

Západočeská univerzita v Plzni

**Fakulta pedagogická
Katedra anglického jazyka**

Diplomová práce

**VYTVÁŘENÍ KRITÉRIÍ PRO HODNOCENÍ MLUVENÉHO
PROJEVU**

Stanislav Rychtařík

Plzeň 2014

University of West Bohemia

**Faculty of Education
Department of English**

Thesis

DESIGNING CRITERIA TO ASSESS SPEAKING SKILLS

Stanislav Rychtařík

Plzeň 2014

Prohlašuji, že jsem práci vypracoval samostatně s použitím uvedené literatury a zdrojů informací.

V Plzni, dne 26. června 2014

.....

Stanislav Rychtařík

ACKNOWLEDGMENTS

I would like to express my thanks to my supervisor, Mgr. Gabriela Klečková, Ph.D., for her professional advice, patience and encouragement, for without her guidance the process of writing would be much less productive and enjoyable.

ABSTRACT

Rychtařík, Stanislav. University of West Bohemia. June, 2014. Designing Criteria to Assess Speaking Skills. Supervisor: Mgr. Gabriela Klečková, Ph.D.

This graduate thesis deals with designing criteria to assess speaking skills. The first section, serving as a theoretical basis for the subsequent research, presents readers with the reasons for selecting this topic and explains the terms speaking, assessment and assessment criteria in the context of assessing speaking skills. The research, conducted by means of teacher questionnaires, maps the teachers' current practices in the field of the assessment of speaking skills. The results of the research revealed the most frequently employed speaking assessment tasks, teachers' preferences in types of assessment and the fact that the criteria teachers select for individual speaking assessment tasks differ quite significantly. Subsequently, the results are commented on, and possible reasons for their occurrence are mentioned. Implications of the results for language teaching are discussed at the end of the thesis.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ii
ABSTRACT.....	iii
TABLE OF CONTENTS.....	iv
LIST OF GRAPHS.....	vi
I. INTRODUCTION.....	1
II. THEORETICAL BACKGROUND.....	2
Speaking as One of Four Skills.....	2
Basic Types of Speaking – Levels of Interaction of Speaking and Listening Skills.....	3
The Most Frequently Used Speaking Tasks.....	4
Speaking as a Part of the Czech Curriculum.....	5
Speaking as a Part of the Maturita Exam.....	6
Assessment of the oral part of the Maturita exam.....	7
The Difference between Assessment and Testing.....	7
What is Assessment.....	8
Basic Types of Assessment.....	8
Formal and informal assessment.....	9
Formative and summative assessment.....	9
Direct and indirect assessment.....	9
Holistic and analytic scoring (assessment).....	11
Validity and Reliability in Assessment.....	12
Subjectivity in the Assessment of Speaking Skills.....	14
Possible Ways of Ensuring Objectivity.....	14
Designing Rubrics.....	16
Designing Criteria to Assess Speaking Skills.....	17
Possible Scoring Criteria.....	17
Grammar.....	17
Vocabulary.....	18
Pronunciation.....	18
Fluency.....	19
Conversational skills.....	19
Sociolinguistic skills.....	20

Non-verbal communication skills.....	20
Content.....	21
Selecting and Weighting Assessment Criteria.....	21
Levels of Performance and Descriptors.....	22
Analytic Rubrics.....	23
III. METHODS.....	25
Introduction to the Research.....	25
Research Tool.....	25
Research Participants.....	27
Research Procedure.....	28
IV. RESULTS AND COMMENTARIES.....	29
The Frequency of Using Individual Speaking Tasks to Asses Speaking Skills.....	29
Teachers' Preferences in Types of Assessment.....	32
The Choice of Individual Criteria for Interview, Picture Description and Oral Presentation.....	34
Interview.....	34
Picture Description.....	35
Oral Presentation.....	37
Overall Results.....	38
V. IMPLICATIONS.....	40
Implications for Language Teaching.....	40
Limitations of the Research.....	41
Suggestions for Further Research.....	42
VI. CONCLUSION.....	43
REFERENCES.....	44
APPENDICES.....	47
Appendix A.....	47
Appendix B.....	48
Appendix C.....	49
Appendix D.....	50
Appendix E.....	53
SUMMARY IN CZECH.....	56

LIST OF GRAPHS

<i>Graph 1.</i> The frequency of using individual speaking assessment tasks for each type of school separately.....	29
<i>Graph 2.</i> The frequency of using individual speaking assessment tasks for all 5 types of schools combined.....	30
<i>Graph 3.</i> Teachers' preferences of types of assessment of speaking skills for each type of school separately.....	32
<i>Graph 4.</i> Teachers' preferences of types of assessment of speaking skills for all 5 types of schools combined.....	33
<i>Graph 5.</i> The importance of individual assessment criteria as for interview.....	34
<i>Graph 6.</i> The importance of individual assessment criteria as for picture description.....	36
<i>Graph 7.</i> The importance of individual assessment criteria as for oral presentation.....	37

I. INTRODUCTION

This thesis focuses on the assessment of speaking skills. Speaking is considered the most important of all skills by many and yet, students' speaking skills are often practised and developed, but not extensively assessed. The question has to be asked: why are teachers reluctant to assess speaking and focus on assessing other skills instead? The issue with the assessment of speaking skills is that, compared to other skills, it is demanding for teachers. Not only is it time consuming, but teachers often take on more roles (interviewer, assessor, etc.) at once, have to evaluate their students' performances and produce an assessment in real time, which puts a lot of pressure on them. This thesis attempts to lessen the pressure as it provides practical advice on how to assess students' speaking performances.

The first section of the thesis is devoted to establishing the theoretical framework for the practical part of the thesis. Terms that are essential to be familiar with are defined. Individual types of speaking as well as assessment are described and, what is more, ways of ensuring objectivity within the assessment are presented.

The following chapter, titled Methods, builds on the theoretical background and by means of questionnaires attempts to map and describe current practices of Czech teachers of EFL in the field of the assessment of speaking skills. In order to do so, 3 research questions were verbalized:

- What speaking tasks do teachers use to create opportunities for students to speak in order to assess their speaking performances?
- How do teachers assess/grade their students' speaking performances?
- What criteria do teachers choose for individual speaking tasks and what importance do they assign these criteria?

The results of the research are presented in the form of graphs and commented on in the chapter called Results and Commentaries. This is followed by a chapter addressing implications for language teaching that stem from the results. Limitations of the research together with suggestions on how the research could be improved, extended and complemented are also outlined. The thesis finishes with the Conclusion chapter that highlights and summarises the most important findings.

II. THEORETICAL BACKGROUND

The theoretical chapter presents readers with an overview of the theoretical background and aims at providing them with the basic knowledge that is needed for understanding the research of the thesis. Various approaches are taken into consideration and individual types of assessment with respect to speaking are explained. The core of this chapter lies in the section that is devoted to designing criteria for assessing speaking and implementing these criteria in rubrics. Being aware of the elements elaborated in this chapter leads one to seeing the assessment of speaking skills in a new light and reduces the pressure some teachers may experience by helping them realize what it is they want their students to master, which determines the content of individual classes and creates transparency as students know what they should focus on when learning and what is expected of them.

Speaking as One of Four Skills

Before getting to the nuts and bolts of the assessment of speaking skills, first it should be explained what the term speaking actually entails. Speaking, sometimes referred to as oral production, is one of the four skills¹ used when learning a foreign language. Since it belongs to the category of productive skills, it requires production on the part of the speaker. Being able to communicate a message orally is a complex process which involves putting the speaker's theoretical as well as practical knowledge of the foreign language in practice. According to Hinkel (2005), that includes phonetics, morphology, syntax, discourse markers and, last but not least, lexis (p. 485). The assessment of this skill is equally complex.

Ur (1991) emphasizes the importance of speaking skills when learning a foreign language as many, if not a majority of foreign language learners, aim at mastering speaking. It is also worth noticing that a person who is able to use a foreign language is referred to as a *speaker*, which suggests that speaking is, so to say, superordinate to other skills. Yet, the difficulties connected with the actual process of assessing these skills are so severe that many language teachers assess oral production only to a limited extent, give it very little importance, or do not assess it at all (pp. 120 - 134).

¹ The four skills comprise of listening, reading, writing and speaking.

One of the possible reasons is, as Hughes (2003) states, “[that] the objective of teaching spoken language is the development of the ability to interact successfully in that language, and this involves comprehension as well as production” (p. 113). Brown (2004), builds on that and claims that oral production is directly bound to aural intake; that is to say that speaking, except some cases, such as monologues, speeches, telling a story or reading aloud, can hardly be assessed separately from listening because of the fact that it is often a reaction to what the interlocutor has heard (p. 140). The speakers' production of language is often conditioned by their ability to understand the message that is being transmitted towards them and as a result of this it becomes difficult for the teacher / assessor to distinguish between assessing oral or aural abilities of the students.

Basic Types of Speaking - Levels of Interaction of Speaking and Listening Skills

In order to be able to isolate speaking from listening and therefore to be able to fulfil the aim of assessment of oral production and interaction, it is necessary to specify individual levels of interaction of speaking and listening skills. Brown (2004) uses a taxonomy in which he distinguishes between five levels; they are as follows: imitative, intensive, responsive, interactive and last but not least extensive (monologue) speaking (pp. 141 – 142). Concisely, imitative speaking aims at practising e.g. pronunciation patterns, such as word and sentence stress, intonation and aspects of connected speech (linking, elision and assimilation), an example of such kind of speaking is *drilling*. In intensive speaking the students are prompted to produce short, expected stretches of language, an example being picture-cued tasks, reading aloud, sentence translation etc. Responsive speaking, as its name suggests, requires students to be responsive to what they have been told and therefore to participate in a short interaction that is, however, limited in both length and the choice of topics, an example being a technique referred to as *question and answer*. Interactive speaking, on the other hand, differs from responsive speaking in terms of length and complexity. Multiple exchanges and multiple participants may be involved in this kind of interaction. Examples of such kind of speaking may be oral interviews, role plays, discussions, conversations and various games. The last type of speaking is referred to as extensive (monologue) speaking. It is usually prepared prior to the oral production itself, which is reflected in the more formal and deliberate language use. An example of such type of speaking may be an oral presentation (Brown, 2004, pp.

141 - 142). Individual types of speaking are directly bound to individual speaking tasks. A list of the most frequently occurring ones follows.

The Most Frequently Used Speaking Tasks

The following list presents several speaking tasks that may be found particularly often in language classes and they are: picture description, sentence translation, question and answer, interview, roleplay, game, and oral presentation. According to Brown (2004), picture description (or as he refers to it picture-cued tasks), involves some kind of visual stimuli that elicits a word, phrase or even longer propositions. The complexity of the visual stimuli may range from very simple pictures containing 1 or 2 items, to more elaborate pictures, such as maps, scenes from a party, a busy street, etc. (pp. 151 – 156). Brown (2004) goes on and claims that sentence translation, often overlooked due to not being in accordance with direct approaches used to create communicative classes, is a useful speaking task in teaching English as a foreign language (EFL) as understanding the lexical, grammatical and syntactic proposition is essential to non-native speakers (p. 159). Question and answer, being another speaking task, is similar to an oral interview; however, an interview is more interactive, allows one to swap roles and the interviewer may become an interviewee and vice versa, just as it happens in everyday communication. Question and answer tasks, on the other hand, are more oriented towards getting a correct answer and it is the teacher (assessor) who asks questions and it is the student that answers them (Brown, 2004, pp. 159 – 160). When roleplaying takes place, students “take on the persona of someone other than themselves” (Brown, 2004, p. 174), which lets them be creative and practise language they would normally not have the opportunity to practise, e.g. being a tourist asking for directions, a customer at a market trying to get a lower price, etc. (Brown, 2004, p. 174). Games, as Brown (2004) specifies them, may not be typical representatives of an speaking assessment task, but with a set of criteria and a reliable scoring method, they may serve as a means of formative assessment (pp. 175 – 176). Last but not least, oral presentation is also a frequently used speaking task. It requires the presenters to work on an assigned or chosen topic and deliver a speech that corresponds with what is expected of them in relation to their level of English (Brown, 2004, pp. 179 – 180).

Speaking as a Part of the Czech Curriculum

Since this thesis focuses on designing criteria for the assessment of speaking skills and reflects upon how Czech teachers of English design and use them, it is fundamental to mention to what extent students' speaking skills are developed and on what occasions they are assessed. The situations when teachers/assessors may want/need to assess students' speaking performances are many. Conversation classes, being one of the examples, are focused on exposing students to situations in which they have to communicate and therefore learn how to interact with their fellow students, or with their teachers. The aim of these classes is to prepare the students for situations in which they may find themselves in their everyday lives and equip them with what it takes to cooperate with other speakers of English successfully.

Another aim is to prepare the students for the speaking part of the new state Maturita exam. The reason for mentioning the Maturita exam and devoting a few paragraphs to explaining its individual parts is that speaking is an inseparable part of the language part of the Maturita exam and all students who choose a foreign language over Mathematics have to go through it. Therefore, not mentioning it would make the thesis incomplete.

At this stage it is important to make reference to the fact that students of English as a foreign language at secondary schools are supposed to reach at least the B1 level of reference. The expected outputs for students of secondary schools as for speaking are summarised in the “Framework Education Programme” document. The outputs are presented in the following manner:

as for spoken production, [students] should be able to:

- formulate [their] opinion in such a way that [they are] understood using correct grammar, spontaneously and coherently;
- reproduce freely and coherently an authentic text with vocabulary and language structures characteristic of a rather demanding text which [they have] read or listened to;
- describe in detail [their] surroundings, interests and activities related to them;
- use a broad general vocabulary to develop argumentation without reducing the content of the communication;

as for spoken interaction, [students] should be able to:

- express and defend [their] ideas, opinions and attitudes using appropriate oral forms;
- comment on and discuss various opinions on non-fiction and fiction texts adequately and use correct grammar;
- react spontaneously and use correct grammar in more complicated, less common situations while using appropriate phrases and expressions;
- communicate fluently on abstract as well as specific topics in less common or specialised situations, respecting the rules of pronunciation;
- begin, carry on and end conversations with native speakers and join in active discussion on various topics concerning more specialised interests

(Balada et al., 2007, p. 17)

Speaking as a Part of the Maturita Exam

Those who choose a foreign language over Mathematics have to go through a complex exam that focuses on examining their writing, listening, reading and last but not least, speaking skills. Out of one hundred and sixty five minutes that the foreign language exam takes, fifteen minutes are devoted to the oral part. Slightly more time (20 minutes) is given to students to prepare for it. During this time they are allowed to use a dictionary and make notes that they are going to use when being tested and assessed.

The tasks students have to accomplish are delivered by means of worksheets that are comprised of four sections during which spoken production as well as interaction are assessed. The oral part of Maturita exam is designed to determine the students' ability to produce larger stretches of language in L2, namely during a picture description and a presentation on a vocational topic when they have to demonstrate that they are able to speak without being prompted as well as the ability to interact with other speakers of English when introducing themselves and being interviewed by the teacher on a general topic (Ministerstvo školství, mládeže a tělovýchovy [MŠMT], n.d.).

Assessment of the oral part of the Maturita exam. Before taking a closer look at assessment itself, it should be briefly outlined how students' speaking skills are assessed during the oral part of Maturita exam. The act of examining and subsequent assessment of students' speaking performance is done by two certified assessors, one being the examiner and the other being the observer (MŠMT, n.d.). The assessors use a special worksheet with pieces of information and instructions on how to examine and assess the students. This worksheet contains a set of four criteria that help the assessors structure what to pay attention to when an assessment is produced and determine how well students perform on individual sections of the oral part of the exam. The criteria are as follows:

- content and the level of presentation skills;
- lexical competence
- grammatical competence and the means of text coherence and cohesion
- phonological competence

(MŠMT, 2013b, p. 6).

Each of the four sections of the oral part of the exam is evaluated in connection to the first three criteria and the student is awarded points on the scale of zero to three; zero meaning incapability of the particular student to meet the criterion and three meaning performing well and meeting the criterion. Phonological competence is given a somewhat lesser significance since this criterion is not applied to individual sections of the oral exam, but to the whole exam at once. Students pass the oral part of the exam provided they obtain at least 18 points out of 39 (which is equal to 44%) (MŠMT, 2013a, p. 9; MŠMT, 2013b, p. 6). It should also be noted that the oral part of the Maturita exam is given only a 25% importance as a majority is assigned to a didactic test (50%) and to writing two compositions (25%). Whether the ratio reflects the importance of individual skills is the question.

The Difference between Assessment and Testing

Every term that a person comes across carries meaning but also a connotation that stands, so to say, outside the word and represents how the word is perceived. The connotation of the word *assessment* is, without a doubt, not a positive one. Regardless of the emotions associated with the term assessment, it is a crucial and inseparable part of any kind of teaching / learning as it provides necessary feedback on the students' strengths

and/or weaknesses not only to the teacher, but also to the students themselves (and sometimes to other students as well).

When speaking of assessing students, one might often run across the term *testing* used in a similar context. An essential question has to be asked: what is the difference between assessment and testing? These two terms are sometimes misunderstood and may be falsely viewed as synonyms. Brown (2004), however, points out, that there is a difference in their meaning frequently overlooked. He goes on to claim that testing is a procedure prepared in advance during which students try to perform their best and are aware of the fact that they are being tested. On the other hand, assessment is a constant, sometimes even subconscious process that takes place whenever the student answers a question, gives a comment or tries to use a new word / phrase, etc. and may be done by teachers, the students themselves, and possibly even by other students (p. 4). In other words tests are only a subset of assessment (i.e. not every assessment is necessarily a test) and teachers may wish to use more forms of assessing their students.

What is Assessment

Concisely, assessment is a process of collecting, analysing and making use of information about students' performances (Paloma & Banta, 1999, p. 4). Astin (1993) adds that “the basic motive for gathering it is to improve the functioning of the institution and its people” (p. 2). Assessment is an omnipresent part of teaching without which the educational process would be incomplete. It may take numerous forms and over time various divisions have appeared. Despite there being quite a large number of them, it is worth bearing each in mind as they may give teachers a greater insight into this issue and therefore help them become more successful in assessing their students. Being able to distinguish between different types of assessment is crucial as each type has its peculiarities that make it suitable for different learning situations.

Basic Types of Assessment

The Common European Framework of Reference (CEFR), being a major influence on language teaching across Europe, the Czech Republic included, presents an extensive list of individual types of assessment, some examples being: formative/summative

assessment, subjective/objective assessment, direct/indirect assessment, holistic/analytic assessment, etc. (Council of Europe, n.d., p. 183). A complete list of all thirteen types of assessment can be found in Appendix A.

It should be noted that there are as many divisions of individual types of assessment as there are authors dealing with this issue. In other words, the layout of the classification overlaps and coincides with others according to the criteria that are used to create the particular classification.

Formal and informal assessment. Nevertheless, Brown (2004) claims that in its broadest sense assessment can be divided into two basic categories: informal and formal. The former one may be characterized as any kind of teachers' feedback aimed at the students, such as a word of encouragement, a smile, a pointed finger, a piece of advice concerning pronunciation, grammar, etc. (p. 5). However, for the purposes of this thesis, the point of interest is formal assessment produced by teachers. It is referred to as formal because of the fact that it is “[a] systematic, planned sampling technique constructed to give teacher and student an appraisal of student achievement” (Brown, 2004, p. 6).

Formative and summative assessment. Another division comprises of the two following categories: formative and summative. According to Hughes (2003), formative assessment serves the purpose of forming both the teaching as well as the learning part of the educational process. In other words that means that teachers use the feedback of such kind of assessment to monitor and modify their teaching plans and techniques and students may use it to see their progress and change their learning strategies accordingly (p. 5). To build on the terminology of the previous division, each and every informal assessment is formative.

Summative assessment, by contrast, sums up what the students have learned over a period of time - usually a unit, semester or the whole year and looks back at how successful the students have been in achieving the objective(s) of the unit, semester or the whole year (Hughes, 2003, p. 5).

Direct and indirect assessment. Direct and indirect assessment, being another category, is often mentioned in reference to assessing oral skills. The difference between the former and the latter one is the fact that when the teachers choose direct assessment, they assess what the students are actually performing (Council of Europe, n.d., p. 186).

Simply said, if the teachers wish to assess speaking, they have the students produce language orally, i.e. speak. An example of such assessment may be when a small group of students discuss something, the assessor observes such action, compares it with the criteria that were set prior to the assessment procedure and produces an assessment (Council of Europe, n.d., pp. 186 - 187). Whereas when indirect assessment is used, the students perform a skill that underlies the skill(s) which the assessor wants to measure. The assessor may, for instance, want to assess speaking by focusing on pronunciation, being a component of this productive skill. Thus, a pronunciation task may be used as an indirect means of assessing speaking.

Taking this example as a basis for further contemplation, there arises a series of questions that need to be answered: is the fact that the students are able to distinguish between two or more allophones conclusive proof of their ability to use it correctly in spoken discourse? Is the connection between, let us say, a written (multiple-choice) test and speaking strong enough? Is such assessment valid enough? When the students know that only one particular subskill (e.g. a phonemic distinction) is to be tested, will negative backwash effect not appear? The fact is that the students may decide to learn only that particular subskill without attempting to learn what the test item actually aims at – speaking. These are some of the questions that may be considered troublesome as they frequently tend to be the cause of debates to which many language teachers have tried to find legitimate answers.

Much alike any type of assessment, there are some pros but also some cons to it. Hughes (2003) argues in favour of indirect assessment as it “seems to offer the possibility of testing a representative sample of a finite number of abilities which underlie a potentially indefinite large number of manifestations of them” (Hughes, 2003, p. 18). As he points out, this may result in more precise evaluation of one's skills (p. 18). However, he also makes the reader aware of the fact that the connection between the indirect test item and the skill that is to be assessed “tends to be rather weak in strength and uncertain in nature” (Hughes, 2003, p. 18). That means that the ability to speak can not be guaranteed by e.g. the ability to pronounce words correctly because there is, of course, more to being able to speak.

Another argument favouring direct testing is that conducting the conditions which elicit the behaviour that the assessor wants to measure as well as the actual process of assessing productive skills are relatively straightforward (Hughes, 2003, p. 17). In other

words, the assessor simply creates the conditions, observes the student's performance and subsequently compares it with the criteria that were set beforehand. He goes on to say that the situations/materials used for direct assessing are more authentic compared to indirect (Hughes, 2003, p. 17).

Holistic and analytic scoring (assessment). According to The Council of Europe (n.d.), in order for assessment to produce valid, reliable and consistent results it is also extremely important to distinguish between scales for scoring (assessing) speaking skills that are used most frequently - holistic and analytic. Brown (2004) argues that it would be more appropriate to replace the term *scoring* with *assessment* “in order to capture its closer association with classroom language instruction than with formal testing” (p. 243). Notwithstanding the nuances in terminology, both of them are based on the fact that individual levels of the students' performances are assessed. However, when choosing holistic (also referred to as “impressionistic”) method of assessment, the students' performances are intuitively evaluated for their overall quality. In other words the performance is assigned only one single score based on the overall impression that the assessor(s) have (pp. 190 – 191). Such assessment has one indisputable advantage over its counterpart – it is much less time consuming than analytic assessment.

Different from this is the analytic method of assessment which looks at individual aspects separately and awards each of them its own score; these are then summed up to create a final score (Hughes, 2003, pp. 94 – 95). According to The Council of Europe (n.d.), the benefit of assessing individual aspects is that “[it] encourages the assessor to observe closely ... [it] provides a metalanguage for negotiation between assessors, and feedback to learners” (p. 190). Considering various criteria when assessing speaking skills separately and being able to debate whether the student fulfilled them helps to prevent the assessors from being biased or subjective. In other words, it enables the assessor(s) to be objective. This is further discussed in the section that deals with subjectivity in the assessment of speaking skills.

According to Hughes (2003), a research has proven that the data gathered when using holistic and analytic assessments show a very high agreement (p. 130). As a result, some may question why to devote more time and energy to using analytic assessment if the easier way is proven to produce similar outcomes. One of the reasons why it is worth putting one's effort into employing analytic assessment is that compared to holistic

assessment, it gives both the teachers as well as the students a more precise feedback on what the students do well and where there is still room for improvement (Knight, 1992, p. 300). Realizing what one can do and what needs more work is key to becoming more professional in using the language, which is, after all, what language classes aim at.

It should be taken into account that by no means is this list of types of assessment finite and therefore the divisions listed above should simply be viewed as the ones that are necessary to be familiar with. Generally said, the decision on which type of assessment to use lies solely on the assessor; however, the decision should always be based on a number of criteria that reflect upon e.g. the context/situation that precedes the assessment, the aim of the assessment as well as the needs of the students.

Validity and Reliability in Assessment

When inquiring into assessment, there are two terms that need to be explained in greater detail as they are of the same importance to effective assessment as oxygen is to life. They are validity and reliability. According to Brown (2004), validity is “by far the most complex criterion of an effective test – and arguably the most important principle...” (p. 22).

In order to better understand the term it might be beneficial to approach it from a different perspective. Angeles (1981) describes validity through a philosophical example in the following way “A deductive argument is valid whenever its conclusion necessarily follows from the premises; if the premises of the argument are true, then its conclusion cannot be false; the conclusion too must be true” (as cited in Hinkel, 2005, p. 795). That is to say that an argument is valid if its premises are in accordance with one another. Taking the fact that it is a philosophical point aside, in its core this statement applies to language assessment, too.

Ur (1991) offers a simplified explanation and says that “a valid test is one which actually tests what it is designed or intended to” (p. 44). If, for instance, the teachers/assessors attempt to measure the students' speaking skills, they should engage the students primarily in that particular skill and not in other ones, such as listening (Ur, 1991, pp. 21 – 22). Following this principle is one of the components that serve the purpose of ensuring that the assessment is effective.

Reliability, being another key aspect of effective assessment, is described as an element of consistency and dependability. Put simply, “a reliable test is one that produces consistent results when administered on different occasions” (Ur, 1991, pp. 44). According to Brown (2004), the level of un/reliability is dependant on many criteria; some examples being “fluctuations in the student, [in the assessor], in test administration and in the test itself” (pp. 20 – 21).

As for the first fluctuation, the students' physical and psychological condition, such as a temporary illness, anxiety, or just a bad day may result in the decrease of reliability of their performance and the assessment of such performance. As for the second one, the level of rater (assessor) reliability, or to be more precise either inter- or intra-rater (assessor) reliability, or possibly both may be the cause of decreased reliability. The former stands for inconsistency between two or more assessors and the latter stands for inconsistency within only one assessor. The reasons for the assessor to be unreliable are numerous. The lack of attention, inexperience, preconceived biases, unclear scoring criteria, being subconsciously harder, or easier on a few students who are assessed at the very beginning, or simply being tired towards the end of the assessment period are, beyond question, some of the most frequently occurring ones (Brown, 2004, pp. 21 – 22).

Brown (2004) goes on mentioning other elements that may prevent testing and assessment from being reliable; specifically they are the conditions in which the test is administered, such as street noise, poor lighting conditions, poor acoustics, the layout of the classroom, etc. In certain situations even the test itself may become an obstacle. Students tend to make more mistakes if the test is too extensive. Some students' performances may be affected by the fact that they have a time limit within which they have to complete the task(s) (pp. 20 – 12). Therefore, assessment is reliable only if these fluctuations are taken into consideration.

The relationship of reliability and validity is somewhat unclear and uncertain. However, it can be said that reliability is directly bound to validity and vice versa. Some argue that reliability should be seen as only a part of validity, which makes validity a superordinate term. As Lado states “[reliability] provides the framework, the structure on which validity depends and builds ... but is somehow independent of it” (as cited in Hinkel, 2005, p. 796). Explained in simpler terms, it means that a test may give consistent

results despite being invalid. To conclude, “reliability is necessary but not sufficient: sufficiency depends on validity” (Hinkel, 2005, p. 796).

Subjectivity in the Assessment of Speaking Skills

Beyond any doubt, assessing students' oral skills involves the teacher's judgement, thus it is said to be subjective. The difference between objective and subjective assessment resides in the fact that “objective assessment is assessment [from] which subjectivity is removed” (The Council of Europe, n.d., p. 188). On most occasions the aim of assessment is that the level of subjectivity is kept to its minimum. With the polarity of the previous statement reversed, teachers should attempt to make assessment as objective as possible. To build on the terminology of the previous division, holistic assessment is said to be more subjective as the students' performances are assessed, so to say, intuitively and the teachers (assessors) go with their instinct, whereas analytic assessment uses individual criteria that help the assessors be more conscious of what they actually assess.

Possible Ways of Ensuring Objectivity

The current findings advise one to establish a set of criteria that are to be addressed during the act of assessment. Generally said, the more criteria there are to choose from, the greater the chance of the assessment being objective and precise there is.

One way of doing that is to break speaking skills down into individual sub-skills that are to be assessed and approach them selectively. Brown (2004) speaks of so called micro- and macro-skills from which teachers / assessors select one or several that they use as an objective(s) of their assessment tasks (p. 142). The former one refers to using smaller units of language, such as “phonemes, morphemes, words, collocations, and phrasal units” (Brown, 2004, p. 142). An example of these micro-skills may be statements such as “[The students] produce differences among English phonemes and allophonic variants, ... reduced forms of words and phrases ... [and] produce English stress patterns, words in stressed and unstressed positions, rhythmic structure and intonation contours ...” etc. (Brown, 2004, pp. 142 – 143).

The latter one, by contrast, stands for applying larger elements, such as “fluency, discourse, function, style, cohesion, non-verbal communication and strategic options”

(Brown, 2004, p 142); some examples being the following statements “[The students] appropriately accomplish communicative functions according to situations, participants, and goals, ... convey facial features, kinesics, body language, and other non-verbal cues along with verbal language, etc. (Brown, 2004, p. 142). It is worth noticing that some of these sub-skills are not limited to speaking only, but can be applied to communication in general. Full list of micro- and macro-skills is presented in Appendix B.

Another way of ensuring objectivity is to create a list of criteria. The Council of Europe (n.d.) presents a list of 14 qualitative categories (criteria) to select from when assessing speaking skills; some examples being turn-taking strategies, fluency, thematic development, precision, vocabulary range and control, phonological control, grammatical accuracy, etc. (p. 193). The saying “less (fewer) is more” is applicable in this context as one of the aims of using criteria is to make assessment more understandable for teachers as well as for students and choosing too many criteria would be counterproductive. The full list is to be found in Appendix C.

As mentioned above, it is of utmost importance that the number of criteria that the assessor decides to address is feasible. According to The Council of Europe (n.d.), it is not in one's capacities to assess all criteria at once. Rather than that it is suggested that the list be approached selectively and only several criteria that are relevant to the particular situation/context are chosen. It is also stated that attempting to assess 7 or more criteria at once may cause a *cognitive overload* as the assessor has to manage to be the interviewer as well as the assessor. Choosing no more than 4 or 5 criteria secures feasibility as well as reliability of the assessment (pp. 192 - 193).

Brown (2007) merges some of the criteria mentioned in the preceding division and creates a shortened and, so to say, more straightforward list with several possible criteria to be assessed in speaking and they are as follows:

- pronunciation
- fluency
- vocabulary
- grammar
- discourse features (cohesion, sociolinguistic appropriateness, etc.)
- task (accomplishing the objective of the task is also closely dependant on comprehension).

(p. 352).

Both of the suggested sets of criteria are very similar in content, but use slightly different terminology. Therefore, following either set of rules mentioned above is one of the prerequisites of ensuring that the assessment fulfils its objective(s). This will be revisited and discussed in more detail later on in this chapter.

Designing Rubrics

The core of this thesis is designing criteria to assess speaking skills; however, the criteria are part of a larger system that needs to be explained first – rubrics (also referred to as rating scales)². Klečková (2010) defines a rubric as “a scoring instrument/guide [with] clearly stated product or performance criteria” (slide No. 3). Many researchers believe that using criteria to assess students, i.e. incorporating rubrics into one's assessment, is the key to objective and transparent assessment. Rubrics also help to reduce the pressure one may feel in relation to assessment, whether it be the teacher trying to decide what to pay attention to, or the student anxious, not knowing what to expect from the assessment; hence it is argued that students should be familiar with the rubric prior to completing the task.

There are two basic types of rubrics, and they correspond with holistic and analytic scoring (Mertler, 2001, *Designing scoring rubrics for your classroom*, para. 2). The peculiarities and pros and cons of both of them have already been outlined in the section dealing with holistic and analytic assessment. However, Mertler (2001) also draws one's attention to the fact that each scoring rubric is suitable for a different purpose. The answer to the question whether the assessor wants/needs to get an overall, general picture of the students' performance or a more detailed one with respect to individual criteria determines which scoring rubric is more appropriate (*Designing scoring rubrics for your classroom*, para. 3 - 4). He goes on to say that “if an overall, summative score is desired, a holistic scoring approach would be more desirable ... in contrast, if formative feedback is the goal, an analytic scoring rubric should be used” (Mertler, 2001, *Designing scoring rubrics for your classroom*, para. 5). It is designing the analytic rubric that will be further discussed.

² Mertler (2001) clarifies that the terms *rubrics* and *rating scales* denote one and the same notion and may be used interchangeably (*Designing scoring rubrics for your classroom*, para. 1).

Designing Criteria to Assess Speaking Skills

As it was outlined in the above sections of this chapter, the importance given to teaching speaking skills and the actual assessment of these skills are in contradiction to each other. Knight (1992) explains that one of the possible reasons for the occurrence of this phenomenon is “[the difficulty with] deciding which criteria to use in making an assessment” (p. 294). Assessors should not only be able to determine possible criteria that could be used to assess students' speaking skills, but also narrow the list down to those that are the most relevant ones with regard to the objective(s) of the course/lesson/etc. and assign them appropriate weight in relation to the remaining ones (Knight, 1992, pp. 294 – 298).

Possible Scoring Criteria

Before immersing into the art of choosing criteria to be used for the assessment of speaking skills, a list of possible aspects of oral production and interaction to be assessed should be compiled. There are many authors and institutions that have been concerned with this issue and produced a number of lists of criteria; however, the most comprehensive and yet comprehensible and for the purpose of this thesis preferable list is one offered by Knight (1992). The list shows a variety of criteria from which teachers may choose the most appropriate ones based on the objective(s) of the assessment and the context in which the assessment is conducted. According to him, the list is a fusion of various sources, and consists of (a) grammar; (b) vocabulary; (c) pronunciation; (d) fluency; (e) conversational³ skills; (f) sociolinguistic skills; (g) non-verbal skills; (h) content (pp. 295 – 296). The reason why each criterion is addressed separately and is devoted so much attention is that once you know what you are dealing with, you are conscious of it and it becomes easier to spot it.

Grammar. Grammar, being the first criterion enumerated in the list above, is often described as a set of rules by which a language is created, or a set of “rules for forming words and combining them into sentences” (Swan, 2005, p. xix; Hornby, 1989, p. 542). Knight (1992) further distinguishes between the range and accuracy of grammar. The

³ Knight (1992) refers to them as conversational skills; however, it is obvious through the context that it is a typo.

former stands for the amount of grammatical rules that students are able to use (quantitative feature) and the latter stands for how accurately, i.e. correctly, students are able to use them (qualitative feature) (pp. 295 – 296).

It is rather curious that grammar (and more specifically grammar accuracy) is frequently considered the most important criterion when assessing students' speaking skills. Knight's (1992) theory is that it has to do with the way non-native speakers of English learn English rather than the fact that it is more difficult and important to acquire than “discourse and sociolinguistic skills” (p. 301). In other words, there is a misleading concept that some non-native speakers of English have adopted – they tend to emphasize mastering grammar and believe that being able to use grammatical structures correctly means mastering the language itself. Knowing how to form words and combine them into sentences is, of course, greatly important; however, there is much more that forms communication than simply following a set of rules. Other scoring criteria follow.

Vocabulary. Vocabulary is another criterion that is frequently chosen to be assessed in relation to speaking skills. The meaning of the term vocabulary is defined as “[the] total number of words that make up a language”, or “[a] body of words ... used in a particular subject, etc.” (Hornby, 1989, p. 1425). Just as in the case of the previous criterion, it is distinguished between range and accuracy again. The definition suggests that the range (amount of words one uses successfully) should be related to a particular situation or topic and the objective of the course. Accuracy, in relation to vocabulary, is not explained in any of the reference books; however, the definition of the adjective *accurate* “free from error” (Hornby, 1989, p. 9) leads one to thinking that the amount of errors made in choosing the right words in particular contexts stands for vocabulary accuracy. Nonetheless, the notion of vocabulary accuracy is not one hundred per cent unequivocal, and some may suggest it partially falls under the sociolinguistic skill (sociolinguistic appropriateness) that will be mentioned later on.

Pronunciation. As for pronunciation, quite a few more detailed criteria are distinguished within this category. Knight (1992) mentions pronouncing individual sounds (phonemic distinction), applying word and sentence stress and rhythm, intonation and last but not least, aspects of connected speech, including linking, elision and assimilation (pp. 295 – 296). When assessing students' pronunciation, the accuracy of the above mentioned

criteria (and possibly some other ones) is usually compared against the native speaker standard.

Hughes (2003) states that this approach has been criticised lately (p. 130). With English becoming a means of international communication, there arises a question which of the English pronunciation standards to apply. This has become a topic so frequently discussed that attempting to uncover even some of its peculiarities would make up for another thesis and a definitive answer would not be found. Therefore, the choice of which pronunciation standard to teach and compare students' performances against during the assessment lies in the competence of the teacher.

Fluency. Fluency is another criterion that may be taken into consideration when assessing students' speaking skills. It is often mentioned in contrast to accuracy and these two terms may be seen as two opposites that are mutually exclusive. Scrivener (2011) supports this statement by saying that, metaphorically speaking, there is a switch in one's head that switches between accuracy and fluency based on the setting one finds him/herself in (pp. 224 – 225). Accuracy has been described in relation to grammar, vocabulary and pronunciation as “free from error”. When focused on assessing fluency, the main concern is for the speech to be fluent, i.e. to flow naturally without worrying too much about being one hundred per cent accurate.

There are various aspects that make speech seem flow effortlessly. Knight (1992) goes so far as to claim that there are 12 individual subcriteria that can be assessed; some examples being the speed of talking (the number of words per minute), hesitation while speaking and hesitation before speaking, etc. (p. 296). However, according to Cohen, it is important to bear in mind that hesitation may not be something undesirable as it helps students find the correct vocabulary items, and “fluency is rather the ability to know how to hesitate, ... stay silent or self-correct thus producing an acceptable and relaxed quality of speech” (as cited in Restrepo et al., 2003, p. 70). That being said, it becomes obvious that without the knowledge of appropriate vocabulary and idiomatic expressions and the knowledge of how to combine them, there would be no fluency as the students would have nothing to say.

Conversational skills. Conversational skills are one of the broadest criteria of all that are enumerated in this list as the skills that are needed to form a conversation are

many. They are described as the ability to have and maintain a conversation. Among some of them belong the ability to maintain cohesion with one's own utterances as well as with the interlocutor's ones, taking turns in conversation, correcting oneself, using pause fillers, asking for clarification and last but not least, to develop the topic of the discussion (Knight, 1992, p. 296). These are just a few of the most relevant examples of what it is that makes having a conversation possible. One may also notice that some of these rules coincide with the rules of etiquette.

Sociolinguistic skills. The following criterion is referred to as sociolinguistic skill (appropriateness). Knight's (1992) explanation lacks enough detail to fully understand its complexity. He speaks of different registers and styles (e.g. formal and informal) and the use of cultural references (p. 296), but that is not all there is to it. Put simply, it means being able to evaluate situations and know what the right thing to say (or do) is. It is easier said than done. It has to be noted that scarcely ever is teaching the sociolinguistic skill one of the objectives of language classes and yet, without this skill, even a brilliant student in terms of grammar, vocabulary, pronunciation, fluency, etc. may be perceived as ill-mannered, rude, hostile, or worse due to not being contextually appropriate (Broersma, 2001, pp. 191 – 207). Thus, mastering the sociolinguistic skill is just as important as mastering other skills needed for speaking. Nevertheless, this criterion is so complex and difficult to grasp that assessing it is a real challenge. What might make the assessment easier to some extent is to realize that the sociolinguistic skill is closely interconnected with vocabulary, since, simply said, it refers to the ability to use the right vocabulary in the right context as it was mentioned when addressing vocabulary accuracy.

Non-verbal communication skills. Despite the fact that the thesis focuses on the assessment of speaking skills, a part of speaking is not only what we say with our words, but also what we say with our bodies. According to Knight (1992), the way we use body language, i.e. posture, gestures, facial expressions and eye contact is extremely important as it complements what we say and our brains put a considerable amount of attention to interpreting signs of non-verbal communication (p. 296). According to some researches, over a half of our attention is paid to non-verbal signs and less than a half to what we say. However, other researches disproved this theory by saying the original research was misinterpreted as the application of the results was very limited, and what is more, was not

supposed to be applied to normal communications. Regardless of these contradictory findings, non-verbal cues are a very important part of communication. An example explaining why it is so relevant is that when one's verbal and non-verbal way of communication are not in accordance with each other, it may be (and usually is) perceived as at least suspicious and not trustworthy because of the fact that the body language contradicts the words (Johnson, n.d.).

Content. The last of all criteria as presented by Knight (1992) is content. By this he refers to the coherence and relevance of arguments (p. 296). It means that the arguments, ideas, comments, etc. should be arranged in such manner that they are logical, consistent, to the point and therefore easy to understand.

As shown on the example criteria above, speaking is argued to be the most complex of all skills, and requires a high level of concentration on the part of the speaker. Language production as well as interaction are a fusion of many aspects, some of which go beyond the scope of linguistics and may be found to be a subject for debate in sociolinguistics, psychology, neurosemantics, etc. The next section focuses on selecting individual criteria and assigning them a certain level of importance in relation to the other ones.

Selecting and Weighting Assessment Criteria

To begin with, it has to be noted that there is not one definite set of best criteria that will work in every situation. In order to be able to form the best list of criteria possible, it is absolutely essential to know for what context the list is designed. According to Knight (1992), among these context variables that can change the choice of relevant criteria and their weighting belong the purpose of the test, the choice of an elicitation technique and other tests in the battery (p. 298).

As for the purpose of the test, different criteria and their weighting will, of course, be chosen for e.g. a final achievement test at the end of a unit, semester or school year (summative assessment), and different for a diagnostic test in the course of the school year with the objective to determine what the students can do and what requires some further teaching (formative assessment).

According to Knight (1992), the choice of elicitation technique greatly affects the sample of language being assessed (p. 298). An interview, a presentation, a role play, a picture description, a game, etc. will each create different conditions and offer different opportunities for the students to demonstrate their speaking skills (Restrepo, 2003, p. 71). The choice of an elicitation technique also affects the number of participants, and as a result it affects “the degree of freedom or control over what the student could say and do ... [and] the extent to which assessor participates in speaking situations” (Knight, 1992, p. 298).

The battery of tests, i.e. other ways of testing students' speaking performances that the teacher may wish to use may also alter the choice and weighting of individual criteria. An example of such occurrence may be a situation when grammatical accuracy was tested previously, thus testing it in spoken production and interaction may be redundant and the assessor(s) can focus on the assessment of other criteria instead (Knight, 1992, p. 298).

Levels of Performance and Descriptors

Considering setting and weighting of individual criteria as the first step on the journey to objectivity, determining how well students have performed within individual criteria is the next step to take. In order to do so, levels that describe students' ability to meet the criteria need to be compiled. As for the number of them, it may vary and there is not an exact number that is correct in every situation. Mueller (n.d.) advises one to start with fewer levels, e.g. as few as three, because of the fact that it is “easier and quicker to administer, easier to explain to students (and others) and easier to expand than [it is for] larger rubrics to shrink” (How Many Levels of Performance Should I Include in my Rubric?, para. 6).

Another component of a scoring rubric is a descriptor. As claimed by Mueller (n.d.), “A descriptor tells students more precisely what assessment looks like at each level ... and what is expected of [them]” (Descriptors, para. 1). It helps students know what is meant by e.g. *above average*, *average*, *below average*, and makes the rubric more transparent and easier to understand. Including descriptors in the rubric is beneficial, though not necessary (Mueller, n.d., Descriptors, para. 2). Klečková (2010) suggests that clear descriptive language be used and presents several possible sets of adjectives to be used when describing individual levels, some examples being:

- explicit → sufficient → limited⁴; or
- clear → considerable → partial → little → no; or
- outstanding → very good → acceptable → somewhat deficient → very deficient
(slide No.16).

From the information above it is obvious that a scoring rubric is a flexible tool that may be adjusted to meet the specific needs of an assessment, but it should never be used in the opposite direction; i.e. the objective(s) of an assessment should never be adjusted in order to meet a rubric created for a different occasion.

Analytic Rubrics

Now, that all parts of the analytic rubric have been explained, it is time to put the pieces together and design one. Analytic rubrics are, simply said, the merging point of individual criteria and the levels of students' performance. To make matters more understandable, one may visualize a grid where the left vertical column states individual criteria, such as grammar accuracy, pronunciation accuracy, conversational skills, fluency, etc. that are being assessed and the horizontal lines specify to what degree students have met each criterion. This helps the assessor determine the level of students' proficiency in relation to each criterion separately, which gives the assessment a much greater precision than assessing students' performance as a whole.

Klečková (2010) sees the process of designing a scoring rubric in five steps. She starts by examining the learning objectives. Realizing what it is that we want the students to be able to do is perhaps the most difficult and also the most important idea to verbalize. This leads to the second step, which is setting individual criteria (grammatical accuracy, fluency, etc.), or even more specific characteristics (the student can say/describe/...). Along with this, individual levels of performance should be determined. The third step is to write basic descriptions (descriptors) for the highest and lowest performance. Next, she states that writing the remaining descriptions, i.e. for levels between the highest and lowest performance, is the fourth step. She concludes by saying that the last step is to revise the rubric and adjust the criteria or the levels of performance and their descriptions if needed (slides No.12 – 17).

⁴ The sets of adjectives are presented in one line in order to preserve their linearity as they would be used in an analytic rubric.

In most learning situations the data gathered from rubrics are converted into grades (marks) regardless of the fact that many teachers know that grades provide little feedback to students and may have a negative effect on the students' attitude towards learning. Rubrics, on the other hand, are a valuable source of information in language assessment, for they inform both the teacher as well as the student about what was done well, and also indicate what the next step in learning (or teaching) is. Comparing rubrics to grades (marks) makes one aware of the benefits that rubrics offer over the limited, though in most learning situations still preferred, grades. Arter (2006) well sums it up in this advice: “Assess a lot; grade a little” (p. 114).

The theoretical chapter is to be left with a few thoughts to ponder. Speaking is a skill that is arguably the most essential when learning a foreign language, hence it should be treated accordingly not only when practising it, but also when assessing it. Assessment in particular may be challenging for a number of reasons. Knight (1992) explains that some teachers may experience difficulties when it comes to “designing productive and relevant speaking tasks ... being consistent (on different occasions, with different testees and between different assessors) ... [and] deciding which criteria to use in making an assessment” (p. 294).

Analytic rubrics make good tools that help reduce the anxiety some teachers may feel in relation to assessing speaking; however, despite the fact that the main aim of incorporating rubrics into one's teaching style is to remove subjectivity from the assessment, Knight (1992) says it should also be taken into consideration that there still remains a great deal of subjectivity in choosing the criteria, determining how well students performed in each of them and what significance individual criteria should be given in relation to other ones as the answers vary from teacher to teacher (p. 299). What is more, even the same teacher when asked again after some time has passed may choose different criteria and their weighting. Therefore, at the end of the day it is important to realize that objectivity is a relative term and achieving it may be attempted, though never fully accomplished.

The theoretical framework, explaining key concepts, such as speaking, assessment, assessment criteria, etc., equips the readers with the knowledge that may take them on the journey of producing valid, reliable and transparent assessment. That said, it is time to convert this theory into practice and a research study follows.

III. METHODS

This chapter describes the practical part of the thesis - the research, presents research questions, explains how the data were gathered, and invites readers to read on and find out whether the research hypotheses were confirmed or disproved.

Introduction to the Research

When working on the theoretical background chapter, it became obvious that the research would attempt to discover current practices in the assessment of speaking skills and ascertain teachers' perception of criteria and how this perception changes in relation to individual assessment tasks. In other words, the research aims at finding an answer to the question how teachers of English assess their students' speaking performances. In order to do so, three research questions crystallized:

- What speaking tasks do teachers use to create opportunities for students to speak in order to assess their speaking performances?
- How do teachers assess/grade their students' speaking performances?
- What criteria do teachers choose for individual speaking tasks and what importance do they assign these criteria?

Research Tool

A questionnaire, being a quantitative method, was chosen as the most appropriate method for the purpose of the research of this thesis. The reason for this was that it enables one to gather a large amount of data from a large number of respondents in relatively little time. Analysing the data gathered from questionnaires is equally time efficient. In order to avoid any misunderstanding, the whole questionnaire was designed in Czech⁵. The questionnaires were anonymous and took seven minutes to complete at the most.

⁵ The Czech version of the questionnaire along with the English translation is included in the Appendices section and can be found under the title Appendix D and Appendix E, respectively.

As for the structure of the research tool itself, it comprised of 11⁶ questions that were presented in two sections. The former section focused on personal data, such as gender, the number of years of teaching experience in general and in teaching English, respectively; the kind of school where the teachers taught, whether they were qualified to teach English and last but not least, whether they had undergone a seminar/workshop focused on any kind of assessment, and if so, what the focus of the seminar was.

The latter section of the questionnaire was designed specifically to provide answers to the three research questions stated above and therefore focused on the frequency of using various speaking assessment tasks, determining how teachers assessed their students' speaking skills and last but not least, which criteria they would select for 3 pre-selected assessment tasks and how important they considered the criteria in relation to other criteria.

A more detailed clarification of the questions follows. First, the teachers were asked to choose how frequently they use individual assessment tasks. The options to choose from were roleplay, oral presentation, picture description, interview, sentence translation, question and answer, game or other assessment tasks if the teachers had some in the battery of tasks they used. The teachers assigned each task how frequently they use it on the scale of 4 to 0; 4 being *very often*, 3 meaning *often*, 2 meaning *sometimes*, 1 meaning *rarely* and 0 being equal to *never*. In the following question they were supposed to choose an option that best described how they assess their students' speaking skills. There were three options they could choose from – holistically, analytically or subjectively (descriptions of what such assessments look like were used instead of the technical terms). The last question was presented in the form of a table⁷ consisting of 4 columns with the first column stating possible scoring criteria, such as grammar, vocabulary, pronunciation, fluency, conversational skills, sociolinguistic skills, non-verbal communication and content and the remaining three columns representing 3 pre-selected assessment tasks – oral picture description, interview between student and teacher and oral presentation. The choice of these assessment tasks was not random, but was based on the presumption that these are the assessment tasks are the most frequently used ones. The respondents' task was to note down how important they considered each criterion with respect to each of the three

⁶ The necessity to answer all 11 questions was conditioned by the answer to question 6 and 9 which were designed in the sense that one of the options required additional information; i.e. question 7 and 10 were follow up questions in case more information was needed. As a result of this, the number of questions answered may range from 9 to 11.

⁷ It is advised that readers consult the questionnaire as presented in the Appendices section as the table is relatively complicated to be fully visualized and understood purely from this description.

assessment tasks on the scale of 3 to 0; 3 meaning the criterion is *very important*, 2 meaning it is *important*, 1 *not so important* and 0 equal to *unimportant*.

Research Participants

In order to uncover the current situation when assessing speaking skills in English classes at Czech schools, it was necessary to ask the people who conduct assessment – teachers. The answers were gathered by means of questionnaires that were given to teachers of English as a foreign language on various occasions. Out of the total number of 66 questionnaires that were used to provide answers to the research questions - over a half of the questionnaires (39) were filled in by English teachers who participated in a seminar focusing on dyslexia in English language teaching held at the Department of English, Faculty of Education, the University of West Bohemia, Plzeň, Czech Republic on April 14, 2014. The rest of the questionnaires (27) were distributed and gathered via email from teachers in the Pilsen and Klatovy regions.

It has to be noted that not all of the questionnaires were returned, and 6 questionnaires out of the total number of 66 had to be eliminated from the study, for they were incomplete, or the answers to individual questions contradicted each other. As a result, the research employs 60 questionnaires that were filled in completely by teachers working in various spheres, including: lower (14) and upper (16) elementary schools, secondary vocational schools (12) and grammar schools (10) as well as teachers working in the private language industry (8).

The fact that 56 of the respondents were female and only 4 of them were male suggests that there is a great disproportion in terms of gender representation. As for the qualification to teach English, 42 teachers were qualified and 18 were not; i.e. nearly one third of the respondents were not qualified to teach English. Qualification aside, during their teaching career, only 11 respondents attended a seminar/workshop/conference focused on assessment (which leaves 49 teachers uneducated in this area). The most frequently occurring topic of the seminar/workshop/conference was the assessment during the Maturita exam, or the assessment of children with special needs. The teachers had been in the profession for 6-10 years on average and the same amount of professional experience applies to teaching English, too.

The above characteristics of the respondents seem to reflect the current situation and some of the problems of the educational process, the social status of the profession and how it is perceived by the public; however, the variety of the teachers' backgrounds diversifies the viewpoints on the issue at hand and helps to provide a fairly representative and reliable sample of answers from which conclusions may be drawn.

Research Procedure

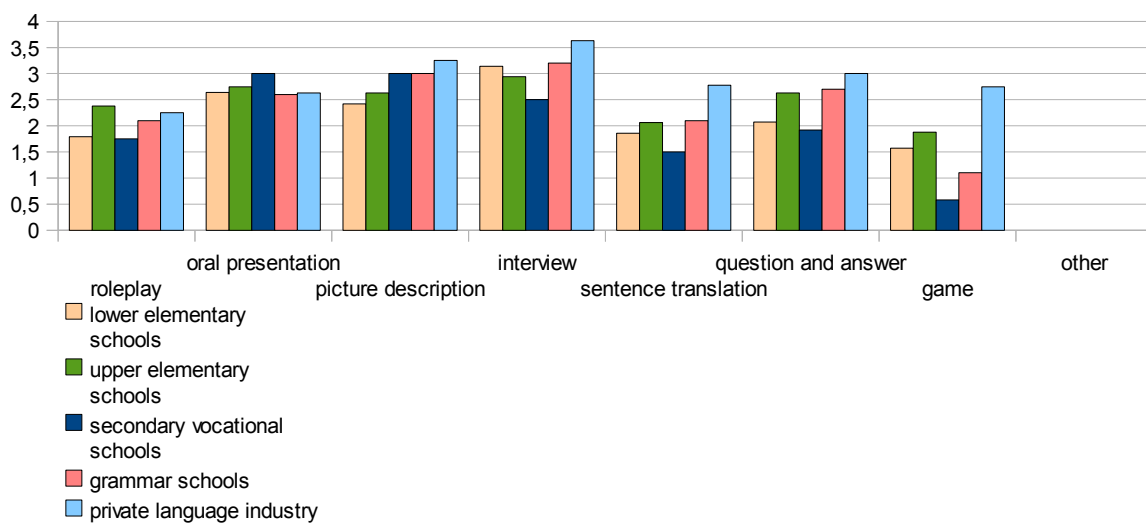
The teachers were each given either a printed, or an electronic version of the questionnaire and were asked to answer the questions based on their preferences, beliefs and professional experience. In nine of the questions they ticked the option that best described their choice and two questions were follow up questions that served the purpose of getting additional information and as a result required the respondents to finish a sentence with a few words. The data gathered from the questionnaires are analysed, presented in graphs and commented on in the following chapter.

IV. RESULTS AND COMMENTARIES

As it was outlined in the previous chapter, the thesis aims at uncovering the current practices in the assessment of speaking skills and ascertaining teachers' perception of criteria for the assessment of these skills. This chapter focuses on analysing the data gathered from the questionnaires on the basis of the theoretical background chapter and presents the results first, with respect to individual types of schools, starting with lower elementary, upper elementary, secondary vocational school, grammar school and private language industry, then the overall results. The results are presented in the form of graphs that are followed by commentaries explaining the results in greater depth and commenting on what they seem to signify. Finally, the results are summarized, confronted with the research questions and conclusions are drawn.

The Frequency of Using Individual Speaking Tasks to Assess Speaking Skills

The first research question focused on determining which speaking tasks were the most frequently used ones for the assessment of students' speaking performances at Czech schools. The following graph presents the results for each type of school separately, starting with lower and upper elementary schools, secondary vocational schools, grammar schools and private language industry. Each speaking task was awarded up to 4 points on a 5-level scale⁸.

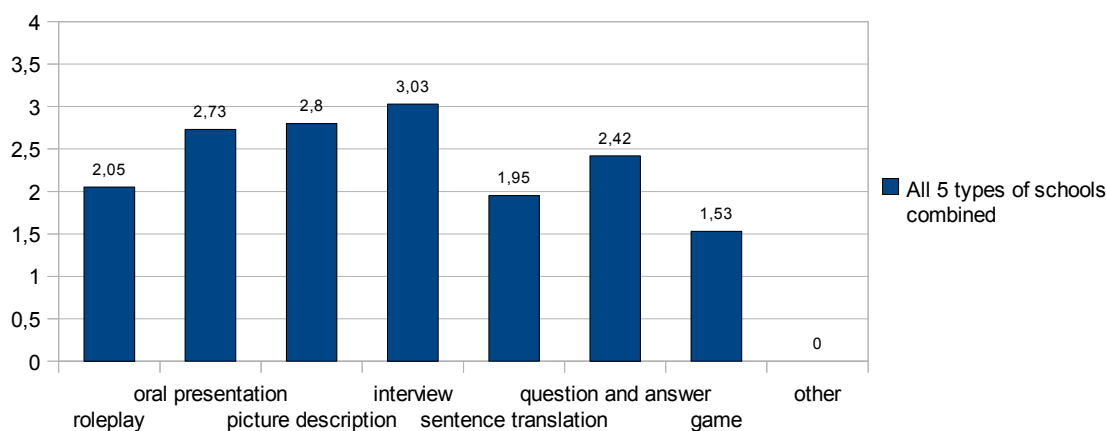


Graph 1. The frequency of using individual speaking assessment tasks for each type of school separately.

⁸ 4 meaning the task is used *very often*, 3 *often*, 2 *sometimes*, 1 *rarely* and 0 *never*

As can be seen in the graph above, the results show that the frequency of using individual speaking tasks at different types of schools is relatively balanced, 2 exceptions breaking the quiet being *an interview* in private language industry with 3.63 points out of 4 and *a game* at secondary vocational schools with mere 0.58 points, which suggests that the former is used between *often* and *very often* in private language industry, and the latter between *rarely* and *never* at secondary vocational schools. It is worth noticing that compared to other speaking tasks, an *interview* is the most frequently used assessment task at 4 out of 5 types of schools and similarly, a *game* is the least frequently used task at 4 out of 5 types of schools. As for games, it is also noteworthy that the difference in the use of games as speaking assessment tasks at secondary vocational schools and in private language industry is over 2 points. None of the respondents stated any other additional forms of speaking tasks that they use to assess their students' speaking skills.

In order to determine which tasks are the most frequently used ones across the Czech curriculum, another, more concise graph that combines the results of individual types of schools together is compiled. Just as in the case of the previous graph, each speaking task was awarded up to 4 points on a 5-level scale.



Graph 2. The frequency of using individual speaking assessment tasks for all 5 types of schools combined.

At first glance, the results suggest that the most frequently employed speaking task is an interview between the teacher and the student, scoring just over 3 points out of 4, meaning that it is *often* used to assess students' speaking skills. A possible reason for this high agreement among the respondents may be that an interview is relatively easy to prepare, conduct and also that it is very close to the way people communicate in their everyday lives. Therefore, it is natural and students do not feel too intimidated as they communicate with their teachers (assessors) in a similar way they communicate with their

peers. Another reason for this occurrence may be that the respondents were in the profession on average between 6 to 10 years, which suggests that they were relatively close with their students in terms of age and as a result they tended to be more upbeat, and chose a speaking task that lets them interact in the most natural way.

The second most frequently used speaking task is picture description with 2.8 points (out of 4), which leaves it less than three tenths behind an interview and it can be said it is also used *often*. Just as in the case of an interview, a reason for this may be that a picture description is relatively easy to prepare and conduct and it also allows longer student talking time and as a result the teachers (assessors) may make a clear and precise assessment of the students' speaking skills as they can focus solely on making an assessment. It is also worth paying attention to the fact that picture description is, to a certain extent, similar to describing a mental picture in one's own mind, i.e. a memory. Talking about memories and experiences is something people do considerably often.

An oral presentation, being the third most frequently used speaking task, scored 2.73 points out of 4, which means it is used rather *often* than *sometimes*. The fact that students have to be active and teachers can enjoy a moment when they do not have to speak may be considered one of the reasons of such high agreement between teachers. Giving a presentation on a chosen or assigned topic, however, unless being particularly captivating, or making the observers active in some way, leads to their decreased concentration. Yet, despite these constraints, teachers ranked it the third most frequently used means of assessment of their students' speaking skills.

Question and answer speaking task, with 2.42 points, is used between *sometimes* and *often*. It is rather similar to an interview in terms of the process itself; however, an interview is more interactive, allows one to change roles, and as a result the interviewer may become the interviewee and vice versa, just as it happens in everyday communication. But question and answer tasks tend to be more oriented towards getting a correct answer and it is the teacher (assessor) who asks questions and it is the student that answers them.

Roleplay and sentence translation achieved remarkably similar overall score and are used *sometimes* as they were awarded 2.05 and 1.95 points out of 4, respectively.

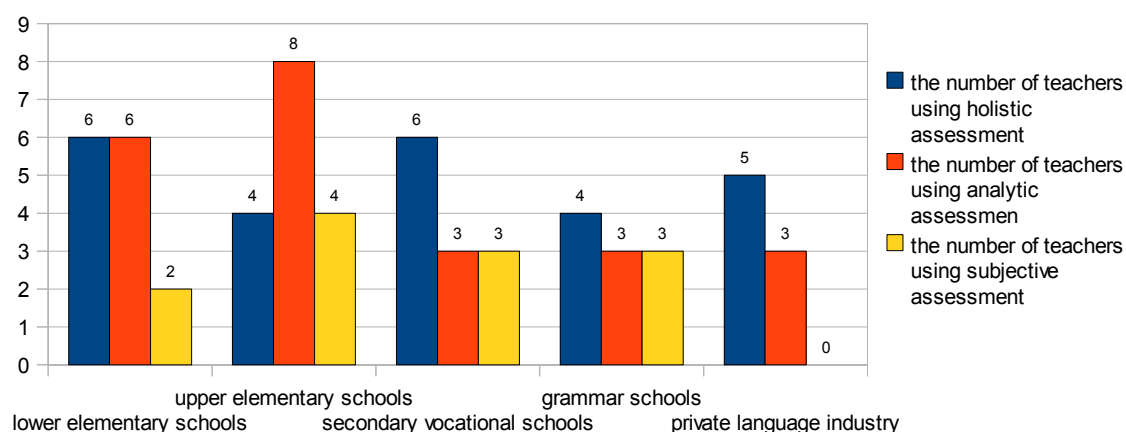
Games proved to be the least frequently used speaking tasks when it comes to assessing speaking. The reason behind this may be that because being formative in nature, they are more suitable for learning and practising certain language phenomena than they are for making an assessment. What is more, assessment is usually associated with a

certain level of seriousness and assessing by means of games might decrease it to a certain extent.

Before attempting to answer the second research question, it might be beneficial to sum up this section. In brief, an interview, picture description and oral presentation belong among the most commonly occurring speaking assessment tasks. Question and answer, roleplay, sentence translation and games, on the other hand, belong to the less frequently chosen options.

Teachers' Preferences in Types of Assessment

The second research question aims at discovering how Czech teachers of English assess their students speaking performances: holistically, analytically, or based on their personal experience, i.e. subjectively. The first graph presents the results for each type of school separately, while the second graph shows the results for all 5 types of schools combined together.

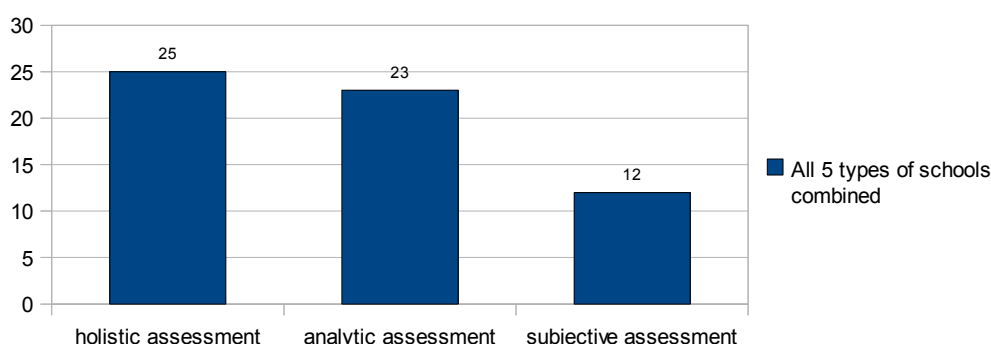


Graph 3. Teachers' preferences in types of assessment of speaking skills for each type of school separately.

As the data in the graph above show, there are some noticeable differences between how teachers assess their students speaking skills at different types of schools. While at lower elementary schools the number of teachers employing holistic and analytic assessment is exactly the same (6) and the number of teachers who assess their students without designing any kind of criteria was relatively low (2), the situation at upper elementary schools was notably different. A majority of teachers (8) set individual criteria and attend to them separately: 4 teachers use holistic assessment and the same number of

teachers assess subjectively. At secondary vocational schools, a half of the teachers (6) use holistic assessment and considerably fewer teachers (3) use analytic assessment. The same number of teachers (3) assess subjectively. Teachers at grammar schools showed balanced results in all three types of assessment: 4 teachers assess holistically, 3 analytically and 3 subjectively. As for the last type of school, 5 teachers assess holistically, 3 analytically and none of the teachers in private language industry chose subjective assessment. It is worth noting that analytic assessment is used as the most frequent means only at upper secondary schools, and compared with holistic and subjective assessment, it is of the same, or lesser importance at the remaining types of schools. This phenomenon may be caused by the fact that analytic assessment is more difficult to prepare as well as produce than holistic or subjective.

In order to get a more general picture of how teachers assess speaking skills at Czech schools, a graph showing the overall results of all 5 types of schools combined is provided.



Graph 4. Teachers' preferences in types of assessment of speaking skills for all 5 types of schools combined.

The graph above uses the same data as graph 3; however, it presents them for all 5 types of schools combined rather than separately. This way it is more transparent and as a result easier to identify how teachers actually assess their students speaking skills. Out of the total number of respondents (60), 25 assessed their students speaking performances holistically, whereas 23 teachers set individual criteria and assessed each of them separately, that is to say analytically. Subjective assessment, i.e. evaluating students' skills based on one's personal preferences was preferred by 12 teachers. The reason why holistic assessment is employed more frequently than analytic may be that it is easier to produce

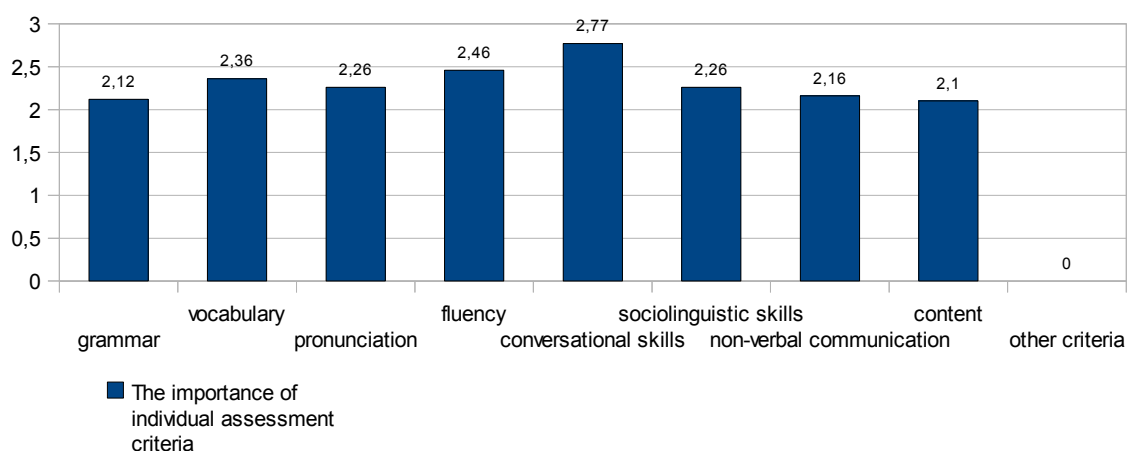
only one score that covers the whole performance than it is to produce scores for individual criteria separately.

The Choice of Individual Criteria for Interview, Picture Description and Oral Presentation

The last research question attempted to discover which criteria Czech teachers of English as a foreign language set for individual speaking assessment tasks. The last section of the research, therefore, analyses how the teachers' choices alter when different assessment tasks are used. Namely the tasks are interview, picture description and oral presentation. These three speaking assessment tasks were not chosen randomly, but the choice was based on the supposition that they would be the most frequently used ones at Czech schools, which was confirmed by the results presented in the first two graphs of the research. Individual criteria were assigned up to 3 points on a 4-level scale; 3 points stand for *very important*, 2 points for *important*, 1 point for *not so important* and 0 points for *unimportant*.

Interview

The first examined speaking task, being the most frequently used speaking assessment task in language classes, is an interview between the student and the teacher (assessor).



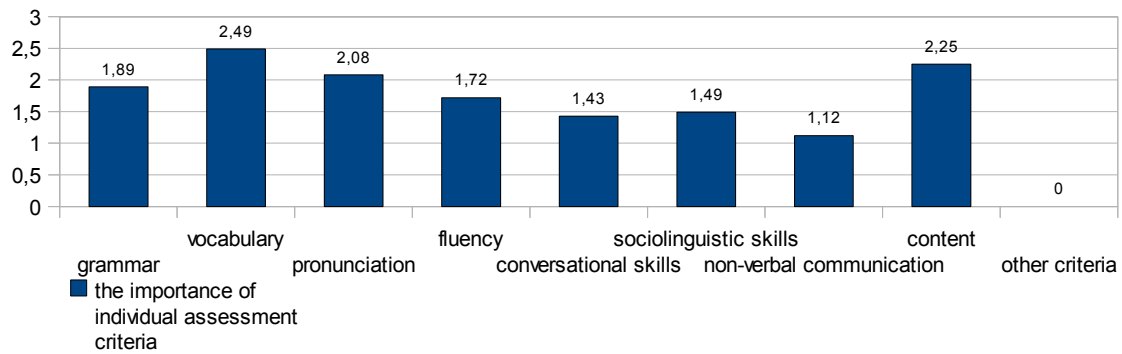
Graph 5. The importance of individual assessment criteria as for interview.

As the graph suggests, the results as for all individual criteria are relatively balanced. The criterion that is considered the most important in relation to the remaining ones is *conversational skills*. The arithmetic mean of all respondents' choices as for this criterion is 2.77 points out of 3, which suggests that the respondents placed it very close to *very important*. The second and third most important criteria are *fluency* and *vocabulary* with 2.46 and 2.36 points, respectively. The remaining criteria may all be considered *important* as for oral interviews concerned. *Pronunciation* and *sociolinguistic skills* both gained 2.26 points. *Non-verbal communication* along with *grammar* and *content* received very similar results, differing only in several hundredths of a point. According to the data, no other criteria to assess speaking skills are used by the respondents.

It is rather surprising that the results show that as for interviews, all criteria were considered *important* or above and teachers chose all of them to be assessed individually. The reason for this may be that an interview is an example of a type of speaking referred to as interactive and therefore requires the interviewees to prove that they have mastered language production as well as interaction, for which numerous skills are needed. However, there is one criterion that may be considered superfluous and it is *content*. The fact that *content* (i.e. the semantic implicature of what the students say), too, is perceived *important* and received only about seven tenths of a point less than *conversational skills* was unforeseen. What is more, this excessive number of criteria may be counterproductive as the assessors will be too occupied asking questions and attempting to pay attention to 8 criteria at once on top of that may, simply said, be too much of a load for the assessor. Selecting 4 or 5 of those that were ranked the most important ones secures validity as well as feasibility.

Picture Description

The following graph presents the results for the second most frequently used speaking assessment task: picture description.



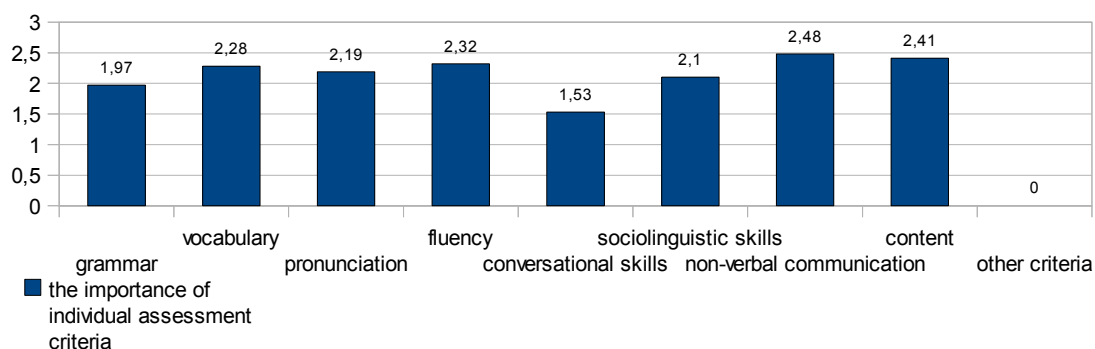
Graph 6. The importance of individual assessment criteria as for picture description.

As for picture description, the importance individual criteria were assigned is more diversified than in the previous case. The results show that the criterion that received the highest score in relation to other ones is *vocabulary*. Scoring 2.49 points, respondents placed it halfway between *very important* and *important*. *Content* is not lagging too far behind as it gained 2.25 points and as a result is the second most important criterion. *Pronunciation* and *grammar*, receiving 2.08 and 1.89 points, respectively, are the third and fourth most important criteria. Fluency received 1.72 points, which suggests it is still close to be considered *important*. *Sociolinguistic skills*, *conversational skills* and *non-verbal communication* are, on the other hand considered the least important when it comes to picture description as they gained results leaning towards *not so important*. None of the respondents stated any other criteria to be assessed.

The order of the importance of the criteria as described above implies that when describing a picture, most teachers considered students' vocabulary range and accuracy along with content and pronunciation the most important assessment criteria. In other words, the results suggest that what the students say, what lexical means they use to describe the visual stimuli and how accurate their pronunciation is determines to what extent the students meet the requirements of the assessment. As for vocabulary and context, these are the two criteria that were likely to be considered *important*; however, it was rather unforeseen that pronunciation was also considered *important*. One might think that grammar (specifically certain grammatical structures, such as phrases with existential *there*, etc.), or fluency, i.e. the ability to produce relaxed and fluent speech would play a more important role in picture description.

Oral Presentation

As the last speaking task that was chosen to ascertain the teachers' perception of criteria in respect to different assessment tasks, oral presentation is analysed.



Graph 7. The importance of individual assessment criteria as for oral presentation.

By looking at the graph above, one may notice that there is not a very big difference between the criteria that scored the most points as they range from approximately 2.5 to 2 points, meaning that a majority of the criteria is considered *important* or above. The data indicate that three of the most essential criteria for teachers to focus on when assessing an oral presentation are *non-verbal communication* together with *content* and *fluency* due to scoring 2.48, 2.41 and 2.32 points, respectively. They are followed by *vocabulary*, *pronunciation*, *sociolinguistic skills* and *grammar* with scores ranging between 2.28 and 1.97. These figures are still regarded as *important*. *Conversational skills*, however, gained the least points of all criteria and as a result are placed between *important* and *not so important*. As in the case of the interview and picture description, the data show that none of the respondents used any other criteria that may evaluate other qualities of the students speaking performances.

The reason for selecting non-verbal communication, content and fluency as the most important criteria in relation to oral presentation reflects what the respondents think makes a good presentation. Provided the content of the presentation is captivating and engages the observers either mentally, or psychically, the presenter delivers the speech in a manner that is relaxed, natural, displays a fair amount of confidence and is accompanied by appropriate non-verbal cues that illustrate and complement what is being said are the prerequisites ensuring the presentation is perceived as one of a high standard. However, the question remains whether e.g. non-verbal skills are developed in language classes, or

whether they are assessed without any preceding training. When the teachers want to assess something, they have to teach and practise it first. Assuming that students are endowed with these skills by nature would be fallacious. It also has to be noted that similarly to the interview analysed in graph 5, the fact that 7 out of 8 criteria were considered *important* may cause more harm than good as it is too many criteria to focus on at once. Selecting 4 or 5 of those that received the highest score helps assessors in making an effective assessment.

Overall Results

The results analysed and discussed in the above graphs provided answers to the 3 research questions and a summary of the most interesting findings follows. As for the frequency of using individual speaking assessment tasks, it was discovered that interview, picture description and oral presentation belong to the three most commonly used ones by Czech teachers of EFL. All of these tasks have several things in common that make them so appealing to teachers: first of all, all 3 tasks are comparatively easy in terms of preparation and realization (practicality aspect), but what is more, all 3 tasks simulate real life situations as these three tasks are presumably those that students have to perform very often in their everyday lives. Being interviewed, or interviewing someone, i.e. having a conversation, is something everybody engages in at least once a day. Describing a real picture may not be done as often; however, describing a mental picture that is stored into one's memory (e.g. a moment from one's life) is, on the contrary, performed rather frequently. Oral presentations are an inseparable part of academic and professional life of many as presenting a report on how a company is doing or a paper at university is also done considerably frequently. It is also worth paying attention to the fact that these 3 tasks are used to assess students' speaking skills during the Maturita exam. Games, on the contrary, are used with noticeably lesser frequency, possibly because of the fact that the mental connection between games and assessment is simply too weak and other tasks are preferred instead.

As for the teachers' preferences when it comes to choosing different types of assessment, nearly half of the respondents preferred to assess holistically, producing only one single score capturing the overall quality of the students' speaking skills. Slightly fewer teachers set individual criteria prior to the assessment procedure and attended to

them separately and the least preferred type of assessment was subjective. The ratio of the teachers' choices was unforeseen and the assumption was that more teachers assessed subjectively. It has to be noted that holistic assessment (similarly to analytic) specifies what the students should be capable of doing, but subjective assessment resides in the teachers' mind and students do not know what it looks like. The difference between these two is rather difficult to grasp and it is possible that teachers confused subjective assessment with holistic. This is further addressed in Limitations of the Research section.

Apropos of the criteria that are considered the most important for 3 of the most frequently occurring speaking assessment tasks, the results differed quite significantly for each type of speaking assessment task. When employing interview, being able to have and maintain a conversation, using a wide range of vocabulary, speaking in a fluent, casual relaxed manner, being contextually appropriate as well as producing an intelligible speech, respecting intonation, word and sentence stress and aspects of connected speech were considered the most important criteria to assess in the descending order. When picture description is used, the choice of criteria was altered and teachers regarded vocabulary, content, pronunciation, grammar and fluency as 5 of the most important ones, respectively. The choice of criteria for oral presentation also varied. The respondents tended to choose the criteria that reflect upon the qualities that make a good oral presentation and according to the data data, among these belong the ability to use appropriate non-verbal cues, talking to the point, speaking fluently, having a wide range of vocabulary and using correct intonation, stress and aspects of connected speech.

What is also worth paying attention to is the fact that in all 3 speaking assessment tasks, too many criteria were considered important, or close to important. This suggests that teachers are most likely to choose more criteria than they are able to attend to at once. Trying to assess 7 or 8 criteria at once is beyond the feasibility of one assessor as one assessor can only assess fewer (approximately 4 or 5) criteria at once.

This chapter attempted to answer the research questions by analysing the data gathered by means of questionnaires. The data, summarised and presented in graphs, were subsequently described, commented on in more detail and possible reasons for such occurrence were mentioned. The most important findings were summarised at the end of the chapter. With the research questions answered, it is now time to discuss how these findings may be relevant to the Czech curriculum, what the limitations of the research are and how the research may be improved by presenting suggestions for further research.

V. IMPLICATIONS

This chapter consists of three parts – implications for language teaching, limitations of the research and last but not least, suggestions for further research. The first part looks at how being aware of the findings of the research may benefit language teachers as well as students. The following section reveals some of the drawbacks of the research and the last section briefly outlines possibilities how the research could be expanded and improved.

Implications for Language Teaching

Based on the results of the research, one of the most important implications for language classes is that teachers should use speaking assessment tasks that resemble everyday life situations and allow students a certain level of freedom (interview/ conversation), let them use the target language as much as possible (picture description and oral presentation), and last but not least, tasks that may be important in the students' academic and/or professional life (oral presentation).

Teachers should use a variety of tasks to help students develop and practise various types of speaking. Employing the tasks that are directly bound to what the students may need for their own personal or professional development increases the students' intrinsic motivation to fully engage in the task and as a result increases the chance of learning and performing well, which is, after all, the very essence of any kind of schooling. This is also known as positive backwash effect.

Another implication that stems from the research is that different speaking assessment tasks require selecting different assessment criteria and teachers should adjust the criteria to meet the particular context in which the assessment is produced. Deciding which criteria the teachers will use lies in their competence; however, simply said, teachers should ask themselves what they personally think is important for the students to do well and what qualities a person who performs well in the particular task has. Asking these questions and devoting a few minutes to contemplate the answers helps the teachers set the most relevant criteria. What is more, it is also important to realize that the number of criteria should be feasible. It is advised that the most appropriate number be approximately 4, or 5 criteria (fewer would violate the validity principle, more would interfere with feasibility).

Knowing what speaking assessment task to use and what criteria to assess also determines the content of individual activities, lessons, units, etc. In other words, verbalizing the goal that teachers want their students to achieve secures that the teachers follow the path that leads their students to achieve the goal. This direct connection between what is done in individual lessons and what the assessment looks like creates transparency and helps the students see why they have to learn what the teacher planned. This is not limited to the assessment of speaking skills only, but can be applied to the assessment of any of the 3 remaining skills.

Limitations of the Research

The research, of course, has its limitations and they are addressed in the paragraphs to come. It should be taken into account that the results reflected choices of a relatively small number of respondents (60). Therefore, the tasks that are the most frequently chosen ones to assess students' speaking skills, as well as the criteria that are considered the most important represent choices of a limited sample of teachers, may be subject to a certain level of cultural bias and thus, should not be approached as a manual that teachers should adopt. Rather than that it should be perceived as a guide for teachers that may give them an idea how other teachers assess and make them think about which assessment tasks and which criteria they use and whether the tasks and the assessment criteria are in accordance with what the classes aim at. In other words, the research simply attempts to map the situation based on the answers of a chosen sample of Czech teachers of English as a foreign language and should make teachers more conscious of how and what they assess when it comes to speaking.

Another possible limitation of the research could be question No. 9 in the questionnaire that asked the respondents about their preferences as for different types of assessment (scoring) scales and presented the them with 3 options to choose from (holistic, analytic, or subjective assessment). It is quite probable that the the difference between holistic and subjective assessment was not stressed enough and the respondents who assess subjectively chose the option holistic. It is due to this fact that the findings may not reflect the reality in ELT at Czech schools and if the question was asked differently, the answers would probably differ. The estimation is that more teachers would say that they assess their students' speaking skills subjectively.

Suggestions for Further Research

Beyond any question, the research could be extended or complemented in many ways. One of the most obvious ones would be to increase the number of respondents. Asking more teachers to participate in the research would surely increase the reliability of the results. The number of teachers representing individual types of schools in this research was not identical and equalizing it would provide more representative and reliable sample of respondents. The ratio of male and female respondents also differed quite dramatically. It would be interesting to observe if and/or how much the results would change if the number and ratio of the respondents changed.

It might also be beneficial to change the organization of the respondents and to realize the research not with regard to individual types of schools, but with regard to the years of professional teaching experience that teachers have and ascertain whether and/or how much their answers about the assessment (scoring) scales they employ, their choices of speaking assessment tasks and criteria they set would change with increasing experience.

The last part of the research focused on finding out what criteria teachers preferred for 3 of the most frequently used speaking assessment tasks. Extending the research and discovering what criteria teachers would choose for the speaking tasks that are not used as often could also provide some more implications for teaching. However, taking into consideration the fact that the questionnaire was a 2-page document already, adding more questions would surely add one or two more pages and the willingness of the respondents to answer all questions would almost certainly decrease.

To conclude, despite the fact that the research brings together some interesting results with implications that may benefit both language teachers as well as students, it is important to note that the results should not be generalized too much as the ways of improving and extending the research are numerous. That said, the section focusing on how the results can enrich and better the educational process, what the limits of the research are and how the research could be expanded and improved is finished. Now the last chapter focuses on summarising and highlighting the main points of the thesis.

VI. CONCLUSION

As the title of the thesis suggests, it investigates and maps the current practices of teachers of EFL across the Czech curriculum in terms of how they assess their students' speaking skills. In order to do so, theoretical basis that explains the characteristics of both speaking and assessment had to be established. Then, three research questions were defined and they are as follows: what speaking tasks do teachers choose for their students to assess their speaking performances, what types of assessment do teachers prefer and last but not least, what criteria do teachers set for the assessment of individual types of speaking tasks? In the process of analysing the data and answering the research questions, interesting findings were made. First of all, the speaking tasks that teachers use the most frequently were those that were presupposed to be very common: interview, picture description and oral presentation. All three tasks show certain similarities that may be generalized to being close to everyday life situations, enabling the students to use as much L2 as possible and being useful for the students' future personal, academic and professional development.

Another, rather surprising result is that holistic and analytic assessment evince similar numbers of choices (differing only in 2 respondents who favoured holistic assessment). It was assumed that more teachers would select holistic assessment as it is easier to prepare as well as conduct. Fewer than one quarter of all respondents assess subjectively, without setting any criteria or levels of performance.

As it was mentioned above, the criteria that are considered important for individual types of speaking tasks differ quite considerably as each type of speaking task has its own peculiarities that determine which criteria are considered important and which superfluous.

Subsequently, limitations of the research as regards the the number and ratio of participants were acknowledged and it was suggested that the research could investigate whether there are any differences in the teachers' answers in relation to the years of their professional teaching experience.

Familiarizing oneself with the elements elaborated in this thesis may benefit teachers of EFL as determining the speaking task and assessment criteria to focus on enables the teachers to see the assessment of speaking skills in a new light. It may also lessen the uncertainty some teachers may feel about their own capabilities when it comes to assessing this, arguably the most essential, skill.

REFERENCES

- Arter, J. A., & Chappuis, J. (2006). *Creating & recognizing quality rubrics*. Boston, MA: Pearson Education.
- Astin, A. (1993). *Assessment for excellence: The Philosophy and practice of assessment and evaluation in higher education*. Phoenix, AZ: Oryx Press.
- Balada, J., Baladová, G., Boněk, J., Brant, J., Brychnáčová, E., Charalambidis, A., ... Zelendová, E. (2007). *Framework education programme for secondary general education (grammar schools)*. Praha: VÚP.
- Broersma, D. (2001). You're so white, so fat, and so hairy!: Developing sociolinguistic competence in a second language. In L. Dickerson (Ed.), *Helping learners develop second language proficiency* (pp. 191 – 207). Colorado Springs: Mission Training International.
- Brown, H. D. (2001). *Teaching by principles: An Interactive approach to language pedagogy*. White Plains, NY: Pearson Education.
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. White Plains, NY: Pearson Education.
- Brown, H. D. (2007). *Teaching by principles: An Interactive approach to language pedagogy* (3rd ed.). White Plains, NY: Pearson Education.
- Council of Europe [COE]. (n.d.). *The common European framework of reference for languages: Learning, teaching, assessment*. Retrieved from http://www.coe.int/t/dg4/linguistic/cadre1_en.asp
- Harmer, J. (2011). *The practise of English language teaching, 4th edition*. Harlow: Pearson Education Limited.
- Hinkel, E. (2005). *Handbook of research in second language teaching and learning*. Mahwah: Lawrence Erlbaum Associates.
- Hornby, A. S. (1989). *Oxford advanced learner's dictionary* (4th ed.). London: Oxford University Press.
- Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge: Cambridge University Press.
- Johnson, C. E. (n.d.). *The 7%, 38%, 55% myth: Blasting away an old NLP myth*. Retrieved from <http://www.neurosemantics.com/nlp-critiques/the-7-38-55-myth>

- Klečková, G. (2010, September). *Assessment of language skills: Designing tasks and developing scoring rubrics* [PowerPoint slides]. Poster session presented at the 7th International and 11th National Association of Teachers of English of the Czech Republic (ATECR) Conference, Usti nad Labem, Czech Republic.
- Knight, B. (1992). Assessing speaking skills: A Workshop for teacher development. *ELT Journal*, 46 (3), 294 – 302. doi: 10.1093/elt/46.3.294
- Kohn, A. (2006). Speaking my mind: The trouble with rubrics. *The English Journal*, 95(4), 12 – 15. Retrieved from <http://www.jstor.org/>
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7, 25. Retrieved from <http://pareonline.net/getvn.asp?v=7&n=25>
- Ministerstvo školství, mládeže a tělovýchovy [MŠMT], (n.d.). Nová maturita. Retrieved from <http://www.novamaturita.cz/ustni-zkouska-1404033215.html>
- Ministerstvo školství, mládeže a tělovýchovy [MŠMT], (2013a). Přílohy, které jsou nedílnou součástí kritérií hodnocení maturitní zkoušky ve školním roce 2012/2013. Retrieved from <http://www.novamaturita.cz/kriteria-hodnoceni-1404035057.html>
- Ministerstvo školství, mládeže a tělovýchovy [MŠMT], (2013b). Úplné znění kritérií hodnocení maturitní zkoušky ve školním roce 2012/2013. Retrieved from <http://www.novamaturita.cz/kriteria-hodnoceni-1404035057.html>
- Mueller, J. (n.d.). *Rubrics*. Retrieved from <http://jfmuller.faculty.noctrl.edu/toolbox/rubrics.htm>
- Palomba, C. A., & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco, CA: Jossey-Bass Publishers.
- Restrepo, A. P. M., Aristizábal, L. D., Orozco, F. C., Monsalve, S. G., Orozco, L. A. L., Urán, M. P. (2003). Assessing spoken language in EFL: Beliefs and Practices. *Revista Universidad EAFIT*, 129, 63 – 73. Retrieved from <http://www.redalyc.org/articulo.oa?id=21512906&iCveNum=580>
- Scrivener, J. (2011). *Learning teaching*. Oxford, UK: Macmillan Publishers Limited.
- Swan, M. (2005). *Practical English usage*. Oxford: Oxford University Press.
- Turner, J. (1998). Assessing speaking. *Annual Review of Applied Linguistics*, 18, 192-207. doi:10.1017/S0267190500003548.

Underhill, N. (1987). *Testing spoken language: A Handbook of oral testing techniques*.

Cambridge: Cambridge University Press.

Ur, P. (1991). *A Course in language teaching*. Cambridge: Cambridge University Press.

APPENDICES

Appendix A Types of Assessment

1	Achievement assessment	Proficiency assessment
2	Norm-referencing (NR)	Criterion-referencing (CR)
3	Mastery learning CR	Continuum CR
4	Continuous assessment	Fixed assessment points
5	Formative assessment	Summative assessment
6	Direct assessment	Indirect assessment
7	Performance assessment	Knowledge assessment
8	Subjective assessment	Objective assessment
9	Checklist rating	Performance rating
10	Impression	Guided judgement
11	Holistic assessment	Analytic assessment
12	Series assessment	Category assessment
13	Assessment by others	Self-assessment

Appendix B

Micro- and Macro-skills of Oral Communication

Micro-skills

1. Produce differences among English phonemes and allophonic variants.
2. Produce chunks of language of different lengths.
3. Produce English stress patterns, words in stressed and unstressed positions, rhythmic structure, and intonation contours.
4. Produce reduced forms of words and phrases.
5. Use an adequate number of lexical units (words) to accomplish pragmatic purposes.
6. Produce fluent speech at different rates of delivery.
7. Monitor one's own oral production and use various strategic devices – pauses, fillers, self-corrections, backtracking – to enhance the clarity of the message.
8. Use grammatical word classes (nouns, verbs, etc.), systems (e.g., tense, agreement, pluralization), word order, patterns, rules, and elliptical forms.
9. Produce speech in natural constituents: in appropriate phrases, pause groups, breath groups, and sentence constituents.
10. Express a particular meaning in different grammatical forms.
11. Use cohesive devices in spoken discourse.

Macro-skills

12. Appropriately accomplish communicative functions according to situations, participants, and goals.
13. Use appropriate styles, register, implicature, redundancies, pragmatic conventions, conversation rules, floor-keeping and -yielding, interrupting, and other sociolinguistic features in face-to-face conversations.
14. Convey links and connections between events and communicate such relations as focal and peripheral ideas, events and feelings, new information and given information, generalization and exemplification.
15. Convey facial features, kinesics, body language, and other non-verbal cues along with verbal language.
16. Develop and use a battery of speaking strategies, such as emphasizing key words, rephrasing, providing a context for interpreting the meaning of words, appealing for help, and accurately assessing how well your interlocutor is understanding you.

(Brown, 2004, pp. 142 – 143)

Appendix C

A full list of possible scoring criteria in relation to the assessment of speaking skills as presented by The Council of Europe:

1. Turn-taking strategies
2. Co-operating strategies
3. Asking for clarification
4. Fluency
5. Flexibility
6. Coherence
7. Thematic development
8. Precision
9. Sociolinguistic competence
10. General range
11. Vocabulary range
12. Grammatical accuracy
13. Vocabulary control
14. Phonological control

(The Council of Europe, n.d., p. 193)

Appendix D

The Questionnaire in Czech

Milí kolegové, milé kolegyně,

provádím výzkum týkající se vytváření kritérií pro hodnocení mluveného projevu a chtěl bych Vás požádat o vyplnění následujícího dotazníku. Dotazníky jsou anonymní a budou použity pouze pro potřeby méj diplomové práce. Jeho vyplnění Vám nezabere více než 7 minut.

Děkuji Vám za Váš čas.

Stanislav Rychtařík

Část 1

1. Jsem muž žena

2. Kolik let se věnujete učitelské profesi?

1 – 5 let 6 – 10 let 11 – 15 let 16 – 20 let 20 a více

3. Kolik let se věnujete učení anglického jazyka?

1 – 5 let 6 – 10 let 11 – 15 let 16 – 20 let 20 a více

4. Na jaké škole učíte?

1. stupeň základní školy 2. stupeň základní školy střední odborná škola
 gymnázium soukromá sféra

5. Máte kvalifikaci na učení anglického jazyka? ano ne

6. Zúčastnili jste se někdy semináře/workshopu týkajícího se hodnocení? ano ne

Pokud jste zaškrtnli „ano“, odpovězte prosím na otázku č. 7, jinak pokračujte na otázku č. 8.

7. Na co byl seminář/workshop zaměřen?

Část 2

8. Zaškrtněte (✓) jak často používáte následující úkoly k hodnocení mluveného projevu Vašich studentů:

	Velmi často	Často	Občas	Zřídka	Nikdy
Roleplay					
Ústní prezentace					
Popis obrázku					
Rozhovor					
Překlad vět					
Otázka a odpověď					
Hra					
Jiné (doplňte jaké)					

9. Vyberte možnost, která nejlépe vystihuje, jak hodnotíte mluvený projev Vašich studentů:

- Hodnotím kvalitu mluveného projevu studenta jako celek (např. aby student dostal 1 - musí vždy používat správné gramatické struktury, mít bohatou slovní zásobu, která je kontextově správná, nedělá více než 1, nebo 2 chyby ve výslovnosti a jeho projev je plynulý a uvolněný.
- Vybírám kritéria – např. gramatická správnost, rozsah slovní zásoby, správnost výslovnosti, plynulost, atd. - a hodnotím každé kritérium zvlášť.
- Hodnotím mluvený projev svých studentů podle zkušeností.

Pokud jste vybrali poslední možnost, odpovězte prosím na otázku č. 10, jinak pokračujte na otázku č. 11.

10. Doplňte prosím následující větu:

Když zkouším mluvený projev svých studentů, hodnotím:

11. Zaškrtněte na škále 3 (velmi důležité), 2 (důležité), 1 (ne příliš důležité), 0 (zbytečné) jak důležité považujete dané kritérium při hodnocení jednotlivých typů ústního zkoušení pro každý ze 3 úkolů.

Kritérium	Ústní popis obrázku	Rozhovor (učitel a žák)	Ústní prezentace
Gramatika	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Slovní zásoba	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Výslovnost	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Plynulost (schopnost mít plynulý, nenucený, klidný projev)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Konverzační dovednosti (schopnost konverzovat a udržet konverzaci – respektování střídání promluv, logická návaznost promluv, sebeopravování se, dožadování se objasnění při nejasnostech)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Sociolingvistické dovednosti (schopnost přizpůsobit svou mluvu dané situaci: formální, neformální)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Prvky neverbální komunikace (držení těla, mimika, gestika, haptika, proxemika, oční kontakt)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Obsah (správnost a relevantnost obsahu)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Jiné (prosím vyjmenujte)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0

Appendix E
The Questionnaire in English

Dear colleagues,

I am conducting a research on designing criteria to assess speaking skills and I would like to ask you to fill in the following questionnaire. The questionnaires are anonymous and will be used strictly for the purpose of my diploma thesis. The completion will take you 7 minutes at the most.

Thank you for your time,

Stanislav Rychtařík

Part 1

1. I am male female

2. How long have you been in your profession?
 1 – 5 years 6 – 10 years 11 – 15 years 16 – 20 years 20 and more

3. How long have you been teaching English?
 1 – 5 years 6 – 10 years 11 – 15 years 16 – 20 years 20 and more

4. What kind of school do you teach at?
 lower elementary upper elementary secondary vocational school
 grammar school private language industry

5. Are you qualified to teach English? yes no

6. Have you ever taken part in a workshop/seminar focused on assessment? yes no
If you ticked "yes", please answer question no. 7, otherwise go to question no. 8.

7. What was the seminar focused on?

Part 2

8. Tick (✓) how often you use the following tasks to assess your students' speaking skills:

	Very often	often	sometimes	rarely	never
Roleplay					
Oral presentation					
Picture description					
Interview					
Sentence translation					
Question and answer					
Game					
Other (please specify)					

9. Choose the option that best describes how you assess your students' speaking performances:

- I assess the overall quality of students' spoken performances (e.g. an A student has to: use grammatically correct structures, have a wide range of vocabulary that is contextually appropriate, make no more than 1 or 2 pronunciation mistakes and his/her speech is fluent and relaxed.

- I choose criteria – e.g. grammatical accuracy, vocabulary range, pronunciation accuracy, fluency, etc. - and assess each criterion separately.

- I assess my students' speaking performance based on my experience.

If you chose the last option, please answer question no. 10, otherwise go to question no. 11..

10. Complete the following sentence:
When I assess my students' speaking performance, I focus on:

11. On the scale of 3 to 0 (3 = very important, 2 = important, 1 = not so important, 0 = unimportant), tick how important you consider individual criteria when assessing your students' speaking performances for all three tasks.

Criterion	Oral picture description	Interview (teacher and student)	Oral presentation
Grammar	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Vocabulary	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Pronunciation	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Fluency (The ability to speak in a fluent, casual, relaxed, manner)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Conversational skills (the ability to have and maintain a conversation – respecting turn-taking, self-correcting oneself, demanding clarification when needed)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Sociolinguistic skills (the ability to adjust one's style of speaking: formal, informal)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Non-verbal communication (posture, facial expressions, eye contact)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Content (correctness and significance)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0
Other (please specify)	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0	<input type="checkbox"/> 3 <input type="checkbox"/> 2 <input type="checkbox"/> 1 <input type="checkbox"/> 0

SUMMARY IN CZECH

Tato diplomová práce se zabývá vytvářením kritérií pro hodnocení mluveného projevu. První část, sloužící jako teoretický rámec pro následný výzkum, předkládá čtenáři důvody, které vedly k vybrání tohoto tématu, a vysvětluje termíny mluvení, hodnocení a hodnotící kritéria v kontextu hodnocení mluveného projevu. Výzkum byl proveden pomocí dotazníků a mapuje současné praktiky učitelů v oblasti hodnocení mluveného projevu. Výsledky výzkumu odhalily nejčastěji používané úkoly pro hodnocení mluveného projevu, preference učitelů ohledně jednotlivých typů hodnocení a fakt, že kritéria, která učitelé pro jednotlivé typy úkolů stanovují, se poměrně značně liší. Výsledky výzkumu jsou okomentovány a možné příčiny jsou zmíněny. V závěru práce jsou prodiskutovány možné implikace pro učení jazyků.