



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

**“Considerations for the teaching and testing of
speaking in Austrian lower secondary EFL”**

Verfasserin

Andrea Stockner

angestrebter akademischer Grad

Magistra der Philosophie (Mag. phil.)

Wien, 2014

Studienkennzahl lt. Studienblatt: A 190 344 299

Studienrichtung lt. Studienblatt: Lehramtsstudium UF Englisch UF Psychologie und Philosophie

Betreuerin: Univ.-Prof. Mag. Dr. Christiane Dalton-Puffer

Acknowledgements

I am indebted to many of my colleagues for supporting me during the writing process. I am thankful to my supervisor, Dr. Christiane Dalton-Puffer, whose encouragement and support from the initial to the final level enabled me to develop an understanding of the subject. Lastly, I offer my regards and blessings to my family and friends who supported me during the completion of the project.

Abstract

This work should serve as an introduction to the teaching and testing of speaking in New Middle Schools in Austria. A theoretical part should form the basis of the thesis. Communicative language teaching, the skill of speaking and some concepts in speech production will be explained for that purpose. In order to make the theory more useful for the teachers in Austria, the Austrian curriculum will serve as a guideline here. The practical part consists of two major issues, namely teaching and testing speaking. In both cases, suggestions for the Austrian classroom will be given, as well as general issues will be discussed. The goal of this thesis is to provide the reader with a theoretical background, so that he or she can use the provided suggestions in his or her professional career.

TABLE OF CONTENTS

Acknowledgements

Abstract

1	Introduction.....	1
2	Communicative Language Teaching.....	3
2.1	Roles of teachers and learners	7
3	Concepts in speech production.....	7
3.1	Articulation	8
3.2	Fluency.....	9
3.3	Managing talk.....	10
4	Speaking as a skill.....	11
4.1	The components of the skill of speaking.....	14
4.2	The Importance of using corpora.....	16
4.3	Speaking versus writing.....	17
5	Some basic considerations about the teaching of speaking.....	19
5.1	Are native speakers better in language teaching?	19
5.2	Listening and speaking – an effective relation?	22
5.3	The significance of Pronunciation in the L2 language classroom.....	24
5.4	Accuracy versus fluency	27
6	The communicative curriculum and the teaching of speaking.....	31
6.1	The CEFR as a basis.....	31
6.1.1	The designing process of the CEFR.....	32
6.1.2	How teachers can use the CEFR in the teaching and testing of speaking.....	35
6.2	Integrating speaking in the curriculum in Austria.....	36
6.2.1	General issues.....	36
6.2.2	The curriculum in Austria	37
6.3	Approaches in teaching speaking	38
6.3.1	Direct versus indirect approach.....	39
6.3.2	Integrated skills	40
6.3.3	The task-based approach.....	42
6.3.4	Classroom organisation.....	43
6.4	The role of dialogues in the teaching of speaking.....	44
6.4.1	Scripted dialogues.....	45
6.4.2	Unscripted language samples	46

6.4.3	Semi-scripted texts.....	47
6.5	Activity types in teaching speaking.....	48
6.5.1	The framework by Burns and Joyce	49
6.5.2	Activities from the E8 standards	50
7	Summative assessment of speaking.....	52
7.1	Test evaluation	53
7.2	The history of second languages speaking tests.....	54
7.2.1	Rating scale development.....	56
7.3	Rating scales	58
7.3.1	Advantages and disadvantages	67
7.3.2	Test anxiety	68
7.4	Specific assessment frameworks	68
7.4.1	E8 standards	68
7.4.2	Performance decision trees.....	72
7.5	Possible rating solution for New Middle Schools in Austria.....	78
8	Conclusion.....	81
9	Bibliography.....	82

1 INTRODUCTION

The following thesis developed from a problem which I observed during my teaching in a New Middle School in Lower Austria. The New Middle School and its curriculum rose manifold problems in the teaching of English as the curriculum demands new strategies from the teachers. One of those problematic areas is that speaking must have the same value in teaching and grading as other skills, for example reading and listening. :

Die Fertigkeitsbereiche Hören, Lesen, An Gesprächen teilnehmen, Zusammenhängend Sprechen und Schreiben sind in annähernd gleichem Ausmaß regelmäßig und möglichst integrativ zu erarbeiten und zu üben. (bmukk 2012: 34)

Furthermore, the curriculum suggests that in the beginning the focus should be on acquiring oral skills and receptive skills (especially listening):

Im Anfangsunterricht allerdings sind die Teilfertigkeiten des Hörverstehens und der mündlichen Kommunikation durch regelmäßige Hörübungen sowie durch ein möglichst häufiges Angebot an Sprechanlässen verstärkt zu fördern. (bmukk 2012: 34)

At first sight that seems to be easily manageable for the teachers, but in fact teaching the skill of speaking is a complex issue as the teacher needs to be clear about certain theoretical as well as practical areas in order to ensure useful teaching sequences. This is the point where this thesis should be of great help for teachers. It should serve as an overview of different theoretical points and, furthermore, suggest different practical solutions for the teaching and testing of speaking.

All the above suggests that speaking has to be included in grading too. During my teaching time I learned that testing speaking gives pupils a chance whose strengths are not in writing but in communicating. Additionally, through testing speaking the grading might become fairer, and it would be more in accordance with the curriculum.

In order to understand the points which will be made, it is important to know that the basic notion of this thesis is that language should be taught communicatively, which is also the case in the Austrian curriculum. That is also a reason for starting this thesis with a rather theoretical part about communicative language teaching.

As stated earlier, this thesis should serve as a basis for the teaching and testing of speaking. It should give a theoretical background, as well as practical guidance to include speaking in the curriculum in Austria. For that purpose, it consists of six major parts, which would be communicative language teaching, concepts in speech production, speaking as a skill, some basic considerations about the teaching of speaking, the communicative curriculum and the teaching of speaking, and summative assessment of speaking. Special attention will be given to the CEFR (Common European Framework of Reference) and the E8 standards as they serve, on the one hand, as a basis for the curriculum and, on the other hand, as a measurement tool for the progress of students in Austrian schools

2 COMMUNICATIVE LANGUAGE TEACHING

Speaking is usually taken for granted. We do not think about it, we just do it. However, when learning a foreign or second language, we are reminded about how hard it is to achieve this ability. So, what needs to be thought through is the question about what is involved in speaking. For this purpose, the communicative approach should be examined in this chapter.

Communicative language teaching is and has been for many years now a main coining term in language classrooms. As the name “communicative language teaching” already suggests, most of the teachers know that the focus should be on communication. However, more points need to be taken into consideration when talking about this approach. For that purpose some of the most important theorists and their views should be analysed here.

Richard and Rogers (1986: 144) give some of the characteristics of the communicative view of language:

1. Language is a system for the expression of meaning.
2. The primary function of language is for interaction and communication.
3. The structure of language reflects its functional and communicative uses.
4. The primary units of language are not merely its grammatical and structural features, but categories of functional and communicative meaning as exemplified in discourse.

One can conclude from that list that the main point in communicative language teaching is that it starts from a theory of language as communication. The ultimate goal is to develop what Hymes (1972) defined as "communicative competence". What Hymes wanted to reach with coining that term is contrasting Chomsky's theory of competence (Richards and Rogers 1986: 142). Chomsky claimed that

linguistic theory is concerned primarily with an ideal speaker-listener in a completely homogeneous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitation, distraction, shifts of attention and

interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance (Chomsky 1965: 3).

Hymes thought that that kind of view of linguistic theory was sterile, furthermore, to him, linguistic theory needed to be seen as part of a more general theory incorporating communication and culture. Hymes says that a person who is communicatively competent acquires both knowledge and ability for language use with respect to

1. whether (and to what degree) something is formally possible;
2. whether (and to what degree) something is feasible in virtue of the means of implementation available;
3. whether (and to what degree) something is appropriate (adequate, happy, successful) in relation to a context in which it is used and evaluated;
4. whether (and to what degree) something is in fact done, actually performed, and what its doing entails. (Hymes 1972: 281)

Another view, which is also favoured in CLT, would be Halliday's functional account of language use. In his works, Halliday elaborated "a powerful theory of the functions of language, which complements Hymes's view of communicative competence for many writers (e.g. Brumfit and Johnson 1979, Savignon 1983)" (Richards and Rogers 1986: 143).

Halliday described (1975: 11-17) seven functions which language performs for children learning their first language:

1. the instrumental function: using language to get things;
2. the regulatory function: using language to control the behaviour of others;
3. the interactional function: using language to create interaction with others;
4. the personal function: using language to express personal feelings and meanings

5. the heuristic function: using language to learn and to discover;
6. the imaginative function: using language to create a world of the imagination;
7. the representational function: using language to communicate information.

That approach to learning a language again underlines that language performs functions and that language is used with a purpose. That needs to be kept in mind when it comes to teaching and learning a second language, as defining a goal in the learning process might help to structure the teaching.

Henry Widdowson is another theorist known for his view on communicative language teaching. In his book *Teaching Language as Communication* (1978), Widdowson "presented a view of the relationship between linguistic systems and their communicative values in text and discourse. He focused on the communicative acts underlying the ability to use language for different purposes" (Richards and Rogers 1986: 143).

As mentioned earlier Hymes states that a person who is communicatively competent needs both knowledge and ability for language use with respect to four dimensions, which would be formality, feasibility, appropriateness and whether something is done or performed and what comes along with its doing. Canale and Swain (1980) also identified four dimensions of communicative competence, which would be: grammatical competence, sociolinguistic competence, discourse competence, and strategic competence. Grammatical competence includes the domain of grammatical and lexical capacity. Sociolinguistic competence consists of the understanding of the social context in which communication takes place, including role relationships, the shared information of the participants, and the communicative purpose of their interaction. Discourse competence refers to "the interpretation of individual message elements in terms of their interconnectedness and of how meaning is represented in relationship to the entire discourse or text" (Richard and Rogers 1986: 143). Strategic competence is built up of coping strategies that communicators employ to initiate, terminate, maintain, repair and redirect communication.

Despite those theories which focus only on the communicative competence as such, there are further principles which need to be explained, as they might also be

important in some teaching situations. Elements of an underlying learning theory are the communication principle which says that activities that involve real communication promote learning. Another element would be the task principle, which mainly means that activities in which language is used for carrying out meaningful tasks promote learning (Johnson 1982). A third element is the meaningfulness principle which sees the importance in language that is meaningful to the learner in the language classroom. Only that would support the learning process. So, the teachers need to be pay special attention to the materials they use, Krashen, for example, stresses the importance of the use of real language and that language learning comes about through using language communicatively. Another learning theory is explained by Littlewood (1984). For him, the acquisition of communicative competence in a language is an example of skill development. Littlewood (1984: 74) describes the theory as follows:

The cognitive aspect involves the internalisation of plans for creating appropriate behaviour. For language use, these plans derive mainly from the language system - they include grammatical rules, procedures, for selecting vocabulary, and social conventions governing speech. The behavioural aspect involves the automation of these plans so that they can be converted into fluent performance in real time. This occurs mainly through practice in converting plans into performance.

So far, only the way to communicative competence has been explained. When it comes to teaching objectives, Piepho (1981: 8) discusses the following levels:

1. an integrative and content level (language as a means of expression)
2. a linguistic and instrumental level (language as a semiotic system an an object of learning);
3. an affective level of interpersonal relationships and conduct (language as a means of expressing values and judgments about oneself and others);
4. a level of individual learning needs (remedial learning based on error analysis);
5. a general educational level of extra-linguistic goals (language learning within the school curriculum).

Those objectives can be seen as general objectives and can be applicable to any teaching situation. The objectives cannot be closely defined as they "assume[s] that language teaching will reflect the particular needs of the target learners" (Richard and Rogers 1986: 146).

2.1 ROLES OF TEACHERS AND LEARNERS

As the emphasis in Communicative Language Teaching is not on the mastery of language forms but on the processes of communication, there are different roles for learners. The learners themselves have preconceptions about what teaching and learning should be like. When those preconceptions stay unrealized, they can lead to learner confusion and resentment. Learners need to realise that failed communication is a joint responsibility and not the fault of the speaker or listener, and also the other way round they need to see that successful communication is an accomplishment jointly achieved and acknowledged. So, it might be helpful to make all this clear to the students before they start to work in groups. Also the goals of the language class should be explained before starting to work. That might reduce the confusion to a minimum.

Not only the learners have different roles to fulfil, also the teachers are exposed to different demands. There is the role of the needs analyst, which means that the teacher needs to find out what the language learners need. Another role is the counselor. According to Richard and Rogers (1986: 151) "[...] the teacher-counselor is expected to exemplify an effective communicator seeking to maximize the meshing of speaker intention and hearer interpretation, through the use of paraphrase, confirmation, and feedback". The function of the group process manager is to organize the classroom as a setting for communication and communicative activities.

3 CONCEPTS IN SPEECH PRODUCTION

In the following paragraphs the concepts of articulation and fluency as well as managing talk should be explained shortly but sufficiently for the purpose of this thesis.

The ability to speak and, furthermore, to speak fluently is very complex. Scott Thornbury (2005: 1) claims that “[f]or a long time it was assumed that the ability to speak fluently followed naturally from the teaching of grammar and vocabulary, with a bit of pronunciation thrown in”. Nowadays, the majority of language teachers know that the situation is different. Skills are crucial, but also different types of knowledge have to be taken into consideration. Bygate claims that in language courses most of the time is devoted to teaching grammar and vocabulary as it is an essential part of using a language successfully. However, he states that it needs more than that to prepare learners to be able to use the language (Bygate 1987: 3). That claim shows that he is of the same opinion as Thornbury when it comes to defining the skill of speaking.

The first point, which needs to be made, is that speaking is linear, it happens in real time. Thornbury (2005: 2) explains this in an intelligible way:

Words follow words, and phrases follow phrases. Likewise, at the level of utterance (that is to say; the spoken equivalent of sentences), speech is produced utterance-by-utterance, in response to the word-by-word and utterance-by-utterance productions of the person we are talking to (our interlocutor).

That statement clearly underlines the importance of spontaneity in speech production. One might deduce from it that speaking is unplanned, but that is definitely not the case. On the contrary, speaking is most of the time “planned”, however, the planning time might be severely limited, so one would not define speaking as a planned activity. It might be the case that the planning of one utterance overlaps with the production of another one. This essential spontaneity explains many of the characteristics of speech production.

3.1 ARTICULATION

Necessarily, what we plan to speak needs to be articulated. In the simplest words, articulation means that the organs of speech are used to produce sounds, more precisely, “a stream of air is produced in the lungs, driven through the vocal cords, and ‘shaped’ by, among other things, the position and movement of the tongue, teeth, and lips” (Thornbury 2005: 4). Over 40 phonemes can be produced by

speakers of English. Those 40 phonemes determine the meaning of a word. One should not think of those 40 phonemes like bricks. The sounds are produced in a continuous stream, with many different vocal organs involved concurrently, such that the articulation of one sound will affect the articulation of its neighbours. Of course other factors contribute to the meaning making of an articulated utterance, such as loudness, pitch direction, tempo, and pausing. As stated above, all this happens at an enormous speed.

3.2 FLUENCY

Most of the time, the ultimate goal in teaching speaking is to make the students speak fluently. That is an easy thing to say, but what makes a speaker fluent? Is it only the ability to speak fast? Of course speed is an important factor, but it is not the only one. Correct pausing is equally important. Every speaker must pause, either to formulate an utterance or to catch up with its formulation, however, frequent pausing is a sign of a struggling speaker. But what about the length of the pauses? The length is not as crucial as the frequency. In terms of how fluent a speaker is, the listener might conclude that more pauses mean that the speaker is not fluent, whereas the length of the pauses does not really make a difference here. The placement of pauses is another point which a speaker has to deal with. Unnatural placement of pauses makes speaking sound unnatural. Thornbury (2006: 7) makes clear that “[n]atural-sounding pauses are those that occur at the intersection of clauses, or after groups of words that form a meaningful unit”.

What do speakers do to give at least the illusion of fluency, and to compensate for the attentional demands involved in speech production? The answer is that they use production strategies. One of those strategies would be to fill the pauses. Some common pause fillers would be “uh” or “um”. Vagueness expressions are also used to keep the illusion of fluency, like “sort of” and “I mean”. A common device for gaining time is repetition, here the speakers uses both fillers and repeats.

As stated earlier, speaking fluently is an important goal in learning a language, but when is a speaker likely to develop fluency? Nation and Newtown (2009: 152-153) name three conditions under which fluency is likely to be developed (in a classroom setting):

- The activity is meaning-focused. The learners' interest is on the communication of a message and is subject to the "real time" pressures and demands of normal meaning-focused communication.
- The learners take part in activities where all the language items are within their previous experience. This means that the learners work with largely familiar topics and types of discourse making use of known vocabulary and structures. These kinds of activities are called "experience" tasks because the knowledge required to do the activity is already well within learners' experience.
- There is support and encouragement for the learner to perform at a higher than normal level. This means that in an activity with a fluency development goal, learners should be speaking and comprehending faster, hesitating less, and using larger planned chunks than they do in their normal use of language. A fluency development activity provides some deliberate push to the higher level of performance often by using time pressure.

Furthermore, there should be opportunities to use both receptive and productive skills, as both are needed for developing fluency in speaking. Another crucial point, which was made by Nation and Newton (2009: 153), is that the issues must be already available for fluent use, the speakers must know the topic which they need to talk about, otherwise the tasks are of little use for the development of fluency.

3.3 MANAGING TALK

Up to this point, only those concepts have been explained which focus on the speaker him or herself. Of course, most speaking is sort of an interaction or a face to face dialogue and not a monologue. Even in a two or three utterance dialogue complex interaction takes place. The speakers might be "jockeying for conversational turns, introducing new topics and engaging in word play [...]" (Thornbury 2005: 8). Turn-taking is a necessary part for holding up a conversation. Thornbury (2005: 8-9) lists some of the rules and skills which should be applied:

- long silences are to be avoided
- listen when other speakers are speaking
- recognizing the appropriate moment to get a turn
- signalling the fact that you want to speak
- holding the floor while you have your turn

- recognizing when other speakers are signalling their wish to speak
- yielding the turn
- signalling the fact that you are listening

In order to keep up the conversation different discourse markers are used, which signal a speaker's conversational intention (for example, "that reminds me", "by the way", "like I say", "yes, but" etc.). Furthermore, speakers use backchannel devices, such as "uh-huh" or "really", to show the interlocutor that one is listening. However, negotiation of speaking is not only a matter of words. Interactional use of eye gaze and gesture might also lead the conversation. Those concepts belong to the field of paralinguistics.

4 SPEAKING AS A SKILL

According to Bygate the first step in the language classroom is to find out if the learners can actually speak the language. To do this, the next logical step would be to make the students say something. In other words, the teachers must give the learners time to practice. That advice by Bygate suggests that there must be a difference between knowledge about a language and the skill in using it. Otherwise, Bygate would not insist on the practice factor so strongly. Bygate (1987: 3) underlines this crucial distinction with the help of an example from real life:

What knowledge does a car driver need? Clearly he or she needs to know the names of the controls; where they are; what they do and how they are operated (you move the pedals with your feet, not with your hands). However, the driver also needs the skill to be able to use the controls to guide the car along a road without hitting the various objects that tend to get in the way; you have to be able to do this at a normal speed (you can fail your driving test in Britain for driving too slowly or hesitantly); you have to drive smoothly and without getting too close to any dangerous obstacles. And it is not enough to drive in a straight line: the driver has to be able to manage the variations in road conditions safely.

With that statement Bygate underlines that practice and spontaneity are essential. The students need to use the language and their knowledge about language in real

situations (as real as possible). Furthermore, for using a language successfully, it is necessary to be able to react to different situations and statements spontaneously. According to Bygate (1987: 3) "[t]his means making decisions rapidly, implementing them smoothly, and adjusting our conversation as unexpected problems appear in our path". That statement shows that Bygate and Thornbury are of the same opinion when it comes to the importance of spontaneity in speaking.

So, how can we define the difference between knowledge and skill? One can say that both of them can be understood and memorized, but only a skill can be imitated and practiced (Bygate 1987: 4). In the pedagogical context, that distinction is important because problems in the two different areas might also need different consequences.

Bygate (1987: 5) describes two different ways in which something a speaker does can be seen as a skill. Firstly, there are the motorperceptive skills and secondly, the interaction skills. The motor-perceptive skills can be seen as the context-free kind of skill. It includes perceiving, recalling and articulating in the correct order sounds and structures of the language. This skill has been recognized in the teaching of speaking for many years now. Much importance is given to doing things in the right way. Bygate mentions W.F. Mackey (1965). His focus was on the motor-perceptive skills and on doing things correctly. So, his exercises mostly consisted of model dialogues, pattern practice, oral drill tables, look-and-say exercises, and oral composition. Bygate tries to critique this kind of approach with the help of the learning to drive a car example: "This is a bit like learning to drive without ever going out on the road" (Bygate 1987: 5). Bygate refers to David Wilkins (1975) who pointed out that there were problems that those exercises could not solve. The most crucial one might be that those exercises cannot ensure that the things that are learned in the classroom will also work out in real life situations, or if they will have any influence on what will happen outside of the classroom. This kind of transition is often called "transfer of skills" (Bygate 1987: 5). Furthermore, Bygate refers to Wilkins (1975) as he states that the teacher guided language use protects us from making our own decisions and choices. This makes clear that there are more skills to be developed than only motor-perceptive ones, in order to be able to communicate in a successful way. Bygate calls those skills interaction skills, which would include choice-making, and controlling one's own language production. Bygate claims that "[i]nteraction skills

involve making decisions about communication, such as: what to say, how to say it, and whether to develop it, in accordance with one's intentions, while maintaining the desired relations with others" (Bygate 1987: 6). At this point, the goal is not to teach ultimate correctness but to teach the students how to make the right choices. What is right or wrong depends on what we decided to say or not to say, how good we have been communicating so far, what we want to achieve and what kind of relation we think we have to our interlocutors.

In addition to that, interaction skills also include the ability to use language in order to satisfy particular demands. One can differentiate between two demands. On the one hand, there are the so-called processing demands. They are related to the internal conditions of speech, which would be the fact that speech is not slow, on the contrary, it takes place under the pressure of time. On the other hand, there are the reciprocity conditions which involve the dimension of interpersonal interaction in conversations.

So, what effect do the processing demands have on speech? If one imagines a well prepared speech and a spontaneous conversation with a friend, the difference becomes easily visible. The time component affects our choice of words and our style. Bygate mentions Brown and Yule (1983) who state that native speakers tend to use shorter utterances than sentences, which are only loosely strung together. Brown and Yule (1983: 26) state that:

If native speakers typically produce short, phrase-sized chunks, it seems perverse to demand that foreign learners should be expected to produce complete sentences. Indeed it may demand of them, in the foreign language, a capacity for forward-planning and storage which they rarely manifest in speaking their own native language.

One can conclude from that statement that processing conditions are crucial for becoming a fluent speaker of a language. Bygate claims that "[t]he ability to master the processing conditions of speech enables speakers to deal fluently with a give topic while being listened to" (1987: 8). Most of the time that is no problem at all in our first or native language, but learners often have problems with that if they only dealt with the language in written form or with heavy emphasis on accuracy. As stated above the time factor is not the only source of problems. The speakers also have to decide what to say, and this is affected by the reciprocity conditions. Those

conditions describe the relation between the speaker and the listener during a conversation. Both need to make sure that the conversation is working. There are at least two addressees and two decision-makers. The speakers might need to adjust his or her speaking to the needs of the listener (use different vocabulary for example) and the listener needs to reply to what the speaker says. All of that requires a lot of flexibility from the speakers.

Bygate finishes his explanation of the skill of speaking with stating that "[t]he main topic of this [work], then is to discuss ways in which speakers effectively use knowledge for reciprocal interaction under normal processing conditions, and to explore ways in which the ability to do this can be developed in foreign-language or second-language learners" (1987: 9). That statement clearly shows that Bygate emphasizes the importance of the use of language in real-life situations and not accuracy or correctness. To sum up, according to Bygate, the focus in the language classroom should be on the ability to communicate. So, seeing speaking as skill entails seeing the importance in the communicative abilities. One can conclude that also in Austria the skill of speaking, which should be taught, is based on communicative abilities. So, what should be taught is using the language and not knowing about the language. However, those communicative abilities are not easily defined, as one can see in the following chapter.

4.1 THE COMPONENTS OF THE SKILL OF SPEAKING

For the purpose of this thesis it is important to define what the skill of speaking really is or what it includes. This should be done with the help of reference to different theorists. All this is important for the central question of this thesis as teachers need to find out what the skill of speaking includes, before they start teaching speaking, otherwise defining goals in a teaching sequence might be hard.

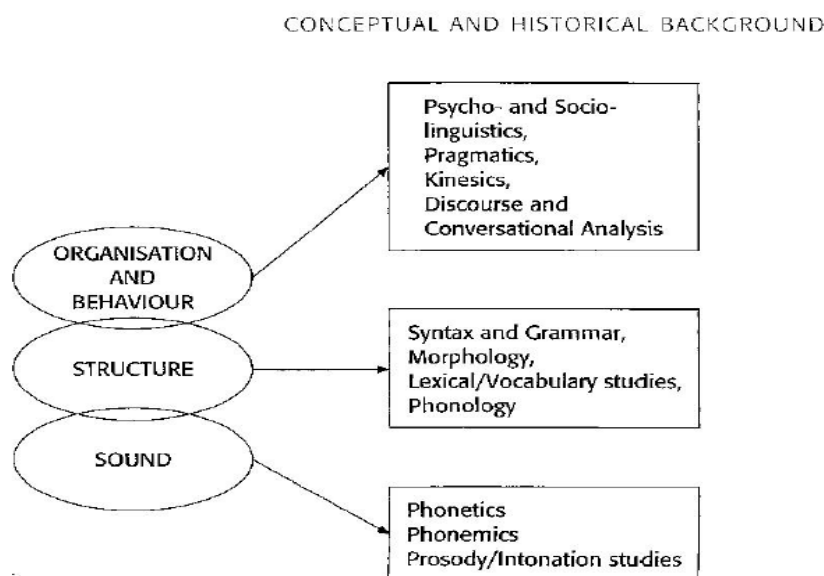
What Rebecca Hughes wants to make clear right at the beginning of her work *Teaching and Researching Speaking* is that speaking is not a discrete skill. It seems that Thornbury, Bygate, and also Hughes are of the opinion that this is the most important thing to be said about the skill of speaking. To underline the importance of that statement she poses some hypothetical questions (2011: 6):

How far, for instance, is the structure of a conversation culturally determined (also dealt with in pragmatics and ethnography)?

How far are the grammar and vocabulary of speech different from other sorts of grammar (which is related also to the fields of syntax and semantics)?

What are the critical factors in the stream of speech that make it intelligible (prosody, phonetics/phonemics)?

She uses the following figure to explain what the skill of speaking includes (2011: 7):



The three areas "Organisation and Behaviour", "Structure" and "Sound" relate to different areas in linguistics. "Organisation and Behaviour" consists of Psycho-and Sociolinguistics, Pragmatics, Kinesics, Discourse and Conversational Analysis, while "Structure" compiles Syntax and Grammar, Morphology, Lexical/Vocabulary studies and Phonology. The area "Sound" includes Phonetics, Phonemics and Prosody/Intonation studies. This figure immediately brings the structure of different rating scales into the mind of the reader as they are, most of the time, also grouped into different areas.

Rebecca Hughes (2011: 7) goes on to say that teaching speaking is not easily separated from other objectives. So, the objective of a speaking activity might not be to make the speaking more fluent or practice spontaneity but to make the pupils

aware of a grammatical rule or an intonation pattern. To Hughes speaking and also teaching speaking consists of more than just speaking.

Hughes also states that too often teaching speaking is not about producing language but learning about language. She (2011: 7) states that

[t]his has had the effect of dislocating the fundamental fabric of spoken mode- fluent intelligibility over a sophisticated range of styles and discourses - from other linguistic features. These are too often taught in isolation from the speaking skills needed to deliver them.

Hughes uses the example of teaching idioms to underline the importance of that statement. She says that idioms are often taught too early when the learner's level of productive speech is too low for them ever to achieve delivery without causing confusion. For Hughes, a much more useful step would be to teach simple conversational strategies, such as using fillers to manage the time pressure in a conversation.

4.2 THE IMPORTANCE OF USING CORPORA

A possible solution for the problem mentioned above would be the use of corpora. Hughes mentions the emerging importance of that device in the researching of teaching speaking as the corpora of natural speech might help the teachers to understand what speaking is actually like (Hughes 2011: 8).

She also states that as a learner of a second language one needs to be aware of a large number of aspects other than grammar and vocabulary, which would relate to areas such as "culture, social interaction, and the politeness norms". The speaker must change or expand his or her identity and needs to learn about the cultural, social and even political factors too make the right choices in a conversation (Hughes 2011: 8-9). Here corpora work might lead to a broader understanding of the concepts.

Rebecca Hughes makes a statement about the difference between written and spoken style and comes to a very complex conclusion. She (2011: 53) states that

there is a sense of 'high' and 'low' register being the main distinguishing feature between the spoken and written forms of

language. The notion of a minimal level of structure and vocabulary, 'slurred' and elliptical forms and commonplace or everyday discourse as opposed to high-flown or literary style being the norm of speech means that it is not something to be taken as a model for correct, acceptable language use in all circumstances.

Hughes comes to the conclusion that although the spoken form should be seen as unique, it should not be regarded as a role model for the correct use of a language and should therefore not be introduced in the classroom as a role model. If we compare these claims to Bygate's view, one immediately realises that there is a difference. Bygate thinks that speaking fluently is the ultimate goal and should, of course, be taught in the classroom. Hughes tries to prevent any misunderstandings. She does not want speaking to be excluded from teaching. On the contrary she thinks that speaking should only be treated differently in researching and teaching, as it has its own rich and diverse grammar (2011: 53). However, what this suggests for the teaching of speaking is that "the spoken form is a neglected source of subtle language choices for the learner, and a form needing to be brought closer to the heart of language descriptions and into the 'menu' of language choices made available to learners" (2011: 53). What Hughes does is that she introduces two different approaches on the importance of teaching speaking one is by Yungzhong (1985: 15) who does not see the teaching of forms that are unique to the spoken as an essential part of a student's structural knowledge (Hughes 2011: 51). Carter and McCarthy have a different view (qtd. in Hughes 2011: 52). They see the diverse language choices which have to be made when speaking as central in a teacher's repertoire to be introduced in the classroom. Hughes mentions the importance of authentic speech data in teaching speaking, which goes along well with Bygate who thinks that the corpora of natural speech might help the teachers to understand what speaking is actually like. That might help the teachers in Austrian Middle Schools too as they are not really exposed to corpora work during their teacher training. It could be a solution to include corpora work in their permanent education.

4.3 SPEAKING VERSUS WRITING

The following chapter should define the differences between spoken and written language. That distinction is crucial in the teaching of speaking as students and

teachers need to be aware of the fact that learning to speak a language is not only reading a text or reading a dialogue from a textbook correctly. They should be aware of the fact that speaking is communicating and that speaking performs different functions.

Bygate makes a very simple but powerful statement, namely "speech is not spoken writing" (Bygate 1987: 10). That makes clear that speech is not simply a verbalised form of written language. It needs totally different decisions and forms. People who speak like a book do not seem normal to us that is because the vocabulary might be too formal and the sentences far too complex and too long

Even reading from a book might not be that easy, as one thinks it should be. Getting the intonation right is hard work, as we are not used to speak in such long and complex sentences. Bygate (1987: 10) describes speaking like a book as "disagreeable and difficult". The reasons for that are that written language is not adjusted to the processing conditions and the conditions of reciprocity (Bygate 1987: 10-11).

As stated earlier the processing conditions are closely linked to the fact that we need to make decisions under time pressure. So, the speaker cannot plan and organize the language as much as the speaker would when writing the language. As an effect, the speakers make mistakes, which sound totally natural to the listener but would not be acceptable in written form, for example "we forget things we intended to say; the message is not so economically organized as it might be in print; we may even forget what we have already said, and repeat ourselves" (Bygate 1987: 11).

Further reasons for not speaking like a book, would be that the reader has the chance to reread while the listener has to ask for repetition or he can have memory problems which could lead to misunderstandings.

Not only the processing conditions make up the difference, also the reciprocity conditions play an important role. Mostly, we speak to a person who is standing in front of us. This can be seen as an advantage as the speakers immediately get feedback from the listener. That means that the speakers can immediately react if the communication did not work. That cannot be provided in writing. Furthermore, Bygate (1987: 12) says that the writer needs to imagine what the reader wants or does not want:

In written communication a considerable part of the skill comes from both the reader's and the writer's ability to imagine the other's point of view. A writer has to anticipate the reader's understanding and predict potential problems. In doing this the writer has to make guesses about what the reader knows and does not know, about what the reader will be able to understand, and even about what the reader will want to read. If the writer gets this wrong, the reader may give up the book or article in disgust before getting far.

When it comes to speaking the situation is different. The speakers need to make sure that communication is taking place, with the help of patience and also imagination. Bygate claims that "they [the speakers] have to pay attention to their listeners and adapt their messages according to their listeners' reaction" (Bygate 1987: 12). With the help of these reactions, the message can be adjusted from moment to moment, understanding can be improved, and the speaker's task is therefore facilitated.

5 SOME BASIC CONSIDERATIONS ABOUT THE TEACHING OF SPEAKING

The following section comprises a selected number of issues which should be considered when teaching speaking. It covers inter alia: teacher identification, the relationship between listening and speaking, the significance of pronunciation and the issue of accuracy versus fluency.

5.1 ARE NATIVE SPEAKERS BETTER IN LANGUAGE TEACHING?

The perception that native speaker teachers are better language teachers than non-native speaking teachers is widely spread. When one has a look at the Austrian school system, this becomes clearly visible as most schools make use of the programs which send native speakers from school to school (for example ABC or foreign language assistants). It somehow raises the value of language teaching in the eyes of headmasters and also parents. But how did this picture evolve? And what are the reasons for thinking like that? For that purpose, two studies and their findings will be compared. On the one hand a study carried out in Austria by Kerstin Kaim for her

diploma thesis "English Language Teaching in Austria: Self-Assessment of non-native speaking teachers" and, on the other hand, a study carried out in Taiwan by Chung Ke and Kun-huei Wu, which is called "Haunting Native Speakerism? Students' Perceptions toward Native Speaking English Teachers in Taiwan".

First of all some basic beliefs about language teachers will be explained in order to understand the findings and intentions of the two studies more easily.

Phillipson (1992: 184) postulates five tenets which "represent influential beliefs in the EFL profession". Those tenets could be summarized as follows:

- English is best taught monolingually.
- The ideal teacher of English is a native speaker.
- The earlier English is taught, the better the results.
- The more English is taught, the better the results.
- If other languages are used much, standards of English will drop.

(Kaim 2004: 101)

Tenet Number two seems to be the most crucial for the purpose of this section. It says that "[t]he ideal teacher of English is a native speaker". So, a native speaker or a teacher whose accent is native-like seems to be the best role model for students.

Phillipson (1992: 194) claims:

The native speaker serves as the model who can personify the native speaker abstracted and reified in works on standard and grammar and vocabulary and in 'received pronunciation', and which teaching materials and sound recordings seek to reanimate.

One could say that the model of the native speaking teacher has not really been criticised for a long time. However, a problem could be that the learners have neither a realistic nor a reachable goal in their minds. Most of the students' motivation is based on passing an exam or another school year. They do not really aim at sounding native-like. It could be the case that learners give up at some stage at the beginning because they cannot find a clearly defined goal for their efforts in language learning (Kaim 2004: 102). Phillipson (1992) contrasts that opinion and gives more advantages. He claims that the language performance of the native speaker is

superior to that of the non-native speaker. Their pronunciation and their range of vocabulary might be better, but do they really perform better in the classroom? If language competence was the only variable for good teaching, then yes. However, teachers, who are practicing their profession, know that there are more factors that are at least as important as linguistic competence. Variables like motivating students, the teacher's character and the teaching methods and approaches play an important role in the learning process. One could say that a lack of confidence or a feeling of inferiority would not help the situation of the non-native speaking teachers (Kaim 2004: 102-3).

All the previous findings show that it is not easy to establish clear criteria providing either the native or the non-native speaker to be the better foreign language teacher. There are advantages and disadvantages on both sides. However, one must say that native speakers must have a sound understanding of the learners' L1, of its system and its articulatory phonetics. Otherwise,

[..] they will be able to inform but not instruct: to do little more than model their own accent and hope that acquisition will follow by some mysterious magic process (Jenkins 2000: 221).

As stated earlier studies in this section a lot expected from native speaking teachers, but what do we expect from non-native speaking teachers. A common conception would be that they are more sensitive towards the whole learning process of students, including language difficulties and development throughout acquisition. All those qualities are highly valuable. Furthermore, Thomas (1999: 12) states:

They not only empathize with their student but they make another very vital contribution to the field, although rarely acknowledged. They bring something unique to the ESOL profession. They are role models, they are success stories: they are real images of what students can aspire to be. Instead of being the exception, NNSs [non-native speakers] need to be the rule, found in every rank and level of ESOL teaching.

To conclude, there is no need to assume that native speaking teachers are worth more than non-native speaking teachers. Non-native speaking teachers have qualities, which the native speakers do not have and the other way round. A non-native speaking teacher is a teacher and a learner, which is definitely an advantage.

The teacher can imagine what the students feel when learning a new language, which is an enormous help.

Now to the findings of the two studies. Kerstin Kaim compares the result of a study carried out by Barbara Seidlhofer in 1996 with the study she did in 2002. The focus of the study is on the self-perception of the non-native speaking teachers. Some of the most striking changes from 1996 to 2002 would be that fewer teachers regard themselves as inferior to their native-speaking colleagues and secondly, that still most teachers believe that being a native speaker of the target language would have helped their career. So, one could say that the teachers think that native speaking teachers are worth more on the job market but that they do not think that they are any better than the non-native speaking teachers themselves.

The study by Wu and Ke (2009) goes into a different direction. The focus is on the students and their perception of the native-speaking English teachers. The information includes questionnaires from 107 students and interviews with three NESTs and 19 students who have filled out the questionnaire. The results show that students expect more from the native-speaking English teachers than from other teachers. They want to have more encouragement and interaction. Furthermore, they expect relaxed activities with fewer assignments and tests. Also an authentic accent plays an important role. That is demanded by a third of the students while a quarter do not care about accents at all. On the side of the teachers, there are other difficulties. They complain about the students' passiveness and lack of responsiveness. One can conclude that students expect the NEST to be interactive, while they themselves seem to give an impression of unwillingness to participate (Wu and Ke 2009: 44).

5.2 LISTENING AND SPEAKING – AN EFFECTIVE RELATION?

In order to teach speaking effectively it is crucial to understand the implications of the relation between listening and speaking as this combination is also highlighted in the Austrian curriculum:

Im Anfangsunterricht allerdings sind die Teilfertigkeiten des Hörverstehens und der mündlichen Kommunikation durch regelmäßige Hörübungen sowie durch ein möglichst häufiges Angebot an Sprechanlässen verstärkt zu fördern. (bmukk 2012: 34)

Consequently, in this part of the thesis, the influence of speaking on listening will be examined. A study by Yan Zhang (2009) about the effects of listening on the oral proficiency of college students will be explained, as well as Krashen's Comprehensive Input Hypothesis, which forms the basis of the study.

The study by Zhang was done in China, as China requires more and more college graduates with higher oral English proficiency. The author (2009: 194) of the thesis posed two hypotheses:

- 1) Students' listening ability and their oral English production ability are correlated.
- 2) Teachers who bring listening and audio-visual materials into oral English class are likely to have better teaching results.

Yan Zhang aims at doing a quantitative analysis on the effects of listening on speaking for college students. Furthermore, a quantitative analysis on the correlation between listening and speaking is also made.

The results of the study show that listening and speaking do not work without each other, on the contrary, the listening and speaking ability are closely related. The study proved that listening does have a positive effect on improving college students' oral English (Zhang 2009: 194).

As stated earlier Krashen's Comprehensive Input Hypothesis forms the basis of the study. What Krashen argues is that SLA is dependent on the availability of comprehensive input before the learners can produce the language themselves. Krashen presents "the case for comprehensible input in the form of the input hypothesis. He argues that for SLA to take place, the learner needs input that contains exemplars of the language forms which according to the natural order are due to be acquired next" (Zhang 2009: 195). But how is input made comprehensible. One way is to refer to what the learners already know. A different way would be the

"here-and-now" orientation which "enables the learner to make use of the linguistic and extra linguistic contexts and his general knowledge to interpret language which he does not actually know" (Zhang 2009: 195). A third way would be the modification of the interactional structure of conversation. Zhang (2009: 195) thinks that a "here-and-now" orientation, together with interactional adjustments, are the main source of comprehensible input. They "ensure that communication proceeds, with exposing the learner to new linguistic material" (Zhang 2009: 195). According to Zhang (2009: 195):

Krashen's code breaking approach to listening became a strong influence on language teachers in the 1980's. It is saying essentially that L2 acquisition depends on listening: decoding is code breaking. It did not, however, lead to a generation of published listening-based main course books.

The study by Yan Zhang proves that teaching speaking does not have to be about production only. "Decoding is code breaking". That might help the teachers out of a dilemma. Most of the time teaching communication is reduced to practice speech production but the study proves that also listening practice has its value and helps fostering the ability to communicate in an L2.

5.3 THE SIGNIFICANCE OF PRONUNCIATION IN THE L2 LANGUAGE CLASSROOM

It is a common belief that learners of a foreign language with good pronunciation are likely to be understood even if they make other mistakes, for example in the field of grammar. Learners with problematic pronunciation might be afraid to talk English because they might have experienced social isolation, employment difficulties or limited opportunities for further study. Most of the time, the first impression we get from a learner of English is an oral one, so we judge people by the way they speak. Speakers with poor pronunciation might be judged as incompetent, uneducated or lacking in knowledge. Unfortunately, pronunciation is not an easy topic when learning a new language, often explicit help from a teacher is needed to acquire that skill. Gilakjani (2012: 96) goes even further and states that "Pronunciation instruction is a prominent factor in foreign language teaching. Since sounds play an important role in communication, foreign language teachers must attribute proper importance to

teaching pronunciation in their classes". In conversation, it is essential that communication does not break down. Gilakjani (2012:96) says that "unless he [the speaker] has sufficient knowledge of the sound patterns of the target language, he can neither encode a message to anybody nor decode the message sent by another person by learning the sounds of the target language within his mother tongue". For Gilakjani pronunciation is the key to gaining full communicative competence.

Pronunciation is normally referred to as the production of sounds that we use to make meaning. That includes (Gilakjani 2012: 96)

attention to the particular sounds of a language (segments), aspects of speech beyond the level of the individual sound, such as intonation, phrasing, stress, timing, rhythm (suprasegmental aspects), how the voice is projected (voice quality) and, in its broadest definition, attention to gestures and expressions that are closely related to the way we speak a language.

It is important to keep in mind that all those components are closely linked. They do not work in isolation. It is the combination that makes good pronunciation.

The more traditional approaches in teaching pronunciation focused mostly on segmental aspects, on the sounds as such, as they are in some way related to writing and therefore easier to work on. However, the more recent approaches focus on the suprasegmental aspects of pronunciation because they are said to have the most effect on intelligibility. It is important that the students master good pronunciation and communication and not perfect accents.

When one looks back in time, one can immediately realise that there was a tremendous development in the teaching of pronunciation. In the 50's and 60's, when the audio-lingual approach was very popular, the goal was the imitation of a native-like accent. From the late 1960's to 1980's the instructional focus was questioned frequently. The native-like pronunciation was viewed as an unreachable goal. So, pronunciation instruction was reduced or eliminated altogether. In the 80's pronunciation became important again. However, the focus then was on supra-segmentals, sound co-articulation and voice quality. One could say that communicative competence came into play. Less emphasis was on correct articulation of L2 specific sounds. In the 90's pronunciation was seen as an essential

part of communication. The focus was on communicating meaning, meaningful practice, and the uniqueness of each individual ESL learner" (Gilakjani 2012: 97). Now the student was in the focus of the teaching process.

Being understood is essential in our global society. So, acceptable English pronunciation is important. When it comes to define acceptable English James (2010: 96) differentiates between 3 basic levels:

Level 1: People often do not understand what the speaker is saying. The speaker uses the wrong sounds when making English words or uses the wrong prosodic features when making English sentences. [...]

Level 2: People understand what the speaker is saying, but the speaker's pronunciation is not pleasant to listen to because he/she has a distracting and/or heavy accent. [...]

Level 3: People understand the speaker, and the speaker's English is pleasant to listen to. [...]

So, level 3 would be the ultimate goal in teaching pronunciation, Gilakjani (2012: 98) calls that level of pronunciation "comfortably intelligible". The aim is not to sound like a native speaker, as some teachers would suppose. Useful models can be found everywhere. Different news channels from all over the world use English. Their pronunciation is easy to understand and pleasant to listen to. To sum up, the learners "should develop their own accent which is close to a standard variety" (Gilakjani 2012: 98). Pronunciation is more than producing sounds correctly, it is a crucial part of communication that should be incorporated into classroom activities. It should be given as much attention as any other part of language learning, such as grammar, writing or vocabulary. To fulfill that requirement, teachers must be trained in teaching pronunciation and materials, testing tools and classroom activities must be integrated. The teachers should actively encourage students to speak and to pay attention to their pronunciation.

The Austrian curriculum suggests the following when it comes to the importance of correct pronunciation:

Dennoch ist insgesamt und in sinnvollem Maße eine möglichst hohe Qualität und zielsprachliche Richtigkeit der fremdsprachlichen Äußerungen anzustreben; lernersprachliche Abweichungen von der

Zielsprache sind dabei stets niveaubezogen und aufgabenspezifisch zu behandeln. (bmukk 2012: 35)

So, one can deduce from that quote that the Austrian curriculum suggests nearly the same for the Austrian classroom. The teachers should always aim for the best but leave space for mistakes. This leads directly to the next issue, namely accuracy versus fluency.

5.4 ACCURACY VERSUS FLUENCY

Brumfit (1984: 51) summarises the problem of accuracy versus fluency clear and intelligible:

From this direct concern with a teaching need came an increasing awareness that, particularly for language work but arguably for most learning, the demand to produce work for display to the teacher in order that evaluation and feedback could be supplied conflicted directly with the demand to perform adequately in the kind of natural circumstances for which teaching was presumably a preparation.

So, what needs to be thought through is the fact that a language is ultimately learned in order to communicate in the real world and not to provide feedback to the teacher. When evaluating the progress of a learner explicit knowledge, problem solving, and evidence of skill-getting used to be in the focus of the teacher. However, authentic language use needs fluency, expression rules, a reliance on implicit knowledge and automatic performance (Brumfit 1984: 51). The teacher must make a decision. That decision is not coming from psychology or linguistics. It is rather a methodological one. He or she must decide how much time is spend on activities focusing on accuracy and how much time is spend on fluency or fostering communicative competence.

When it comes to the Austrian English classroom, the following must be kept in mind, as the curriculum has its clearly defined tenets:

Der funktionale Aspekt der Grammatik hat Vorrang gegenüber dem formalen Aspekt. Generell sind die situative Einführung und ein induktives Erschließen grammatischer Sachverhalte aus kommunikativen Zusammenhängen und Textbeispielen anzustreben. Grammatische Teilsysteme dürfen sich keineswegs verselbstständigen und wegen ihrer leichteren Überprüfbarkeit indirekt zum eigentlichen Lernziel des Fremdsprachenunterrichts werden. Wo es sinnvoll ist, sind grammatische Strukturen besser ohne Regelformulierung als lexikalische Einheiten zu vermitteln. (bmukk 2012: 34-35)

Die Bereitschaft der Schülerinnen und Schüler, neue sprachliche Strukturen in den Bereichen Lexik und Grammatik anzuwenden und dabei Verstöße gegen zielsprachliche Normen zu riskieren, ist im Sinne des übergeordneten Zieles der kommunikativen Kompetenz von zentraler Bedeutung und bei der Evaluation der Schülerleistungen dementsprechend einzubeziehen.

Im Fremdsprachenunterricht ist weiters auf allen Lernstufen zu berücksichtigen, dass sich Schülerinnen und Schüler der Zielsprache über lernersprachliche Zwischenschritte annähern und dass Fehler ein selbstverständliches Merkmal des Sprachenlernens sind. Dies ist in Übungsphasen und bei der Fehlerkorrektur zu berücksichtigen.

Dennoch ist insgesamt und in sinnvollem Maße eine möglichst hohe Qualität und zielsprachliche Richtigkeit der fremdsprachlichen Äußerungen anzustreben; lernersprachliche Abweichungen von der Zielsprache sind dabei stets niveaubezogen und aufgabenspezifisch zu behandeln. (bmukk 2012: 35)

To sum up, the Austrian curriculum does not really give a choice in this matter. It is clearly stated that accuracy should not be the main goal in the language classroom. Of course, correct language use is a goal which needs to be aimed for but communicative competence is in the centre of attention: "Für den gesamten Fremdsprachenunterricht steht Fertigungsorientierung im Vordergrund" (bmukk 2012: 36).

However, there are still interesting issues which need to be discussed as there are situations in the daily routine of a teacher when he or she needs to decide for him- or herself if accuracy or fluency is what should be aimed for.

If one wants to define the term “accuracy”, one could think of it as teaching, which will result in usage, rather than use of language in the classroom (Brumfit 1984: 52). However, there are still some points which need clarification (summarised from Brumfit 1984: 52-53):

1. Fluent language can also be accurate language. The focus of the teacher or the learner is on fluency and not on accuracy, but, of course, the language that is produced is, in the ideal case, also accurate.
2. There is no ultimate right or wrong. Both directions are important. However, one has to keep in my mind that their functions in language teaching are quite different and that an overuse of accuracy-based teaching might not lead to successful language development.
3. Accuracy can refer to listening and reading, but also to speaking and writing. That is also the case when it comes to fluency. Any language activity can have its focus on accuracy or on fluency as it refers to a focus of the user.
4. The criterion is always the intended mental set of the user. The quality of the outcome is always irrelevant to the distinction.
5. Monitoring can also take place during fluency work, but it must be used in the same way that native speakers use it. However, “it is recognised that the value of the distinction for teachers should not lead them to prevent learners, particularly at intermediate and advanced levels, from combining a concern with language use with worry about formal accuracy in terms of specific language items[.]” (Brumfit 1984: 53).

The term “fluency” is not as easily definable as “accuracy”. While “accuracy” has a meaning close to the common one, “fluency” is used with slightly different implications. C.J. Fillmore (1979) distinguishes four different kinds of fluency, with exclusive reference to production. The first would be the ability to talk without significant pauses for an extended period. In that case, monitoring must be unconscious or automatic and the focus should be on quantity and not on quality. In the second kind coherence is important. The sentences must be coherent and syntactically dense. Fillmore’s third kind underlines the importance of vocabulary knowledge. The learners should be able to have appropriate things to say in a wide range of contexts. The last areas, which are important for Fillmore, are creativeness and imagination. So, the decisions the speaker makes must be spontaneous and

most beneficial for the conversation. To conclude all this, a speaker who has all those abilities is a “maximally gifted wielder of language” (C.J. Fillmore 1979: 93).

All the characteristics mentioned above relate to speed and continuity, coherence, context sensitivity and creativity (Brumfit 1984: 54). However, one has to keep in my mind that these are abilities that language users possess to varying degrees. Another question is the extent to which the abilities can truly be considered linguistic abilities, for with the exception of the first they all require capacities, which we see in people who are not linguistically fluent. Brumfit (1984: 54) states that

The ability to marshal arguments cogently and present them with maximum skill may exist in someone who can do this only after successive redraftings; the ability to respond sensitively and appropriately to varying situations and circumstances may be possessed, and demonstrated, by people who are not verbally fluent but who express themselves primarily through non-verbal means, by sympathetic expressions and gestures, by subtle judging of how much or how little physical contact to make, and so on; and creativity in language use has some relationship to the ability to establish significant relationships between concepts, visual and aural patterns and systems of thought – the creativity is expressed through the language, and not merely within the linguistic system.

To sum up, linguistic qualities are not the only areas, which make a speaker fluent, personal qualities and characteristics, play an equally important role. What Fillmore’s discussion draws attention to is the interaction between language and knowledge of the world in the development of fluency. Leeson (1975: 136) takes up a completely different view. To him fluency is “the ability of the speaker to produce indefinitely many sentences conforming to the phonological, syntactical and semantic exigencies of a given natural language on the basis of a finite exposure to a finite corpus of that language”. He only focuses on the fluency described by linguists, as he leaves out all other categories from authentic language. To Brumfit (1984: 55), “[i]t is clear that such a formulation ignores precisely those ‘performance’ elements which will contribute to acquisition if we take the creative construction hypothesis seriously”. Brumfit wants to make clear that activities based on such a formulation are not sufficient for any language learner.

Fluency must be regarded as “natural language use” (Brumfit 1984: 56). That assumption brings along several other issues, which need to be kept in mind (summarised from Brumfit 1984: 56-57):

1. Language that is produced should have been processed by the speaker, without interruption from the teacher.
2. The speaker should decide what her or she wants to say, of course, in respect to the certain task or as an answer to what has already been said.
3. The speaker has to react spontaneously in different situations, which would include improvising, paraphrasing, repair and reorganisation.
4. The language is always a means to an end. The objective of the activity should be quite distinct from the formation of appropriate or correct language.
5. Students should see the teacher rather as a communicator during the performance and not as someone who judges their performance. Brumfit claims that “[t]his has implications for the power relations in the class, but the crucial point is that the teacher’s unavoidably greater power to determine what is or is not appropriate behaviour should not affect students’ freedom to hide or reveal their own intimate feelings, or personal information, in the same way as they would be free to choose in a non-pedagogic environment”. A further implication affects the attitude towards error. There should be no or only little place for correction in fluency work, as it might distract from the purpose of the task, or may even be perceived as rude. One has to accept that making errors is an essential part of second-language acquisition.

According to Brumfit (1984: 57), fluency “can be seen as the maximally effective operation of the language system so far acquired by the student”. The students must be put in situations, where they need to use language as fluently as possible. Then the process of creative construction will be assisted.

6 THE COMMUNICATIVE CURRICULUM AND THE TEACHING OF SPEAKING

6.1 THE CEFR AS A BASIS

As stated in the introduction, the CEFR is one of the most influential guidelines for teachers in Austria. It was designed to provide a clear, coherent and comprehensive basis for the elaboration of language syllabuses and curriculum guidelines, the design of teaching and learning materials, and the assessment of foreign language proficiency. For all those reasons, it is interesting to see how the CEFR was designed.

6.1.1 THE DESIGNING PROCESS OF THE CEFR

The work by North (1995, 1996) might be the most important developing of scales through the scaling of descriptors. That was done in the context of developing a Common European framework of reference (CEFR) for reporting language competency. The CEFR claims that “all human competences contribute in one way or the other to the ability to communicate and may therefore be regarded as aspects contributing to communicative competence” (Council of Europe 2001: 101). In the CEFR communicative competence consists of linguistic competences, sociolinguistic competences and pragmatic competences, which are further divided into subsections. The design method is empirical, it is user- and assessor-oriented, involves holistic scoring, and the focus is real-world in that the final descriptors are framed as "can do" statements. Fulcher explains the approach in stating that “[i]n this approach scale developers collect large numbers of 'stand-alone descriptors' from as many different scales as possible, and re-sequence descriptors that can be calibrated onto new scales” (Fulcher 2003: 107). The measurement model underlying this process is multi-faceted Rasch analysis. The Rasch analysis was mostly used for the post hoc analysis of existing rating scales to address validity questions. North (1996) followed a systematic process of four consecutive phases of analysis: (a) intuitive; (b) qualitative; (c) quantitative, and (d) replication. Fulcher (2003: 109) explains the development of the CEFR as follows:

Phase A: Intuitive analysis

Step 1: North (1993) collected some 30 rating scales, the content of which was pulled apart into sentence-length descriptors, and placed into six proficiency levels.

This provided a 'pool' of 2000 band descriptors drawn from the language testing literature, and historical and operational tests.

Step 2: The 'pool' was classified into different types of communicative activities, and different aspects of strategic and communicative competence. New descriptors were written to fill perceived gaps in the descriptive scheme.

Phase B: Qualitative analysis with informants

Step 3: In a rolling series of workshops, pairs of teachers are given an envelope of band descriptors on confetti-like strips of paper and asked to sort them into 4 or 5 given, related categories. For each descriptor teachers mark those that they find particularly clear and useful. Several pairs will sort the same descriptors. Descriptors may be re-edited according to teacher comments (e.g. removing double negatives, or ambiguities.).

Step 4: The teachers read through the same band descriptors again, putting a circle around the ticks of those band descriptors that are relevant to their own teaching.

Step 5: In other workshops, pairs of teachers are give the band descriptors for particular related categories, and asked to put them into three piles: 'low', 'middle' and 'high'; and then to divide each of these piles into two subdivisions, leaving six piles of descriptors. The performance of each descriptor in Steps 3-5 is recorded with codes in a detailed 'item history'.

Step 6: The descriptors interpreted most consistently are then used to construct overlapping questionnaires of approximately 50 descriptors each, each questionnaire being linked to the one immediately above and below by anchor items (items that they have in common). The result is a 'chain' of questionnaires, with balanced content, targeted at each of the levels, linked by the anchor items. These are then used for the main data collection.

PHASE C: Quantitative analysis with questionnaire data

Step 7: A rating scale is attached to each descriptor on each questionnaire. The rating scale is reproduced from (1995: 451).

0 This describes a level which is definitely beyond his/her capabilities. Could not be expected to perform like this.

1 Could be expected to perform like this provided that circumstances are favourable, for example if he/she has some time to think about what to say, or the interlocutor is tolerant and prepared to help out.

2 Could be expected to perform like this without support in normal circumstances.

3 Could be expected to perform like this even in difficult circumstances, for example, when in a surprising situation or when talking to less cooperative interlocutor.

4 This describes a performance which is clearly below his/her level. Could perform better than this.

Step 8: A (preferably large) group of teachers are asked to rate a small sample of learners from their classes on the rating scale for each of the descriptors on the questionnaire.

Step 9: Multi-faceted Rasch analysis is conducted to construct a single scale from the descriptors on the chain of questionnaires covering levels from beginner to very advanced. The analysis discovers which items 'misfit' the model; this amounts to discovering which descriptors cannot be placed onto a unidimensional scale. These descriptors are removed from the scale. In linking the questionnaires together, various corrections need to be made for distortions arising from the statistical model and exaggeration by the teachers (North 1996/2000: 208-22)

Step 10: Descriptors are identified that have statistically significant difference of difficulty across different language groups, or educational sectors. Such differences in difficulty could be caused by variation among teachers in the way they interpret the descriptors, or differences in curriculum. Some difference of interpretation in different contexts may be appropriate for profiling grids and checklists, but not for holistic scales.

Step 11: Determine the cut-offs between levels of attainment on the arithmetical scale according to

- Difficulty estimates, in order to have equidistant bands
- Natural gaps and groupings on the vertical scale of descriptors
- Comparing the pattern of gaps and groupings to levels on the source scales

PHASE D: Replication

Step 12: Repeat the entire process with different teachers - and in this case also adding other languages (French and German as well as English) and other skills (Listening and Reading as well as Speaking). North (1996/2000: 339) reports a correlation of 0,99 in the scale values produced in the original study and those resulting from the replication study.

The result of this process is claimed to be a linear, equal interval, proficiency scale, based on a theory of measurement.

All this was produced within a specific context. The Council of Europe's Common European Framework for language teaching and learning required scaled definitions. The main purpose of the scale should be to assess learners of various

first languages, learning a variety of second languages, spread across a wide geographical area, with different educational systems and curricula. The results that are produced by the scale, therefore, need to be consistent and comparable. However, there is still a disadvantage. North and Schneider (1998: 242-3) address the issue that the method is essentially a-theoretical in nature. It is not based on any empirically validated descriptions of language proficiency or a model of the language learning process. It is, more or less an attempt to provide a working framework for the special context in Europe. As North and Schneider (1996: 242, qtd. in Fulcher 2003: 113) state:

The purpose of descriptors of common reference levels is to provide a metalanguage of criterion statements which people can use to roughly situation themselves and/or their learners, in response to a demand for this. It is widely recognised that the development of such a taxonomy entails a tension between theoretical models developed by applied linguists (which are incomplete) on the one hand and operational developed by practioners (which may be impoverished) on the other hand.

6.1.2 HOW TEACHERS CAN USE THE CEFR IN THE TEACHING AND TESTING OF SPEAKING

As it is clear now, how the CEFR developed and why it was developed, it is interesting to elaborate how it can be used in the teaching and testing of speaking, since this is one of the main issues of this thesis.

The CEFR provides not only the common reference levels, it also gives a descriptive scheme of definitions, categories and examples that language professionals can use to better understand and communicate their aims and objectives. That is the first point which contributes to the teaching of speaking. The CEFR helps to define learning goals, which is not often an easy step to take. Defining learning aims is essential in planning teaching sequences. Otherwise the activities would not be focused on a specific skill which the students should acquire. Furthermore, referring to the CEFR when choosing learning aims is advisable because the teachers are then able to justify their choices with the help of a legal document.

Additionally, it provides an orientation for assessing student's performances. It states what a student in the first form should be able to do.

So, for many teachers, the CEFR is the most important help for designing speaking activities or speaking tests. This matter will be discussed in chapter 7.4.1 when it comes to the E8 standards, as they are closely linked to the CEFR.

6.2 INTEGRATING SPEAKING IN THE CURRICULUM IN AUSTRIA

The following chapter should serve as an introduction on what decisions need to be made when planning a curriculum. In the first part, more general issues, like finding the right weighting, will be examined. The second part of this chapter should focus on the curriculum in Austria.

6.2.1 *GENERAL ISSUES*

When planning a curriculum or a syllabus, several decisions need to be made. One needs to decide how much emphasis should be given to speaking, how this emphasis will vary according to such factors as the level of the learners and their learning context, if speaking should be taught separately or integrated into the teaching of other aspects of linguistic competence and, finally, how speaking will be assessed.

When it comes to weighting the importance of the different skills, one has to find out about the learner's needs. Learners with an ESL background will be highly motivated to improve their speaking skills, whereas students in an EFL context might not feel as much urgency to improve their oral skills. Learners who need English for academic purposes might feel the need to work mostly on their written language than on spoken while learners who are learning English as an international language are more likely to prioritize intelligibility over accuracy, as they will have to deal with speakers of English from all over the world. So, what can be done before planning a programme is a needs analysis. A needs analysis can also function as help to find the best balance between accuracy and fluency in a course. The analysis can be done formally, through questionnaires or through interviews. The last step would be to do a placement test. Those are used for an initial assessment of the candidates' speaking skills. A placement test should include a range of interaction test in order to find out about the candidates knowledge and skills.

In planning a syllabus the question of accuracy and fluency is not really a question of weighting but of order. Should a focus on accuracy precede a focus on fluency, or should it come later? In earlier days, accuracy was the focus and only after mastering accuracy, fluency came into play. Often language learners are set objectives that most native speakers would find hard to meet. Quite obviously this approach to language learning frustrated many learners because in most cases complete mastery of a L2 is an unrealistic goal (Thornbury 2005: 115). So, what should be the focus is fluency and not accuracy.

However, as stated earlier in the chapter “the skill of speaking”, speaking does not only include oral production of grammar and vocabulary, so a large list of items like that is not a speaking syllabus. Multi-layered syllabuses, which specify not only on grammar and vocabulary components, but also on the skills to be taught, might serve as a solution.

Some examples of such syllabuses would be (Thornbury 2005: 117):

in unit one ...	in unit two ...	in unit three ...
speaking <ul style="list-style-type: none"> - talk about conversation topics - talk about jobs - discuss hopes and plans - start a conversation with a stranger - how to ... keep a conversation going 	speaking <ul style="list-style-type: none"> - talk about travelling - discuss different forms of transport - decide what makes a good holiday - how to ... make a complaint 	speaking <ul style="list-style-type: none"> - describe your perfect day - describe a famous actor - talk about your childhood - talk about your interest in arts - how to ... talk about your past

6.2.2 THE CURRICULUM IN AUSTRIA

All the issues discussed above would be most influential in a course for adults or at a higher level of competence. In our case, it is important to look at the Austrian

curriculum and what it says about the implementation of speaking in the English classrooms.

In Austria, the curriculum clearly underlines that speaking must be taught equally to other skills, for example listening or reading.

Die Fertigkeitsbereiche Hören, Lesen, An Gesprächen teilnehmen, Zusammenhängend Sprechen und Schreiben sind in annähernd gleichem Ausmaß regelmäßig und möglichst integrativ zu erarbeiten und zu üben. Im Anfangsunterricht allerdings sind die Teilfertigkeiten des Hörverstehens und der mündlichen Kommunikation durch regelmäßige Hörübungen sowie durch ein möglichst häufiges Angebot an Sprechanschlüssen verstärkt zu fördern. (bmukk 2012: 34)

What this suggests for the planning of lessons and school years, is that the teacher must devote time to the teaching of speaking. That time can be used in different forms, which will be explained in chapter 6.5, when different speaking activities will be examined. Last but not least, the CEFR levels can serve as a help for planning speaking times and activities as they state clearly defined goals. Additionally, the structure of course books can help to find the right balance between the skills. However, what the Austrian Curriculum clearly underlines is the importance of choosing manifold topics, which can also be found in the CEFR:

Die Schülerinnen und Schüler sind durch die Einbindung der sprachlichen Mittel in vielfältige situative Kontexte mit verschiedenen Themenbereichen vertraut zu machen (wie Familie und Freunde, Wohnen und Umgebung, Essen und Trinken, Kleidung, Körper und Gesundheit, Jahres- und Tagesablauf, Feste und Feiern, Kindheit und Erwachsenwerden, Schule und Arbeitswelt, Hobbys und Interessen, Umgang mit Geld, Erlebnisse und Fantasiewelt, Gedanken, Empfindungen, Gefühle, Einstellungen und Werte, Umwelt und Gesellschaft, Kultur, Medien, Literatur). Dies entspricht den „vertrauten Themenbereichen“ in den Kompetenzbeschreibungen des GER. (bmukk 2012: 36)

6.3 APPROACHES IN TEACHING SPEAKING

Teachers in all subjects have got their tenets and beliefs, which they want to follow. That is the same with teaching speaking in a foreign language. One needs to make decisions on what one wants to teach and how this should be taught. The following

section should explain some of the most important approaches and methods in the teaching of speaking. The difference between the direct and the indirect approach as well as two other approaches, namely the integrated skills approach and the task-based approach will be explained. Furthermore, the possibilities in classroom organisation and the advantages of group work will be discussed. However, what needs to be kept in mind is that the Austrian curriculum clearly favours the integrated skills approach:

Die Fertigkeitsbereiche Hören, Lesen, An Gesprächen teilnehmen, Zusammenhängend Sprechen und Schreiben sind in annähernd gleichem Ausmaß regelmäßig und möglichst integrativ zu erarbeiten und zu üben. (bmukk 2012: 34)

6.3.1 *DIRECT VERSUS INDIRECT APPROACH*

Anne Burns draws a clear distinction between two major currents of thinking that have informed contemporary debates on the teaching of speaking. Burns (1998: 103) explains that

[t]he first current focuses upon the development of skills for the accurate production of speech forms (phonological patterns, lexis/vocabulary, grammatical form and structure), while the second centers upon enhancing fluency through communicative tasks (Nunan 1989) which, in turn, enable opportunities for developing functional language use through non-controlled activities.

It is the teacher's decision what teaching approaches best support the development of the ability to communicate fluently. Burns states that one has to decide between "direct" or "indirect" approaches. In the "direct" or "controlled" approaches "the learners mainly focus on specific elements of communicative ability which are isolated and practiced" (Burns 1998: 103). Activities in that approach might involve

analyses of the typical structures of spoken genres, the learning of formulaic lexical phrases and institutionalized routines, discussions of the use of feedback devices and backchannelling in conversation, learning activities where learners construct their own grammatical awareness inductively, and the development of metalinguistic knowledge (Burns 1998: 103).

In contrast to that the indirect approach focuses on the production of more authentic and functional language use. Activities that fit in that approach can be "discussions, information gaps, project work, role plays, simulations, and talking circles involving discussion of personal experiences based on narratives, anecdotes, news, personal events, and so on" (Burns 1998: 103). Those activities give the learners a chance to practice their communication skills, the focus is on the "skill-using", which would include "for example, identifying the purpose of the communication, controlling and understanding its generic shape, clarifying key words and concepts through rephrasing or questioning strategies, initiating topics, expressing opinions, and agreeing or disagreeing" (Burns 1998: 104).

The relation between those two approaches is, of course, not an easy one. On the one hand, the direct approach is not enough, also practice is needed. On the other hand, only practice without any knowledge about the underlying concepts also might not lead to enhancing fluency.

6.3.2 *INTEGRATED SKILLS*

There is a tendency in language teaching to focus on only one skill at a time. One part of a lesson might be devoted to writing, while another part is devoted to writing and so on. Sometimes that reflects the needs of the learners, however, often it is a "pedagogical convenience" (Byrne 1986: 130), rather than a useful necessity and most of the time it just reflects the structure of a certain course book. That kind of teaching does not depict language in an authentic way. In real life, one does not use language in any set order. Byrne (1986: 130) gives an example:

For example, if we see an interesting advertisement in the paper for holiday, we may discuss it with somebody and then perhaps ring up or write for more information. This nexus of activities, which so far has involved reading – speaking/listening – either speaking/listening or writing, may continue or stop at that point. It can provide a model for integrating skills in a realistic way and is especially useful at a post-elementary level.

Another simple but effective way to integrate skills, is to make learners collaborate with each other, either in group or pair work. The teacher could give the students

different roles. One student could be working in a tourist office while the other one is a tourist who wants to gather special information. Both get an instruction sheet, which they need to read beforehand. That might be a way to combine speaking, reading and maybe also writing in one single activity.

Finally, simulation and project work provide a natural framework for integrating skills. Byrne (1986: 130) lists reasons why integrated skills activities are important in the learning process:

1. They provide opportunities for using language naturally, not just practicing it.
2. Many pair and group work activities call for a variety of skills, sometimes simultaneously, in order to involve all the learners.
3. Students seem to learn better when they are engaged on activities involve more than one skill.

However, Byrne admits that also single-skill activities will sometimes be useful in the classroom and that there might be several occasions when they are more appropriate than integrated skill activities. Still, the teacher should look out for opportunities to knit skills together.

Project work is another great opportunity for integrating skills. Project work might be defined as an individual or group research over a period of time. It can result in a presentation or a piece of written work but does not necessarily have to result in a product. The focus should be on the process and not on the product. Ultimately, it should be something that is satisfying for the purpose of language acquisition but also purposeful and fun for the learners. Some skills (Byrne 1986: 133) that are required in group work are:

- communication skills (when interviewing and reporting back);
- research skills (when reading);
- social skills (when discussing, collaborating, meeting people etc.)

Of course, projects need to be carefully planned by the teacher, however, their setting up, as well as the guidance and monitoring that is needed in support, is largely a matter of common sense. Like in other activities, things might go wrong, but that should not be a source of frustration, as so many things might have been

learned. The role of the teacher is crucial (Byrne 1986: 133) “in identifying or helping to identify project areas; in providing adequate support and in motivating the learners”.

6.3.3 *THE TASK-BASED APPROACH*

In task-based language teaching the focus is on the use of authentic language and tasks which are meaningful to the learners and make them use the target language. The task-based approach has several obvious advantages which suggest it should be included in the teaching of speaking as one of the approaches which are available to the teacher. A tremendous advantage of the task where the essential content is clear is that the output by the learner is directly comparable to each other. Each speaker has to face the same task with equal difficulty. That is different in free tasks. One speaker might have an advantage because he or she has to talk about playing football, while the other one must talk about sky diving.

One example of eliciting talk from a pupil is to ask him or her to give a short talk on a particular topic. Anderson (1984: 44) explains some teaching examples:

In the tasks we use there is a controlled input and the speaker is asked to draw on that input to give information to a hearer who needs the information for a particular purpose. Simple examples of such tasks would be (i) where the speaker has a diagram in front of him and he has to instruct the hearer how to replicate the diagram and (ii) where the speaker has a series of photographs showing a traffic accident in front of him and has to report the accident to the hearer so that the hearer can fill out an insurance claim.

In tasks which follow the task-based approach, ideally, the teacher knows what needs to be mentioned in the talk to fulfil the task. However, there can be difficulties as well. For example, sometimes it is not clear to the students what a talk about the student's favourite activity needs to include. Such a lack of clearly defined content is a source of problems.

When it comes to diagnostic assessment the task-based approach has several further advantages. As the input can be analysed beforehand, an objective assessment procedure can be devised which lists the information which the speaker has to mention to complete the task successfully. Besides that, it is much easier to

assess other aspects of talk than information transferring ability, for example fluency or range of vocabulary.

All those advantages play an important role in the E8 standards. The tasks in the E8 standards mostly follow the task-based approach as it is easier to assess the talk if it is clear what the talk should ultimately include.

6.3.4 CLASSROOM ORGANISATION

In communicative language teaching and also in teaching speaking it is important to give the students time to practice their communicative abilities. One way of producing such opportunities is group work. For that purpose, it is important to have a closer look at what decisions need to be made when teachers organise their classroom.

According to Bygate (1987: 96), “practice in interaction skills requires the participants to negotiate meaning and manage the interaction jointly. This by definition implies certain kinds of activities.” Of course, for an interaction more than one person is needed. So, there are various ways of grouping learners in the classroom. Long and Porter (1985) claim that small-group activities allow more talk for each of the students. Another advantage of small-group work is that the students are not as often corrected as they would be in teacher-fronted situations. The students spend more time on negotiating and checking on meanings. In addition to that, Long and Porter (1985) found out that mistakes are no more frequent in small-group work than in teacher-fronted interaction. Further advantages (Brumfit 1984: 77) would be that in small group work each student is more involved in the conversation. That is one reason why the quality of language practice is higher in small groups. Furthermore, feedback and monitoring might also be more effective. When it comes to naturalness, a conversation in a small group seems more natural than a conversation held with the whole class. That also influences the stress factor as speaking in front of the whole class is often very intimidating to some students. Another issue would be that group work makes it easier to work individually with two or three students. That is, one could say, a basic requirement in the Austrian New Middle Schools as the teachers are confronted with students who are most of the time not on the same level, when it comes to language skills. Thinking about the most helpful classroom organisation in speaking activities beforehand might be a reasonable step here.

Many different possible sizes of groups exist. Some of them would be: dyads (pairs), triads, groupings of four, groupings of six, groupings of ten, half the class, any of the preceding performing before the rest of the class and the whole class without the teacher.

Brown and Yule (1983: 35) explain the disadvantages of performing alone in front of the whole class: Brown and Yule claim that it is not an easy task to tell the whole class about the last weekend,

[u]nless, he [the student] has been provided with very clear models of what is expected from such as task, this is going to be very difficult for the student. He has to extract from his mass of experience over the weekend some chunk, which can have some structure or meaning attributed to it, and in order to give an account of it, he has to imagine how much background knowledge of the circumstances is shared by the teacher and the other members of the class.

Of course, it can also be useful to combine various groupings for different stages of the same activity. Roberts (1981) states that debates can be difficult for speakers, especially if the topic is not clear to them. So, he suggests that some planning should be included before doing the task.

Finally, Bygate (1987: 100) claims that group work is not something that will solve all the problems. He sees it as a starting point, which should be used more often in the development of oral interaction skills. He states that “[t]here is plenty of scope for exploration in the use of group activities” (1987: 100).

6.4 THE ROLE OF DIALOGUES IN THE TEACHING OF SPEAKING

An important part of communicative language teaching is the use of real or authentic teaching materials and those which foster the development of oral communication skills seem to make up a crucial part of those real materials. However, Burns (1998: 105) states that "representations of authentic spoken interaction in teaching materials are generally hard to find, even in such ostensibly communicatively-oriented publications (Cathcart 1989, Slade and Gardner 1993), since the majority of currently available materials draw extensively on models of

grammar which are rooted in descriptions of written English (Carter and McCarthy 1995)". In addition to that it is not easy to transfer speech acts from one language to another, for example requests, apologies, expressions of gratitude, compliments, invitations, and complaints. Burns also mentions Slade (1986; 1990) who states that "most commercially produced materials rely on introspected versions of the (usually native speaker) writer's intuitions and thus 'deauthenticate' speech". This is also the case with scripted dialogues. Furthermore, it is underlined that often the underlying motivation for using speaking activities and also dialogues is not speaking or a communicative act as such but bringing across a certain grammar point. Burns cites Carthcart (1989: 105) who says that "[t]he dialogues in most current 'survival texts, even after years of so-called communicative language teaching, still tend to be thinly veiled excuses for the presentation of a grammar point". All that and the fact that most of the speaking activities in course books in Austria have the form of a dialogue makes clear that it is worth having a look at the advantages and disadvantages of this kind of speaking activity. In this section the advantages and disadvantages of using scripted dialogues, unscripted language samples, as well as semi-scripted texts will be discussed.

6.4.1 SCRIPTED DIALOGUES

What needs to be mentioned at the beginning is that the teacher has to be aware of the fact that scripted or semi-scripted dialogues might not prepare students to speak to native speakers outside the classroom. They are often an idealized version of interaction. However, if communication outside the classroom is the intended goal, the materials must be based on authentic spoken language (Burns, Joyce 1997: 82).

In the history of language teaching, speaking has often been taught with the help of dialogues. They are usually invented or scripted. One has to admit that scripted dialogues are much easier for students to deal with, especially for beginners. According to Burns and Joyce (1997: 84-85) the teacher has to be aware of the characteristics of scripted dialogues. They tend to be strongly influenced by the writer and his or her assumptions about what spoken interactions should be like. Spoken language is not something that is always the same and fully correct. However, that is what scripted dialogues suggest. They represent conversations as predictable and fully formed. Another problem with scripted dialogues would be that they tend to

make use of the grammar of written English and not spoken English. Furthermore, they often leave out the grammatical features which native speakers use. Dialects are also underrepresented in scripted dialogues as they mostly use standard pronunciation. Another point would be that most of the time scripted dialogues are used to present a special grammar point. That is a reason why they tend to repeat special grammatical structures with unnatural frequency and use full sentences instead of short utterances. When it comes to turn-taking features, scripted dialogues seem to be the wrong role models as there are hardly any interruptions or repetitions, besides that they give equal speaking time to the speakers which make the conversations unnaturally slow. Backchannelling, which is a crucial part of natural conversation, is almost completely ignored in those dialogues. Furthermore, as the dialogues are often embedded into a complete teaching sequenced the vocabulary is often reduced to a minimum. That is also not the case in natural conversations.

Scripted dialogues can be very useful for beginners but as their language development progresses they need to be introduced to authentic spoken interactions. Spoken language has got different functions than written language and those functions need to be discussed in class. Burns and Joyce (1997: 85) suggest that “[i]ntroducing learners to real as well as scripted samples of spoken language may help them to understand and prepare for unpredictable speaking situations outside the language classroom“.

6.4.2 UNSCRIPTED LANGUAGE SAMPLES

As stated above, in most English classes, the main goal is to prepare students to use and understand the language outside the classroom. If that is the main goal, the teacher needs to present students with authentic spoken texts in the classroom. However, the teacher needs to know how to use the materials effectively. He must be aware of the differences between authentic texts and scripted or semi-scripted materials. Typical features of unscripted texts are:

- fragmented utterances which are difficult to set out as sentences
- utterances which vary greatly in length
- varied grammatical structures, some of which are incomplete
- overlappings and interruptions rather than distinct turns

- informal and idiomatic language
- reference to shared knowledge and understandings of locations and processes
- implied context (Burns, Joyce 1997: 86)

If the students are only exposed to scripted texts in course books, it might be the case that they will not be able to participate in spoken interactions in social contexts, as they might develop an unrealistic view of the features of spoken language. The teacher must introduce authentic language gradually in the classroom. As authentic texts are more difficult, the tasks around them should be carefully controlled by the teacher (Burns, Joyce 1997: 86). Burns and Joyce (1997: 86) suggest,

[f]or example, beginning students could be asked to listen to a short fragment of authentic spoken language to identify the number of speakers and whether they are male or female. They could also practice features of authentic discourse such as contracted verb forms. Intermediate students could be asked to identify the stages of particular spoken genre after they have been introduced to the stages in another text.

Recordings and tapescripts can be used in a number of ways, in different settings and for different groups of learners. For example, they might be used to analyse features of spoken interaction, from overall text structure to specific discourse features and strategies. The most important use might be that they represent an opportunity to listen to native speakers in real conversations. However, the teacher needs to find out beforehand, if the transcript is too difficult and if the students have sufficient literacy skills to deal with spoken language which is written down.

6.4.3 *SEMI-SCRIPTED TEXTS*

Semi-scripted texts represent a mixture of both scripted as well as unscripted texts. They are a reaction to the problem that scripted texts generally present learners with unrealistic models of spoken interaction, and authentic texts are difficult to gather and present to students. Semi-scripted texts are created by “asking two or more people to perform a particular spoken language interaction which is recorded as they improvise” (Burns, Joyce 1997: 88). However, there are some facts which need to be

considered. When producing a semi-scripted text, one must set a context, identify a purpose for the interaction and identify aspects of authentic discourse, which should be included. Semi-scripted texts represent an opportunity to introduce students to the aspects of authentic speech in a controlled way.

6.5 ACTIVITY TYPES IN TEACHING SPEAKING

If the teacher wants to contribute to the communicative competence of the learners, he or she has to develop teaching-learning sequences which will foster the students' ability to use spoken language. Some decisions the teacher needs to make are (Burns 1997: 82):

- what spoken texts they will present to the students
- what activities they will use or design
- what roles the teacher and the students will play in the teaching-learning processes of the classroom.

In order to reach speaking objectives in teaching different activities and methods can be used. However, classroom activities should mostly focus on completing tasks that are mediated through language or involve negotiation of information and information sharing. Johnson (1982: 151) explains that

[t]hese attempts take many forms. Wright (1976) achieves it by showing out-of-focus slides which the students attempt to identify. Byrne (1978) provides incomplete plans and diagrams which students have to complete by asking for information. Allwright (1977) places a screen between students and gets on to place objects in a certain pattern: this pattern is then communicated to students behind the screen. Geddes and Sturtridge (1979) develop "jig-saw" listening in which students listen to different taped materials and then communicate their content to others in the class. Most of these techniques operate by providing information to some and withholding it from others.

A crucial distinction is the distinction made by Littlewood (1981) who differentiates between "functional communication activities" and "social interaction activities". On the one hand, functional communication activities would include such tasks as discovering missing features in a map or working out a likely sequence of events in a set of pictures. Social interaction activities, on the other hand, include conversation

and discussion sessions, dialogues and role plays.

There are a variety of sources from which to gather speaking activities. Course books provide ready-to-use activities but also the teachers can design their own activities. Whatever the source is, one should keep in mind that the activities should always aim at a particular goal. Most of the time that goal is to help students develop skills in using spoken language. What needs to be kept in mind when designing or choosing an activity is not only the goal but it is also important that students, where possible, understand the aim of the activities before taking part in any of them. According to Burns and Joyce (1997: 93), “[t]his will help them to develop awareness of the learning process and give them a sense of where the lesson is going. It also assists learners to take responsibility for their own learning”. In this section some activities or activity frameworks will be introduced, firstly, a framework suggested by Burns and Joyce (1997) and, secondly, two activities from the E8 standards.

6.5.1 THE FRAMEWORK BY BURNS AND JOYCE

A possible framework for speaking activities is the framework by Burns and Joyce (1997: 92). It is based on scaffolding for categorizing the activities, which is a way of systematically sequencing activities which move students towards independence. The five categories would be (Burns, Joyce 1997: 92):

- preparation activities
- activities which focus on language awareness and skills
- activities which focus on discourse awareness and skills
- interaction activities
- extension activities

The first part of a speaking activity would be the preparation phase. The purpose of preparation activities is to make students aware of the type of spoken interaction they will be producing and the context which is needed. These activities need to draw on student experience, elicit language, develop motivation and encourage students to participate in spoken interaction. They should represent a help for the students. According to Burns and Joyce (1997: 93) the time spent on such an activity depends on a number of factors, such “the level of the students, the class profile, class size, the type of interaction” and so on.

The next step would be to expose the students to tasks which focus on language awareness and skills. The task is again preparation for the activity and should focus on the necessary skills and the language necessary for effective communication, which would include vocabulary and control of grammatical structures to produce spoken language.

After that, discourse awareness and skills should be brought along with the help of a task. The students should be made familiar with knowledge about how participants build a spoken text together when they are speaking. In other words they must know how to begin a spoken interaction, how to maintain it and how to end it. Furthermore, the students must realize that there are certain rules and strategies which govern the way discourse is developed.

All the activities explained so far aim at preparing the students for the activity. Now the students should get the opportunity to practice these spoken interactions inside the classroom. However, it would be particularly useful if the students get the chance to practice the conversations in real life outside the classroom.

In order to stabilize the new knowledge and skills, they need to be recycled throughout the language programme. Furthermore, in order to become confident, the new skills and knowledge need to be extended. According to Burns and Joyce (1997: 96-97) extension can mean:

- introducing similar spoken interactions in different contexts
- increasing the complexity of the spoken interaction through:
 - o increasing the length
 - o increasing the number of interactants
 - o introducing unpredictable or problematic elements into the discourse
 - o increasing the number of outcomes which have to be achieved through the discourse (eg having to post a number of different items going to different destinations)

6.5.2 ACTIVITIES FROM THE E8 STANDARDS

First of all it is worth discussing what the E8 standards are and what purpose they serve. The BMUKK wanted to conduct standardised tests for German, English and Mathematics. Those tests should focus on the competences which the students should have acquired up to the 8th grade in the Austrian school system. The basis for

those tests is the CEFR. Primarily, the tests should give feedback to the students, the teachers, and the Austrian government. The E8 standards should show the teachers and the government areas of improvement and, furthermore, give the students a chance to find out about their strengths and weaknesses. Additionally, it should make comparisons between students from all over Europe easier. A tremendous amount of practice material was published to make it easier for the students to get used to the format of the test. Those practice materials are not only a great source of ready-made activities but also a great source of inspiration, as the tasks can easily be altered, in respect to level of competence and also topic.

The first example, which will be analysed, works with the help of pictures (bifie 2012: 3):

Speaking Task E6/E7: Monologue My favourite animal



Think about the topic for a minute. Then try to talk for about 2 minutes.

Tell me as much as you can about **your favourite animal**:

- your favourite animal
- what it looks like
- why it is your favourite animal
- what you can do with it

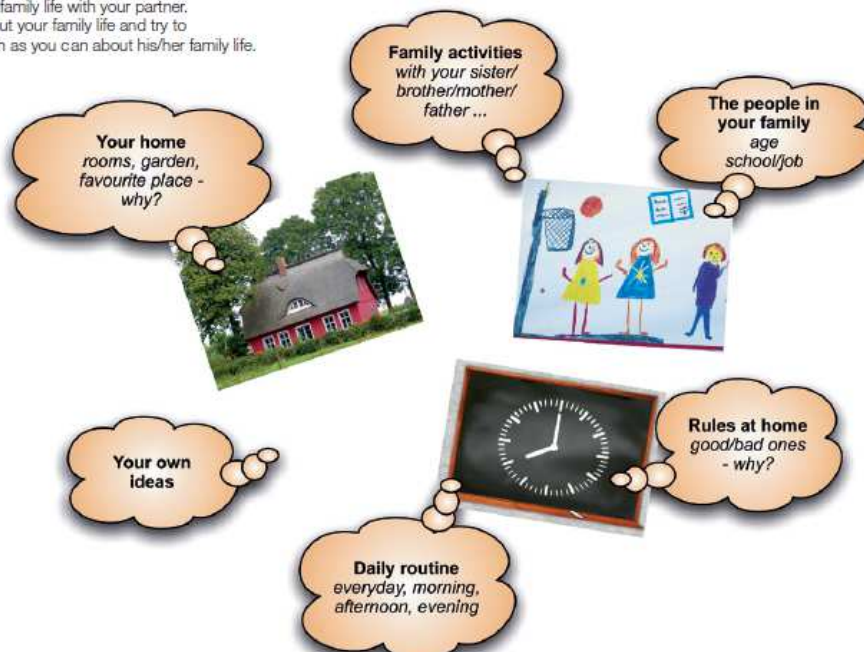
In this example, it is easily visible that the task is completely open for change. If the teacher wants the pupils to talk about hobbies or sports, then only little changes make up a completely different task. Additionally, the time frame might give the teachers a good orientation.

The next activity is designed for two speakers (bifie 2012: 5):

Interactive Speaking Task E6/E7 Family life

Try to talk with your partner for about 3 minutes.

Talk about your family life with your partner.
Tell him/her about your family life and try to
find out as much as you can about his/her family life.



This example makes clear that little help, through some key words might make it much easier for the students to talk. Furthermore, not much work needs to be done to design a new task, only new picture and keywords are needed.

7 SUMMATIVE ASSESSMENT OF SPEAKING

So far, some basic concepts in teaching speaking have been explained, such as communicative language teaching, concepts in speech production, the skill of speaking and some aspects in the teaching of speaking. Having established those implications for the teaching of speaking in the language classroom, the possibility to assess these learning processes seems to be equally important. As stated earlier,

the teachers in the New Middle Schools in Austria have to include speaking in the final grade. This section should help to find and examine the possibilities which are at the teachers' disposal. For that purpose, some basic terms in test evaluation and some factors affecting test scores should be explained. The history of second language speaking tests will form an entry to the topic. As rating scales are of tremendous importance in schools and official tests, they should be explained in greater detail. Afterwards different rating scales, namely the E8 standards and the performance decision trees, will be discussed. The last point should summarise the findings and give practical input. It will propose a suggestion for including speaking in the final grade in the first form in a New Middle School in Austria.

7.1 TEST EVALUATION

Before one can understand the different kinds of rating scales, some basic terms in test evaluation need to be explained. The process of establishing the general validity of a test procedure is called validation. It depends not only on the situation in which the test is used but also on the test itself. Validation is not an absolute process. It is a relative one as the degree of validity of a test relates only to the particular circumstances in which it was established.

Face validity wants to find answers to the following question: On the face of it, does it look like a reasonable test? In order to research this form of validity one can question the different people who come into contact with the test. Learners generally produce very informative and objective comments about tests, irrespective of their own personal performance. If one has any doubts this can be easily checked by comparing the general tone of the comments with the scores (Underhill 1987: 105-106).

Construct validity wants to find out if the test matches the theory behind it. As almost all tests are part of a programme and every programme usually makes some basic assumptions about the purposes and processes of language learning, a test should obviously share the same assumptions and the same philosophy as the programme. Underhill (1987: 706) claims that “[f]or example, if the teaching programme aims to give learners a limited competence in particular professional areas, using a lot of exposure to authentic language and documents, then the test procedure should follow the same basic approach”. According to Underhill (1987:

106) construct validity is not a straightforward idea to work with and most of the time it has no real value outside language testing research. Furthermore, there may not be a crucial difference between construct and content validity. Underhill (1987: 106) summarises the concept of construct validity in one single question: "Does the test match your views on language learning?".

This means that, in this case, in testing speaking, we need to find out what needs to be tested when we want to test speaking. So, again it is important to be clear about what the skill of speaking really includes. Hopefully, the teacher is of the opinion that speaking is a skill. If that is the case, the test should find out if the speaker is able to use the language communicatively. The test should not test the knowledge about the language but the ability to use it. Of course, further categories and subsections need to be formed but for that purpose, it is crucial that the teacher has a picture of what the skill of speaking includes, in order to test the student's progress.

7.2 THE HISTORY OF SECOND LANGUAGES SPEAKING TESTS

In order to understand the current situation of speaking tests, it is important to know about the history and the origins of those tests. For that purpose, this section should give a short overview of the most crucial tendencies and changes in the area of oral speaking tests.

The first point which needs to be made is that the ACTFL (American Council on the Teaching of Foreign Languages) and its ACTFL Oral Proficiency Interview strongly influenced foreign and second language instruction and assessment (Turner 1998: 192). The guidelines of that association see the "students' abilities to use language in performing particular functions and tasks" as crucial and not what they know about the language (Turner 1998: 192). That is one reason why investigations of different features of oral proficiency interviews were of growing interest, including "their reliability and construct validity, their comparability to other test formats, and their concurrent validity" (Turner 1998: 192).

Turner goes on in reconstructing the history of that tendency. She states that in 1988 a special issue of *Studies in Second Language Acquisition* was published,

which included the main issues and agendas at the time. Much criticism and many problems are addressed in that issue, for example Lantolf and Frawley (1988: 191) state that "the one place where we cannot properly study what people do with or through language is the oral interview", as it is, unlike many non-test interactions, being directed by the interviewer. Consequently, what needs to be analysed are the examinee, as well as, the interviewer in oral proficiency tests. Furthermore, there is a difference between interview and conversation. Most speaking proficiency tests are interviews, but that is not as close to real conversations as they should be. Lazaraton (1992) states that "[t]he most that can be said about the question 'interview or conversation?' is that the encounters share features with conversations, but they are still characteristically instances of interviews".

One thing that happens in interviews is that the interlocutor supports the test-taker in many ways. As in real life conversations between a native speaker and a non-native speaker, there are instances of priming topics and slowed speech.

Although the ultimate goal of language proficiency tests is to test the ability of the test-taker, it is known that various factors play an important role, for example the features of the test, the environment, and the participants. The features of a test also include the planning time. Wigglesworth (1997) investigated the impact of planning time on test performance using a tape-mediated test called ACCESS (Australian assessment of communicative English skills). In two sections of the test there was planning time, while in the other two there was none. Turner (1998: 197) states that "[a]lthough there were no significant differences in the scores of the performances with planning time and without planning time, Wigglesworth reported that planning time seemed to interact with examinee proficiency level and task difficulty." It seemed that only higher ability examinees benefited from planning time on some tasks, while it made no difference with the examinees at lower levels of ability. A further factor which could influence test performance is the number of interlocutors. Fulcher (1996) did a study where forty-seven examinees completed three tasks. Fulcher found out that most of the examinees preferred the group task because of various reasons, for example the examinees stated that it seemed like a more natural conversation and it provoked less anxiety.

Turner (1998: 199) makes a final statement about the current state of oral proficiency test. She states that

[t]he investigations to date into the nature of examinee, interviewer, and rater performance and variables that affect their performance are revealing but inconclusive. They offer new perspectives on test validity, further evidence of the variability of test performance and test scores. Important questions regarding who should be involved in the development of rating scales have been posed, and general suggestions for more effective approaches to scale development have been made.

7.2.1 RATING SCALE DEVELOPMENT

Although Turner states that rating scales are generally accepted as methods to train raters and to establish reliability and validity, it is still not clear how those scales should be created. One big point of criticism is the fact that the ACTFL guidelines and the ACTFL-OPI lack of an empirical basis for the descriptors that form a kind of foundation. Turner cites Fulcher (1996: 228) who states that "[u]ntil test researchers and developers take seriously the validity of tests at the development phase rather than a post hoc notion, the problem of the indeterminacy of validation studies and the interpretability of test scores will remain serious". One way of solving that problem would be that "research into the empirical derivation of rating scale criteria should be carried out into the description and operationalization of constructs for language testing, reinforcing the necessary the link between applied linguistics, second language acquisition research and language testing theory and practice" (Fulcher 1996: 228).

An empirical approach for building rating scales is proposed by Upshur and Turner (1995), that approach should be suitable for the use in the classroom. Existing proficiency scales do have some limitations in that sector, for example they tend to have rather broad, imprecise descriptors, which consequently do not reflect teachers' more narrowly defined aims closely enough to provide meaningful measurement of student learning. Furthermore, they are based on descriptors of features that may not co-occur in actual student performance. The approach by Upshur and Turner (Turner 1998: 201) should work as follows:

[I]nstead of attempting to match test performance to verbal descriptors from scales, the scale is derived from a hierarchical set of binary questions formed by teachers using a subset of examinee performances. These scales are developed for particular tasks and

are intended to show how well examinees perform only on these types of tasks.

For example, if a teacher wants to test if a student can manage to post a letter at the post office. The approach by Usphur and Turner would pose the following questions: Did the student start the conversation with an acceptable greeting? Did he or she clearly state that she wants to post a letter? Did she ask in a polite way? Did she end the conversation in an acceptable way?

Those questions show that the rating scale would only work in this specific task and in no other one. It is much easier to handle as it consists of yes/no questions which are answered by the examiner in a very fast way.

Those kinds of scales should assess students' learning more accurately than the less focused proficiency scales. Usphur and Turner called those scales EBB scales (empirically-derived, binary-choice, boundary-definition scales). Turner (1998: 201) ends in pointing at the advantages such EBB scales bring along:

The authors claimed that an EBB-scale approach addresses the feasibility, reliability, and validity limitations of existing proficiency scales for use in specific educational contexts by describing boundaries that are concrete, simple, and precise.

Another crucial difference is that the scales are derived empirically through a consensus among teachers and not through scaling proficiency levels. So, it might be helpful for language teachers, also in Austria, to think about creating such EBBs for their speaking tasks or exams as they might be a tremendous help in making the scoring more feasible to teachers, the parents and, of course, the students.

Turner also mentions Haggstom (1994) who described another solution to measuring students' learning. The assumption was that when we teach communicative competence, the tests should also include communicative tasks. Haggstom (1994) also noted that "tests should allow measurement of students' progress and mastery of specific material as well as being accurately scoreable and feasible in terms of administration and grading time" (qtd. in Turner 1998: 203). Haggstom's testing method included the video-taping of students' performances in communicative activities that were typical of her classroom activities. An advantage of that method is that on-going frequent assessment was possible because the

testing was integrated with classroom activities. The method worked as following: "In this 50-minute test procedure, the teacher moved around the room with a camcorder and videotaped each student on three occasions as he or she participated in a small group activity. Two example activities were described and criteria for selecting tasks and scoring responses were provided" (Turner 1998: 203). The drawback might be that this method is not easily feasible or appropriate in all classrooms, but the description offers useful insights into integrating instruction and assessment and designing tests for specific learning environments (Turner 1998: 203).

7.3 RATING SCALES

Underhill presents a very traditional description of a rating scale. To him (1987: 99), the level aspect is of great importance. Each level should describe as briefly as possible what the "typical" learner should be able to do on each of the levels. So, it should be easy for the assessor to decide what level he or she should assign the student to. To conclude, the use of a rating scale should be to offer a series of prepared descriptions to the assessor from which he or she can pick the description which fits best. What is interesting here is that Underhill calls the learner "typical". This can be seen as a typical disadvantage of those kinds of rating scales as it is not clear what a "typical" learner is. Furthermore, one can even doubt if there is something like a typical learner.

Fulcher gives another definition of what a rating scale is (Fulcher 2003: 88-89). Fulcher's definition is quite close to the one presented by Underhill. Fulcher sees a rating scale as a "scale for the description of language proficiency consisting of a series of constructed levels against which a language learner's performance is judged. The scales range from zero to an end-point. That end-point should represent "the well-educated native speaker", which is again a point of criticism. It is the same problem as in Underhill's explanation. What is "the well-educated native speaker"? What Fulcher mentions in contrast to Underhill is that the descriptors do not work without each other. Furthermore, he states that "[s]cales are descriptions of groups of typically occurring behaviours; they are not in themselves test instruments and need to be used in conjunction with tests appropriate to the population and test purpose". Fulcher points out that raters or judges need to be trained beforehand, otherwise the scores will not be comparable or reliable.

Besides using rating scales for scoring speech samples, and to guide test developers in the selection of tasks, there are also other uses for scales (Alderson 1991, qtd. in Fulcher 2003: 89):

- User-Oriented Scales: used to report information about typical or likely behaviours of a test taker at a given level.
- Assessor-Oriented Scales: designed to guide the rating process, focusing on the quality of the performance expected.
- Constructor-Oriented Scales: produced to help the test constructor select tasks for inclusion in the test

In those different types of scales, different details might be thought about. In User-Oriented Scales it might be necessary to phrase the band descriptors (level descriptors), in Assessor-Oriented Scales construct definition might be important "but it may need to be expressed in a way that can be processed in the limited time available to award a rating in a face-to-face speaking test" (Fulcher 2003: 89). The Constructor-Oriented scale may be the most detailed one, it can contain references to the types of task that are most likely to elicit the language sample required for the scores to be meaningful.

A main distinction in the types of rating scales used for scoring speech samples is the distinction between holistic and analytic rating scales. Holistic assessment is a procedure which does not count features. Analytic assessment is the opposite. In an analytic assessment the examiner counts incidents. Holistic scoring can be divided into "holistic scoring", "primary-trait scoring" and "multiple-trait scoring". Fulcher (2003: 90) defines the three constructs as following:

Holistic scoring: A single score is given to each speech sample either impressionistically, or guided by a rating scale. This single score is designed to encapsulate all the features of the sample, representing "overall quality". [...]

Primary-trait scoring: This approach assumes that one can only judge a speech sample in its context, and so rating criteria should be developed for each individual task. [...]

Multiple-trait scoring: Providing multiple scores for each speech sample, with each score representative of some features of the performance, or construct underlying the performance. [...] (Alderson, 1981, qtd. in Fulcher 2003: 90)

The type of scale which an examiner uses may depend on the purpose of the test. Fulcher gives two examples. He gives the example of a test of speaking for hotel receptionists. Here the test developer may choose to use primary trait scoring with "real-world" descriptions in the rating scale. So, the scoring would have a direct link to the very specific functions of a hotel receptionist. A different kind of scale would be chosen for assessing a test in English for Academic purposes, namely a multiple-trait scale. It would be more useful because it is necessary to generalise the "ability to study in the second language", but also because diagnostic feedback might be important and useful for the test-taker (Fulcher 2003: 91). Fulcher summarises his findings about the description of rating scales in the following framework (2003: 91):

Orientation:

- User
- Assessor
- Constructor

Scoring:

- Analytic Approach
- Holistic Approach
 - Holistic Scoring
 - Primary-trait scoring
 - Multiple-trait scoring

Focus:

- Real World
- Construct

(Fulcher 2003: 91)

There are two basic approaches on rating-scale design. One would be to use "intuitive" methods, the other one would be to design upon some kind of empirical data. Those two approaches can be divided into three-subcategories, which are explained by Fulcher as following (2003: 92):

- Intuitive methods

- Expert judgments. An experienced teacher or language tester writes a rating scale in relation to existing rating scales, a teaching syllabus, or a needs analysis. [...]

- Committee. As for expert judgments, but with a small group of experts who discuss and agree on the wording of the descriptors and the levels of the scale.

- Experiential. Perhaps starting with expert judgment or committee design, the rating scale evolves and is refined by those who use it, so that over a period of time the users intuitively "understand" the meaning of the levels in relation to sample performances. [...]

- Empirical methods

- Data-based or data-driven scale development. This approach requires the analysis of performance on tasks and the description of key features of performance that can be observed to make inferences to the construct.

- Empirically derived, binary-choice, boundary definition scales. Expert judges are asked to take speech or writing samples and divide them into better or poorer performances. The reason for the categorisation is recorded, and used to write a sequence of yes/no questions that lead the rater to the score.

- Scaling descriptors. In this approach many band descriptors are collected in isolation from a scale, and experts are asked to rank them in order of "difficulty". They are then sequenced to create the scale.

Fulcher explains some widely spread rating scales in greater detail. The first scale which is analysed is the FSI (Foreign Service Institute) family tradition, an intuitive and experiential scale development. Wilds (1975: 35) states that the FSI is

very much an in-house system which depends heavily on having all interviewers under one roof [...] It [the system] is most apt to break down when interviewers are isolated by spending long periods away from home base, by testing in a language no one else knows, or by testing so infrequently or so independently that they evolve their own system.

Many other rating scales evolved from the FSI scale, namely the ILR (Interagency Language Roundtable - standard grading scale for language proficiency in the

Federal service- formally the FSI) and the ACTFL. The FSI scale is assessor- and user-oriented, and it has been widely used in the development of curriculum as well as in assessment contexts. The scores are holistic, and they have a real world focus. According to Adams and Frith (1979: vi, qtd. in Fulcher 2003: 93) the FSI scale descriptors cannot be assessed against any external criteria. Each level within a scale is defined in relation to the other levels. The only key reference point is "the ultimate standard, the ultimate criterion reference, is the proficiency of the educated native speaker" (Fulcher 2003: 93). That was also the starting point for the designing of the scale. However, Perren (1968: 11, qtd. in Fulcher 2003: 93) criticized that the definition was not clear enough, he asked "What kind of native, speaking about what, and to whom?". Despite all that criticism Wilds (1975: 36, qtd. in Fulcher 2003: 93) argued that the linguistic ability of a "well educated native speaker" was the "absolute standard" upon which the FSI rested. Through the principle of internal consistency the other levels were defined. One advantage of the test is again that it judges the students' production in relation to "language as it is spoken by well-educated native-speakers". However, there are also some drawbacks. The use of the concept of the educated native speaker for scale development has increasingly come under attack. The most crucial problem is that the ability of an educated native speaker is not stable, it shows considerable variation. It is argued (Fulcher 2003: 94) that "the ACTFL scale definitions are firmly rooted in the misconception that we can clearly identify native speakers and their standard of language performance". The scale still contains "the notion of a monolithic group of native speakers", which can be seen in phrases from the band descriptors such as "can be understood by native interlocutors", and "using native-like discourse". They only deal with "THE" native speaker, whereas in reality only types of native speakers exist. Lantolf and Frawley (1988) identify four types:

1. Idiolectal, or informants.
2. Statistical, or typical speakers.
3. Normative, or expert speakers.
4. Former, or speakers from historical records.

One could say that expert speakers, are also only experts in certain areas. So, one could conclude that the test developers are right when talking about "the ideal native speaker" and not about "the real native speaker". Studies (Davies 1990) have demonstrated that the concept of the "native speaker" is one which is not consistent, and no researchers have sufficiently defined the term to make it useful in a testing context.

A common problem faced by the scale designers is to develop a testing procedure that generates sufficient evidence to be scored, and a rating scale that describes the constructs to be measured (Fulcher 2003: 95). Furthermore, Jones (1981) goes on in stating that most oral testing procedures do not relate the elicitation technique to the scoring systems in any specific way.

The intuitive approach, as the name already suggests, has led to vagueness and generality in the descriptors used to define bands. Fulcher goes on in listing the disadvantages, which clearly outnumber the advantages, he states that "the disadvantages of general, vague terminology are lack of clarity and the possibly 'meaninglessness' of the band for users, assessors or test constructors. Further, if the sequence of band descriptors is meant to reflect progression in second language acquisition it is important that the descriptors clearly reflect theory if validity studies are to address score meaning" (Fulcher 2003: 96).

Vague terms like "good" or "fluent" do not make the situation any better, on the contrary, Matthews (1990: 119) notes, that bands "are described in only vague and general terms and abound in qualifiers, so that only gross distinctions can be made with any confidence". Of course, all this vagueness does not contribute to the fairness of an oral proficiency test. Hieke (1985: 137) states that tests cannot be fair "as long as they [the descriptors] hinge upon prose statements to delineate levels while these are peppered with notions that cannot withstand close scrutiny".

However, terms which seem vague to any other person, may be meaningful for experienced evaluators who have been trained and socialised in the use of the scale. So, according to Fulcher (2003: 97) again the issue of experience is provided as the most important reason for rating scales appearing to be meaningful and providing reliable results. Emphasis must be put on rater training and socialisation

over time. However, the problems still exist when the scales are separated from the training which raters must receive in order to become certified raters.

In contrast to the intuitive approach, a scale can also be developed with the help of empirical data. In that kind of approach the descriptors must be based on observed learner behaviour as opposed to postulated or normative notions of how learners ought to behave. There is a close relationship between linguistic behaviour, the task and the scoring procedure. That relationship should be transparent. The scales in this approach are typically assessor-oriented, require holistic or multiple trait scoring, and have a construct focus.

Now to the analysis of speech samples, or the coding of speech. Fulcher uses the example of fluency. Speech samples were recorded from a range of speaking tests and transcribed for analysis, then the transcriptions were coded for features of fluency. The six speech phenomena that were investigated were:

1. Fillers such as 'er(m)'
2. The repetition of the first syllable of a word or a full word.
3. The negotiation of reference indicated by the re-selection of referring devices.
4. The re-selection of lexical items.
5. Anacolouthon (not expected shift in grammar).
6. Longer pauses, indicated in the transcripts and examples as two or three colons. (Fulcher 2003: 99)

However, there was still a problem with counting only pauses and repetitions and seeing those instances as a reduction of fluency, as the number of pauses did not automatically translate into a perception of reduced fluency. It was necessary to introduce the explanatory categories that would attempt to take account of why a pause occurred, or why there was word repetition. However, Fulcher (2003: 99) names the disadvantages of that development. He states that "firstly, there is likely to be no one-to-one relationship between a speech phenomenon (like pausing) and the explanatory category (the suggested reason for the occurrence of the speech phenomenon). Secondly, the hermeneutics of classifying a speech phenomenon into an explanatory category is likely to require a high degree of inference".

Another drawback would be that the focus is not on the "why" of the use of a phenomenon, but just on its existence. So, the including of explanatory categories, into which individual observations of speech phenomena can be coded, is essential. Furthermore, a methodology like this requires a level of "interdeterminacy", as there is no one-to-one mapping between speech phenomena and the explanation for their occurrence. So, methods need to be employed that control the quality of data coding and interpretation. Those could be double or (even triple) blind coding and discriminant analysis. In the first method, the double (triple) blind coding, two or more individuals use the same transcripts to code surface phenomena into the explanatory categories. Finally, the degree of agreement is a measure of the reliability of the coding. In contrast to that, the second method requires all the speech samples to be drawn from speaking tests in which the test takers received scores. Fulcher (2003: 103) explains that "[e]ach of the test takers therefore 'belongs' in a category, such as 'band 4' or 'band 5'. Once the coding into categories has been completed, discriminant analysis can be used to ask the question: can we predict the score a test taker received only from the coding of speech phenomena into explanatory categories?"

A further approach is being investigated namely, the Empirically derived, binary-choice, boundary definition scales (EBBs). This approach was named and developed by Usphur and Turner (1995). It is assessor-oriented, uses primary-trait scoring, and has a real-world focus. It tries to discover what features distinguish between responses in different piles, and then placing them on a scale. The EBB scale development relies on expert judgement. The aim of the EBB is not to make any assumptions about a theoretical, linear, process of second language acquisition. The EBBs rely upon how sample performances are sequenced, and how these can be scored by asking raters to make a series of binary (yes/no) choices about features of performance that define the boundaries between score levels.

The following section describes the design procedure for 6-level EBB areas (Usphur and Turner, 1995, qtd. in Fulcher 2003: 105):

Step 1. Select performances to be rated, so that the range of performances covers the ability range that is to be tested.

Step 2. The individuals from a team of experts divide the performances into equal numbers of 'better' and 'poorer' performances impressionistically.

Step 3. The team of experts discuss why they placed the samples into the two piles, reconciling any differences they may have. As a team they are asked to write a single question, the answer to which would result in a sample being placed in the 'better' or 'poorer' team

Step 4. The individual members of the team rank order the 'better' samples and score them as '4', '5', and '6' impressionistically.

Step 5. The team of experts discuss their rankings and reconcile any differences. The team then write critical questions to distinguish level 6 performances from level 4 and 5 performances, and then level 5 performances from level 4 performances.

Step 6. Steps 4 and 5 are repeated for the 'poorer' performances.

Usphur and Turner (1995: 10, qtd. in Fulcher 2003:106) state that the most crucial difference between traditional scales and EBBs is that instead of having a descriptor that attempts to define the 'midpoint' of a band, the questions on an EBB

describe the 'boundaries' between categories. Ratings are based upon the perception of differences rather than similarities. In this way the EBB scales are more like familiar measuring instruments.

Another advantage is the simplicity of this approach because they rely on only one judgment in answer to a critical question, it is claimed that it reduces the problem in other scales of having co-occurring features in the band descriptor. The decision is made easier. Fulcher (2003: 106-7) names some more advantages:

The EBB process clearly has a number of advantages. The first is that it is simple, and can be used to generate rating scales by teachers for locally produced speaking and writing tests. The second is that it is easy to use, and probably results in an increase in reliable scoring when used by teachers who have been on the team that developed the rating scale. Thirdly, as a primary trait scale, each scale is linked to a specific task. In a pedagogic context, especially if teachers are using a task-based methodology, these rating scales can provide a rich source of information on student progress.

However, the drawback is that it is only specific to one special task, so it cannot be used on another one. That is one reason why we might not see that kind of approach provided by the large testing agencies. However, it is still a great tool for primary trait scoring.

7.3.1 *ADVANTAGES AND DISADVANTAGES*

Underhill presents some of the disadvantages of traditional rating scales. One problem is that rating scale descriptors see the learners as typical learners and it is a fact that not all learners are typical. In contrast to length, language cannot be measured on some kind of ruler as language ability is something very complex and combines a number of different skills and factors. The second major problem would be that one needs to find out how detailed the profile for each learner should be. Underhill (1987: 99) claims that “[t]he more information you give, the easier it will be for an assessor to find something that seems to match the learner sitting in front of her. At the same time, the more detail at each level, the more likely it is that some of it will be contradictory or that statements in different categories will seem to place a learner at different levels”.

Underhill (1987: 99) claims that the most important point is to maintain a continuity between the descriptors at different levels. However, that is what Fulcher, for example, questions. He thinks that language development is not a linear development and should therefore not be tested with the help of a linear scale. According to Underhill (1987: 100) vague expressions, such as occasional errors, confidently and accurately, most, often and sometimes are not a major problem. The only requirement is that they need to be directly related to the learner’s performance and not to rating scale descriptions at other levels. If for example a descriptor says “Makes fewer errors but is only slightly more fluent”, this only makes sense in connection to other descriptors and that would lead to a rating scale with circular, self-defining descriptions. It is also important to use only the number of levels one needs and not more than that. If you want to categorize learners into elementary, intermediate and advanced, three levels are enough. It is important to keep the rating scale as simple as possible. The simpler the test the easier it is to assess, and the higher the reliability will be. According to Underhill (1987: 100), “[a] rating scale will only work well if the assessor can hold it in her mind while listening or talking to the learner, and does not have to keep referring to a large manual to tell her what to look for”. Furthermore, when using a rating scale it is important to look for a range not a point on a scale. Occasional flashes of brilliance which are not sustained should be ignored, as should a single serious error; only consistent patterns of strength and

weakness should be compared against the rating scale to produce an assessment (Underhill 1987: 100).

7.3.2 TEST ANXIETY

Test anxiety is a crucial factor, which definitely influences the outcomes and test validity of oral exams. In this section, the factor of test anxiety and its influence on test outcomes should be explained shortly. A study by Huang and Hung (2013) will be used to explain the concepts. In the study a total of 352 students studying English as a foreign language took two independent tasks for which they spoke without input support, performed two integrated tasks for which they orally summarized the reading and listening input. Path analyses of the data show that test anxiety strongly influenced integrated performance, test anxiety affected independent performance and integrated performance in a statistically equivalent way, and that there were no effects dependent on the topic. One could conclude from those findings that

the advantage of integrated tasks over independent tasks might not relate to the reduction of test anxiety or to its impact on test performance and that integrated tasks suffer the construct validity threat posed by test anxiety as much as independent tasks (Huang and Hung 2013: 244).

7.4 SPECIFIC ASSESSMENT FRAMEWORKS

In this section, two extremely different assessment frameworks will be explained, on the one hand, the more traditional E8 standards and, on the other hand, the more experimental performance decision trees. Advantages and Disadvantages of both testing tools should be examined, in order to provide teachers inspiration and ideas for designing their own testing tool.

7.4.1 E8 STANDARDS

The testing of speaking for the E8 standards is based on the concept of communicative competence. The goal is to test if the test taker is able to communicate successfully. Moreover, attention should be paid to the fact that spoken

English is not verbalized written English. All that has its roots in the Austrian National Curriculum for Foreign Languages:

Ziel des Fremdsprachunterrichts ist die Entwicklung der kommunikativen Kompetenz in den Fertigkeitsbereichen Hören, Lesen, An Gesprächen teilnehmen, Zusammenhängend Sprechen und Schreiben.

Als übergeordnetes Lernziel in allen Fertigkeitsbereichen ist stets die Fähigkeit zur erfolgreichen Kommunikation – die nicht mit fehlerfreier Kommunikation zu verwechseln ist – anzustreben (bmukk 2009: 1–2).

The emphasis should be put on successful communication and not on completely correct communication. Fluency, and not accuracy, should be the goal. So, the teachers should adopt the role of a guide. They should leave correcting behind and support and encourage speech processes instead. Furthermore, there is an obligation to assess spoken interaction and oral production regularly and reliably.

Die Fertigkeitsbereiche Hören, Lesen, An Gesprächen teilnehmen, zusammenhängend Sprechen und Schreiben sind in annähernd gleichem Ausmaß regelmäßig und möglichst integrativ zu erarbeiten und zu üben. (bmukk 2009: 2)

Da aber die Erfassung der mündlichen Kompetenzen in der Gesamtbeurteilung vom Lehrplan im Sinne der Gleichwertigkeit der Fertigkeiten explizit gefordert wird, muss ein GERS-orientierter Unterricht mündliche Prüfungs- und Übungsformen beinhalten, die sowohl monologische als auch dialogische Sprechkompetenzen verlässlich abbilden. (Brock et al. 2008: 12)

The need for a testing system like E8 standards is underlined. It is mentioned that in Austrian schools testing speaking “hardly ever happens in a systematic way” and that the students’ speaking ability hardly ever influences their final grade in English (Mewald et al. 2012: 5).

The E8 standards should describe what the speakers should be able to do in spoken interaction and oral production. That is done in can-do descriptors and CEFR levels. The basis of this rating scale lies in the works about communicative competence, for example Hymes (1972), Brumfit & Johnson (1979), Widdowson

(1978), Canale & Swain (1980). However, the CEFR is the most crucial influence as the E8 standards are strongly oriented towards the CEFR levels.

As stated above, the E8 standards are strongly influenced by the CEFR, which sees linguistic competences, sociolinguistic competences and pragmatic competences as the most prominent elements of communicative competence. Those three sections are divided into further subsections. However, the E8 does not use all of the categories. In linguistic competence the E8 standards include lexical, grammatical and phonological competence. When it comes to sociolinguistic competence the E8 standards focus on the linguistic aspects of sociolinguistic competence, it is mostly restricted to “linguistic markers of social relations” and “politeness conventions” (Mewald et al. 2012: 10). The pragmatic competences include design competence, functional competence and discourse competence.

It is also essential that authentic language takes place in a limited time span and so it is mostly unplanned. This should also be the case in the E8 standards. Although the test takers in the exam are given a short time to prepare, the performances cannot be called planned. Both vocabulary and grammar can be limited in their range because of the factors named above. This should be acknowledged in the E8 standards (Mewald et al. 2012: 12).

The descriptors of the analytic assessment scale have been linked with the construct to report about the test takers' abilities in four dimensions. The Assessment scale is used during the speakers' performance in the exam. That was one reason for making the scale even shorter and more user-friendly. If that is really the case might lie in the eye of the user, not all raters will find the scale easy to use. The four dimensions which are used in the scale are task achievement & communicative skills, clarity & naturalness of speech, grammar and vocabulary. The three parts of the exam, namely the monologue, the short dialogue and the long dialogue are assessed holistically. The reason for that approach is that the rater has to do his or her job while the student is speaking, so, there is not enough time for counting. Consequently, the test takers get one single score in each of the four dimensions. Rater training is essential here. The construct would not be valid, if there was not sufficient training beforehand. One reason for that is that the expressions in the descriptors are not really clear and rather vague (Mewald et al. 2012: 42)

Speaking Assessment Scale (Jan 2011)

	Task Achievement & Communication Skills	Clarity & Naturalness of Speech	Grammar ¹	Vocabulary
7	<ul style="list-style-type: none"> detailed information communicated reliably description or narrative with main points expanded by relevant, detailed information and examples effective turntaking through initiating maintaining and/or closing discourse sometimes using stock phrases 	<ul style="list-style-type: none"> fluent and spontaneous at a fairly even tempo with natural pauses longer stretches of language clear, natural pronunciation and intonation 	<ul style="list-style-type: none"> good range of structures relatively high degree of grammatical control and few inaccuracies which do not impair communication message clear 	<ul style="list-style-type: none"> good range of vocabulary communicating clear ideas; formulations sometimes varied to avoid repetition generally accurate vocabulary
6	<ul style="list-style-type: none"> clear and concrete information of immediate relevance with main points communicated comprehensibly 	<ul style="list-style-type: none"> some degree of fluency with some pausing for repair or grammatical and lexical planning 	<ul style="list-style-type: none"> generally sufficient range of structures occasional inaccuracies which can impair communication message clear 	<ul style="list-style-type: none"> sufficient range of vocabulary communicating clear ideas occasionally inaccurate vocabulary; major errors possible when expressing more complex ideas
5	<ul style="list-style-type: none"> straightforward description or narrative basic turntaking through initiating maintaining or closing discourse sometimes using stock phrases 	<ul style="list-style-type: none"> connected stretches of language in a clear, linear sequence of points clearly intelligible pronunciation and intonation, sometimes with a foreign accent; occasional mispronunciations 	<ul style="list-style-type: none"> generally sufficient range of structures occasional inaccuracies which can impair communication message clear 	<ul style="list-style-type: none"> sufficient range of vocabulary communicating clear ideas occasionally inaccurate vocabulary; major errors possible when expressing more complex ideas
4	<ul style="list-style-type: none"> limited information on familiar and routine matters communicated in a simple and direct exchange 	<ul style="list-style-type: none"> noticeable pauses, hesitation or false starts, sometimes causing breakdown of communication 	<ul style="list-style-type: none"> limited range of simple structures frequently inaccurate with basic mistakes, generally without causing breakdown of communication message usually clear 	<ul style="list-style-type: none"> limited range of vocabulary mostly communicating clear ideas frequently inaccurate vocabulary controlling a narrow lexical repertoire
3	<ul style="list-style-type: none"> description or narrative in a simple list of points on sentence and word-group level effective questioning in information exchange 	<ul style="list-style-type: none"> short contributions and exchanges linked with some simple connectors intelligible pronunciation; foreign accent or mispronunciations which sometimes impair understanding 	<ul style="list-style-type: none"> limited range of simple structures frequently inaccurate with basic mistakes, generally without causing breakdown of communication message usually clear 	<ul style="list-style-type: none"> limited range of vocabulary mostly communicating clear ideas frequently inaccurate vocabulary controlling a narrow lexical repertoire
2	<ul style="list-style-type: none"> very little information even in simple everyday situations 	<ul style="list-style-type: none"> much hesitation frequently causing breakdown of communication 	<ul style="list-style-type: none"> extremely limited range of simple structures limited control causing frequent breakdown of communication message seldom clear 	<ul style="list-style-type: none"> extremely limited range of vocabulary communicating few clear ideas mostly inaccurate vocabulary frequently causing breakdown of communication
1	<ul style="list-style-type: none"> basic statements or negations on word or word group level attempted questioning to get information 	<ul style="list-style-type: none"> very short, isolated, mainly pre-packaged utterances frequent mispronunciations; only understood by speakers of English with some effort 	<ul style="list-style-type: none"> extremely limited range of simple structures limited control causing frequent breakdown of communication message seldom clear 	<ul style="list-style-type: none"> extremely limited range of vocabulary communicating few clear ideas mostly inaccurate vocabulary frequently causing breakdown of communication
0	<ul style="list-style-type: none"> no task achievement 	<ul style="list-style-type: none"> no assessable language 	<ul style="list-style-type: none"> no assessable language 	<ul style="list-style-type: none"> no assessable language

¹ Descriptors referring to range and control reflect the features of the task and the nature of grammar and vocabulary in unplanned speech

In the category task achievement and communication skills “the information the test takers provide (propositional precision, in all parts), the quality of the narrative (thematic development, primarily in the monologue part) as well as the ability to interact with a partner (turntaking, primarily in the dialogue part) are assessed” (Mewald et al. 2012: 28). In clarity & naturalness of speech a performance is considered natural and clear if “the pronunciation is intelligible and the intonation makes it sound natural” (Mewald et al. 2012: 28). The category grammar includes descriptors for range, control and the clarity of the message. So, what is assessed is the use of a range of grammatical structures, the level of their accuracy as well as their impact on the message. In order to assess vocabulary the raters

look at content words (nouns, ‘full’ verbs, adjectives, adverbs), collocations and chunks of language that a speaker uses to fulfil a communicative task. They assess the range of lexis that creates meaning and manages to accomplish successful communication and control, i.e. the level of accuracy (Mewald C., Gassner O., Brock R., Lackenbauer F., Siller F. and K.: 2012: 33).

One can easily see that the E8 standard rating scale is based on the assumption that language acquisition happens in a linear form, every band is somehow based on the lower band. That is criticized by other researchers (for example Fulcher, see Performance Decision Trees, chapter 8.4.2).

The developers of the E8 standards claim that it is a useful tool to identify strengths and weaknesses in the speaking competence of Austrian pupils at the end of year 8. One could say that this is not really always the case, as the scale is not easily interpretable for non-trained teachers and pupils, as the expressions and bands are not clear enough. However, a major advantage is that the results are comparable across all the test takers and that the scale can be used on nearly all speaking exercises.

7.4.2 PERFORMANCE DECISION TREES

Glenn Fulcher, Fred Davidson and Jenny Kemp (2011) constructed a different rating scale, namely the performance decision tree (PDT). The developers’ goal was to

create a scale which is more practical and authentic than other popular rating scales. There are two popular approaches for rating scale design: the measurement-driven approach and the performance data-driven approach. As in the CEFR or in the Bildungsstandards, measurement-driven approaches try to order descriptors onto a single scale. Meaning does not depend on authentic language samples but it is derived from the scaling methodology and the agreement of trained judges as to the place of any descriptor on the scale. In the performance decision trees the situation is different. That rating scale is performance data-driven, which means that it places primary value upon observations of actual language performance, and attempts to describe performance in sufficient detail to produce descriptors that have a direct relationship with the original observations of language use. Meaning is not derived from scaling methodology, but from the link between performance and description. Fulcher, Davidson and Kemp (2011: 5) argue that “measurement-driven approaches generate impoverished descriptions of communication, while performance data-driven approaches have the potential to provide richer descriptions that offer sounder inferences from score meaning to performance in specified domains”.

An important fact is that most of the current rating scales have the underlying assumption that the construct increases in a linear fashion, as they describe it in their levels or bands. That is a big disadvantage because this assumption is too simple and does not reflect how humans learn a second language. Not only measurement-driven approaches assume that their levels represent a “ladder to be climbed” (Fulcher, Davidson, Kemp 2011: 6), this is also the case in most data-driven approaches. To solve that problem, Fulcher, Davidson and Kemp (2011: 6) claim that

performance data-based scales need to evolve into a new type of rating instrument, which we call Performance Decision Trees (PDTs). PDTs represent an improvement on performance data-based scales in that they escape from the illusion of linear development in language use. They are based in a thorough analysis of the context of performance and the nature of interaction in specific communicative situations.

The constructors of the PDTs use the example of travel agency discourse, to illustrate the method. That is done to show the importance of context and interaction in the assessment of speaking.

One of the most influential rating scales nowadays is the CEFR (Common European Framework of Reference) by North. The six levels (A1 to C2) are often

used in Europe to guide assessment and learning. Fulcher, Davidson and Kemp (2011: 7) claim that the CEFR is not the ideal. They state that

[a]lthough the scale is empirically derived, it is not based on performance data, as there is no reference to the performance of learners or test takers on specific tasks, or even perceptions of the value of performances. Rather, this methodology depends on the ability to use a measurement model to place band or level descriptors drawn from disparate sources onto a single scale using teacher estimates of descriptor difficulty as data. The measurement model – Rasch in this case – is seen as an external arbiter that decides what does and does not survive in the scale development process.

There is most of the time no relation to authentic language samples. The scale does not make any difference between buying chips or a car. That is an enormous disadvantage. Only sometimes there is reference is to real world scenarios, but very inconstantly:

Some descriptors refer to specific situations, while others do not. Level B2, for example, refers to getting a traffic (parking?) ticket, damaging property, and dealing with being blamed for an accident. Other levels are less specific. When a context of language use is mentioned, it is not necessarily referred to in other descriptors. Dealing with travel agents is specifically mentioned in Level B1, but not at other levels, despite references to travel. We are therefore left with the question of whether 'dealing with travel agents' is something that is suddenly possible at level B1. Furthermore, participant roles are mixed within the same level. At A2 for example, the learner can 'ask for and provide' goods and services. This seems to imply that an A2 learner would be able to function as a service provider as well as a server seeker. At level B2 would this mean that a learner could explain to a client how to seek compensation, as well as ask for compensation as a customer? The distinction between levels is unclear, with descriptors referring to the vague concept of complexity' at each level (Fulcher, Davidson, Kemp 2001: 8).

Another disadvantage is that the expressions are very vague, for example "most transactions", "less routine situations". There is no clear definition of those terms. One could conclude that scales like that are inadequate in their descriptions. They are not sensitive enough to communicate context and the interactional complexities of language use. For Fulcher, Davidson and Kemp (2011: 9) the solution is a "richer

description of contextually based performance”, in order to minimize the gap between the score and its meaning.

When creating a PDT the first step is to describe the nature of interaction in the specific communicative context of interest. In this case, the elements of interactional competence in service encounters need to be explained, which would be (Fulcher, Davidson and Kemp 2011: 9-10) discourse competence (“ability to understand and utilize knowledge of the structure of a service encounter to provide a service, or get the service needed”), discourse management and pragmatic competence. Of course, all those competences need to be explained in greater detail and examples from authentic situations need to be examined (see Fulcher, Davidson, Kemp 2011: 11-16). The outcome of that analysis of real language looks as follows (Fulcher, Davidson, Kemp 2011: 16):

Interactional Competence in a Service Encounter

A. Discourse Competence

1. Realization of service encounter discourse structure
2. The use of relational side-sequencing

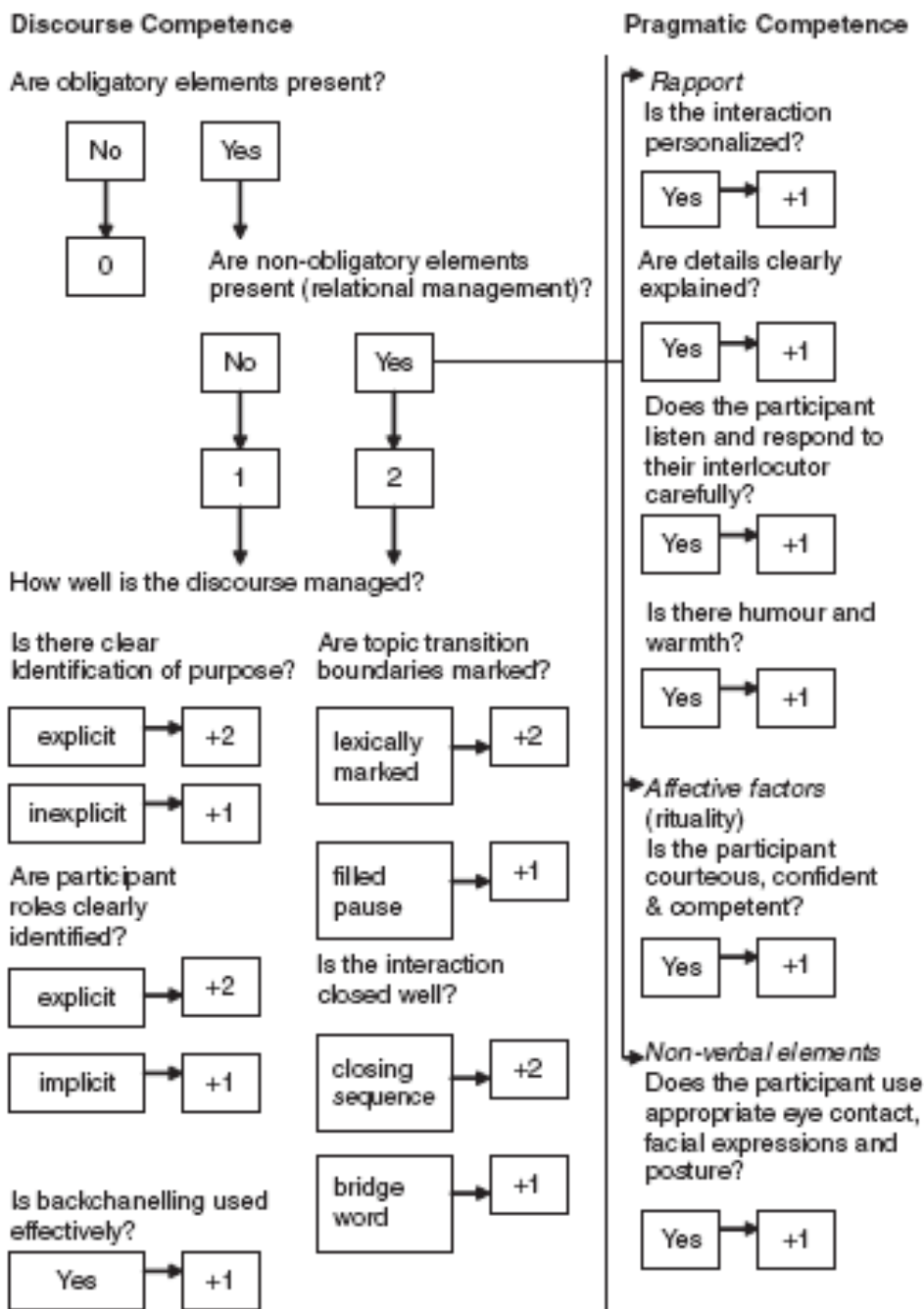
B. Competence in Discourse Management

3. Use of transition boundary markers
4. Explicit expressions of purpose
5. Identification of participant roles
6. Management of closings
7. Use of backchannelling

C. Pragmatic Competence

8. Interactivity/rapport building
9. Affective factors, rituality
10. Non-verbal communication

Those elements form the basis of the performance decision tree, which is shown here. The PDT provides a straightforward picture of the competencies and skills that are needed to master complex service encounters.



Ideally, the testing situation is a dialogue in which one participant is the service provider and the other one is the customer. The elements on the left side of the PDT could be used to score the performances of both test takers, whereas the right side is more useful for the scoring of the service provider. However, scoring both test takers is not easy because the two roles need different knowledge, so a separate score for each participant is needed to reflect how they contribute to the co-construction.

That again underlines the fact that the focus is on interactional competence in PDTs. One can see that communication does not take place between two equal participants with the same rights and power.

If one has a close look at the PDT above, one can see that the score can be between 0 and 20, depending on how well the discourse and pragmatic competence in the interaction was. Again a scale is produced, but not a traditional linear one.

Fulcher, Davidson and Kemp (2011: 22) state that

there is no implication that ability is uniform across descriptors, or that a particular score is arrived in a uniform manner. The PDT brings together the description of performance-data based rating scales and the EBB methodology in a system that offers rich description behind the scale, but provides raters with a much simpler set of binary decisions that may be much easier to use in live rating.

Another advantage of the PDT is that it allows the rater to design a diagnostic profile of the test taker. One could easily find out about the strengths and weaknesses of the test taker in each specific area. It gives the learners a chance to find out what exactly needs to be improved for a higher score to be awarded. Furthermore, it gives the teacher useful hints on what he or she needs to focus on in class. Fulcher, Davidson and Kemp (2011: 23) claim that “[t]he PDT may therefore bring us a step closer to integrating the outcomes of classroom assessment into more targeted instruction”.

Although the constructors of this method argue that raters will find the PDT easier to use than other complex performance data-based rating scales and traditional rating scales, they admit that further empirical investigation is needed.

However, they (2011: 23) still argue that

an analysis of how people use language in actual communicative contexts can form the basis for more dynamic and contextually sensitive approaches to rating that help to define the nature of interactional competence in context. Performance Decision Trees are more flexible and do not assume a linear, unidimensional, reified view of how second language learners communicate. They are also pragmatic, focusing as they do upon observable action and performance, while attempting to relate actual performance to communicative competence.

7.5 POSSIBLE RATING SOLUTION FOR NEW MIDDLE SCHOOLS IN AUSTRIA

The following section tries to find a solution for grading a student's speaking performance in a first grade in a New Middle School in Austria. It should somehow summarise what has been said so far about summative assessment and represent a creative solution which can be altered for different purposes.

To start, I would like to refer to the Austrian Curriculum and what it claims about grading:

Bei der Einschätzung und Bewertung von Schülerleistungen sind der individuelle Lernfortschritt und das Bemühen um die Optimierung von Arbeitsergebnissen mit zu beachten. Verstöße gegen die Sprachrichtigkeit sind nur eines der Bewertungskriterien und sind für die Gesamtleistung nicht alleine ausschlaggebend. Weitere Gütekriterien wie Verständlichkeit der Äußerungen, soziolinguistisch und pragmatisch angemessene Situationsbewältigung sowie Differenziertheit der verwendeten sprachlichen Mittel sind mit von Bedeutung. (bmukk 2012: 35)

All those criteria should be kept in mind, when creating a rating method. This is also the case for this special method. In my opinion, it is worth creating a new scale for every speaking activity. That sounds like a tremendous amount of work. However, that is not the case if one develops his or her standard binary-choice grading method. For me this would look as follows:

All necessary information given:

Underpoint 1: Yes/No

Underpoint 2: Yes/No

Underpoint 3: Yes/No

Used acceptable range of vocabulary: Yes/No

Pronunciation was intelligible: Yes/No

Plus Point 1: Yes/No

Plus Point 2: Yes/No

This system is easy to handle and, additionally, leaves space for better performances. It is possible to gain plus points. The system is not restricted for the use in only one single rating situation, on the contrary, it can be easily changed for different tasks and purposes. So it tries to combine easy handling with sufficient feedback.

If one tries to apply this method to a task from the E8 standards from section 6.5.2, this would look the following:

Speaking Task E6/E7: Monologue My favourite animal



Think about the topic for a minute. Then try to talk for about 2 minutes.

Tell me as much as you can about **your favourite animal**:

- your favourite animal
- what it looks like
- why it is your favourite animal
- what you can do with it

All necessary information given:

Talked about favourite animal: Yes/No

Talked about what it looks like: Yes/No

Talked about why it is his/her favourite animal: Yes/No

Talked about what he/she can do with it: Yes/No

Used acceptable range of vocabulary: Yes/No

Pronunciation was intelligible: Yes/No

Plus point1: Gave more information than necessary: Yes/No

Plus Point2: Used very complex vocabulary: Yes/No

Now if we want to really grade the talk, one needs to find a scale. I would suggest this one:

4 Points – okay

5 Points – good

6 Points – very good

Of course, every teacher has his or her own system here. This one should only serve as an example, which tries to avoid the disadvantages, which are in the E8 rating scale and use the advantages from the experimental rating scales (PDT). That means that this rating scale does not use vague terms, on the contrary, it is as straightforward as possible. Furthermore, it is not linear and gives every student a new chance to show what he or she can do. To sum up, with the help of this paper it should be easier to create a tool for grading speaking. Of course, this can have different forms but speaking has, finally, found its place in teaching and grading which was in some way the aim of this paper.

8 CONCLUSION

After finishing this thesis, I can say that teaching and testing speaking is not an unproblematic field as manifold opinions exist about this issue. However, this work should show that it is worth diving into the topic since the new demands from headmasters and the government will ultimately focus on English as a means of communication. Unfortunately, what this thesis could not provide is an ultimate solution for the problem. It can only serve as an introduction or guideline as the creation of a solution is in the hand of the teacher him- or herself. This work should provide the basis and help the teachers find their own solution. In my case, a meeting with my colleagues will be held to find a common solution for our school in order to reach a fair rating system for all students.

9 BIBLIOGRAPHY

- Adams, M. L.; Firth, J. R. 1979. *Testing Kit: French and Spanish*. Washington DC: Department of State and the Foreign Services Institute.
- Alderson, C. J. 1991. "Language testing in the 1990s: How far have we come? How much further have we to go?" In Anivan, S. (ed.). *Current developments in Language Testing*. Singapore: Seameo Regional Language Center.
- Anderson, A. 1984. *Teaching Talk*. Cambridge: Cambridge UP.
- Bifie. 2012. *IKM Speaking Tasks E6/E7*. Available at:
<https://www.bifie.at/downloads?dokumenttyp%5B%5D=20&Zielgruppe%5B%5D=13>
- Bmukk. 2009. *Lehrplan Lebende Fremdsprache*. Available at:
http://www.bmukk.gv.at/medienpool/17135/lp_hs_lebende_fremdsprache.pdf
- Bmukk. 2012. *Lehrplan der Neuen Mittelschule*. Available at:
https://www.bmbf.gv.at/schulen/unterricht/lp/lp_nms.html
- Brock, R.; Horak, A.; Lang-Heran, H.; Moser, W.; Schatzl, Z.; Schlichtherle, B.; Schober, M. 2008. *Leistungsfeststellung auf der Basis des Gemeinsamen Europäischen Referenzrahmens für Sprachen (GERS)*. Praxisreihe: Heft 8. Graz: ÖSZ
- Brown, G.; Anderson, A.; Shillock, R.; Yule, G. 1984. *Teaching Talk*. Cambridge: Cambridge UP.
- Brown, G.; Yule, G. 1983. "Some factors in the selection and implementation of communication strategies". In Faerch, C.; Kasper, G. (eds.). *Strategies in Interlanguage Communication*. London: Longman.
- Brumfit, C. J.; Johnson, K. (eds.). 1979. *The Communicative Approach to Language Teaching*. Oxford: Oxford UP.
- Brumfit, C. J. 1984. *Communicative Methodology in Language Teaching*. Cambridge: Cambridge UP.
- Burns, A. 1998. "Teaching Speaking". *Annual Review of Applied Linguistics* 18 (1998): 102-123.
- Burns, A.; Joyce, H. 1997. *Focus on Speaking*. Sydney: NCELTR.
- Bygate, M. 1987. *Speaking*. Oxford: Oxford UP.
- Byrne, D. 1986. *Teaching Oral English*. London: Longman.
- Canale, M.; Swain, M. 1980. "Theoretical bases of communicative approaches to second language teaching and testing". *Applied Linguistics* 1 (1) (1980): 1-47.
- Carter R.; McCarthy, M. 1995. "Grammar and the spoken language". *Applied Linguistics* 16 (1995): 141-158.
- Cathcart, R. 1989. "Authentic discourse and the survival English curriculum". *TESOL Quarterly* 23 (1989): 105-126.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. The Hague: Mouton.

- Council of Europe (ed.). 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press.
- Davies, A. 1990. *Principles of Language Testing*. Oxford: Basil Blackwell.
- Dinçer, A.; Yeşilyurt, S. 2013. "Pre-Service English Teachers' Beliefs on Speaking Skill Based on Motivational Orientations". *English Language Teaching* 6 (2013): 88-95.
- Fillmore, C. J. 1979. *Individual differences in language ability and language behavior*. New York: Academic Press.
- Fulcher, G. 1996. "Testing tasks: Issues in task design and the group oral". *Language Testing* 13 (1996): 23-52.
- Fulcher, G. 2003. *Testing Second Language Speaking*. Harlow: Pearson.
- Fulcher, G.; Davidson, F.; Kemp, J. 2011. "Effective rating scale development for speaking tests: Performance decision trees". *Language Testing* 28(1): 5-29.
- Gilakjani, A. P. 2012. "The Significance of Pronunciation in English Language Teaching". *English Language Teaching* 5 (April 2012): 96-107.
- Haggstrom, M. 1994. "Using a videocamera and task-based activities to make classroom oral testing a more realistic communicative experience". *Foreign Language Annals* 27: 161-175.
- Halliday, M. A. K. 1975. *Learning How to Mean: Explorations in the Development of Language*. London: Edward Arnold.
- Hedge, T.; Andon, N.; Dewey, M. (eds.). 2009. *English Language Teaching*. Volume 1. London and New York: Routledge.
- Hieke, A. E. 1985. "A componential approach to oral fluency evaluation". *Modern Language Journal* 69 (2): 135-142.
- Huang, H-T. D.; Hung, S-T A. 2013. "Comparing the Effects of Test Anxiety on Independent and Integrated Speaking Test Performance". *TESOL Quarterly* 47 (2) (June 2013): 244-269.
- Hughes, R. 2011. *Teaching and Researching Speaking*. Harlow: Longman.
- Hymes, D. 1972. "On Communicative Competence". In Pride, J. B.; Holmes, J. (eds.). *Sociolinguistics*. Harmondsworth: Penguin, 269-293.
- James, R. B. 2010. *Teaching Pronunciation Gets a Bad R.A.P: A Framework for Teaching Pronunciation*. Hankuk: University of Foreign Studies.
- Jenkins, J. 2000. *The Phonology of English as an International Language*. Oxford: Oxford UP.
- Johnson, K. 1982. *Communicative Syllabus Design and Methodology*. Oxford: Pergamon.
- Jones, R. L. 1981. "Scoring procedures in oral language proficiency tests". In Read, J. A. S. (ed.). *Directions in Language Testing (Anthology Series 9)*. RELC: Singapore University Press, 100-107.
- Kaim, K. 2004. "English Language Teaching in Austria: Self- Assessment of Non-native Speaking Teachers". Diploma Thesis. U of Vienna.
- Lantolf, J. P.; Frawley, W. 1988. "Proficiency: Understanding the construct". *Studies in Second Language Acquisition* 10: 181-195.

- Lazaraton, A. 1992. "The structural organization of a language interview: A conversation analytic approach". *System* 20: 373-386.
- Leeson, R. 1975. *Fluency and Language Teaching*. London: Longman.
- Littlewood, W. 1981. *Communicative Language Teaching*. Cambridge: Cambridge UP.
- Littlewood, W. 1984. *Foreign and Second Language Learning: Language Acquisition Research and Its Implications for the Classroom*. Cambridge: Cambridge UP.
- Long, M. H.; Porter, P. A. 1985. "Group work, interlanguage talk and second language acquisition". In *Working Papers* 4/1: 103-137. Department of English as a Second Language: University of Hawaii at Manoa.
- Lumley, T.; McNamara, T.F. 1995 "Rater characteristics and rater bias: Implications for training". *Language Testing* 12 (1995): 54-71.
- Nation, I. S. P.; Newton, J. 2009. *Teaching ESL/EFL Listening and Speaking*. New York and London: Routledge.
- Mackey, W. F. 1965. *Language Teaching Analysis*. London: Longman.
- Matthews, M. 1990. "The measurement of productive skills: doubts concerning the assessment criteria of certain public examinations". *English Language Teaching Journal* 44 (2): 117-121.
- McCarthy, M.; O'Keeffe, A. 2004. "Research in the Teaching of Speaking". *Annual Review of Applied Linguistics* 24: 26-43.
- Mewald, C.; Gassner, O.; Brock, R.; Lackenbauer, F.; Siller, K. 2012. *Testing speaking for the E8 standards. Technical Report*. bifie.
- Norris, Timothy. 1991. "Nonnative English-Speaking Teacher Assistants and Students Performance". *Research in Higher Education* 32(4): 433-448.
- North, B. 1995. "The development of a common framework scale of descriptors of language proficiency based on a theory of measurement". *System* 23 (4): 445-465.
- North, B. 1996. "The development of a common framework scale of descriptors of language proficiency based on a theory of measurement". In Huhta, A.; Kohonen, B.; Kurki-Suonio, L.; Luoma, S. (eds.). *Current Developments and Alternatives in Language Assessment*. Jyväskylä: University of Jyväskylä, 423-427.
- Perren, C. S. 1968. "Testing spoken language: Some unresolved problems". In Davies, A. (ed.). *Language Testing Symposium*. Oxford: Oxford University Press, 107-116.
- Phillipson, R. 1992. *Linguistic Imperialism*. Oxford: Oxford UP.
- Piepho, H. E. 1981. "Establishing objectives in the teaching of English". In Candlin C. (ed.). *The Communicative Teaching of English: Principles and an Exercise Typology*. London: Longman.
- Richards, Jack C.; Rodgers, T. S. 1986 [2009]. "Communicative language teaching". In Hedge, T.; Andon, N.; Dewey, M. (eds.). *English Language Teaching*. Volume 1. London and New York: Routledge.

- Roberts, J. 1981. "Discussion through decision making". *Modern English Teacher* 9 (2): 31-34.
- Savignon, S. 1983. *Communicative Competence: Theory and Classroom Practice*. Reading, Mass.: Addison-Wesley.
- Shohamy, E.; Gordon, C. M.; Kramer R. 1992. "The effect of raters' background and training on the reliability of direct writing tests". *Modern Language Journal* 76: 27-33.
- Slade, D. 1986. "Teaching casual conversation to adult ESL learners". *Prospect* 2 (1) (1986): 68-88.
- Slade, D. 1990. *Teaching casual conversation*. Paper presented at the National Centre for English Language Teaching and Research Spoken Discourse Project Workshop. Macquarie University Sydney, September 1990.
- Slade, D.; Gardner, R. 1993. "Teaching casual conversation: The issue of simplification". In Tickoo, M. (ed.). *Simplification: Theory and application*. Singapore: SEAMEO Regional Language Centre. 82-100. [Anthology Series 31.]
- Thomas, J. 1999. "Voices from the Periphery: Non-Native Teachers and Issues of Credibility". In Braine, G. (ed.). *Non-native Educators in English Language Teaching*. London: LEA, 5-13.
- Thornbury, S. 2005. *How to teach Speaking*. Harlow: Pearson Education Limited.
- Turner, J. 1998. "Assessing Speaking". *Annual Review of Applied Linguistics* 18 (1998): 192-207.
- Underhill, N. 1987. *Testing Spoken Language*. Cambridge: Cambridge UP.
- Upshur, J. A.; Turner, D. E. 1995. "Constructing rating scales for second language tests". *ELT Journal* 49: 3-12.
- Widdowson, H. G. 1987. *Teaching Language as Communication*. Oxford: Oxford UP.
- Wigglesworth, G. 1997 "An investigation of planning time and proficiency level on oral test discourse". *Language Testing* 14 (1997): 85-106.
- Wilds, C. 1975. "The oral interview test". In Jones, R. L.; Spolsky, B. (eds.). *Testing Language Proficiency*. Arlington, VA: Center for Applied Linguistics, 29-44.
- Wilkins, D. A. 1975. *Second Language Learning and Teaching*. London: Edward Arnold.
- Wu, K.; Ke, C. 2009. "Haunting Native Speakerism? Students' Perceptions toward Native Speaking English Teachers in Taiwan". *English Language Teaching* 2 (September 2009): 44-52.
- Yungzhong, L. 1985. "Writing versus speech in foreign language teaching". *Wai Guo Yu* 3 (37): 12-15.
- Zhang, Y. 2009. "An Experimental Study of the Effects of Listening on Speaking for College Students". *English Language Teaching* 2 (September 2009): 194-204.

Abstract

This work should serve as an introduction to the teaching and testing of speaking in New Middle Schools in Austria. A theoretical part should form the basis of the thesis. Communicative language teaching, the skill of speaking and some concepts in speech production will be explained for that purpose. In order to make the theory more useful for the teachers in Austria, the Austrian curriculum will serve as a guideline here. The practical part consists of two major issues, namely teaching and testing speaking. In both cases, suggestions for the Austrian classroom will be given, as well as general issues will be discussed. The goal of this thesis is to provide the reader with a theoretical background, so that he or she can use the provided suggestions in his or her professional career.

Zusammenfassung

Die Neue Mittelschule und die geänderten Rahmenbedingungen bringen immer neue Herausforderungen für die Lehrer mit sich. Eine dieser Herausforderungen ist es der Kompetenz Sprechen exakt dieselbe Aufmerksamkeit, sowohl im Lehren als auch im Bewerten, zukommen zu lassen, wie zb Lesen oder Hören. Diese wissenschaftliche Arbeit soll als Einführung in das Lehren und Bewerten der Kompetenz „Sprechen“ in Neue Mittelschulen dienen. Ein theoretischer Teil dient als Basis der Arbeit. Das Kommunikative Klassenzimmer, die Fertigkeit „Sprechen“ und verschiedene Konzepte in der Sprachproduktion sollen aus diesem Grund erläutert werden. Um den Inhalt für die Lehrer/innen in den österreichischen Mittelschulen noch zugänglicher zu machen, wird der österreichische Lehrplan als Orientierung genützt. Bei den Themen Lehren und Testen werden sowohl praktische als auch allgemeine Themen behandelt. Das Ziel dieser Arbeit ist es den Leser mit theoretischem Wissen auszustatten, damit er/sie dieses in seiner/ihrer professionellen Karriere bestmöglich nützen kann.

Lebenslauf

- Schulbildung: 4 Jahre Volksschule Rosenau
4 Jahre Hauptschule Rosenau
5 Jahre HLW Amstetten
- Karriere als Pädagogin: SS 2013 Englisch-Trainerin in 4 Kindergärten in NÖ

Seit WS 2013 Englisch Lehrerin in der NMS
Allhartsberg und in der HAS Waidhofen/Ybbs
- Universitäre Laufbahn: WS 2008 Start des Studiums (Lehramt Englisch und
Geschichte)

WS 2009 Wechsel des Studiums (Lehramt Englisch
und Psychologie/Philosophie)

SS 2011 Abschluss des 1. Abschnitts

Herbst 2014 Abschluss des Studiums