“No, I will I do my homework I did, I was... did...my homework.”)

In practice and mostly in testing, the grammatical competence is still “a big wall” for test takers. First, they think about the possible ways how to say things correctly. It was supposed that the grammatical competence would test future expressions in relation to present tense, the usage of prepositional phrases, imperative and interrogative moods or testing fall-rise intonation, the rhythm or sentence stress. However, learners did not succeed in all stated grammar issues and in some cases they had even problems starting. Second, they were shy to say things wrong and it is a common problem among learners. Therefore, the grammatical accuracy achieved was 2.625, vocabulary (2.0) and pronunciation (1.25). In total the average for grammatical competence was 1.958 which seemed to be a strong feature for learners as well. To illustrate the average results of criteria better, see Table 4:

Table 4: The average of criteria in the role-play technique



To sum up the CC components, this technique has high pros of sociolinguistic and discourse competences. The grammatical competence is among the strong elements

as well, and lastly is the strategic competence. In my opinion, test takers became more successful if they were tested in pairs which they had selected themselves, and thus the communicative competence can be shared by both candidates.

Analyses of the learners' questionnaires

From the point of a learner, we must analyse and evaluate learners' questionnaires (see Appendix 7.4.1.). At the beginning, candidates responded that five out of eight learners were not nervous during the task. Instead, the task excited them. To analyse the questions in relation to learners' responses will follow. In the first question we learn:

→ Why learners liked or did not like the activity?

Six learners liked the activity and explained why: "It was interesting.", "I tried to give questions in order the answers would be YES.", "Because it was quite enjoyable.", "It was a fight.", "My arguments were better.", "I could win over my friend." Through the answers, we can conclude that learners achieved a success in persuading the other learner and therefore, they liked the activity.

Next question can give answers to:

→ Why the task was interesting?

For six test takers, the task was interesting because: "I tried something new.", "The quarrel was rather interesting.", "I would get into a role easier if George was with me.", "I have the same discussion with my parents over and over." Mostly, learners enjoyed the quarrel between the parent and the child because learners were influenced by this topic through their own experiences and it will always be a topic where they can easily express their ideas, views, and suggestions.

→ Whether it was difficult to persuade the other learner or not and why?

Only two learners found the activity difficult and here are their explanations why: "Because I did not know how to answer to his questions.", "I wanted to be a child instead of a parent." Next group of test takers viewed the activity not difficult because: "It was not difficult at all to get myself into a role of child.", "I was quite successful to use and remember the right expressions." When evaluating the difficulty of the task, learners realized their own mistakes and in some cases, they rather preferred to be someone else. They thought if the role had been different, they would have been more successful. This is important to know when testing another role-play.

Mainly, the responses corresponded to the learners' individuality and to their language competence when evaluating the task. The purpose for the evaluation of these techniques or tasks was to design a technique which would be similarly appropriate for all learners: nevertheless, every learner is unique and as a result we have to design different kinds of tasks which would be suited to heterogeneous learners' needs.

Analyses of the rating scale

Through the rating scale, I can conclude that learners' interactive comprehension, one of the criteria in the rating scale, was well realized. Learners were discussing, persuading each other, and quarreling to reach the aimed goal. The communication among test takers can also be increased by the fact that learners matched in pairs by themselves and there was not any division, for example as strong and weak learners as in the previous pair-work task. The average of learners' total marks was 2.062. Based on re-listening to the performances, I can conclude that learners searched for vocabulary during their speech and were aware of their mistakes. When they realized a mistake, some of the test takers repeated the structure or word and fixed it in a correct way. Other participants tried to deliver the message without any worries that they might say it incorrectly. Learners were persuasive to support their reasons so that they would have won the "family struggle".

Conclusion

As we have seen, the role-play can be very diverse. It depends on the content of the task, on the interaction pattern of test takers, and also on the purpose of the task. In my opinion, the role-play was practical and authentic due to the task content. Learners were enthusiastic and got involved in their prescribed roles, and they used simple language to be comprehensive, used simple strategies to avoid the unknown words or grammar, and repeated themselves to be fluent.

In some cases, it was difficult to warn examinees of the time limit because they were playing good actors. For further recommendation, I would like to add that learners needed some hints which they could follow. According to the preparation project, I compiled a list of instructions which was a helpful tool for test takers. It is important for learners to imagine the situation, in order that they would perceive the task first. Therefore, we must prepare a good plan of the test specification which should be based on learners' syllabus and on validity.

3.5. Re-tell a story

Description

The testing technique, re-telling a story, contained a combination of the description and the re-telling of a task. At the beginning, learners were acquainted with the criteria and the aim of the test. Further, they were given the written instructions. The technique was conducted in pairs where the first test taker described his sequence of pictures and the second test taker listened to the story without seeing the pictures. Then, the roles switched and by the end, test takers could make an idea of the complete story. Instead of a plain description of the story, I used an information gap method, so that learners could figure out the story by themselves without looking at the pictures. The last task was to re-tell the complete story individually. This technique was chosen because it contained the criteria for the selection of testing techniques in consideration to the language level, and to the age of test takers. To get an idea of learners' performances, see Appendix 7.5.

Skills to be assessed

Learners were assessed by describing simple pictures using their limited vocabulary, to construct the story completely and at the end, to re-tell the story by themselves. According to their syllabus, they should be able to re-tell the story which was based on their previous lessons, where they re-told similar stories.

Authenticity

Concerning the authenticity of the task, I must conclude that language authenticity was appropriate to the test takers' level. The instructions contained vocabulary and structures learners had known. Even though pictures were similar to real-life situations, the content was prescribed; therefore, I did not consider the content of this technique authentic. If the story was told from the test taker's perspective or their experience, it would be sufficient for the authenticity. On the other hand, it was better to select the ready-made picture story for the testing technique which interested and motivated learners in generating good results.

Preparation project

While designing the preparation project for this special technique, the preparation itself took eight and half hours. The most time-consuming part was devoted to the consultation and researching the literature. Regarding consulting literature, there

are tips which we can follow or avoid while testing the appropriate testing technique. Briefly, the helpful point was to find the attractive and enjoyable topic for learners. We know that when learners enjoy the activity, they are also good at it and hence, they can receive good marks. On the other hand, the most difficult to keep in mind was the fact that when evaluating the learners' performances we have to avoid the prejudices of learners and be consistent.

Due to the materials use, I can state that the preparation of written instructions for test takers enhanced their performances. Learners concentrated on the tasks they were supposed to do and were not confused. Moreover, before learners were given written instructions, they were given instructions orally by a tester, and it enhanced their understanding. In conclusion, the preparation project showed some problematic parts in planning the testing technique. There were difficulties to find the suitable content of the story for test takers, which would attract their attention and would interest them. From my point of view, this topic content was relevant and enjoyable in this use of testing technique.

Reactions to the test specification

The task's interaction was a pair-work where test takers selected their partners for testing. The picture story also corresponded to the learners' age and language level, since the structural range contained the present and past tense, the use of pronouns, complex sentences, transitive action verbs, descriptive adjectives, sentence stress, and accent. Further, the vocabulary was appropriate to the learners' level and to the picture story as well. Concerning the test structure, test takers are obliged to complete two tasks. This is partially typical for the whole testing of the speaking skill; however, to conduct the picture story, I made use of the information gap in the technique, so that test takers had to negotiate the story first and then re-tell it. This had to be done in ten minutes; however, during the testing I found out that the time was not sufficient for all learners and the combination of three tasks prolonged the testing an additional five minutes. In the scoring procedures, the methods used were the same for each testing technique; it lasted three hours to complete the evaluation of the performances. The test specification was enhanced by the preparation project construction that contributed to the testing of re-telling a picture story.

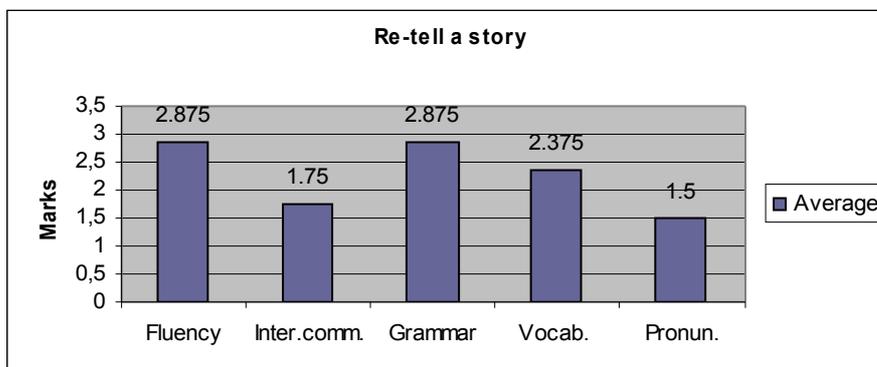
Analysis regarding the components of CC

Regarding the analyses of the components of communicative competence, we can look at the testing technique from four different angles. The grammatical competence was partly mentioned in the analyses of the test structure. It contained parts of the language structures, and it gives an idea to the evaluator to focus on grammar overall. Looking at the rating scale, the criteria as pronunciation, vocabulary, and grammar would be corresponding to the concept of grammatical competence which is evaluated as 2.25 in total. The weakest criterion in this technique was grammar accuracy (2.875) and on the other hand, the strongest criterion was pronunciation (1.5) and vocabulary with the average 2.375. In my belief, the grammatical competence still causes problems for learners; even though, the practice was sufficient. But the most appreciative is the fact that learners attempt to communicate and transfer their communicative messages.

In order to show the interactive communication, we can evaluate the sociolinguistic competence. The total average of interactive communication is 1.75. This result is a strong element in the re-telling of a story because the sociolinguistic competence used 'Student – Teacher' interaction where a student is submissive and a teacher dominant. Further, the interaction changed and was followed by the discussion between two students. In this case, the level of interaction was appropriate and informal than in the first case. Moreover, in the discourse competence a student produces a story using descriptive language on his/her own. We can say that the combination of interactions towards the sociolinguistic competence was successful and well selected for the technique.

In the discourse competence, we can see a close relationship to the sociolinguistic competence. Learners attempted to persuade the other partner in order to achieve his/her goal. Though the average of fluency criterion was 2.875 high which in overall evaluation of the testing technique was weak as the grammar accuracy in this case. To compare the results of each criterion in the technique, see Table 5:

Table 5: The average of criteria in re-telling a story



As stated, the strategic competence was very weak according to the usage of test takers. Learners not only used co-operative strategies between each other, but between the tester as well, (for example: a teacher asking: “Are they happy or unhappy?” – Michael answers: “Are they happy and unhappy.”). Furthermore, compensatory and avoidance strategies were used as well: (“They go ..tickets.”, “They are café.”). They did not enhance their communication by paraphrasing strategies but used self-correction and repetition devices: (“And they ...they are they was in director’s box and watching football match.”). Learners avoided or dropped the word rather than to paraphrase it or describe it in a different way.

In conclusion to the communicative competence, the testing technique totally estimates the strong results that contain less strategic competence and more sociolinguistic competences. Through discourse competence, learners attempted to succeed in communication of the prescribed tasks; even though, the discourse competence was a weak component. The stress was also on the grammatical competence; however, its average was not successfully managed in this technique.

Analyses of the learners’ questionnaires

At the beginning of the questionnaires, I wondered whether learners were also nervous during testing the re-telling of a story. Half of the learners were not nervous which surprised me. I thought that the seating of two learners together would not have caused any nervousness for learners; besides, it should have increased their communicative skills, but the interaction between ‘Student – Teacher’ could have become more problematic during the testing.

Next question was aimed at:

→ Why learners liked or did not like the activity?

Three learners liked the activity because: “It was quite enjoyable.”, “I liked the story.”, “Because, whenever I speak I feel more and more confident.” Three learners did not really know why they liked the task and the next part of learners did not like it at all because: “I don’t know what to say.”, “I was nervous and I didn’t know some vocabulary.”, “I am disoriented in grammar and this makes the difficulty for me to speak. Therefore, I did not like the activity.”

As learners pointed out, there are still obstacles for them to speak in a freer way. They thought it was caused by their lack of vocabulary or grammar knowledge, and then, they became nervous or hesitant. On the other hand, strong learners enjoyed the task because he is confident or he enjoyed the story. The likeness of the task is closely connected to the learner’s knowledge of the language.

Further, I wondered:

→ Why the task was interesting for test takers?

Three quarters of test takers found the activity interesting because: “It seemed to me that all pictures were interesting.”, “It was interesting to learn the other half of the story.”, “The activity followed and it was a different story than we have in our books.” On the other hand, learners claimed that the activity was not interesting because: “I don’t know if the task was interesting but I’m afraid to say nonsense when I am speaking.”, “The pictures could be colorful.”, “I was confused to hear only a half of the story from my partner. I did not understand him what he said.”)

Under the circumstances, I learned that learners appreciated the innovations of the test structure and the combination of the tasks and the story itself as well. Nevertheless, the fact that learners did not find the task interesting because they did not understand his partner or their belief that the pictures were not colorful, was surprising. It gives me an idea what to focus on in the next testing the re-telling of a story.

Last question concerned:

→ Whether it was difficult to describe and re-tell a story or not. And why?

Only two candidates found the activity difficult because “I could not manage to say sentences in a chain, in order to re-tell a story.”, “Because I did not remember the story.” It is advisable to mention that the story for re-telling was too long for some test takers.

To conclude the learner's evaluations of the testing technique, we can recognize that for weak learners who were troubled with their background language knowledge, the task was extremely hard. In some cases, learners did not remember the story and to re-tell a story was too demanding than to describe the pictures. Despite this, learners tried to grasp some positive attributes of the technique. The combination of tasks in this testing technique was interesting and motivating for learners to speak better.

Analyses of the rating scale

In the same way, we can see learners' results in this particular technique where the total average of marks was 2.437. To analyse the individual criteria, it is important to look at the chart of marks where fluency and grammatical accuracy became very weak criteria in evaluating the learners' performances. It is obvious that learners are not skilled in communicative strategies yet and still, they miss the main grammatical rules; even though, it had been practiced many times. In addition, pronunciation and vocabulary criteria increased learners' average marks in most of the cases. Similarly, the interactive communication became an advantageous mark in evaluating the tasks.

Conclusion

Learners try to get the message across, so that the tester would understand and it is a main goal for a teacher to teach communicative competence. This testing technique was prepared and more stressful for learners and for the teacher as well. There may be many reasons, for example: the combination of two testing techniques in one activity or the length of the story or its content.

Still, I was satisfied with the results and I can only find a few mistakes which I would avoid the next time. First, the interaction could have been changed, so that learners could put the story together and not individually. Second, test takers would get more time to spend on the preparation where they can prepare thoroughly. These ideas could increase learners' motivation in testing and the technique could be evaluated more positively.

Conclusion of the research

By testing speaking ability and through evaluating the testing techniques, I have been persuaded that testing speaking skill is an essential procedure for a teacher and for a learner as other testing skills and subskills.

Due to the description of each testing technique, I can conclude that all tests were achieved as designed. Through Littlewood's division of communicative activities, the testing techniques were mostly focused on functional communicative activities. I designed the testing techniques according to the need of the information gap activities. As an example, in the map-direction technique, which lacked the information gap itself, I added two extra tasks, so that learners transmitted their message to the tester.

Concerning the reliability and validity of the tests, I tried to use the material which had been proved to be reliable and valid. In addition, I managed to show that the importance of validity and reliability of evaluative instruments must be considered as well. I have piloted the questionnaires, the self-designed rating scale, the preparation project, and the test specification. To see the reliability of testing techniques, other research would be required; however, I focused on different principles in the research, such as: the analyses and evaluation of communicative competence in relation to the rating scale criteria.

As mentioned earlier, the evaluative instruments such as the preparation project, the test specification, and the practical checklist, enhanced the process of testing. Since I had little prior testing experiences, I had to design these lists in order to follow the structure of the testing and to evaluate the testing techniques. Furthermore, I gathered qualitative and quantitative data, as in the example of measuring time in the preparation project. The most time-consuming testing techniques became "re-telling a story" with eight and half hours taken for the preparation. The long amount of time included studying the relevant literature and selecting material adequate for the learners.

In terms of practicality, testing techniques in this paper were considered to be practical and useful for a teacher and for learners, except the process of scoring that was found to be as time-consuming.

After the data gathering, which were related to learners marks, I must conclude that the testing technique: "role-play", which received an average 1.8, can be compared to the best achievable technique among all the learners' performances (see Appendix 8, Table 1). However, it surely depends on many factors which could have influenced the

average, for example: the choice of a topic, the difficulty of the task, the classroom atmosphere, the interaction patterns etc.

Through gathering data qualitatively, learners concluded that the testing technique: “discussion about the future”, is the most interesting of all; and therefore, they enjoyed talking about this topic. In Appendix 8, Table 2 illustrates what testing techniques were achievable from the learner’s point.

Since the testing techniques from the viewpoint of a learner were evaluated through the anonymous questionnaires; test takers came to the conclusion that through testing speaking, they are able to express their attitudes, likes or dislikes, and their feelings. In addition, through the questionnaires (Appendix 7.4.1.), which were completed after testing, learners also self-evaluated themselves and found out why the techniques were interesting or difficult for them. Therefore, the questionnaires also had another effect for learners as well. The questionnaires enabled me to find out whether test takers view techniques as appropriate or useless. In my opinion, the tests and techniques were positively perceived by learners; thus, the tests contained positive face validity.

Because of data, materials and the process of testing itself, we can conclude that the choice and evaluation of testing techniques is based on criteria mentioned earlier. After evaluating tests and analysing data, the testing technique: “discussion about the future”, showed positive results in learners’ performances and as well positive learners’ perceptions of the questionnaire data.

Moreover, the evaluation of selected testing techniques from the viewpoint of a teacher and a learner reached in this paper can be helpful for beginner teachers who can get positive ideas and attitudes toward testing speaking. In my opinion, the research of the evaluated testing techniques also had a good impact on learners who realized their strong and weak sides of their language communicative competence.

To conclude, I believe that the evaluation of testing techniques of the speaking skill is still a “hot issue” that needs to be examined, re-proved and practiced. As mentioned, communication and speaking are important features of learning the target language and teachers should be consistent in testing the speaking skill.

I can state that I was successful in gathering data of the evaluation of testing techniques. Moreover, the results of the research were beneficial to get and view the

testing procedures. In addition, learners were positively influenced by testing speaking because their speaking results have been improved by the end.

Resumé

Tématem mé práce je dovednost mluvení v cizím jazyce, zejména jak tuto dovednost ověřovat. Vzhledem k tomu, že dané téma je velmi obsáhlé, obracím svou pozornost především na hodnocení jednotlivých ověřovacích, testovacích technik. Hodnocení se týká jak žáků, tak i učitele. Proto cílem této práce je zhodnotit testovací techniky dovednosti mluvení z pohledu učitele a žáka.

Protože si myslím, že znalost komunikace v cizím jazyce je důležitá, musí ji provázet také ověřování neboli testování této dovednosti. Komunikovat v cizím jazyce je velice náročný proces, který se utříbí a zdokonalí jedině tím, že člověk bude komunikovat pravidelně a aktivně. Zároveň je však důležité ověřovat znalosti, aby žáci byli schopni tuto dovednost mluvení praktikovat a rozvíjet ji. Ověřování musí ale probíhat za určitých znalostí a okolností napodobujících situaci reálného života, proto se věnuji i teoretickému seznámení se s znalostmi a jejich definování.

Úvodní část práce se zabývá popisem a analýzou dovednosti mluvení, dále pak rozborem výsledné komunikativní kompetence, jež se stala náplní výzkumné části. S tím úzce souvisí i komunikativní aktivity, kde kladu důraz hlavně na účel komunikace. Uvádím zde též demonstrativní rozbor komunikativních aktivit podle W. Littlewooda.

Další kapitola se teoreticky soustředí na definici, druhy testů a testování dovednosti mluvení. Čtenář se zde seznámí s kritérii, které by měl test splňovat, a poukáže na praktičnost, autentičnost a přípravu testu. Zmiňovaná kritéria jsou poté vybraná pro výzkumnou část.

Následující kapitola vysvětluje výraz, jenž se objevuje v názvu diplomové práce, a to ověřování dovednosti mluvení. Poukazuje na smysluplnost otázky: „Proč je důležité hodnotit tuto dovednost?“ a směřuje k definování testovacích technik, ve kterých se věnuji v samotném výzkumu. Testovací techniky obsažené v teoretické části jsou vybrány na základě kritérií, která jsou dále vysvětlena a jsou poté zhodnocena. Současně se také zabývám způsobem ověřování dovednosti mluvení při zkoušení, zda upřednostnit zkoušení individuální, ve dvojicích nebo ve skupinách. Uvádím zde výhody a nevýhody těchto metod ověřování. V závěru této kapitoly charakterizují způsoby a metody, jak hodnotit testovací techniky.

Závěrečná kapitola se zabývá obecně hodnocením výkonů žáka a poukazuje na nástroje, podle kterých je hodnocení žáků, ale i testovacích technik, použitelné

vzhledem k věku a jazykovým schopnostem žáků. Z teoretických poznatků se dozvíme, jak pomocí nástrojů (např. dotazníků, hodnotících škál a používání nahrávek) ověřovat dovednosti mluvení žáků. Na úplný závěr této teoretické části se seznámíte, jak by teorie měla prolínat výzkum.

V úvodu výzkumné části podávám základní informace k danému výzkumu, který má přiblížit testovací techniky. Jedná se např. o to, kde a kdy výzkum probíhal, a dále přibližuje postupy realizace výzkumu a popisuje kroky, které vedly k vyhodnocení výzkumu. Jak již bylo uvedeno, stěžejním cílem této práce je zhodnotit a zanalyzovat vybrané testovací techniky z pohledu učitele a žáka.

Druhy testovacích technik byly vybrány vzhledem k věku a jazykové způsobilosti žáků. Na základě rozboru komunikativních aktivit podle W. Littlewooda, jsem upřednostnila „funkční komunikativní aktivity“, které byly také předpokladem pro výběr testovacích technik. Do výběru testovacích technik jsem zařadila: popis mapy s pohovorem, popis obrázků, popis a převyprávění příběhu, drama ve dvojici a diskuse na téma „budoucnost“. Všechny testovací techniky byly připraveny a naplánovány předem se zřetelem na věk a délku studia cizího jazyka. Žáci byli zkoušeni ve dvojicích nebo individuálně, záleželo na druhu testu i kontextu.

Z pozice učitele se analýza testovacích technik týká hodnocení různých aktivit v rámci popsané komunikativní kompetence, s níž jsme se seznámili v teoretické části této práce. Pomocí M. Canala a M. Swaina jsem vybrala komunikativní kompetenci, která obsahuje čtyři typy dílčích kompetencí: gramatickou, sociolingvistickou, vzájemnou a kompetenci projevu. V každé testovací technice by měly být zahrnuty všechny tyto čtyři dílčí kompetence a mým záměrem bylo zjistit, do jaké míry byly právě na základě hodnocení testovacích technik tyto kompetence docíleny. Pro tyto a další účely byla navržena hodnotící škála, podle které byly projevy žáků vyhodnoceny. Hodnotící škála se skládala z pěti kritérií: gramatická správnost, výslovnost, slovní zásoba, vzájemná komunikace a plynulost projevu. Pro přehlednější porovnání testovacích technik byly výkony žáků vyhodnoceny pomocí školní klasifikace (známku 1 až 5). Tato kritéria dále sloužila k porovnání výsledků komunikativní kompetence. Nejprve byla kritéria hodnotící škály zařazena pod jednotlivé typy komunikativní kompetence s ohledem na jejich zaměření a popis. Gramatická správnost, slovní zásoba a výslovnost byly zahrnuty do gramatické kompetence. Vzájemná kompetence byla součástí

sociolingvistické kompetence a plynulost projevu částečně spadala do kompetence projevu. Jedině strategická kompetence byla analyzována na základě výstupů žáků a porovnána se strategiemi podle M. Bygate.

Z důvodů objektivitu ověřování a nedostatku další kvalifikované osoby (kdy jedna osoba vede celý proces ověřování a druhá hodnotí výkony žáků) jsem výkony všech žáků nahrávala na magnetický pásek. Jelikož jsem oba postupy prováděla sama, hodnocení žáků probíhalo mimo proces ověřování, tedy mimo vyučovací hodinu. Metoda hodnocení probíhala následovně: přehrání a zdokumentování výkonů žáků, vyhodnocení výkonů na základě kritérií z hodnotící škály a zprůměrování výsledné známky. Žáci obdrželi výsledky v následující hodině s tím, že jsem jim zpravidla přehrála jejich výkony pro získání zpětné vazby.

Numerické výsledky zmiňovaných kritérií byly sečteny a zprůměrovány. Z těchto výsledků můžeme vyvozovat závěr, která testovací technika dosáhla nejlepšího průměru a která z technik byla pro žáky zvládnutelná, nebo naopak. Žáci dosáhli nejnižšího průměru v testovací technice „drama ve dvojici“. Z toho vyplývá, že způsob provedení, výběr tématu a vyhodnocení testovací techniky se projeví jako nejúspěšnější. Naopak testovací technika „popis mapy“ dosáhla nejvyššího průměru mezi zkoušenými, a tím se ukázala jako nejméně úspěšná.

Dalšími nástroji hodnocení testovacích technik z pohledu učitele byly přípravný projekt a specifikace testu. Tyto nástroje měly napomoci samotnému testování a zjistit, zda přípravné fáze testovacích technik byly dostatečně propracované a neměly vliv na hodnocení výkonů žáků.

V případě přípravného projektu jsem např. počítala čas, který byl vynaložen na přípravu testu. Podle dostupných materiálů mohu zkonstatovat, že testovací technika, která byla časově nejnáročnější na přípravu, se ukázala technika „popis a převyprávění příběhu“ s naměřeným časem osm a půl hodiny, přičemž potřebný čas byl věnován prostudování materiálu a výběru vhodného tématu tak, aby žáky motivoval svojí zajímavostí.

Dokumentace ke specifikaci testu sloužila současně jako osnova samotného testu, ve kterém jsem mohla naplánovat celou jeho strukturu (např. brát v úvahu dovednosti, které jsme měli hodnotit v procesu ověřování, vymezit čas, vybrat téma, naplánovat dané úkoly do částí a uvést, jaké poznatky budeme ověřovat, atd.).

Ověřování dovednosti mluvení z pohledu žáků bylo dosaženo pomocí známek dle hodnotící škály a dále z dotazníků. Tyto dotazníky byly navrženy pro každou testovací techniku zvlášť a žáci je obdrželi k vyplnění hned po skončeném testu. Hlavním záměrem dotazníku bylo zjistit, z jakých důvodů byla testovací technika zajímavá nebo obtížná. Z tohoto kvalitativního výzkumu bylo zjištěno, že testovací technika „diskuse na téma: budoucnost“ splňovala očekávání žáků. Tato technika byla zajímavá díky diskutovanému tématu. Dále zkoušení na toto téma ve dvojici umožnilo žákům volně hovořit o dané situaci. Vedlejším produktem dotazníků bylo sebehodnocení žáků. Žáci uvedli například, že v procesu ověřování dovednosti mluvení měli slabou slovní zásobu na to, aby mohli úspěšně komunikovat v modelové situaci.

Na základě těchto poznatků můžeme porovnat, zda se výsledky ověřování této dovednosti ztotožňují s osobním pohledem žáků na danou testovací techniku. Podle mého názoru, žáci si uvědomují své úspěchy i neúspěchy a jsou schopni sebekritiky. Např. technika „diskuse na téma: budoucnost“ dosáhla jak úspěšných výsledků na základě hodnotících škál, tak i pozitivních ohlasů z pohledů žáků.

Z dostupných materiálů, příkladů a poznatků jsme se seznámili s postupem a výsledky při ověřování dovednosti mluvení. Myslím si, že je potřeba provádět více výzkumů pro ověřování této specifické dovednosti. A abychom mohli tento požadavek splnit, potřebujeme více praktických zkušeností a poznatků, které následně můžeme předávat svým kolegům.

Závěrem bych ráda podotkla, že zkoušení neboli ověřování dovednosti mluvení je pro žáky, kteří se učí cizímu jazyku, velmi důležité. Dovednost mluvení je nutná jak pro samotnou komunikaci, tak pro porozumění dialogu, rozhovoru, atd. Zkoušení by mělo žáky motivovat a napomoci dosáhnout jejich co nejlepší komunikativní dovednosti. Výsledky vyhodnocení testovacích technik by mohly napomoci nejen učitelům, kteří na jejich základě při ověřování dovednosti mluvení svých žáků mohou upřednostnit výzkumem prokázané úspěšné testovací techniky, ale i žákům v motivaci k aktivní snaze dosáhnout optimální dovednosti mluvení.

Bibliography:

- Alderson, J. Charles, and Clapham, Caroline, and Wall, Dianne. *Language Test Construction and Evaluation*. Cambridge: CUP, 1995.
- Bachman, Lyle, F. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press, 1996.
- Brown, Douglas H. *Language Assessment: Principles and Classroom Practices*. Longman, 2004.
- Brown, Gillian, and Yule, George. *Teaching the Spoken Language: An Approach Based on the Analysis of Conversational English*. Cambridge: CUP, 1983.
- Bygate, Martin. *Speaking*. Oxford: Oxford University Press, 1987.
- Celce-Muria, Marianne. *Teaching English as a Second or Foreign Language*. Waldsworth: Heinle and Heinle Publishers, 1991.
- Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: CUP, 2001.
- Fulcher, Glenn. *Testing Second Language Speaking*. Great Britain: Pearson Education Limited, 2003.
- Halliwell, Susan. *Teaching English in the Primary Class*. Longman, 1992.
- Heaton, J. B. *Classroom Testing*. Longman Group UK Limited, 1990.
- Hopkins, David. *A Teacher's Guide to Classroom Research*. Open University Press, 2002.
- Hughes, Arthur. *Testing for Language Teachers*. Cambridge: CUP, 2003.
- Johnson, Keith. *An Introduction to Foreign Language Learning and Teaching*. Pearson Education Limited. 2001.
- Lado, Robert. *Language Testing: The Construction and Use of Foreign Language Tests*. Longmans, 1961.
- Luoma, Sari. *Assessing Speaking*. Cambridge: CUP, 2004.
- Mariani, Luciano. "Developing Strategic Competence: towards Autonomy in Oral Interaction", *Perspectives, a Journal of TESOL-Italy* – Vol. XX, No. 1 June 1994. <<http://www.learningpaths.org/papers/papercommunication.htm>> (6 February 2006).
- Madsen, Harold, S. *Techniques in Testing: Teaching Techniques in English as a Second Language*. Oxford American English, 1983.

- O'Malley, J.M., and Pierce, L.V. *Authentic Assessment for English Language Learners – Practical Approaches for Teachers*. Addison-Wesley Publishing Company, 1996.
- Savignon, Sandra J. *Communicative Competence: Theory and Classroom Practice*. Addison- Wesley Publishing Company, Inc., 1983.
- Seliger, H. W, and Shohamy, E. *Second Language Research Methods*. Oxford Universtity Press. 1989.
- Underhill, Nic. *Testing Spoken Language: a handbook of oral testing techniques*. Cambridge: CUP, 1987.
- Ur, Penny. *A Course in Language Teaching Practice and Theory*. Cambridge: CUP, 1996.
- Ur, Penny. *Discussions that Work: Task-Centered Fluency Practice*. Cambridge: CUP, 1981.

Appendices

Appendix 1: Analyses of testing techniques with regard to communicative competence

Testing techniques:	Grammatical competence	Sociolinguistic competence	Discourse competence	Strategic competence
<p>Re-telling a story: - Introduction – general questions S x T - Task 1: Student A describes the sequence of pictures for student B and vice versa. - Task 2: Student A re-tells the complete story and then, student B</p>	<ul style="list-style-type: none"> - present and past tense, S-V order - pronouns, action verbs, descriptive adjectives - complex sentences - prepositional phrases - sentence stress, word stress 	<p>This technique enables S – T interaction – formal and informal register and S – S interaction – appropriate level, informal register</p>	<p>S produces a story using descriptive language</p>	<ul style="list-style-type: none"> - co-operative strategies, - avoidance strategies, - compensatory strategies repetitions, hesitation devices
<p>Map direction: contains 3 tasks: - recommending a shop - giving directions to the movie theatre - the favorite film of the test taker</p>	<ul style="list-style-type: none"> - prepositions, directions, -imperative mood. , “there is /are” - concrete and countable nouns, superlatives, adverbs lexis- shops, town - rhythm, accent, pronunciation of words 	<p>S – T interaction - formal register</p>	<p>Ss must make their directions clear for a foreigner, for example: directions to someone who is lost.</p>	<ul style="list-style-type: none"> - avoidance strategies - compensatory strategies repetitions, hesitation markers, simplifying structures
<p>Picture comparison: This technique tests the ability to compare 2 different pictures in order to find the best holidays for the test taker.</p>	<ul style="list-style-type: none"> - existential phrases – “there is/are...”, the use of “would”, - descriptive adjectives, comparatives and superlatives - lexis: summer sports 	<p>S – T interaction - formal register - appropriate level</p>	<p>Ss express their ideas about summer holidays connected with a lexical cohesion which evokes the relevant words.</p>	<ul style="list-style-type: none"> - reduction strategies - avoidance strategies - compensatory strategies, hesitation devices,

	- sentence stress			false starts, self-corrections, simplifying structures
Discussion about “Future”: combination of 3 tasks: - Task 1: Ss compare present and future life - Task 2: What do they think future will look like? - Task 3: Describe your own future.	- future expressions, past and present tense, complex sentences, conjunctions - stative verbs, abstract words, descriptive adjectives - accent, sentence stress, pronunciation of individual words	S – S interaction - appropriate level - informal T = a tester and a prompter - informal register	Ss complete a speech about imaginary future.	- reduction strategies - co-operative strategies - avoidance strategies false starts, repetitions, hesitation devices
Role-play: Ss through the roles are persuading each other to reach their own goal: Student A: a role of a child Student B: a role of a parent They are discussing the allowance to the B-day party which starts late in the evening.	- future expressions and past tense - imperative and interrogative mood, - prepositional phrases - fall-rise intonation, accent, rhythm	S (a parent) - S (a child) informal, familiar register This interaction enables to change the context.	Possible attempts to find ways how to persuade the other S. - A parent claims that a child should study and do not go out so late. - A child claims that his friend will celebrate his B-day party.	- compensate strategies - avoidance strategies - paraphrase strategies self-corrections, repetitions and rephrasing, formulaic expressions

Appendix 2

Recalling the practical checklist (after testing the speaking ability)

1) Were the instructions, materials and criteria to testing prepared before the tests?

Learners knew the purpose of the test, the criteria and the way of test procedure. Further, they were given special tasks to accomplish in a lesson before being tested. The learners were introduced to the use of the rating scale. Also, the instructions were explained before the test and the students were given 1 or 2 min. before the start of the test.

Also, they received questionnaires to evaluate the testing technique. The administrative details were prepared and thought before the test.

2) Could students finish the tests during the relevant time?

From my point of view, the students could. However, in some cases, students answered simply and after delivering the information, they were done with the task. In order to evaluate the learner's performance, I had to ask further questions.

3) Could the tests be administered smoothly, without procedural irregularity?

Through the testing, I had to intervene in testing: giving test takers further questions, reminding them the task. I used the scoring system after the recordings were re-played. Since I was the only person who was a teacher, and at the same time a tester, I could not pay attention to student's performance or errors and at the same time to the evaluations. My task was to interact in order to test student's speaking ability.

4) Was the scoring system practical in teacher's time frame, so that learners got the marks right away?

I spent approximately 3 hours by evaluating all the test takers in this method. The scoring system was practical in a way that the rating scale was ready. What is more, I think that the rating scale uses the practicality because it is designed at the specific group of learners in my class.

However, to re-record the testing and to find the adequate measure for student's performance was troublesome. Because, it was complicated to evaluate it, learners did not get their marks after being tested but received their marks next lesson.

Appendix 3

The self-designed rating scale:

Vocabulary:

- 1 broad and general vocabulary to deal with complex problems and with different social situations
- 2 adequate vocabularies for discussion of special interests and other uncomplicated situations
- 3 vocabulary limited to basic areas like *time, food, family...*
- 4 the choice of words sometimes inaccurate which limits the discussion
- 5 inadequate vocabulary for the simple conversation as well

Pronunciation:

- 1 no errors in pronunciation and no foreign accent, (but considering the age of the primary students, the pronunciation can have a foreign accent)
- 2 occasional mispronunciation which does not interfere with understanding
- 3 mispronunciation lead to occasional misunderstanding
- 4 frequent errors and for heavy accent, it is difficult to understand the meaning
- 5 pronunciation is not recognizable, errors cause the complete misunderstanding

Grammar accuracy:

- 1 no more than 2 errors during the speaking activity
- 2 few errors, fixed patterns of tenses
- 3 occasional errors but the weakness of them does not cause the misunderstanding
- 4 frequent errors showing some incomplete patterns and occasional misunderstanding
- 5 grammar totally inaccurate, not understandable

Fluency:

- 1 speech is smooth, especially in speed
- 2 speech is occasionally hesitant and rephrased
- 3 speech frequently hesitant having incomplete sentences
- 4 speech is slow and uneven except the short or routine sentences
- 5 speech is so stuttered that conversation is impossible

Interactive communication:

- 1 understand everything in normal conversation, except for less-used items, take turns and know where to start the speech and where to enclose it
- 2 understand quiet well normal speech but occasionally require repetition, take turns with a speaker/listener but occasionally they hesitate whether they can start or listen to the talk
- 3 understand simple speech directed towards them but sometimes repeated and rephrased; are awkward in taking turns; their ability to switch turns is slow
- 4 understand only slow simple speech on social topics, require repetition and rephrasing; wait for their turn since they are not sure when they can talk
- 5 understand too little for even simple conversation, do not take turns and they do not know where to start the speech and where to enclose it.

Appendix 4

The Preparation project

Testing technique: Discussion about the topic: FUTURE

Starting date of planning: 6th May 2005 – 10th May 2005

Date of planned testing: 11th May 2005

Consulting Literature:

Fulcher, *Testing Second Language Speaking*. 2003, 76 - 78.

Underhill, Nic. *Testing Spoken language*. 1987, 49 – 51.

O'Malley, J.M. and Pierce, L.V. *Authentic Assessment for English Language Learners*. 1996, 79 – 80.

Allocate time spent on literature: 6 hours

Tips to follow (through the relevant sources):

- practice free discussions before testing
- encourage learners
- make a list of factors along with the instructions

Types of materials:

Self-made materials:

instructions

an examiner outline

questionnaires

Tools:

a tape-recorder

Instructions (Oral x Written, Czech x English)

Designing the instructions:

- written (including tips for further discussion)
- in English

Designing the examiner outline:

-oral

- in English

- Read your cards, work in pairs and try to react to “How do people live nowadays? And how will they live in the future?”
- In the second task, think about “What will you do in the future? “

(Remind test takers’ tasks and help to proceed them through the misunderstanding of the tasks.)

Time for preparing materials including instructions: 2 hours

Total planning time: **8 hours**

Appendix 5

Test specification – “map-direction” technique:

(selected from Hughes, Arthur. *Testing for Language Teachers*. CUP, 2003.)

Content

Skills to be assessed: ability to instruct a tester to get to a certain place, ability to answer simple questions concerning the topics, ability to be fluent, ability to react in interaction student-teacher (S – T).

Types of interaction: pair-work between S - T

Topics: a town, shops, films

Structural range: the use of prepositions, directions: imperative mood, the existential phrase: “there is, there are”, pronouncing words, accent

Vocabulary range: buildings and shops in a town, film’s types (action, fairy-tale, documentary...)
concrete and countable nouns, superlatives, adverbs of place

Structure, timing

Test structure:

- introduction
- warm-up: recommend a favorite shop or a building in a town and explain why
- task: to give directions from one place to another through the map
- conclusion: the attained place (a movie-theatre) is discussed, a favorite film...

Number of tasks: 3

Timing expeditious: 5 min.

Techniques: map-direction where the focus is on speaking fluency, grammar, vocabulary, pronunciation and interactive communication

Scoring procedures:

- recordings of the individual performance
- re-listening the recordings, making notes
- the evaluation through the rating scale: fluency, vocabulary, pronunciation, grammar accuracy, interactive communication
- the summation of all criteria to make a total to each individual

Appendix 6

Discussion about future

You have 2 min. to read your instructions. Choose and think of topics you want to talk about, and work in pairs.

1) How does life look like *now*? x How will life look like *in the future*?

You can choose out of these topics:

- *travel*
- *social life*
- *food*
- *jobs*
- *nature*

2) What do you think about your future?

I think

I believe that

People will

I will

Appendix 7: The collection of transcriptions to specific testing techniques

Appendix 7.1.

Transcription of the testing technique: “Map direction”

Date: 14th April 2005

Names of test takers were changed.

George

T³: Hello, George.

G: Hello.

T: George, can you choose a place for me on the map and tell me why you chose it?

G: A place...a pet shop because ...are.. ehm.. ... animals.

T: What's your favorite animal?

G: My favorite...a mouse.

T: Your 2nd task is to give me directions from the bus station to the movie theatre. How do I get there?

G: I ...I take 2nd street. Turn right. Go /stringing / ..straight and turn left. Oxford Avenue, go straight and turn left and there is ..a movie theatre.

T: What's your favorite film or movie? And what is it about?

G: “Hrátky s čertem”. It's about...is he's a devil ..it's a devil and ...a girl.

T: Thank you.

Tom

T: Hello, Tom.

Tom: Hello.

T: Here's your map, can you recommend me a store and tell me why?

Tom: Sporting good store. In the sporting good store...was bikes and sporting...things.

T: Do you like sports?

Tom: Yes, I do. I like ..I play volleyball, floorball.

T: All right. And next, can you give me directions from the bus station to the movie theatre, please?

Tom: I go from the bus station to 2nd street....Go right to the 2nd street. From the 2nd street...turn left to the Prospect Avenue, from the Prospect Avenue turn right to Third Street to the movie theatre.

T: When we came to the movie theatre, do you like any films?

Tom: I like comedy or thriller or action.

T: Tell me something about your favourite film, please?

Tom: This story “Skryté Hlasy” ...about human..which he listen ghost. Hm, ghost..were was dead.

T: Great, thank you.

John

T: Hello, John. Can you recommend me a shop and tell me why you recommend it ?

J: Coffee shop. Here is very good coffee and very nice....very nice...

T: And anything else?

J: Yes, women. It's very nice shop and small.

³ T= a tester/ a teacher

T: Great, you are close to bus station now. How do I get from the bus station to the movie theatre?

J: You go ..you go down on straight. You go on the Oxford Avenue and you are in movie theatre.

T: What is your favourite film? And what is it about?

J: My favourite film is "Horem Pádem". Here are ..some man and woman. Hm, they're they weren't...they were having some child and some man was buying some children.

T: How did it end up? Was there a happy end?

J: They are...they are on the police station and they are on the police station and it's happy end.

T: Well done, thank you.

Bill

T: Hello.

B: Hello.

T: A map, here. Bill, can you please choose a shop and recommend it to me, why did you choose it?

B: I like restaurant....because they cooking good....good food.

T: What food do they cook there? Italian, French...

B: Czech.

T: All right. I'll go there. Now, how do I get to the movie theatre from the bus station?

B: You going hm, in Third Streetgoing to Main Street and turn left to bus station.

T: But I thought from the bus station to the movie theatre.

B: You go 2nd street and turn left to Oxford Avenue and then you going to movie theatre.

T: What is your favorite film?

B: "Taxi Taxi"

T: What is it about?

B: This film aboutbetter taxi and about man...who is ..a driver. And he helps to ... Police.

T: Great, thank you.

Larry

T: Hello.

L: I'm Larry.

T: Hello.

L: Hello.

T: Can you recommend a shop and tell me why?

L: ???

T: What's your favorite shop?

L: My shop....I was ..shop...rabbit, snake.

T: Then, I should go to the pet store because there are...

L: ...there are snakes in the store.

T: What's your favorite animal?

L: My favorite animal... rabbit.

T: Now, how do I get to the movie theatre from the bus station?

L: Bus station...hm....and 2nd street. Go 2nd street and First Avenue and go street and ... movie theatre is...on the ...left....Movie theatre.

T: Thank you.

Mike

T: Hello, Mike.

M: Hello.

T: Here's a map. Look at it and choose a shop for me. What can you recommend it to me?

M: I was..recommend...ice-cream shop.

T: Why?

M: We...they were....fifty tips ice-cream.....chocolate, strawberry, cherries, ananas.

T: Thanks. Now, how do I get from the bus station to the movie-theatre?

M: Go straight and...turn right, go straight. Go Prospect Avenue and ...turn and right there's movie theatre.

T: Great. What is your favorite film?

M: My favorite film...is "Slunce, seno, strawberry."

T: What is it about?

M: This film is funny and ..good actors.

T: Fine. Thank you.

Jessica

T: Hello, Jessica.

J: Good morning.

T: Can you choose a shop and tell me why I should go there?

J: Hm, you ...you can go ..in ice-cream shop ..because because ...there are ...I like it.

T: Thank you. Now, How do you get from the bus station to the movie theatre?

J: Go....go ..turn left. Go Prospect Avenue, turn right...go down...and there is...movie theatre.

T: Do you like films?

J: Yes.

T: What is your favorite film?

J: My favorite film is.....????

T: Do you like comedies?

J: Yes.

T: Thank you.

Ann

T: Hello, Ann.

A: Hello.

T: Look at the map, please and tell me what shop can you choose for me?

A: Ice-cream shop.

T: Why?

A: Because...I like it ...ice-cream.

T: What ice-cream do you like?

A: I like it ..ice-cream...strawberry, chocolate.

T: Great, now how do I get from the bus station to the movie theatre?

A: Go straight, four ...take left and ...go straight...and there is ice-cream shop.

T: Do you like films?

A: Yes.

T: What films do you like?

A: I like ...action.

T: Thank you.

Appendix 7.1.1.

Evaluation of learners' performances (according to the rating scale):

George:

vocabulary : 3
pronunciation: 3
grammar accuracy: 3
fluency: 3
interactive communication: 2

Complete mark: **3**

John:

vocabulary : 2
pronunciation: 1
grammar accuracy: 1
fluency: 2
interactive communication: 1

Complete mark: **1-**

Larry:

vocabulary : 3
pronunciation: 2
grammar accuracy: 4
fluency: 3
interactive communication: 4

Complete mark: **3 -**

Jessica:

vocabulary : 3
pronunciation: 2
grammar accuracy: 3
fluency: 3
interactive communication: 3

Complete mark: **3**

Tom:

vocabulary : 2
pronunciation: 1
grammar accuracy: 2
fluency: 2
interactive communication: 1

Complete mark: **2+**

Bill:

vocabulary: 2
pronunciation: 1
grammar accuracy: 2
fluency: 2
interactive communication: 1

Complete mark: **2**

Mike:

vocabulary: 3
pronunciation: 1
grammar accuracy: 3
fluency: 2
interactive communication: 2

Complete mark: **2-**

Ann:

vocabulary: 3
pronunciation: 2
grammar accuracy: 3
fluency: 4
interactive communication: 3

Complete mark: **3 -**

Appendix 7.2.

Transcriptions from the testing technique: “Picture Comparison”

Date: 28th April 2005

Names of test takers were changed.

John

T: Can you look at these pictures and describe them. Then, tell me where would you like to go on holidays?

J: (Holding picture no.1)

This is sea and beach, houses, hotels, trees and more..... On the 2nd photo is church, corns /koruns/ , trees, houses, village.

T: Excellent. Where would you like to go on holidays, please?

J: I will be on the beach. In the Carribean.

T: Why?

J: Here is nicest, eh...better. Here is....I can ...swimming on the sea and be on the hotels, and next. I will want go to the sea because on the village I know and I live on the village.

T: Yes, excellent. Thank you, Michael.

Mike

T: Look at the picture and try to describe them. Then, choose where you would like to go on holidays?

M: I watch in picture sea, ..stones and house. I watch in the picture village and castle.

T: Where you would like to go on holidays?

M: I was... I go ..in the sea. I like sea.

T: Why?

M: Because ..I like it there. ..I was playing volleyball and swimming. It is fun.

T: Why do you like sea?

M: To tell truth I don't like sea because I like pool.

T: Why do you like pool and not sea?

M: In the sea....salt water and pool ... it's clean water.

T: Great , thank you.

Jessica

T: Look at the picture and try to describe them. Then, choose where you would like to go on holidays.

J: In this picture is ehm.... my village.

T: What else can you see? Can you see nature, countryside?

J: Nature.

T: And 2nd picture?

J: And in this picture is sea and beach and stone.

T: Where would you like to go on holidays?

J: I would like to go ...in my village.

T: Why?

J: Because in village....is my friends and we have fun. We go in bicycle. Go we go in swimming pool.

T: Why wouldn't you like to go to the sea?

J: Because eh...I don't know. Maybe, my friends here in village and not in sea.

T: All right. Thank you, Jessica.

Tom

T: Look at the picture and try to describe them. Then, choose where you would like to go on holidays.

Tom: I will go to the Czech country to the mountain ...on the holiday. Eh, I will go to České Středohoří.....because is a lot of volcanoes there ...and eh, a lot of look-out towers.

T: Can you first describe the pictures?

Tom: I can see church in the Dubenec and woods and trees and corn. I can see sea and stones. I can see beach with hotels.

T: Where would you like to go on holidays?

Tom: I will go to the countryside.

T: Why?

Tom: Because it's near and touristic stamps.

T: Great. Thank you for your ideas.

Tom: Bye.

George

T: Look at the picture and try to describe them. Then, choose where you would like to go on holidays. What can you see?

G: I see this picture...village and forest, eh and house.

T: And the 2nd picture?

G: The picture...sea and beach and ...hotels.

T: Where would you like to go on holidays? Here or there? (Pointing to each pictures.)

G: I would like to go to see...eh America. Beaches, hotels.

T: Why?

G: Because....American beach very nice. Hotel ...because nice hotels. I play beach volleyball and swimswimming and ride.....scooters.

T: All right. Thank you.

G: Thank you.

Larry

T: Look at the picture and try to describe them. Then, choose where you would like to go on holidays. What can you see?

L: In Carribic is ...is big big city and eh big sea and beaches and eh... Picture Two is Dubenec.

T: What can you see there?

L: Picture and church and large.

T: Where would you like to go on holidays?

L: I will go holidayswas ...to Caribic.

T: Why?

L: Caribic was.....eh, I like swimming and beach volleyball.

T: What else?

L: ?

T: All right. Thank you.

Bill

T: Look at the picture and try to describe them. Then, choose where you would like to go on holidays. What can you see?
 B: This picture is church and village Dubenec andand many house. On this picture is nature ...a lot of nature.
 T: And the 2nd picture?
 B: On this picture is sea and small town and beach.
 T: Where would you choose to go on holidays?
 B: I want to go ...eh, sea. I like swimming.
 T: Why would you like to go to sea?
 B: I like small town because I like sea and I like swimming. I don't like Dubenec.
 T: Why?
 B: Because here is ...boredom.
 T: Great. Thank you.
 B: Thanks.

Ann

T: Look at the picture and try to describe them. Then, choose where you would like to go on holidays. What can you see?
 A: I look two pictures, sea and village.
 T: Describe it more.
 A: There is.....There is ...sky and house and
 T: And beach?
 A: And beach.
 T: And the second?
 A: There is village and field....and castle.
 T: O.k. Where would you like to go on holidays?
 A: I will go holidays.....beach.
 T: Why?
 A: (No answer.)
 T: Because?
 A: Because sea and don't,We don't have sea and swimming and beach volleyball and..
 T: Would you like to go with your friends or with your family?
 A: I go eh with friends.
 T: Why wouldn't you like to go to village?
 A: I know village because...I ...
 T: live here?
 A: I live in village.

Appendix 7.3.

Transcription of the testing technique: “Discussion about future” in pair-work pattern

Date: 11th May 2005

Names of test takers were changed.

John x Mike

T: Read your cards and try to react to “How do people live nowadays? And how will they live in the future?” And 2nd task is to talk about “What will you do in the future?”

J: I think that peoples will travel in the.... peoples will travel in theflying cars. Eh, and it’s very ..It’s will very expensive....expensive. And people will rich and because because ...because!!

T: How does life look like now and how it will look like? Are there many cars?

J: People will have got ...many cars, many cars. Children will go to school in flying bus. And....!

M: I think people traveling in sky cars and plane because plane doesn’t ‘t a lot. Plane won’t cost money. ...People will rich and laziness. Everywhere...everywhere will factory and doesn’t will tree.

T: Thank you and now, your future.

J: I willPeople will live in the big cities and we will. I will live in a big city, I will have a sky car. I will be in some company. I will....I will live in the skyscraper. I will be president and I’ll be very rich and I will have got many childrens and.... many wives.

M: I will live in village and I will big money and....I will live in....live in palace. I will drive sky car and plane. I will.....classic village life.

T: What will you do?

M: I doing hm.... I will director bank.

T: Anything else?

M: Everything people live in big cities and doesn’t in village.

Tom x Bill

T: Read your cards and try to react to “How do people live nowadays? And How will they live in future” And 2nd task is to talk about “What will you do in the future?”

Tom: I think that food will ...food will in a tin, in microwave. People will eh..buy on the Internet and food will ...food will be ..frozen, people will cook on the . They will ..they won’t cook... food will instant and people will... they put water in it.

B: I think that...I think that people will be popular and people will.eh... more going on travel. I think that people will travel ..will travel to space and and.

T: Thank you and now, your future.

Tom: I think I will go on a sky car to my job. My job will on the universe and I will be more rich because I will invent new new..rockets.

T: And your family?

Tom: I have ..I will have got a family. I will live in the Czech Republic and I will live in big house with big garden.

B: I think that I play popular football players and...better and will better than ...these players ..and I will million dollars will have million dollars.

T: Where will you live?

B: I will live in the Czech Republic.

T: What will you have?

B: I will six cars, five house.

T: And family?
 B: Yes, I will five children. ... (Laughing) I will have two children.
 T: And animals?
 B: I will five dobrmans.
 T: Anything else?
 B: I will have plane, small and biggest plane.
 T: Two planes?
 B: And I will big yacht and .
 T: I am glad I met a millionaire.

George x Jessica

T: Read your cards and try to react to “How do people live nowadays? And how will they live in the future?” And 2nd task is to talk about “What will you do in the future?”
 G: I think that people ..people will be clever and I will that...I think that people will be eat frozen food and people will not have many jobs and people will not travel. And in nature ...nature will be destroyed.
 J: I think that people will kill each other. Nature ..I believe...I believe that that there nature will be destroyed.
 T: Thank you and now, your future.
 G: I think that my life /li:f/ I will my life will be interesting and I will traveling and I hope...I hope that that... I will be....happy.
 J: I will kill.
 T: Will you kill or will YOU be killed?
 J: Yes. I killed. I....like...action films and I...will be ..actor. And travel ..I will...be ... away.
 T: Very well. Thank you.

Larry x Ann

T: Read your cards and try to react to “How do people live nowadays? And how will they live in the future?” And 2nd task is to talk about “What will you do in the future?”
 L: I think that people will travel in Mars. Mars will in people. Car will for people... ..People will destry the Earth and go to Mars and Jupiter.
 A: Hm...I ...think ...we will ...always...be . People... go...by space car...and we not will ...cook., .hm...we eat ...we will eat....some “blé”food.
 T: Thank you and now, your future.
 L: I will small food. We will have little food, we will die.
 A: No,...hm we will...die...not.
 T: And your future, please?
 L: I ...hm., I...go...I go...to the ..zoo. I....my favorite ...animals. I will....I ..work ...zoo. And I ...have no family, I ...will ...animals.
 A: Hm., I think....I ...I will be ...dance...I dance. I ..have...family and ..friends. We ... hm ...go ..parties.
 T: Great. Thank you.

Appendix 7.4.

Transcription of the testing technique: “Role-play between ‘a parent x ‘a child’ “

Names of test takers were changed.

Date: 23rd May, 2005

John x Tom

J: Hi dad, can I go on the ..on the birthday party ..on the night at half past nine p.m.?

T: Half past 9 p.m? Tonight? Too late.

J: No, I will come back early.

T: You...you know you you will come too late and you you don't go to the B-day party.

J: No, I will come back early.

T: No, I ..you will ..you come too late and school tomorrow, you will ...you must go to school tomorrow.

J: Hm, here is my best friends..,on the B-day party will be my best friends.

T: You see best friend on the school tomorrow.

J: Hm, I will I won't see my best friends on the B-day party.

T: No, you don't go to the B-day party.

J: It's B-day party of my friend.

T: Hm, you...you doing homework to the, hm, school tomorrow.

J: No, I will I do my homework I did, I was... did... my homework.

T: No, you don't you didn't do homework, your homework. You hm, you drink alcohol on the B-day party.

J: I won't, won't drink alcohol, really! /reli/

T: What will you drink on the B-day party?

J: I drink lemonade.

T: You are kidding.

J: No, I am not kidding.

T: Yes, you will drink alcohol.

J: I won't drink alcohol I drink, will drink... lemonade. Can I go to the B-day now?

T: Hm, you must go to home at eleven p.m.

J: Yes, I will go. Thank you.

T: All right.

Bill x Mike

M: Hello, how are you?

B: I am fine and you?

M: All right, dad. Can I go please? Can I please go party?

B: When hm, it is ...is it?

M: I must be there. It's tonight at nine thirty.

B: But this is too late.

M: No, it isn't.

B: But you ..you will ..you go to school tomorrow.

M: No, we are go cinema.

B. Hm, and you have homework.

M: No.

B: You you drink alcohol there.

M: I will not drinking alcohol.

B: What will you drinking?

M: I drinking lemonade. I must be there.

B: No, he..no you don't have to.

M: I have friends.

B: You don't go party.

M: Please?

B: You won't go.

M: Please?

George x Larry

L: Can I have go B-day party?

G: No, no because you was drinking alcohol.... last time.

L: no drinking and no alcohol

G: What time ..this is start the party?

L: I can ..nine thirty.

G: It's too late.

L: It's too late. I come back early.

G: But tomorrow is school.

L: But we ...holidays.

G: But you have homework.

L: But it's homework when come back early.

G: When you come, it will be too late.

L: Can I go B-day party?

G: Yes, but you must making homework.

L: I homework – it's come. I will do homework, then B-day party.

Ann x Jessica

A: Can I go ..is birthday party?

J: No, it's too late.

A: But ... it starts at half past nine.

J: Hm, it's school tomorrow.

A: I must go ...B-day party ...because my best friend.

J: You ...you have.. has homework.

A: I ...no homework.

J: You drinking alcohol.

A: I don't know ...I don't ...no alcohol.

J: You can go to party at eight o'clock.

A: No, I must go B-day party. Mum, please. Please!!!

J: No, you are young. and you are home.

Appendix 7.4.1.: A Questionnaire Sample

Dotazník pro žáky 8. ročníku

Cílem tohoto dotazníku je zjistit jaké metody při zkoušení "Mluvení v anglickém jazyce" se žákům líbí nebo nelíbí a co může napomoci zkoušejícímu v dalším ohodnocení.

Vyplňte, prosím, čitelně!

Vaše jméno: *Tomáš Baumel*

Věk: *14 let*

Zakroužkujte prosím odpovědi následně: 1 – vůbec nesouhlasím
2 – nesouhlasím
3 – nevím
4 – souhlasím
5 – úplně souhlasím

1) Během zkoušení jsem byl/-a nervózní. 5 4 **3** 2 1

2) Aktivně jsem se zapojil/-a do této aktivity. 5 **4** 3 2 1

3) Líbilo se ^{mi} prosazovat patřičné důvody. 5 **4** 3 2 1

Proč: *Protože to byl souboj.*

Premýšlel jsi o svých dalších nápadech/ ~~důvodech~~? Proč ano/ proč ne!!!

Musel jsem ho málem přesvědčit, takže ano.

4) Bylo zajímavé se vžít do jedné z rolí. 5 **4** 3 2 1

Proč: *Zajímavé bylo hlavně hádání.*

5) „Role-play“ byla těžká. 5 4 **3** 2 1

Proč: *Kocela ano.*

6) Pokud by sis měl/-a pro zkoušení MLUVENÍ vybrat následující aktivity, jak bys je seřadil/-a podle své oblíbenosti.

- 6*...Diskuse na dané téma ve skupině
- 2*...Popis obrázku ve dvojici
- 4*...Dítě x rodiče – dosažení cíle
- 5*...Interview (žák – učitel)
- 1*...Interview (žák – žák)
- 2*...Převyprávění příběhu
- 7*...Mluvení ve dvojici – otázka x odpověď

Appendix 7.5.

Transcription of the testing technique: “Description and re-telling a story”

Date: 3rd June 2005

Names of test takers were changed.

1st pair: Jessica x Ann

T: Hello! How are you?

J + A: Hello, I'm fine.

T: Where are you from, Jessica?

J: I'm from Dubenec.

T: And how old are you?

J: I'm thirteen.

T: Where are you from, Ann?

A: I'm from.....in Velký Vřeš'tov.

T: And how old are you?

A: I ...thirteen. I'm thirteen.

T: Look at pictures. Then, you will describe them.

After two minutes:

A: We're going to watch team.

T: Who we?

A: Ann and Steve

J: And they are go to the café. They are drinking. Then, she lost.....left her bag at station café.

She go for tickets but she lost her bag. She go....take it.

T: How do you think Ann felt in these pictures? Happy or unhappy?

J: Unhappy.

T: Yes. Thank you. And next pictures.

A: They arelate. So, ..bus by car.....They go by car to station....and sitting in director's box.

T: How do you think they feel in picture 6 and 9?

A: Scared.

Jessica retelling the story

J: They go to the café. They are drinking. They are left bag. They go ..tickets. She go to station café. They are.....running for bus. Then, ..left, so they go by car. And ..they are match football team.

T: Yes. Excellent.

Ann retelling the story

A: Ann and Steve are going to watch their team. They are café.

T: Why?

A: Becausetime. They are drinking. She left ..a bag. He go to tickets, she go to station ... café for bag. They late , hm.....and they go....a car....station and they are director's box ...sitting. They are happy.

T: Yes. Thank you.

2nd pair: Mike x John

T: Hello. How are you, Mike?

M: I'm fine. Thank you.

T: And you, John?

J: I'm fine too.

T: Where are you from?

M: I'm from the Czech Republic?

J: I'm from Velký Vřeš'tov.

T: And how old are you?

J: I am fourteen.

T: What are your hobbies?

J: My hobbies are football, floorball....some sport.

T: And Mike, how old are you?

M: I am thirty.

T: Thirty?

M1: Thirty? Thirteen!!

T: And what are your hobbies?

M1: Jo, My hobby ... play football.

T: Thank you. Look at pictures now and describe them, please.

After two minutes:

M: They are going to football match.

J: On picture two they were good time. They were going to a café. On the picture three they are ...were good time. They were speaking. On the picture four they were going on a bus. She left her bag on a café. She....she was going to a café. ...and give her a bag.

T: What do you think Ann feels in picture 4?

J: She was scared.....sad because she left ..your bag.

M: They are go train. Train went away. They were stopping a car and asked for lift on football match. They are a good place in director's box.

T: How do you think they feel in picture 6 +7 ?

M: They are.....

T: Are they happy or unhappy?

M: Are they happy and unhappy.

John retelling the story

J: Hm,...they are going ...were going on a football match. They were having a good time because they were going to a café. They were having a good time and they were speaking about football. Hm,... they were going to the bus when she left....lost her bag.... She was going to a café and give her bag. Hm..... They were running to the train and train went away. On the picture 7 they were stopping some car on a football match. On the picture 8 they were on the stadium. They were on *director's* box. And go on a *match*.

Mike retelling the story

M: They were go on a football match....and they were a good time and go for a café. And....she leave bag and go...go oooo to train station. She ...find out her bag. He came back and go train station. But....train goes away because they're late. They..... stopping a car and asked for a lift and they are good place watching football match.
T: Very well. Thank you.

3rd pair: Tom x Bill

T: Hello, how are you?

Tom: I'm fine. Thank you.

T: Where are you from?

Tom: I'm from Lanžov.

T: What are your hobbies?

Tom: My hobbies are volleyball, floorball and paper's models.

T: How are you, Bill?

B: I'm fine. Thank you.

T: And how old are you?

B: I...oh, how old are you? I am fourteen.

T: What are your hobbies?

B: floorball, volleyball, athletics...

T: Look at pictures and describe them.

After two minutes:

Tom: Teenagers are on a railway station and are talking about football match. They are going...hm, to the football express. They are late. They are running on a railway station to the football express. Teenagers are stopping on the road and ..and they are stopping for a lift. Then, they are on a car park near stadion and ..and they are going to a football match.

T: And picture 9?

Tom: They are on a director's box and they have got ..have got a best....best place.

T: How do you think they feel in picture 7?

Tom: On the....railway station are sad becausefootball express went away and.... on the road they are happy because ..car stopping.

T: Yes. Excellent.

T: And Bill?

B: They are....watching on a train list when she wants go station café. Hm...., they are...talking about football match in a station café. When.., hm...., she forgot ...his / her bag and eh, she ...goes station café for bag.

T: How do you think she feels in picture 4?

B: She feels horrible.

T: O.k. Thank you.

Tom retelling the story.

Tom: They are on a railway station and they are....they are talking about football match and they are happy. Then, they are going to the station café and they are..hm, they are drinking cola.And they are going to the railway station and she ...she's forgot.....her bag on the station café and she is ..go back to the station café for a bag. But on the railway station football express went away....because....they are goinggoing late. They are, they are going to a road and stop. Hm, stop a car. And they have...have got lucky. Driver....driver is going on a stadium and they are going to the director's box and they are...have got....the best place.

Teacher: And they're lucky to watch football. O.k. Thank you.

Bill retelling the story:

J: They are going to a railway station. Oh...when ...they ..reading train list. Hm, she wants ...Ann want go to station café. They are drinking and hm, talking about football

match. And when she...they going to train, she...left... left her bag. Hm,...they came back to the station café but theylate football express. They.....stop a car and they are going to football match stadium. And they ...they are they was in director's box and watching football match.

T: Excellent. Thank you.

4th pair: Larry x George

T: Hello, how are you?

L: O.K. Thanks.

T: And you? (looking at another learner).

G: I am fine. Thank you.

T: Where are you from?

L: From... Czech Republic.

T: And where are you from?

G: I'm from Czech Republic too.

T: What are your hobbies?

L: My hobbies...is ...animals.

G: And my hobby is swimming and playing computer games.

T: Great. Thank you. Now, look at your picture, and read the instructions. You have two minutes to go through the instructions and pictures.

After two minutes:

T: Please, describe your pictures to each other.

L: Boy and...a girl are in the station. They aregoing and ...train ..away. Picture seven....a car... and they go....by car to football. They.....watchingfootball in directors' box.

T: How do you think they feel in picture 7?

L: Happy. They ...want football...and a car.....go to football.

T: Great. Can you describe your pictures, Joe?

G: A boy and a girl are looking at....the train times andgo to the café. In the café... .they talking...about football and drinking. Then,....The boy wants....tickets....and the girl ..has her bag...in the café. She go back to café.... and take ...bag.

T: How do you think she feels in picture 4?

G: She is not happy but....horrible.

T: All right, you could both hear a part of the story. Now, put it together and re-tell the whole story.

Larry retelling the story:

L: They look at ...time andgocafé. In café....they talkingfootball and have ...coffee. They not ...have ...tickets and she....gocafé. She ...take bag and ...in platform, ...the train ..is not ...waiting. They ...stop...car and go ...football and look at football.

T: Very well. Thank you. And now, George!

George retelling the story:

J. Ann and Steve want go tofootball but...they are wait..on the bus. They go to the café and drinking café and lemonade. Hm,...they talk about football and hm,....go... to the station. The boy wants ...tickets ..but the girl ..left bag in the café. She go....goes

back to café andtakesa ..bag. In the station, the bus....train goes away. They stop...a car and he goes football too. They sit in directors' box and watch football.
T: Thank you.

Appendix 8:

Table 1: Results of the average of all criteria to the testing techniques

Average of all test. techniques criteria

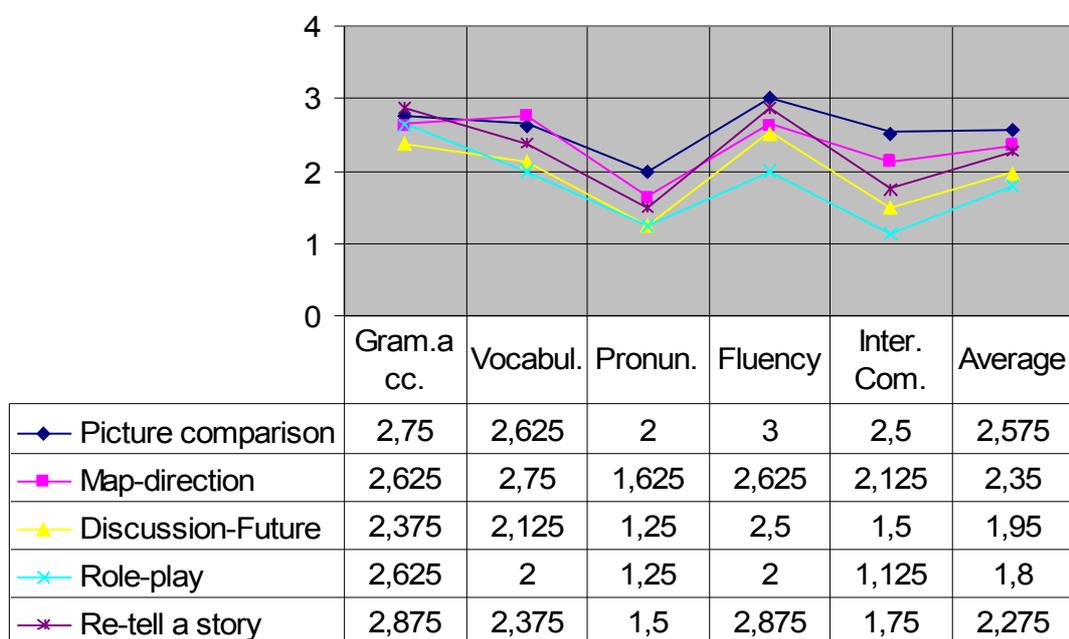


Table 2: Learners' results through questionnaires:

The well-perceived technique by learners

