



universität  
wien

# DIPLOMARBEIT / DIPLOMA THESIS

Titel der Diplomarbeit / Title of the Diploma Thesis

„Alternatives in assessment in Austrian EFL classes“

verfasst von / submitted by

Michael Schneller

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of  
Magister der Philosophie (Mag. Phil.)

Wien, 2017 / Vienna, 2017

Studienkennzahl lt. Studienblatt /  
degree programme code as it appears on  
the student record sheet:

A 190 344 456

Studienrichtung lt. Studienblatt /  
degree programme as it appears on  
the student record sheet:

Lehramtsstudium UniStG  
UF Englisch UniStG  
UF Geographie und Wirtschaftskunde UniStG

Betreut von / Supervisor:

Mag. Dr. Helen Heaney, BA



## **Acknowledgements**

First and foremost, I would like to thank my parents for believing in my dreams and abilities. Thank you for your love and support, your encouragement and for piquing my interest to attend university. I wouldn't be where I am now if it wasn't for you!

I would also like to thank my partner Melanie, with whom I have shared my life for quite some time now. Thank you for being at my side whatever the circumstances are, for your patience and your kind words during this diploma thesis project. Love you!

I would like to express my sincerest gratitude to my supervisor Dr. Mag. Helen Heaney, BA, for her encouragement and support throughout the process of carrying out this diploma thesis project. I hope that I was not too much of a burden.

Last but not least, I want to thank Summer and Finn for cheering me up on a daily basis.



# Table of contents

<b>Abbreviations .....</b>	<b>i</b>
<b>List of figures .....</b>	<b>ii</b>
<b>List of tables.....</b>	<b>iii</b>
<b>1. Introduction .....</b>	<b>1</b>
1.1. Defining the research area and questions .....	2
1.2. Chapter overview .....	3
<b>2. Theoretical framework .....</b>	<b>4</b>
2.1. Attempting to define educational assessment .....	4
2.2. Purposes of educational assessment .....	6
2.3. Educational assessment in Austria .....	8
2.3.1. Legal regulations .....	9
2.3.2. Quality criteria and language assessment .....	14
2.4. Alternatives in (language) assessment – an assessment trend .....	19
2.4.1. Traditional testing vs. alternatives in assessment .....	20
2.4.1.1. Changes in learning and teaching philosophies .....	21
2.4.1.2. Standardized testing .....	22
2.4.1.3. Testing Situations .....	24
2.4.1.4. Quality criteria.....	25
2.4.2. Formative assessment .....	25
2.4.2.1. Formative assessment in EFL/ESL contexts .....	28
2.4.2.2. Characteristic features of formative assessment .....	31
2.4.2.3. A European approach to formative assessment: Assessment for learning (Afl) .....	35
2.4.3. Alternative assessment tools in EFL contexts .....	37
2.4.3.1. Portfolios .....	37
2.4.3.2. Self- and peer-assessment .....	40
2.4.3.3. Journals.....	42
2.4.3.4. Observations and rubrics .....	43
2.4.3.5. Performance assessment .....	45
2.5. Alternatives in (language) assessment in Austria .....	46
<b>3. Empirical study .....</b>	<b>48</b>
3.1. Research questions and aims .....	48
3.2. Research methodology .....	49
3.3. The sample .....	50

3.4.	Questionnaire design .....	51
3.5.	Procedure .....	53
3.5.1.	Digitalization .....	53
3.5.2.	Piloting the questionnaire .....	53
3.5.3.	Analyzing the data .....	53
4.	Results .....	54
4.1.	Sociodemographic data .....	54
4.2.	EFL assessment use and know-how .....	57
4.3.	Assessment and teacher education .....	62
4.4.	Teachers' beliefs and attitudes .....	64
4.5.	Final comments .....	73
5.	Discussion .....	74
5.1.	EFL assessment use & know-how .....	75
5.2.	Teachers' attitudes towards alternative assessment .....	78
5.3.	Factors contributing to teachers' decisions about assessment procedures .....	83
6.	Conclusion .....	84
7.	Bibliography .....	87
	Appendix 1: Abstracts .....	94
	Appendix 2: Test statistics .....	96
	Appendix 3: Questionnaire .....	103
	Appendix 4: Answers to the open ended questions .....	110

## Abbreviations

AFL .....	Assessment for learning
AHS .....	<i>Allgemein bildende höhere Schule</i> ‘Academic secondary school’
BIFIE.....	<i>Bundesinstitut für Bildungsforschung, Innovation &amp; Entwicklung des österreichischen Schulwesens</i> ‘Austrian institute of educational research, innovation and development of the Austrian school sector’
BHS .....	<i>Berufsbildende höhere Schulen</i> ‘Colleges for higher vocational education’
BMB .....	<i>Bundesministerium für Bildung</i> ‘Austrian ministry of education’
BMHS .....	<i>Berufsbildende mittlere und höhere Schulen</i> ‘Schools for intermediate vocational education and colleges for higher vocational education’
CEFR .....	Common European Framework of Reference
EFL .....	English as a foreign language
ELP .....	European language portfolio
ESL .....	English as a second language
IKM.....	<i>Informelle Kompetenzmessung</i> ‘Self-evaluation tools for diagnostic purposes provided by the BIFIE’
LBVO.....	<i>Leistungsbeurteilungsverordnung</i> ‘Austrian act for assessment’
LOA .....	Learning-oriented assessment
NMS .....	<i>Neue Mittelschule</i> ‘New secondary school’
OECD.....	Organization for Economic Co-operation and Development
ÖZEPS .....	<i>Österreichisches Zentrum für Persönlichkeitsbildung und soziales Lernen</i> ‘Austrian center for personal development and social learning’
PH/PÄDAK .....	<i>Pädagogische Hochschule/Pädagogische Akademie</i> ‘College of teacher education’
PIRLS.....	Progress in International Reading and Literacy Study
PISA .....	Programme for International Student Assessment
SchUG .....	<i>Schulunterrichtsgesetz</i> ‘Austrian school education act’
TIMSS .....	Trends in International Mathematics and Science Study
TOEFL .....	Test of English as Foreign Language

## List of figures

Figure 1:	Agents and key features of formative assessment (Wiliam & Thompson 2008; quoted in Wiliam 2011: 12) .....	31
Figure 2:	Answers to item no. 5 “Please indicate the three most important purposes of language assessment in your English teaching in general.” .....	58
Figure 3:	Answers to items no. 6-22 “Select the response which best describes how well you know the term. ” .....	59
Figure 4:	Answers to item no. 23 “I’m well aware of the legal regulations as they are described in the SchUG and the LBVO concerning a assessment in Austrian schools.” .....	60
Figure 5:	Answers to items no. 24-37 “How often do you use the listed types of assessment procedures in your EFL classroom?” .....	61
Figure 6:	Answers to item no. 67 “Assessment you would have liked to have discussed in more detail during teacher education or in-service training.” .....	64
Figure 7:	Answers to items no. 86/97 “... they are obligatory to determine pupils' grades.” .....	67
Figure 8:	Answers to items no. 87/98 “... they provide information about pupils' achievement.” .....	67
Figure 9:	Answers to items no. 88/99 “... they provide information about pupils' strengths and weaknesses.” .....	68
Figure 10:	Answers to items no. 89/100 “... they are essential in fostering learning processes.” .....	68
Figure 11:	Answers to items no. 90/101 “... they provide information about the quality of teaching.” .....	69
Figure 12:	Answers to items no. 91/102 “... they are required by the LBVO and the SchUG.” .....	69
Figure 13:	Answers to items no. 92/103 “... they are practical (easy to design, easy to score, ...).” .....	70
Figure 14:	Answers to items no. 93/104 “... they provide information for improving future instruction.” .....	70
Figure 15:	Answers to items no. 94/105 “... they motivate pupils to learn English.” .....	71
Figure 16:	Answers to items no. 95/106 “... they provide the opportunity to assess the pupils' performance in real-world tasks and exercises.” .....	71
Figure 17:	Answers to items no. 96/107 “... pupils benefit from the assessment process.” .....	72
Figure 18:	Answers to item no. 108 “Would you consider using alternatives in assessment in your future EFL classrooms?” .....	73



## List of tables

Table 1:	Answers to item no. 1 “What university/teacher training college (PH) did you graduate from?” .....	55
Table 2:	Answers to item no. 2 “What subject(s) do you teach?” .....	55
Table 3:	Answers to item no. 3 “How long have you been teaching?” .....	56
Table 4:	Answers to item no. 4 “Which sort of school are you teaching at?” .....	57
Table 5:	Teachers working at NMS and other school types .....	57
Table 6:	Answers to items no. 55-58 “Agreement/disagreement with statements concerning teacher education.” .....	62
Table 7:	Answers to items no. 59-66 “I feel comfortable doing the following in regard to language assessment.” .....	63
Table 8:	Answers to items no. 68-85 “Agreement/disagreement with statements concerning language assessment.” .....	65



## 1. Introduction

Recent Austrian developments in educational assessment draw a picture of standardization processes topped with quality assurance measures. Several issues have dominated the educational assessment research landscape since the turn of the millennium. Assessment has become first and foremost a political matter of quality management, fueled by results on international performance tests like the Programme for International Student Assessment (PISA). One monumental development constitutes the establishment of educational standards in conjunction with the orientation towards a competence-based teaching methodology. Most recently though, the standardization of the Austrian school-leaving exam, that is Austria's most important high-stakes test in secondary education, has been thoroughly discussed by all stakeholders and the media. All in all, recent trends in Austria's teaching and assessment can be described as fulfilling a paradigm shift from input- to output-oriented quality management characterized by policy decisions in favor of standardization processes on various levels.

However, teaching and assessment research have given rise to other trends, which have been somewhat implemented in the Austrian educational landscape. While the current public discourse about educational assessment can be characterized by quality measures and standardization, developments in modern society demand an emphasis on different aspects of assessment (Dochy 2001: 11). Dochy (2001: 12-13) states that learning and assessment needs to be adapted to changing educational objectives. While he argues that gaining knowledge was the "core issue" (Dochy 2001: 12) in education for past decades, nowadays the main goals in education are shifting towards developing "reflective practitioners" (Dochy 2001: 13). Active learning, constructivist learning or student-centered learning approaches are promoted. Problem-solving skills are of major importance. In addition, independent knowledge acquisition and a reflective stance on learning processes are acknowledged as being crucial to successful learning (Dochy 2001: 13). Due to these developments, teachers and researchers argue that a new learning culture based on the aforementioned characteristics also requires a new assessment culture. One trend that answers to most of those needs has been described as a "hot topic" (Brown & Abeywickrama 2010: 16) in (language) assessment and has been captioned alternative assessment or alternatives in assessment (Brown & Hudson 1998: 657). However, the

concept is not closely defined, which means that different interpretations and working definitions can be found in the literature. What most have in common is a clear distinction to testing and a general attitude that all stakeholders should benefit from the assessment process. In addition, institutions like the Organization for Economic Co-operation and Development (OECD) argue in support of alternative ways of assessment in contrast to traditional testing, as “[f]ormative assessment has been shown to be highly effective in raising the level of student attainment, increasing equity of student outcomes, and improving students’ ability to learn” (OECD 2005: 2).

Based on research dealing with different types of alternative forms of assessment and their effects on learning and teaching, personal experience as well as recent Austrian assessment trends raised the question of whether alternative assessment tools have actually found their way into Austrian classrooms, specifically EFL classrooms.

### **1.1. Defining the research area and questions**

Assessment in its various forms in Austrian EFL classrooms are addressed in the course of this diploma thesis. The theoretical part is dedicated to describing the educational situation in Austria in which assessment takes place, defining characteristics of alternative assessment in order to highlight the differences to traditional assessment methodologies, and reviewing a variety of alternative assessment procedures which might be beneficial to Austrian EFL classrooms. The empirical part of this thesis aims at identifying whether or not alternative forms of assessment are implemented in EFL classes in secondary schools and how teachers feel about alternative assessment models. Therefore, the following specific research questions have been defined:

1. What types of assessment are used by current EFL teachers?
2. What are teachers' attitudes towards the use of alternative assessment in Austrian EFL classrooms?
3. What factors contribute to teachers' decisions about assessment procedures?

A quantitative empirical study answers these questions and gives an insight into an area of assessment which has so far not received thorough attention in Austria.

## **1.2. Chapter overview**

Chapter 2 aims at exploring the relevant theoretical framework in which language assessment in Austria is administered. Section 2.1 tries to define the concept of educational assessment. While different purposes of assessment are discussed in 2.2, Section 2.3 deals with the legal conditions that regulate assessment in Austrian schools and assessment quality criteria. Section 2.4 emphasizes alternative assessment as a particular assessment trend. Subsections 2.4.1 and 2.4.2 examine two particular approaches to this rather loosely defined concept. First, alternative language assessment is juxtaposed with traditional language testing in order to identify the differences between both concepts and the shortcomings of traditional testing. Second, the approach to alternative assessment is presented by contrasting summative assessment, which is associated with traditional testing, and formative assessment, which is usually associated with alternative forms in language assessment. The benefits for all assessment stakeholders is uncovered by discussing the features of formatively functioning assessment. Subsection 2.4.3 moves on to a variety of alternative assessment methods that have been explicitly labeled as alternatives in language assessment. Section 2.5 explains common Austrian assessment trends and examines whether or not alternative assessment is currently of particular interest.

Chapter 3 introduces the empirical part of this diploma thesis. First of all, the study objectives and research questions are reiterated. Next, several sections deal with the research methodology and certain aspects of conducting the study. The concept of quantitative research and other questionnaire-related issues like the sample and questionnaire design are central to this chapter. Chapter 4 analyzes the collected information and summarizes the findings quantitatively. The chapter is divided into sections according to the sectioning of the questionnaire. Afterwards, Chapter 5 interprets the data in relation to the posed research questions and, finally, Chapter 6 arrives at a conclusion.

## 2. Theoretical framework

### 2.1. Attempting to define educational assessment

Educational assessment is a diverse concept that is difficult to define due to its complex nature (Joughin 2009: 1). Nowadays, educational assessment in general and educational language assessment in particular assume a range of purposes which are accompanied by a variety of assessment forms, models and tools. Hamp-Lyons mentions in this regard that due to ongoing developments in language teaching, the language that deals with assessment evolves likewise and, therefore, is “becoming more nuanced” (2016: 13). This contributes to a considerable number of terms and definitions in the discourse about educational assessment. Different definitions emphasize different aspects of assessment. Nevertheless, to give the reader a first impression of the different emphases in educational assessment definitions, this chapter opens with examples of definitions and common misconceptions in the field of assessment terminology.

A rather concise definition for the umbrella term *educational assessment* is provided by Joughin who argues that “[t]o assess is to make judgements about students’ work, inferring from this what they have the capacity to do in the assessed domain, and thus what they know, value, or are capable of doing” (2009: 16). Assessment in educational settings can therefore be considered to be a judgment of pupils’ knowledge, skills and capabilities in a certain domain. According to Mousavi, assessment is “appraising or estimating the level or magnitude of some attribute of a person” (2009: 36). In addition, assessment is depicted by Brown and Abeywickrama as an “ongoing process that encompasses a wide range of methodological techniques” (2010: 3). Black and William on the other hand summarize assessment as all efforts, methods and tools used by teachers and/or students “that provide information to be used as feedback to modify teaching and learning activities” (1998: 7). Finally, from a teacher’s perspective, the assessment of pupils can be defined as a task. It is part of a teacher’s routine, something that is constantly being done by teachers all day, every day. Besides actual teaching, it is probably the most time-consuming task teachers have to fulfill.

A very common misconception amongst people who deal with educational settings is founded on them not being aware of the differences between the concepts of testing and assessment. The concepts behind both terms cannot be treated as equal according to Brown and Abeywickrama (2010). They argue that

[t]ests [...] are a subset of assessment, a genre of assessment techniques. They are prepared administrated procedures that occur at identifiable times in a curriculum when learners muster all their faculties to offer peak performance, knowing that their responses are being measured and evaluated. (Brown & Abeywickrama 2010: 3)

Accordingly, tests are one variety of assessment techniques. The general term assessment, on the other hand, describes a much broader concept. Historically and linguistically speaking however, the term test and its denotative meaning is regarded as the “root term” (Hamp-Lyons 2016: 13) in the field of educational language assessment. According to Montgomery (1965, quoted in Hamp-Lyons 2016: 13), the first appearance of examinations dates back to 19<sup>th</sup> century England. English universities introduced entrance examinations in order to select students on the basis of their “proficiency or achievement in a language or a body of knowledge” (Hamp-Lyons 2016: 13). The term test on the other hand spread at the beginning of the 20<sup>th</sup> century when it was used in the military context of mental testing in the years of World War I. The term assessment became significant rather recently. Nowadays, it refers to “all activities teachers use to help students learn and to gauge students progress” (Black and William 1998: 143). As a consequence, it follows that tests are types of assessment; however, assessment does not have to be some sort of test.

Other terms and concepts that could be confusing in this area of education are, according to Brown and Abeywickrama (2010: 4), measurement and evaluation. Measurement in an educational setting is defined as “the process of quantifying the observed performance of classroom learners” (Brown & Abeywickrama 2010: 4). For language teaching contexts, Bachman explains that measuring language performance constitutes the “assigning of numbers” to observed abilities of language learners based on certain “rules and procedures” (Bachman 1990: 19-20). On the other hand, evaluation is the process of decision making that is based on the information gathered by administering all kinds of assessment methods (Brown & Abeywickrama 2010: 5). Thus, Bachman “distinguishes the information-providing function of measurement from the decision-making function of evaluation” (1990: 23).

At this point, it can be concluded that educational assessment is a significant part of every teacher's routine. It deals with the judgment of pupils' work, their skills and capabilities in specific domains. Assessment does not necessarily mean testing and it should be a way to facilitate learning experiences.

## **2.2. Purposes of educational assessment**

There is a general consensus about assessment having a variety of purposes (Jang 2014: 11; Hamp-Lyons 2016: 13). This section looks at educational assessment from different perspectives in order to reveal and distinguish a range of functions that are fulfilled by assessment in educational settings.

Educational assessment serves selective and allocative purposes (Stern 2010: 31), which can be discussed from a societal and an individual angle. According to Wiater (2009: 70), schools have to fulfill a purpose for society. Besides socialization processes and the communication of societal norms and standards, it is the purpose of schools to provide professional chances and opportunities for upward mobility (Stern 2010: 31). Due to society's meritocratic structure and its limited capacity for high positions, educational assessment is the tool that proves performance or achievement and provides the integration of the individual into the hierarchically structured society. Therefore, Wiater (2009: 70) argues that the selective and allocative function of assessment is bound to preserve social hierarchies. In addition, Fend (2006: 44) emphasizes the effects of assessment on the individual due to the significant influence of achievement and performance on social status, prestige, and income. Thus, assessment is highly relevant to the future lives of children because, for instance, grades as a direct result of assessment processes affect future social and professional chances and choices. The selective and allocative function is the responsibility of schools and educational assessment and cannot be neglected.

Educational assessment purposes can be classified differently because certain assessment tools are applied for different reasons. Hamp-Lyons (2016: 14) as well as Brown and Abeywickrama (2010: 9) differentiate between several crucial motives for language assessment. Following their arguments, assessment can be used in order to measure achievement and proficiency; thus tests are either called achievement tests or proficiency tests. Tests that measure achievement are usually held when a unit, course or school grade, etc. is completed and pupils are tested to ascertain whether they met the set objectives (Brown & Abeywickrama 2010: 9). Proficiency tests on the contrary do not focus on any



objectives of teaching units or the curriculum but on “overall ability” (Brown & Abeywickrama 2010: 11). Another test type used in order to make decisions about placing pupils in courses of certain levels is called a placement test. However, while placement tests are very common in the United States of America, this sort of test has no relevance to Austria’s education. Tests for diagnostic purposes are administered in order to find out about shortcomings, difficulties or potentials. In this regard, diagnostic language assessment is crucial because, as Law and Eckes (2007: 19) note, pupils come from different backgrounds, with divergent previous school careers and therefore dissimilar knowledge bases and skills in a language. In order to adjust instruction to individual students, teachers need information about their abilities, skills and knowledge. Appropriate and effective teaching will not be possible without a diverse set of information, gathered via multiple assessment methods (Law & Eckes 2007: 19).

Gathering information about pupils’ knowledge and their capabilities is central in educational assessment. In most countries students need to be graded at the end of a module, term or year. In order to do so, teachers assess what pupils have learned and what they are capable of doing by considering different kinds of performances and by administering a combination of different achievement assessment tools. This type of assessment has been described as assessment of learning (Wiliam 2011: 10). However, gathering information, as already mentioned, can also have a diagnostic purpose, meaning that assessment results can be used to improve learning processes as well as future teaching. This quite different purpose of assessment has been emphasized by Black and Wiliam (1998) for instance. They argue that the main focus of recent assessment practice has been on tests “which are only weakly linked to the learning experiences of students”, while a development of assessment procedures away from traditional tests towards learning enhancing methods could be a reason for improved learning processes and constant long-term learning success (Black & Wiliam 1998: 7). Assessment approaches that focus on the modification or improvement of learning have been captioned formative assessment (Black & Wiliam 1998: 7-8), learning-oriented assessment (Purpura 2004: 212), or assessment for learning (Wiliam 2011: 9-10).

Additionally, assessment serves as an evaluation of teaching success and quality assurance as has been described by Stern (2010). According to Stern (2010: 31) the results of assessing pupils are a valuable resource for teachers. That is, students’ achievement is an indicator of teaching quality and assessment results can be used to evaluate the instructional program (Stern 2010: 32). From a broader Austrian perspective, the evaluation of success

and quality of teaching in particular and the school system in general has recently become a matter of assessment due to the implementation of educational standards in 2009 (Dalton-Puffer, Faistauer & Vetter 2011: 201). The results of international student assessment surveys like PISA or the Trends in International Mathematics and Science Study (TIMSS) and the demand for comparability of pupils' achievement led to the implementation of "quality assurance measures" (Eller & Schratz 2009: 228) which simultaneously led to a paradigm shift in quality assurance from input- to output-orientation. Thus, periodic assessment of basic competences which pupils should possess after the years 4 and 8 of schooling are now expected to ensure quality of instruction and success in learning (Eller & Schratz 2009: 229).

Summing up, it should be emphasized that the demands on assessment are vast and, as Stern notes, probably discordant (Stern 2010: 31). He argues that educational assessment is in an awkward predicament, because the different functions or purposes of assessment might be incompatible (Stern 2010: 31). One contributing aspect to this notion is the fact that assessment is also used for disciplinary actions (Stern 2010: 31). In other words, fulfilling all the purposes of assessment at once is probably impossible and therefore always subject to compromise (Stern 2010: 32). One possibility to tackle the dilemma of conflicting requirements of assessment is to outsource the assessment process in the sense of implementing external testing (Stern 2010: 31). However, external testing has advantages and disadvantages. While arguments for external, standardized testing emphasize that teachers could focus on the teaching and training of pupils without the task of assessing and grading achievement especially at high stakes moments (BMBWK 2002: 13), external, standardized testing might lead to teaching-to-the-test, a phenomenon that can be perceived worldwide wherever external, standardized tests are used (Brown & Abeywickrama 2010: 98). However, at present, Austrian teachers have to find a way to satisfy the various demands as well as possible.

### **2.3. Educational assessment in Austria**

Before this diploma thesis goes into detail about alternatives in assessment, the status quo of the research in this field and their role in Austrian education, the regulatory framework within which educational assessment is administered is outlined. The focus is on Austria's policies and the legal regulations that shape educational assessment due to the Austrian setting of this thesis. In addition, quality criteria are presented.

### 2.3.1. Legal regulations

Assessment in Austrian schools is regulated by law. Legal regulations protect the weak, in this case the students, from arbitrary assessment on the part of the teachers (Neuweg 2014: 1-2). Two important federal acts regulate the assessment of pupils in Austria. The first is the *Schulunterrichtsgesetz (SchUG)* ‘Austrian school education act’, in particular §18 to §23, and second is the *Leistungsbeurteilungsverordnung (LBVO)* ‘Austrian act for assessment’, which is a ministerial order that was adopted in 1974 and has been amended several times, the most recent amendment being put forward in December 2016. Both federal acts are binding for teachers as well as pupils and describe what is subject to assessment in school, how assessment is supposed to take place and to what extent assessment should be undertaken. Additional federal acts regulate the final examinations at the end of secondary school (BMB 2017a). One act deals with the school-leaving exam at *Allgemein bildende höhere Schulen (AHS)* ‘academic secondary schools’ and the other one regulates the school-leaving exam at *Berufsbildende mittlere und höhere Schulen (BMHS)* ‘schools for intermediate vocational education and colleges for higher vocational education’. The *Informationsblätter zum Schulrecht Teil 3: Leistungsfeststellung und Leistungsbeurteilung* ‘Fact sheets on school laws part 3: performance assessment and performance evaluation’ provides explanations of and comments on various legal aspects of assessment in school (Feigl 2007: 4). Finally, a commentary by Neuweg (2014) on legal regulations was published for the first time in 2000 and provides counsel for teachers and pupils in all imaginable kinds of assessment matters.

In terms of content, generally speaking, the LBVO distinguishes between two assessment stages: The §§ 2-10 discuss the so-called “*Leistungsfeststellung*” (LBVO 2016), which can be summarized as the measurement of pupils’ learning results and the §§ 11-17 discuss the so-called “*Leistungsbeurteilung*” (LBVO 2016), which is the evaluation of those learning results that eventually leads to the assigning of grades (Neuweg 2014: 4).

The term *Leistungsfeststellung* refers to methods teachers may use to collect evidence about the pupils’ achievement of learning goals (Neuweg 2014: 22). The § 3 (1) LBVO distinguishes between certain forms of *Leistungsfeststellung*, containing a variety of subforms (LBVO 2016). Three forms are particularly important for language assessment:

- § 4 LBVO “*Mitarbeit der Schüler im Unterricht* [pupils’ participation in class]” (LBVO 2016)

Teachers should collect information on pupils’ participation in class. This means that teachers need to chart pupils’ achievements during class time, be it in oral or written form. In addition, homework as extended individual work as well as participation in pair or group work have to be considered (LBVO 2016). It must be noted however, that instances of exceptional or unacceptable behavior of pupils are not regarded as evidence of class participation in terms of *Leistungsfeststellung* (Neuweg 2014: 28).

- §§ 5-6 LBVO “*Mündliche Leistungsfeststellung* [oral performance assessment]” (LBVO 2016)

Teachers are obliged to provide the opportunity for pupils to show their achievements in a “*Mündliche Prüfung* [oral examination]” (LBVO 2016), if there is a need for further examination or pupils ask for it (Neuweg 2014: 36). Depending on the school type and age of the students, different lengths of exams and the number of relevant topics are legitimate. Nevertheless, oral exams must comprise two questions that are independent from each other and must be scheduled at least two days in advance (LBVO 2016). “*Mündliche Übungen* [oral exercises]” (LBVO 2016), on the contrary, need to be announced a week in advance and the topics must also be determined in advance (LBVO 2016). A typical example of an oral exercise would be a presentation.

- §§ 7-8 LBVO “*Schriftliche Leistungsfeststellung* [written performance assessment]” (LBVO 2016)

The LBVO distinguishes two important written examination types. In § 7 the so-called “*Schularbeiten*” (LBVO 2016) and in § 8 the “*Schriftliche Überprüfungen*” (LBVO 2016) are specified. *Schularbeiten* are a specific type of test that differ from *Schriftliche Überprüfungen* due to their length and range of relevant topics. They are usually administered for the duration of one lesson, although in § 7 (1) it is stated that subject curricula can modify regulations in this respect (LBVO 2016). *Schriftliche Überprüfungen* on the other hand are shorter with a maximum of 25 minutes, depending on school type and the age group of the pupils (Neuweg 2014: 54). Teachers are not allowed to administer as many written examinations as they want. The duration of all written examinations in a specific subject within a semester or school year is limited, which is again contingent on the pupils’ age and the school type (Neuweg 2014: 55-56).

The second prominent term mentioned in the legal regulations is *Leistungsbeurteilung*, which is based on the various aforementioned forms of *Leistungsfeststellung* (Neuweg 2014: 72). *Leistungsbeurteilung* refers to the teacher's judgment of the pupil's achievement compared to a criterion, which is usually an educational norm or standard that is characterized in the curriculum. Three relevant norms of reference can be distinguished, to wit the curricular norm, social norm and individual norm.

- Legislatively speaking, the most important norm is the curricular norm because it is manifested in § 11 (1) LBVO: “Maßstab für die Leistungsbeurteilung sind die Forderungen des Lehrplanes unter Bedachtnahme auf den jeweiligen Stand des Unterrichts. [The criteria that are determined in the curriculum are the benchmark for assessment, although the actual progress of instruction must be considered]” (LBVO 2016). That is, assessment is based on a comparison of the pupil's achievement with the criteria or learning objectives that are determined in the appropriate curriculum. Brown and Abeywickrama (2010: 8) refer to this type of assessment as criterion-referenced assessment, which is supposed to provide feedback to pupils about their achievement of specific learning objectives. Assessment based on curricular norms or learning objectives has advantages and disadvantages. Stern (2010: 31) notes that on the one hand, it helps measure the achievement of certain learning objectives but on the other hand individual preconditions, for instance the involvement of parents or so-called shadow learning that may affect learning outcomes, are not taken into account.
- Assessment according to a social norm evaluates pupils' achievements compared to their peers. In other words, the assessment results of all pupils within a peer group are ranked and interpreted in relation to their classmates. This type of assessment is known as norm-referenced assessment (Brown & Abeywickrama 2010: 8). An advantage of this type of referencing is that it is obvious who achieved above-average or below-average results. Nevertheless, Stern (2010: 29) notes that assessing pupils to a social norm has disadvantages too; for instance, it cannot be inferred from the grading whether or not learning objectives have been met (Neuweg 2014: 76). Weak students will always be below average and the grade for a specific achievement will vary according to the performance of the whole class (Stern 2010: 29). Even though norm-referencing has been a very prominent language assessment, Bailey (1998: 36) concludes that it is “often not appropriate for classroom use”. In addition, Neuweg (2014: 77) argues that a good or bad performance depends on the random affiliation to a certain peer group.

- Assessment based on an individual norm means comparing the current and past achievements of one individual pupil. The major advantage of this type of referencing is that the learning progress becomes visible (Stern 2010: 31). Therefore, the combination of the curricular and the individual norm has been suggested in order to provide feedback on individual progress in addition to the achievement of learning objectives (Neuweg 2014: 83-84).

The final stage of *Leistungsbeurteilung* in the Austrian school system involves grading the pupils' achievement (Neuweg 2014: 72). That is, teachers are to categorize the pupils' achievement according to a grading scheme on a five-point scale:

- *Sehr gut (1)* 'very good'
- *Gut (2)* 'good'
- *Befriedigend (3)* 'satisfactory'
- *Genügend (4)* 'sufficient'
- *Nicht genügend (5)* 'insufficient'

Definitions of each grade are provided by the §14 LBVO (LBVO 2016). The grading process is carried out in two stages. Oral and written examinations, as covered by the LBVO, are subject to grading. Every written and oral examination as well as oral exercises need to be evaluated by the teacher and the pupils' achievement has to be expressed in one of the five grades. In a second step, these evaluations in combination with the evaluation of in-class participation result in a grade for each subject for a module, semester or school year. Austrian schools provide reports that show the numerical grades for each subject.

However, not every type of assessment is graded. For instance, single instances of class participation are not graded. In addition, the LBVO clearly distinguishes between forms of *Leistungsfeststellung* with different purposes. So-called "*Informationsfeststellungen*" which have an exclusively diagnostic purpose and serve as information to teachers about whether or not pupils have reached parts of or entire learning objectives are not regulated by the LBVO and thus not subject to grading, as indicated in § 1 (2) LBVO (LBVO 2016).

The current manifestations of assessment in schools are criticized for various reasons. Due to the focus on alternative assessment in this thesis, the most obvious problem is the lack of encouragement for the use of alternative assessment. As Stern notes, the LBVO covers assessment *of* learning mainly. Most paragraphs explain what can be tested, how it can be tested and how test results have to be graded (Stern 2010: 81). Assessment that fulfills

diagnostic purposes, assessment that fosters learning processes or alternative assessment tools are not part of Austria's most important act that deals with assessment. In addition, while Austria's grading system in general is frequently the focus of criticism, Neuweg (2014: 122) strongly suggests accompanying grades with feedback, even though the LBVO does not indicate feedback in any form. He emphasizes that grades are not solely employed for selection purposes but provide qualitative information about pupils' achievements. Besides, qualitative feedback that is processed over time provides possibilities to address shortcomings as well as identify talents and, additionally, explains in detail how teachers decide on individual grades (Neuweg 2014: 122).

Moreover, whether or not assessment and grading can and should be a motivating factor for learning is a debated issue. While learning in order to receive good grades is associated with extrinsic motivation, research shows that it is intrinsic motivation that is said to cause meaningful, ongoing learning that prepares learners for life (Crooks 1988: 468). Whether or not certain types of assessment are actually capable of fostering intrinsic motivation will be dealt with in Section 2.4.2.2.

Another critical issue is the execution of the prescriptive acts. The LBVO does not value a single form of *Leistungsfeststellung* more than another as indicated in § 3 LBVO (LBVO 2016). However, common practice shows different scenarios. Neuweg (2014: 29), for instance, claims that written and oral examinations are usually devoted more attention than classroom participation. That view is underlined by observations made by Buschmann and Thonhauser (2000; quoted in Eder, Neuweg & Thonhauser 2009: 252) that assessment procedures like *Schularbeiten* or *Schriftliche Überprüfungen* are traditionally given exceptional weight in the grading process.

Finally, the LBVO is criticized by Austrian experts (cf. Stern 2010: 81, Neuweg 2014: 3) for being too vague. Neuweg (2014: 3) states that some of the prescriptions are explicitly detailed, while some parts are not, like the definitions of grades or the assessment of in-class participation. For instance, as Stern (2010: 82) claims, the wording of the definitions of the grades is so abstract that it leaves exceptional room for interpretation to the teachers. However, at the same time, it can be argued that vague definitions result in a considerable flexibility. Morgan (2006: 61) states that, in contrast to England, Austrian teachers enjoy much more freedom in administering alternative forms in assessment. It seems to be a balancing act for the legislator between providing flexibility and providing detailed guidelines (Neuweg 2014: 2-3), which could impair a teacher's autonomy, an undeniably

vital characteristic of teacher professionalism (Hoyle & John 1995: 77). In other words, administrative barriers like overregulation could hinder the pedagogic professional practice of teachers, especially if it comes to assessment.

### **2.3.2. Quality criteria and language assessment**

So far, it has been established that educational assessment deals with collecting information about pupils, their skills and competences. The task of gathering information about pupils requires teachers to abide by the law. Generally speaking, principles of language assessment are proposed in language assessment research (Bachman & Palmer 1996: 17-42, Brown & Abeywickrama 2010: 25-47, Xi & Davis 2016). According to Bachman and Palmer (1996: 18), six principles, namely practicality, reliability, validity, authenticity, washback and interactiveness, express what they call “test usefulness”. This section will outline the six assessment qualities.

Assessment needs to be practical. That is, using specific assessment tools requires certain resources (Bachman & Palmer 1996: 35). In general, schools and teachers have to deal with certain available or unavailable resources. Those resources that are associated with assessment are categorized by Bachman and Palmer (1996: 36-37) in three groups, namely human resources, material resources and time. Human resources could involve test writers, administrators or raters or scorers (Bachman & Palmer 1996: 37). In terms of classroom-based assessment in Austria, all those professions are usually assumed by the teacher. Material resources are space and equipment (Bachman & Palmer 1996: 37), for instance, if scoring needs to be undertaken via a computer, computers must be available. Time is an issue because designing a test and scoring the results need to be doable within usually limiting time constraints. Mousavi (2009: 518, quoted in Brown and Abeywickrama 2010: 26) associates practicality with characteristics like time, costs and ease of administration of assessment tools. Bachman and Palmer (1996: 37) consider the costs of assessment because allocating resources is a matter of money. Another budgetary concern could be expenditure in regard to externally administered testing, which should not be a vast concern within the Austrian school system since the only mandatory external test is the *Reife- and Diplomprüfung* ‘Austrian school-leaving exam’, and language proficiency tests like the Cambridge First Certificate are paid by the individual test taker. Nonetheless, if assessing exceeds any of the available but required resources, the assessment tool in question is considered impractical and should be revised or exchanged for another technique (Brown



& Abeywickrama 2010: 26). In comparison to the other assessment qualities, practicality indicates assessment usefulness on the basis of resources while the other criteria deal with the usefulness of assessment results (Bachman & Palmer 1996: 35).

Another concern for assessment administrators is that of reliability, which is defined as “consistency of measurement” (Bachman & Palmer 1996: 19). Hughes (2003: 45) argues that reliability is a matter of “achieving consistent performances from candidates” and “scorer reliability”. It is a matter of test scores because only reliable test scores provide teachers with the essential information about their pupils’ achievement (Bachman & Palmer 1996: 20). In the realm of assessment reliability, four factors are usually discussed, namely student related reliability, rater reliability, test administration reliability and test reliability itself (Brown & Abeywickrama 2010: 27-29). In addition, reliability is also considered to be a precondition for validity.

In terms of student-related reliability, personal, mental and physical conditions are of importance, for instance anxiety or fatigue (Brown & Abeywickrama 2010: 28). In this respect, providing distraction-free conditions (Hughes 2003: 47-48) and input of equal quality to all pupils might help avoid issues of unreliability (Brown & Abeywickrama 2010: 41). Moreover, familiarity with test or item design affects reliability (Hughes 2003: 47). Students apply a variety of strategies to tackle a certain assessment situation. In this respect, reliability can be increased by clear and distinct test taking instructions and by ensuring that test takers are familiar with test formats or item design (Hughes 2003: 47-48).

Two kinds of rater reliability are defined by Brown and Abeywickrama (2010: 28) as intra-rater reliability and inter-rater reliability. Intra-rater-reliability deals with very common problems in the design, administration and scoring process of a single teacher (Brown & Abeywickrama 2010: 28). Prominent problems are insufficiently described scoring criteria as well as rater bias towards exceptionally bad/good or likeable/unlikeable pupils. Item design that yields objective scoring can improve intra-rater reliability. For the same reason, Hughes (2003: 49-50) suggests a “detailed scoring key” as well as identification of test-takers via numbers instead of names. Thus, bias towards a specific person or a group of people can be avoided more easily. Inter-rater reliability is of concern when, for instance, more than one rater is involved in the assessment process. If several raters arrive at the same result, inter-rater-reliability is given.

The third type of reliability is test administration reliability. Unreliability in this regard might be problematic because of environmental conditions. Hence, the room, light,

malfunctioning equipment as well as improperly designed or printed test materials could be the source for unreliable assessment (Brown & Abeywickrama 2010: 28-29).

Other issues concerning reliability may appear due to the characteristics of the assessment itself and its design (Bachman & Palmer 1996: 29), which is discussed as test reliability by Brown and Abeywickrama (2010: 29). According to them (2010: 29), ambiguity in test item design or the test format have repercussions on test reliability. In terms of classroom-based assessment, test unreliability might be caused by assessment based on open questions, which means that scoring the results involves judgment by the teacher. In addition, ambiguous or non-discriminating test items can be a source of unreliability (Brown & Abeywickrama 2010: 29).

Validity is another quality of appropriate assessment (Brown & Abeywickrama 2010: 29). Valid tests measure what they are supposed to measure. In other words, assessment possesses validity if the results allow teachers to draw conclusions about test takers' abilities according to the intended test purpose. Bachmann and Palmer (1996: 21) term this type of quality construct validity. They define a construct as "the specific definition of an ability that provides the basis for a given test or test task and for interpreting scores derived from this task" (Bachmann & Palmer 1996: 21). Thinking about the validity of assessment instruments is exceptionally important because too often, teachers intend to measure some specific competence but the scoring, however, includes additional, different, actually irrelevant criteria.

In terms of classroom-based assessment, content validity is of major importance. It is achieved if a language assessment procedure asks pupils to perform relevant language competences which are supposed to be evaluated (Hughes 2003: 22). In this regard, Brown and Abeywickrama (2010: 30) argue that assessment possesses content validity if the requested performance or achievement can be observed. Therefore, they distinguish between direct and indirect testing (Brown & Abeywickrama 2010: 31). Language assessment can be designed to measure achievement directly by asking pupils to do exactly what is defined in a learning objective or indirectly by asking pupils to show abilities which are only partly linked with those objectives. Thus, in order to attain content validity is to administer assessment that measures learning objectives directly (Brown & Abeywickrama 2010: 32). Another option to ensure content validity is proposed by Hughes (2003: 22), who states that specifications of skills, knowledge or control of language structures should

be formulated before choosing the type and content of assessment. Adjusting the assessment content to the specifications ensures content validity (Hughes 2003: 22).

Another validity criterion deals with the consequences of applying certain assessment instruments, and is therefore called consequential validity (Brown and Abeywickrama 2010: 34). Bachmann and Palmer (1996: 29-35) use the term washback and impact for the same concept, Hughes (2003: 44) on the other hand, uses the term backwash. All these terms describe an aspect of assessment quality that is about the “impact on society and educational systems and upon the individuals within those systems” (Bachmann & Palmer 1996: 29). That is, assessment has an effect on the individual test taker and future test administration as well as on society.

Another aspect of validity is typically called face validity. This aspect is mainly associated with the students’ perspective of assessment fulfilling its intended purposes and what is often called fairness. However, face validity is considered to be a notion and not scientifically verifiable (Hughes 2003: 27; Brown & Abeywickrama 2010: 35). On account of this, face validity is regarded as being a very subjective matter that is “too dependent on the whim of the perceiver” (Brown & Abeywickrama 2010: 35). Nevertheless, it cannot be denied that, although being a notion, assessment being fair in the eye of the assessed person has consequences. Hughes argues that pupils would not be able to show their true abilities if they perceived an assessment tool as unfair or inappropriate (Hughes 2003: 27). Possible methods for evaluating assessment consequences are surveys or focus groups, as has been suggested, for instance, by Chudowsky and Behuniak (1998). Brown and Abeywickrama (2010: 35) propose a series of actions to increase face validity, for instance sticking to the course work as a basis for assessment or keeping tasks and exercises reasonably challenging.

Bachman and Palmer (1996: 23) argue that useful assessment requires authentic tasks. In order to call a language task authentic, it has to share characteristics with situations where the target language is used outside the English classroom (McNamara 2000: 131). However, the topic of authenticity is quite critically debated. Lewkowicz (2000) conducted a study on how test takers perceive authentic tasks in tests and whether or not authenticity helps them perform better in the test situation. The obtained results show that authenticity “may be of theoretical importance for language testers needing to ensure that they can generalize from test to non-test situations, but not so important for other stakeholders in the testing process” (Lewkowicz 2000: 60). Nevertheless, in order to increase authenticity,

assessment instruments and their items should be “contextualized [...] meaningful [and] relevant” (Brown & Abeywickrama 2010: 37). In this regard, the relevance of tasks for pupils’ lives is discussed frequently because it has been argued time and again that test items in particular often do not resemble real-world tasks (Chun 2006: 304).

In classroom-based assessment, interactiveness is defined as the engagement of pupils’ individual characteristics by assessment tasks. Individual characteristics are defined as language knowledge, strategic competence, topical knowledge and affective schemata (Bachman & Palmer 1996: 25). Those characteristics are crucial factors in language use. Highly interactive assessment tasks engage all characteristics. If tasks only involve one characteristic, for instance, language knowledge, teachers could evaluate the pupils’ achievement in the area of language ability but they would not know about how well pupils can use the language.

Hughes (2003: 1) defines backwash or washback as “[t]he effect of testing on teaching and learning”. Bachman and Palmer (1996: 29) argue that language tests have a direct impact on two levels. They affect the individual pupil and teacher on a micro level and they affect the educational system and/or society on a macro level (Bachman & Palmer 1996: 29-30). Nowadays, impact, washback or backwash are often used synonymously. Nevertheless, Bachmann and Palmer (1996: 30) describe washback as a facet of impact, that is, assessment is supposed to affect future instruction as well as future learning processes on the part of the assessed people. After research into washback increased and its complexity was revealed, the research focused on how it emerges and how it is elicited. All over the world, researchers came up with different washback models (cf. Hughes 1989; Bailey 1996; Green 2007). Messick (1996: 241) argues that washback can go both ways, meaning that it can have a positive or a negative impact. Washback in terms of classroom-based assessment is often associated with the micro level of impact, that is, with benefits to pupils due to test preparation and post-processing that provides pupils with helpful diagnostic feedback about their abilities (Brown & Abeywickrama 2010: 38). Therefore, using assessment that provides the opportunity for appropriate, useful washback can be very challenging.

In summary, these characteristics or quality criteria help teachers of all kinds of school subjects to evaluate whether or not assessment is “effective, appropriate [and] useful” (Brown & Abeywickrama 2010: 25). Thus, assessment that abides by these criteria is considered good assessment. The literature tries to provide guidelines on the basis of criteria that can be used for optimizing the design process of applicable, useful and efficient

assessment as well as evaluating the quality of existing tests or exams. Bachman and Palmer (1996: 18) add three guiding principles for consideration when dealing with test qualities. First and foremost, “overall usefulness” (Bachman & Palmer 1996: 18) of assessment is to be prioritized over single qualities. Second, the individual principles need to be evaluated in combination in terms of their “effect on overall usefulness” (Bachman & Palmer 1996: 18). Third, overall test usefulness cannot be generalized but must be evaluated for single assessment procedures in their unique testing situation (Bachman & Palmer 1996: 18). That is, teachers should strive for maximizing the overall usefulness of assessment but have to consider the specific context the assessment tool is administered within.

#### **2.4. Alternatives in (language) assessment – an assessment trend**

While previous sections summarized the key concepts of assessment in Austria and assessment usefulness, the subsequent sections investigate alternative assessment, which is an international trend in assessment that increasingly attracts attention in Austria (cf. Stern 2010; Schmidinger, Hofmann and Stern 2015).

Brown and Hudson (1998: 656) pointed out that alternatives in language assessment have been enthusiastically discussed in the language testing literature since at least the 1990s. In Section 2.1, it was argued that educational assessment is difficult to define. Similar to that, alternative assessment is rather a broad term encompassing different theories and fulfilling different purposes and is therefore rather loosely defined. In other words, the terminology used across the profession is ambiguous and different researchers and educators work with the concept from different perspectives. Within the scope of this diploma thesis, the emphasis will be on two particular perspectives, which are very common across the field of educational assessment. One approach towards defining alternative assessment is contrasting traditional forms of assessment with alternatives. Therefore, as a first step, it is necessary to examine what is commonly considered traditional assessment and how alternatives differ from that definition. As a second step, it is investigated why there is a call for alternatives in assessment. In order to do so, three perspectives shall be assumed. First, changes in cross-curricular learning and teaching philosophies are considered. Second, problems with traditional, standardized testing are discussed and third, traditional and alternative assessment is compared in terms of quality criteria. A second possibility to approach the topic of alternatives in assessment is examining formative assessment in contrast to summative assessment.

### 2.4.1. Traditional testing vs. alternatives in assessment

Generally speaking, the boundaries between traditional assessment and alternatives in assessment are characterized as rather fuzzy. Brown and Abeywickrama (2010: 18) claim that many assessment methods cannot be classified as one or the other but feature characteristics from both concepts. Additionally, due to these fuzzy boundaries, the definition of alternatives in contrast to traditional assessment is a very subjective matter. While some professionals would classify certain forms of assessment as being traditional, others would argue that they might already be alternative to some extent, because some characteristics of alternative assessment might fit actual traditional approaches or methods. However, characteristics have been identified and utilized to differentiate alternative from traditional assessment. According to Brown and Abeywickrama (2010: 18) and Huerta-Macías (2002 [1995]: 339), traditional language assessment is associated with:

- All forms of standardized tests, including achievement or proficiency exams
- Assessment tools administered within neat time constraints
- Decontextualization, which is therefore considered inauthentic
- Summative assessment
- Product-oriented assessment
- Not including interaction
- Items which ask for discrete, closed answers (e.g. multiple choice)
- Motivation via incentives, which is therefore considered extrinsic

On the other hand, several attempts have been made to define the characteristics of alternative language assessment. Brown and Hudson (1998: 653) merged several, distinguishing features of alternative language assessment that researchers, educators and teachers developed in the 1990s. The idea was to investigate which features of alternative language assessment distinguish them from traditional assessment methods. They combined the ideas from Aschbacher (1991), Herman, Aschbacher and Winters (1992) and Huerta-Macías (2002 [1995]), which add up to 12 characteristics in total (Brown & Hudson 1998: 654). Alternatives in assessment (Brown & Hudson 1998: 653):

1. require students to perform, create, produce, or do something;
2. use real-world contexts or simulations;
3. are nonintrusive in that they extend the day-to-day classroom activities;
4. allow students to be assessed on what they normally do in class every day;
5. use tasks that represent meaningful instructional activities;
6. focus on processes as well as products;

7. tap into higher level thinking and problem-solving skills;
8. provide information about both the strengths and weaknesses of students;
9. are multiculturally sensitive when properly administered;
10. ensure that people, not machines, do the scoring, using human judgment;
11. encourage open disclosure of standards and rating criteria;
12. call upon teachers to perform new instructional and assessment roles.

After examining the contrast between traditional and alternative assessment, the attentive reader might believe that traditional assessment can be equated with what Dochy addressed as a “testing culture” (cf. Dochy 2001: 14), while alternative forms are associated with assessment tools that set themselves apart from traditional tests and examinations. Brown and Abeywickrama (2010: 122) claim that “[e]veryone wants a test for everything, especially if the test is cheap, quickly administered, and scored instantaneously”. Decades of research on language testing quality criteria, item design, etc. as well as powerful lobbying from the influential industry of standardized testing in countries like the United States or Great Britain have led to a certain impeccable status of tests. Nevertheless, at the end of the 20<sup>th</sup> century, language assessment experts began to doubt their impeccability (Bailey 1998: 204; Dochy 2001: 14; Brown & Abeywickrama 2010: 122-123).

#### *2.4.1.1. Changes in learning and teaching philosophies*

From a cross-curricular perspective, Dochy (2001: 13) argues that traditional assessment arises out of traditional instruction. According to Dochy, “[t]he traditional instructional approach viewed learners as passive recipients of information” (2001: 13), believing that knowledge can be transferred from one person to another, which is also associated with the behavioral learning philosophy (Shepherd 2000: 5). In further consequence, the traditional assessment process resembled “testing of basic knowledge, supposedly acquired through tedious drill and practice experiences, rehearsals and repetitions of what was taught in class or in the textbook” (Dochy 2001: 13-14). Current learning philosophies emphasize different instructional approaches (Dochy 2001: 15). Stern (2010: 25) adds that alongside competence orientation, a so-called “neue Lernkultur [new learning culture]” is slowly gaining the upper hand in Austria. That means that pupils learn autonomously, work on complex topics on their own, do research on their own and create some sort of publicly relevant product. According to Stern (2010: 23), this so-called new learning culture becomes apparent in regard to three aspects:

- content – towards a broad variety of topics and teaching/learning methods
- learning – away from simple reproduction of facts towards the development of problem solving skills and autonomous learning
- teaching – towards individualization and student-centered teaching, which demands different types of teacher roles

Teaching and assessment, thus, has to answer to those changes in teaching philosophy. In terms of language teaching, Communicative Language Teaching (CLT) which is the current language teaching approach adopted in Austria, already answers those demands. While for a long time, traditional language teaching pursued a structuralist, grammar-focused approach (Richards & Rodgers 2001: 71), CLT emphasizes the communicative function of languages and the communicative competence of language learners, which requires the integration of language skills and communication in context (Buck 2001: 83). Developments in language teaching methodologies also requires changes in assessment. Testing professionals criticized first and foremost the lack of authenticity concerning language testing tasks because they usually did not resemble real-life application of language (Brown & Abeywickrama 2010: 14). Bachman and Palmer argue in this regard that “for a particular language test to be useful for its intended purposes, test performance must correspond in demonstrable ways to language use in non-test situations” (1996: 9). It is argued that alternatives in assessment might be the better way to answer the demands of a new learning culture and CLT. Huerta-Macías (2002 [1995]: 339) emphasizes that “[a]lternative assessment is different from traditional testing in that it actually asks students to show what they can do. Students are evaluated on what they integrate and produce rather than on what they are able to recall and reproduce”. Therefore, the developing learning culture demands a new assessment culture, away from all sorts of tests and towards alternative assessment tools, which likewise emphasizes the active construction of learning, real life learning situations and a change in teacher role (Stern 2010: 25).

#### *2.4.1.2. Standardized testing*

Apart from dissatisfaction with traditional instruction and ensuing assessment approaches, problems have been voiced especially in countries that have a rich tradition of standardized testing. However, in an Austrian context, where standardized testing has received attention rather recently, only some points of criticism are considered crucial at the moment. The standardized Austrian school leaving exam at the end of upper secondary school cannot be



easily compared to US large scale tests for instance. Nevertheless, standardized testing bears negative potential which led experts of countries with an exceptional history in standardized testing to argue for the pursuit of more learner-centered alternatives (Gallagher 2003: 97, Scott 2004: 36).

Generally speaking, standardized testing has a long tradition. In western countries, standardized tests were encouraged in the 19<sup>th</sup> century as an answer to the industrial revolution and other developments that urged people to attend schools. As a consequence, the educational system was required to assess a rapidly growing number of students (Gallagher 2003: 88). A country that has a rich tradition of standardized testing is the USA, where standardization as a requirement for receiving objective test scores was needed due to the high stakes those tests were related to (Dochy 2001: 14). Law and Eckes (2007: 23) argue that nowadays, standardized tests “carry ever more weight” in terms of access to and graduation from schools, colleges or other educational institutions. Gallagher (2003: 95) claims that in 21<sup>st</sup> century USA, high scores on standardized tests are crucial for decisions about graduation, placement or qualification for financial aid or scholarships. However, the stakes are high for other assessment stakeholders too. Since the middle of the 20<sup>th</sup> century, schools in the USA have been required to submit test scores in order to receive federal funding, a measure that has become known in education as accountability (Scott 2004: 31). Scott (2004: 36) argues that on the basis of accountability measures, a multi-billion dollar testing industry developed in the USA because single states had to implement large-scale tests that were able to provide quantified results.

Criticism of standardized testing as practiced in the USA is manifold (Scott 2004). Standardized testing as an accountability measure is criticized by Scherer (2005: 9), who believes that the unintended effect of large-scale testing downgrades what should be the main function of school, namely the learning process and its outcome. Instead, teaching often degenerates into teaching to the test within what Scott (2004: 30) called a “canned curriculum” that is narrowed due to the preparation of pupils for standardized tests. In this regard, Flinders (2005: 8) states that “[w]hat is tested now determines what is taught” instead of the other way round. In addition, Birenbaum (1996: 5) argues that standardized tests being required to answer to “objectivity and fairness” demands result in the utilization of item formats that require “low level cognitive competencies” without tapping into higher order thinking skills. Another problem with standardized testing is that most scores cannot be used for further teaching. Law and Eckes (2007: 23) claim that scores are usually displayed in the form of percentage points, which is why teachers do not know where

exactly the pupils had difficulties. That is, the results are barely usable for program evaluation, adaption of instruction or the investigation of pupils' strengths and weaknesses. In addition, where standardized tests are administered for selection and allocation purposes, pupils frequently face norm-referenced tests that are designed so that some students fail no matter what (Law and Eckes 2007: 23). The purpose is to rank pupils by comparing their results with the results of their peers (Brown & Abeywickrama 2010: 8). Another essential feature of standardized tests are cut-scores (Law and Eckes 2007: 24). Cut-scores are usually arbitrarily set but have the potential of having tremendous impact on the pupils' individual lives. Where test scores decide the future of individuals, it has been claimed that decisions about cut-scores are crucial.

To conclude, the described shortcomings of traditional, standardized tests were determinative for the creation of the concept of alternatives in language assessment (Brown & Abeywickrama 2010: 122). The idea was to gather additional records of pupils' performances (Brown & Abeywickrama 2010: 123) via instruments that set themselves apart from traditional tests, for instance portfolios, journals, observations, etc. (Huerta-Macias 2002 [1995]: 339-340). While Austria does not have a tradition in standardized testing, the proposed alternative assessment instruments are promoted in Austria in order to answer to the demands of the new learning and assessment culture (cf. Stern 2010; Schmidinger, Hofmann and Stern 2015).

#### *2.4.1.3. Testing Situations*

Huerta-Macías (2002 [1995]: 338) claims that test-situations bear potential problems in comparison to other forms of assessment. Shortcomings of test-situations have been gathered in various sources (Huerta-Macías 2002 [1995]; Dochy 2001; Stern 2010). For instance, one of the drawbacks of tests is that there is always the possibility of pupils failing because they are not familiar with the format of the test, even though they are prepared in terms of knowledge and proficient in terms of required competences (Huerta-Macías 2002 [1995]: 338). In addition, unforeseeable circumstances may compromise the chances of passing tests, for instance having a bad day, being fatigued, ill or having anxiety. Thus, it is argued that there are many conditions which might influence a pupil's performance, even though neither the pupils themselves nor the teachers might be able to affect those positively. As a consequence, Huerta-Macías states that "problems associated with traditional testing often mask what the student really knows, or, in the case of ESL, what

the student can do in her second language” (Huerta-Macías 2002 [1995]: 338). Dochy (2001: 14) states that traditional testing often requires students to work “under stressful conditions and unrealistic constraints”. According to Stern, test situations might also result in anxiety and aversion to learning, especially where high-stakes tests are concerned (Stern 2010: 21). Dochy adds that test results achieved under somewhat less than perfect and inauthentic circumstances affect the future lives of pupils tremendously. Especially when it comes to high-stakes tests, proving language proficiency and the achievement of learning goals with one single performance at one specific point in time is not fair (Dochy 2001: 14).

#### *2.4.1.4. Quality criteria*

Subsection 2.3.2 above was dedicated to a variety of quality criteria against which test usefulness is evaluated. Both approaches, namely traditional testing and alternative language assessment, should be evaluated against those quality criteria. Both approaches usually achieve high scores in different categories. Brown and Abeywickrama (2010: 125) argue that tests and other forms of formal assessment usually lack authenticity and positive washback. Considering the merits and demerits of alternative assessment on the other hand, Brown and Abeywickrama (2010: 124) argue that washback and authenticity are strong qualities while reliability and practicality are perceived as problematic. Thus, alternatives in assessment are frequently criticized, because it is claimed that they lack very important quality criteria that promise useful, high quality assessment (Huerta-Macías 2002 [1995]: 340). Brown and Hudson (1998: 655) believe that overlooking criteria like reliability, validity or authenticity might lead to questionable, possibly irresponsible decisions, thus they argue that teachers have to ensure that alternatives in assessment meet the same appropriate thresholds of assessment usefulness like every other assessment procedure (Brown & Hudson 1998: 655-656). However, objections to either assessment approach based on lacking quality do not have to be accepted as a given. Experts argue that either form of assessment can be optimized in terms of quality due to a broad range of strategies (cf. Brown & Abeywickrama 2010: 125-126).

#### **2.4.2. Formative assessment**

The development of alternatives in assessment can also be examined from a different perspective, which is based on the different assessment purposes that were presented in

Section 2.2. While for a long time the major purpose of assessment was gathering information about pupils' achievement via testing (Wiliam 2011: 3; Turner & Purpura 2016: 255), one rather recently acknowledged purpose of assessment is the diagnosis of pupils' strengths and weaknesses with the objective of enhancing and facilitating their learning processes. Therefore, two concepts are usually distinguished (Rea Dickins & Gardner 2000: 216): summative assessment, which is associated with traditional forms of language assessment (Huerta-Macías 2002 [1995]: 339), and formative assessment, which is associated with alternatives in assessment (Brown & Hudson 1997: 654). This section aims at discussing the concept of summative assessment in contrast to formative assessment.

As described in Section 2.2, one purpose of assessment is the measurement and evaluation of the pupils' achievement of curricular objectives. This function of assessment is called summative and aims to "measure, or summarize, what a student grasped and typically occurs at the end of a course or unit of instruction" (Brown and Abeywickrama 2010: 7). Therefore, common summative assessment methods are achievement tests which are administered at the end of teaching sequences in order to find out whether or not the pupils managed to master the intended learning objectives (Brown & Abeywickrama 2010: 9). Summative assessment can be either classroom based or conducted as large-scale standardized testing, depending on the educational system of the country concerned.

Formative assessment that aims at "informing teaching as a means to supporting learning" (Rea-Dickins 2004: 164) on the other hand is more or less exclusively classroom based. The distinction between summative and formative assessment dates back to the 1970s (Wiliam 2006: 283), when Scriven (1967: 43, quoted in Wiliam 2006: 283) suggested the terms for curriculum evaluation. Two years later, Bloom (1969: 48, quoted in Wiliam 2006: 283) applied the terms to evaluating pupil learning. Until the 1960s, it was assumed that "the causes of any failures to learn lay within the individual learner" (Wiliam 2011:3). It was Bloom's work that acknowledged the significance of instruction and its quality for pupils' achievement or failure. Wiliam (2006: 284) points out that both Scriven and Bloom defined formative assessment by the necessity that it must have some sort of effect on the instruction process. He argues that "the information generated is used to make changes to what would have happened in the absence of such information" (Wiliam 2006: 284) and "[t]he crucial feature is that evidence is evoked, interpreted in terms of learning needs, and used to make adjustments to better meet those learning needs" (Wiliam 2006: 285).

Black and Wiliam (1998) took a similar view. They reviewed about 250 studies on formative assessment from various areas of education in order to collect evidence of its positive effects on learning processes. They reviewed research studies about a variety of assessment procedures “which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged” (Black and Wiliam 1998: 7-8). Besides the general result that effective formative assessment is capable of enhancing learning processes, the review boosted the status of formative assessment’s potential in general pedagogy (Turner & Purpura 2016: 256). However, definitions of formative assessment vary across the field (Wiliam 2011: 9). For instance, Bloom defines formative assessment as “a combination of feedback and instructional correctives” (Wiliam 2011: 9). Black and Wiliam emphasized that assessment “becomes formative assessment when the evidence is actually used to adapt the teaching to meet student needs” (1998: 140). Drawing on Black and Wiliam’s review of formative assessment (1998), Wiliam and Thompson (2008, quoted in Turner & Purpura 2016: 256) emphasized three prevalent processes in order to embed the concept in a theoretical framework:

- Establishing where the learners are in their learning (assessment: teacher/self/peer);
- Establishing where they are going (shared curriculum and intermediary goals);
- Establishing what needs to be done to get them there (instruction, self-/peer-study).

Similarly, Brown and Abeywickrama (2010: 7) define formatively functioning assessment as “evaluating students in the process of ‘forming’ their competencies and skills with the goal of helping them to continue that growth process”. Kahl (2005: 11) claims that formative assessment should detect “misconceptions” about topics or skills while teaching and must therefore be integrated with instruction. Finally, Rea-Dickins and Gardner define formative assessment by saying that it focuses on “improving teaching and learning and to being generally responsive to learner needs” (2000: 217) which, according to them, explains its general attractiveness to teachers (Rea-Dickins and Gardner (2000: 217).

Other researchers doubt that a strict differentiation between formative and summative assessment is essential. Already Scriven (1967) and Bloom (1969) agreed that any form of assessment can have formative functions, even if they are designed for summative purposes. Bennet (2009: 5) claims that assessment that is supposed to function summatively might have formative implications too and vice versa. Wiliam (2011: 10-11) proves these

views by researching seven scenarios where assessment tools were designed to be summative but the involved teachers managed to use some of the results formatively.

In all, what seems to be a shared opinion in the field is that formative assessment serves a distinctive purpose, which is utilizing the elicited information in order to induce change, affecting future instruction and learning processes.

#### *2.4.2.1. Formative assessment in EFL/ESL contexts*

This section explores research on formative assessment in the realm of EFL/ESL teaching. Research that investigates assessment carried out to improve learning was still rare at the beginning of the 21<sup>st</sup> century (Rea-Dickins 2004: 249). Rea-Dickins and Gardner (2000: 215) indicated that the formative, pedagogical function of assessment was neglected for years in EFL/ESL contexts, especially in contrast to the amount of research on language testing.

Rea-Dickins and Gardner (2000) conducted a case study on the construct of formative assessment in nine schools in England and Wales, in which English is learnt as an additional language. They administered questionnaires and interviewed staff in order to learn about the quality of assessment procedures and their purposes. They found that teachers use a wide variety of assessment procedures which serve formative, summative and evaluative functions. However, they argue that the results suggest that the distinction between formative and summative assessment might not be as clear-cut as usually presented (Rea-Dickins & Gardner 2000: 217). In addition, a variety of inconsistencies were encountered in terms of assessment quality, especially in terms of reliability and validity. Finally, while Rea-Dickens and Gardner (2000: 239) acknowledge the appeal of formative assessment, they also emphasize the need for awareness of potential pitfalls.

Leung and Mohan (2004) conducted a case study in two schools near London. The observed two year 4 classes, in which a large number of pupils with different ethnicities learnt English as a second language. Leung and Mohan (2004: 336) wanted to examine how classroom-based formative assessment is accomplished, based on the premise that formative assessment is realized interactionally as part of teaching by teacher-pupil discourse. They observed how formative assessment can affect the pupils' decision making processes while doing certain tasks, a characteristic that, according to the authors, cannot be achieved by traditional summative testing (Leung & Mohan 2004: 342). They discovered that teachers used scaffolding techniques in what they call a "formative

guidance stage” of assessment, in which they guided pupils to correct answers by engaging pupils in discussions (Leung & Mohan 2004: 355).

Cheng, Rodgers and Hu (2004) administered a study consisting of questionnaires for language instructors in three tertiary ESL/EFL settings; Canada, Hong Kong and Beijing. They investigated assessment purposes, methods and procedures used by instructors at University level. As a significant finding, they report that assessment is influenced by the “teaching and learning context where assessment takes place” (Cheng, Rodgers & Hu 2004: 380), for example, the researchers attributed decisions about assessment purposes and instruction on obligatory standardized testing programs in the respective country (Cheng, Rodgers & Hu 2004: 378). Methods and procedures varied across all settings, which can be explained by different teacher beliefs, attitudes and experiences (Cheng, Rodgers & Hu 2004: 378).

The vast complexity of the concept of formative assessment and its relation to summative assessment was examined by Rea-Dickins (2006). By applying an interaction-analytic approach, Rea-Dickins investigated the learner’s role in assessment as well as the role of scaffolding and the impact of teacher attitudes on classroom assessment. The results emphasized that assessment definitely provides language-learning opportunities, for example, language learners evidently showed “some kind of personal progression in their understanding and language development” (Rea-Dickins 2006: 181). However, there was no significant distinction measurable between summative and formative assessment. Both show a potential to enhance language learning. Even if assessment is planned to fulfill solely summative purposes, the collected information could be used formatively. Therefore, she inferred a “complexity of the construct of formative assessment and in particular its relationship with summative assessment” (Rea-Dickins 2006: 183). Her main conclusion was that an even balance between summatively and formatively functioning assessment is crucial (Rea-Dickens 2006: 184).

Colby-Kelly and Turner (2007) investigated formative assessment in a pre-university EAP program in Canada. Their goal was to study the use and effectiveness of formative assessment. They found that only specific forms of formative assessment were used frequently, even though several types of formative assessment were required by the curriculum. However, as a main result, they suggest an assessment for learning approach, that is, “making assessment a primary factor in learning” (Colby-Kelly & Turner 2007: 33). In other words, they consider their research as “a call in the language testing community”

for adopting the paradigm that every language assessment needs to enhance learning (Colby-Kelly & Turner 2007: 33).

Increasing recognition of the importance of formative assessment and assessment for learning in EFL/ESL contexts has resulted in contributions in contemporary language assessment handbooks. Most recently, the “Handbook of second language assessment” (Tsagari & Banerje 2016) recognizes the value of formative assessment respectively assessment for learning in a variety of contributions. In said contribution, Turner and Purpura propose a “learning-oriented approach to language assessment” (Turner & Purpura 2016):

Unlike other assessment approaches, however, LOA [learning-oriented approach to language assessment] prioritizes the interpretation of L2 performance evidence on both learning outcomes and L2 processes, so that goal-referenced decisions can be made by individual classroom agents to further L2 processing and achieve target-like performance. Thus, an LOA approach highlights learning goals, performance evaluation and feedback, and the role they play in developing individual learning progressions (Turner & Purpura 2016: 260).

The authors propose a working framework for implementing learning-oriented language assessment.

Another contribution in the “Handbook of second language assessment” (Tsagari & Banerje 2016) addresses the formative function of assessment as being part of dynamic assessment (Poehner & Infante 2016), an assessment framework that promotes the integration of instruction and assessment in order to grasp learner abilities and student development (Poehner 2008; quoted in Poehner & Infante 2016: 278). The theoretical framework of dynamic assessment is driven by Vygotsky’s sociocultural theory of mind (cf. Vygotsky 1997; Lantolf & Poehner 2014) and the Zone of Proximal Development (ZPD) (cf. Vygotsky 1978). Poehner and Infante emphasize that the distinction to other assessment approaches is that “it intentionally aims to provoke change in the abilities being assessed” (2016: 275). Moreover, for EFL/ESL contexts, they propose a dynamic assessment approach based on interaction between mediators and learners to “examine problems, propose and implement solutions, and reflect on outcomes” (Poehner & Infante 2016: 280) together.



#### 2.4.2.2. Characteristic features of formative assessment

The previous sections revealed the differences between summative and formative assessment and insights on the effectiveness of formative assessment based on recent research. This section aims at displaying major characteristics of formative assessment that might help to fully grasp the concept, namely stakeholder roles, feedback and motivation.

The roles which stakeholders occupy in an educational language assessment setting are one of the prevalent discussions in regard to formative assessment (Turner & Purpura 2016: 257). On the basis of key strategies to formative assessment developed by Leahy *et al.* (2005), Wiliam and Thompson (2008; quoted in Wiliam 2011: 12) defined that teachers, learners and peers are agents that take part in formative language assessment settings as can be seen in Figure 1.

	Where the learner is going	Where the learner is right now	How to get there
Teacher	Clarifying, sharing, and understanding learning intentions	Engineering effective discussions, activities, and tasks that elicit evidence of learning	Feedback that moves learning forward
Peer		Activating students as learning resources for one another	
Student		Activating students as owners of their own learning.	

**Figure 1: Agents and key features of formative assessment (Wiliam & Thompson 2008; quoted in Wiliam 2011: 12)**

Teachers need to provide opportunities for effective formative assessment to take place under conditions that support learning. Wiliam states that “the teacher’s role is [...] to elicit evidence of achievement, and undertake appropriate action to direct, or re-direct the learning in the intended direction” (Wiliam 2011: 12). Turner and Purpura (2016: 258) argue that formative assessment requires “an effective learning environment” which should be created by the teacher. In order to do so, teachers need to possess know-how and a variety of skills. Important knowledge and skill domains have been identified by Heritage (2007). Domain or subject knowledge is required (Heritage 2007: 142). That is, teachers must have knowledge and skills in a specific domain in order to set objectives, know about successful performances and what an efficient “framework to guide assessment and instruction” (Heritage 2007: 142) might look like. The second knowledge domain is pedagogical and didactic knowledge (Heritage 2007: 143). As Pellegrino, Chudowsky and Glaser (2001; quoted in Turner & Purpura 2016: 258) already suggested, it is important for teachers to know how learning actually takes place. Understanding of cognition and learning in a specific subject is vital. In the domain of language teaching, sufficient knowledge about language teaching approaches and methods are indispensable. Another

knowledge domain deals with students' previous encounters with the subject content, the development of student achievements and the learning strategies individual students usually apply. Teachers need to be familiar with pupils' previous learning in order to build on it (Heritage 2007: 143). Finally, assessment knowledge belongs to the essence of administering successful formative assessment. That is, teachers must be aware of the importance of quality criteria and possible sources of interference when they are obtaining valuable data by applying formative assessment (Heritage 2007: 143; Law and Eckes 2007: 37-38).

Peer assessment is considered to be a key strategy towards formative assessment (Leahy *et al.* 2005: 23). Peers can take up teacher-similar roles (Wiliam 2011: 13). According to Wiliam, peers might have "unique insights into learning" (2011: 13). In addition, due to distinct relationships between peers, their insights or comments might have different effects on pupils compared to those of the teachers (Wiliam 2011: 13).

Finally, pupils themselves are agents of formative assessment. Turner and Purpua (2016: 258) argue that pupils need to internalize that they are responsible for their learning. Leahy *et al.* (2005: 23) described this process as activating pupils as "owners of their learning". In this respect, Wiliam (2011: 13) emphasizes the importance of self-regulated learning, which is an aspect promoted by self-assessment. Additionally, in a study in which teachers worked with Leahy *et al.* in workshops about assessment that promotes learning, teachers generally confirmed the effectiveness of self-assessment. They claim that it produced accurate evaluation and provided distinct approaches to learning (Leahy *et al.* 2005: 23).

On the whole, it seems most important to acknowledge all stakeholders, teachers, learners and peers as agents of formative assessment. The most crucial task might be to engage learners in the learning process, making them responsible for their development and achievements.

Feedback is a central component of formative assessment (Black and Wiliam 2012: 18). According to Wiliam (2011: 4), feedback became an important factor in assessment when Bloom (1984) used the term to describe immediate error identification by tutors in one-to-one tutor-learner situations combined with clarification on those errors. In the late 1980s Sadler (1989: 121; quoted in Wiliam 2011: 4) modified Bloom's definition of feedback by building on Ramaprasad's (1983) work, arguing that the information about a learning gap (e.g. a mistake in language use) becomes feedback if it is used to alter teaching instruction, which, in further consequence, helps to close the learning gap. In other words, information

about the learning gap alone is not considered feedback. In their review of AFL research in second language learning, Colby-Kelly and Turner indicate that Tunstall and Gipps (1996; quoted in Colby-Kelly & Turner 2007: 13) describe two different kinds of feedback, which are evaluative and descriptive feedback. However, Rea-Dickins (2006: 168) argues that Tunstall and Gipps emphasize a “greater link between descriptive types of feedback and formative assessment” (Rea-Dickins 2006: 168). A similar distinction between feedback functions has been proposed by Black and Wiliam (1998), who suggest two functions of feedback, namely directive and facilitative. Shute (2008: 157) summarizes Black and Wiliam’s descriptive feedback as being very specific, telling the pupils explicitly what needs to be worked on, and facilitative feedback as more general remarks, engaging pupils in reflective, self-regulated improvement processes. In further consequence, Shute (2008: 154) talks about formative feedback and defines it as the information given to pupils with the intention of positively enhancing learning. However, she emphasizes that how feedback is delivered to the learner is crucial for the effectiveness of feedback (Shute 2008: 154). Hattie and Timperly (2007: 84) propose a “continuum of instruction and feedback”, because, as Kulhavy states, the feedback process “takes on the forms of new instruction, rather than informing the student solely about correctness” (Kulhavy 1977: 212). Wiliam (2010: 21) adds in this respect that effective feedback does not necessarily have to be given in retrospect about learning that already took place. He argues that the most effective feedback might focus on the next step in the learning process. The consequence is that feedback would not be related to mistakes or bad performances (Wiliam 2010: 21).

Quality is essential for feedback to be perceived as effective. Weeden and Winter (1999: 11; quoted in Colby Kelly & Turner 2007: 13) ascertain that feedback can’t be effective if the quality is not appropriate. In particular, they found that feedback is often not specific enough and therefore considered as insufficient by pupils. Feedback specificity is also discussed by Shute (2008: 157), who gathers research results that suggest a need for specific feedback in contrast to general feedback. She summarizes that detailed and clear information is valuable and serves as a “general guideline” (Shute 2008: 158).

Summing up, feedback is acknowledged to be an influential factor in language learning and assessment that can work positively as well as negatively. Most importantly, feedback that is considered formative works by engaging pupils in the feedback process and as a tool to achieve change in the pupils’ learning and development (Rea-Dickins 2006: 167-168).

A major cornerstone of formative assessment is the value of motivation for learning. Motivation has been described as a significant factor for human behavior and learning. In general, Stiggins (2001) argues that motivation is the most crucial aspect responsible for learning to take place. In terms of language learning, Gardner claims that motivation “orient[s] the student to try to acquire elements of the second language, and includes the desire the student has for achieving a goal, and the amount of effort he expends in this direction” (Gardner 1979: 197). In other words, how much effort put into the learning of a language is affected by motivation. In addition, Dörnyei (2001: 5) argues that

in the vast majority of cases learners with sufficient motivation can achieve a working knowledge of an L2, regardless of their language aptitude or other cognitive characteristics. Without sufficient motivation, however, even the brightest learners are unlikely to persist long enough to attain any really useful language.

Thus, it is crucial for teachers to know if and how motivation for learning can be generated or maintained. According to Stiggins (2001: 36), assessment is the most powerful tool for teachers to modify the extent of effort students put into learning a language.

In order to apply concepts of motivation to language learning, it is important to be aware of associated concepts of motivation. In this regard, the most useful distinction of motivation types has been suggested by Deci and Ryan (1985; quoted in Harlen 2012: 172) as part of their self-determination theory. They distinguish two relevant types of motivation based on different incentives that are involved in the process. On the one hand, intrinsic or internal motivation is associated with joy and satisfaction evoked by the learning process itself. Extrinsic motivation on the other hand is associated with potential benefits from learning, regardless of the actual content that has been learned (Harlen 2012: 173). In the latter case, those benefits are the determinative factors for the learning process but, according to Dörnyei, they “do not increase the inherent value of the task or the task outcome, and neither do they concern other important learning aspects such as the learning process, the learning environment or the learner’s self concept” (Dörnyei 2001: 128).

There is research that links the concept of motivation with the distinction between summative and formative assessment. As Harlen notes, “studies on summative assessment provide strong evidence of negative impact on students’ motivation for learning” (2012: 175) which, does not however, mean that there are no positive effects on motivation at all. While extrinsic motivation is said to undermine intrinsic motivation (Deci *et al.* 1999), extrinsic motivation can promote good performances. Nevertheless, teachers need to be

aware that there is a possibility that frequently administered summative assessment tools like tests might contribute to the notion that learning serves the purpose of performing well instead of being satisfied by the learning process and outcome itself (Harlen 2012: 177). In further consequence this is associated with looking for the “easiest route to the necessary outcome” (Harlen 2012: 176) because a good performance results in high scores on tests.

Formative assessment is “strongly linked to increased intrinsic student motivation” (Moss & Brookhart 2009: 16), which is considered to be the preferred type of motivation that drives learning (Harlen 2012: 173). Learning for the sake of interest instead of profiting from certain rewards like grades, prizes or money is therefore regarded as more sufficient, effective and valuable. According to Moss and Brookhart (2009: 16), formative assessment might increase motivation for learning in terms of self-efficacy, self-regulation, self-assessment and self-attribution. A factor which highly influences the components for motivation for learning is feedback. Teachers must be careful though because feedback might promote motivation positively as well as negatively, for instance, feedback that derives from summative assessment is mostly “judgmental and rarely formative” (Harlen 2012: 176).

To summarize, it can be concluded that assessment can have positive and negative effects on motivation and learning in general. For assessment that is supposed to grow beyond the result of extrinsically motivated pupils or teaching to the test, a high amount of support and formative feedback are crucial factors.

#### *2.4.2.3. A European approach to formative assessment: Assessment for learning (Afl)*

From a European perspective, research on formative assessment is the groundwork of a movement in Great Britain that gave rise to the Assessment Reform Group (ARG) in 1989. The ARG was a group consisting of university professors that aimed at influencing assessment policy and practice by researching assessment. As stated by Wiliam (2011: 10), the group acknowledged the power of assessment to enhance the learning process. However, due to the broad interpretation of the term formative assessment, they pleaded for a new term that shall be used in this regard:

The term ‘formative’ itself is open to a variety of interpretations and often means no more than that assessment is carried out frequently and is planned at the same time as teaching. Such assessment does not necessarily have all the characteristics just identified as helping learning. It may be formative in helping the teacher to identify areas where more explanation or practice is needed. But

for the pupils, the marks or remarks on their work may tell them about their success or failure but not about how to make progress towards further learning. (Broadfoot *et al.* 1999: 7)

In order to react to the variety of interpretations of formative assessment, the ARG proposed the term *assessment for learning* in contrast to *assessment of learning*. The result of the group's work, building up on the research into formative assessment, is a series of publications. Following in the footsteps of formative assessment research, the ARG emphasizes that “[a]ssessment is one of the most powerful educational tools for promoting effective learning” on the one hand, and might be able to “raise pupils’ achievements” and, therefore, standards on the other (Assessment Reform Group 1999: 2). Although, the terms *formative assessment* and *assessment for learning* are closely related, the differentiation is based on the emphasis on analyzing elicited information in order to find out “where the learners are in their learning, where they need to go and how best to get there” (Broadfoot *et al.* 2002: 2-3). In addition, Wiliam (2011: 10) points out that the term formative assessment refers to the function of assessment, while assessment for learning already implies the purpose of the assessment procedure. In order to ease the adaption of assessment for learning into educational policy, the ARG published pamphlets which summarize the concept, research results and practical advice. Among those publications promoted by the ARG is a collection of vital principles explaining how *assessment for learning* can be administered (Broadfoot *et al.* 2002):

1. [Afl] should be part of effective planning of teaching and learning.
2. [Afl] should focus on how students learn.
3. [Afl] should be recognized as central to classroom practice.
4. [Afl] should be regarded as a key professional skill for teachers.
5. [Afl] should be sensitive and constructive because any assessment has an emotional impact.
6. [Afl] should take account of the importance of learner motivation.
7. [Afl] should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed.
8. Learners should receive constructive guidance about how to improve.
9. [Afl] develops learners’ capacity for self-assessment so that they can become reflective and self-managing.
10. [Afl] should recognise the full range of achievements of all learners.

Broadfoot *et al.* (2002) emphasize that these principles can be used as a framework to design assessment procedures that exploit the potential benefits of assessment for learning.

### **2.4.3. Alternative assessment tools in EFL contexts**

Until now, the discussion about alternative assessment has not solely referred to second or foreign language teaching. Regardless of the subject taught, traditional tests will have the same shortcomings and alternative assessment might usually have the same characteristics. However, several alternative language assessment methods are proposed in the literature which fulfill special language learning needs. While most of them can be adapted to a variety of subject matters, they are repeatedly reviewed in language teaching contexts. Language assessment differs from other disciplines “not only because of the complexity of the domain [...] but also because of the different types of tests” (Brown & Hudson 1998: 657). Additionally, Turner and Purpura (2016: 257) argue that language teaching and assessment is particular due to the “often non-linear nature of L2 acquisition”. Huerta-Macías (2002 [1995]: 339-340) briefly mentions ESL-specific alternative assessment instruments, for example “journals, [...] roleplays, [...] self-evaluation questionnaires, [...] and teacher observations” (Huerta-Macías 2002 [1995]: 340). Brown and Hudson (1998: 657-667) classify assessment methods within three categories, which are selected-response assessments, constructed-response assessments and personal-response assessments. Looking at the criteria of alternative assessments they collected, personal-response assessments feature most of the criteria because “personal-response assessments are beneficial in that they provide personal or individualized assessment, can be directly related to and integrated into the curriculum, and can assess learning processes in an ongoing manner throughout the term of instruction” (Brown & Hudson 1998: 663). Among personal-response types of assessments, Brown and Hudson emphasized three particular ones, which are portfolios, self- and peer assessment and conferences. All three have been discussed in a variety of handbooks about alternative language assessment (cf. Brown and Abeywickrama 2010: 126-153) In the following, alternative language assessment tools shall be outlined as deemed essential by Brown and Abeywickrama (2010).

#### ***2.4.3.1. Portfolios***

Brown and Abeywickrama (2010: 130) classify language portfolios as the most popular alternative language assessment in the research literature on alternative language teaching. Yin emphasizes that the key idea behind portfolio assessment is probably that “teaching can be done through the process of assessment, and assessment can be done through the process of teaching” (Yun 2014: 674). It is a means of bringing teaching, learning and

assessment together, processes which have formerly been considered separate aspects of language teaching and learning (Yun 2014: 659).

Paulson, Paulson and Meyer (1991: 60) defined a portfolio as

a purposeful collection of student work that exhibits the student's efforts, progress, and achievements in one or more areas. The collection must include student participation in selecting contents, the criteria for selection, the criteria for judging merit, and evidence of student self-reflection.

In terms of language teaching and assessment, portfolios have been defined as purposeful collections of student work that display language skills, abilities and development in language learning or specific areas thereof (Genesee & Upshur 1996: 99, quoted in Brown & Abeywickrama 2010: 130; Brown and Hudson 1998: 662). Different types of portfolios have been described in educational contexts. For instance, O'Malley and Valdez Pierce (1996: 37) discuss three types of portfolios. First, a portfolio can be a collection of the most outstanding works of students, hence called a showcase portfolio. The second type is called a portfolio collection. It provides information about learning processes and products and is compiled over a certain predefined time period. The third type is called an assessment portfolio, and has the purpose of providing evidence about a student's progress in terms of learning objectives. In addition, besides typical student work, assessment portfolios contain self-reflection and self-assessment tasks (O'Malley & Valdez Pierce 1996: 37).

Studies about portfolio assessment in language teaching contexts disclose significant advantages. For instance, Brown and Hudson (1998: 664) describe three advantages of portfolio language assessment. First, portfolios might enhance the learning processes of pupils. Second, incorporating portfolios into teaching and assessment alters the role of the teacher and third, it might lead to the improvement of assessment in general (Brown & Hudson 1998: 664). Beneficial aspects of portfolios can be found in various publications (Paulson, Paulson & Meyer 1991; Brown & Hudson 1998; Brown & Abeywickrama 2010; Stern 2010). According to Paulson, Paulson & Meyer (1991: 61), portfolios provide the possibility of assessing student work "in a broader concept: taking risks, developing creative solutions, and learning to make judgments about their own performances" instead of assessing merely the achievement of learning objectives. In addition, they argue that portfolio assessment supports students in becoming self-reliant learners (Paulson, Paulson & Meyer 1991: 61). Regarding this aspect, Brown and Abeywickrama (2010: 131) argue that portfolios foster pupils' responsibility for their learning and intrinsic motivation. Brown and Hudson (1998: 664) explain that portfolios offer the opportunity to foster pupil-



teacher collaboration. By doing so, teachers assume an alternative role in the teaching and assessment process of being a coach or facilitator to the pupils (Brown & Hudson 1998: 664). Stern emphasizes the difference between tests and portfolios, which are administered over a period of time (2010: 69). Therefore, Stern argues that portfolios provide an appropriate, diverse sample of data considering pupils' work, abilities and skills. Brown and Hudson (1998: 664) claim that portfolios provide an opportunity to assess students in "using meaningful language to accomplish authentic tasks in a variety of contexts and situations". Finally, Stern (2010: 69-70) suggests that by replacing achievement tests, portfolios might lead to the disappearance of test anxiety.

Portfolios are usually granted high ratings in terms of authenticity and washback (Brown & Abeywickrama 2010: 134). Tasks and assignments can be designed to build on the pupils' lives. Considering the assessment of writing, Weigle argues that a portfolio can include "writing samples that were written for some authentic purpose other than the evaluation of writing *per se*" (Weigle 2002: 203), which ensures authenticity. Beneficial washback is achieved, for instance, by incorporating self- and or peer-assessment and collaborative feedback sessions with the teacher. O'Malley and Valdez Pierce (1996: 35) note that pupils are required to "reflect on their work, analyze their progress, and to set improvement goals". Therefore, it can be argued that portfolios have an impact on the learning process.

Portfolios count as a personal-response assessment method (Brown & Hudson 1998: 663), a characteristic that holds potential drawbacks for pupils as well as teachers. Especially in terms of practicality, portfolios compare unfavorably to other assessment techniques. Yun argues that the implementation is often tricky and not without obstacles if teachers and pupils are not used to portfolio assessment (Yun 2014: 668). In addition, portfolios take a considerable amount of time to finish and to evaluate (Brown & Hudson 1998: 665). Aydin (2010: 199), for instance, revealed that some pupils experience portfolio assessment as "tiring, and [that] it takes too much time".

Besides practicality, reliability issues are frequently voiced (Brown & Hudson 1998: 665). Stern (2010:70) argues that problems with reliable grading or teacher bias can be problematic. Yun (2014: 673) summarizes that some scholars emphasize the alternative nature of portfolio assessment, which is why it might require altered concepts of assessment usefulness characteristics altogether.

The implementation process is crucial for successful portfolio assessment because it is the “determinant of [its] usefulness” (Yun 2014: 660). Therefore, incorporating portfolios in language teaching and assessment requires several teacher decisions about the portfolio’s purpose, content, pupil involvement, etc. (cf. Yun 2014: 668f.). However, useful guidelines for successful implementation have been published, for instance by O’Malley & Valdez Pierce (1996). Besides, Brown and Abeywickrama add that several steps are very important in order to successfully implement portfolio assessment (2010: 131). What they deem very important are clear objectives. The pupils must know the purpose of the portfolio and the criteria against which it will be evaluated. In addition, they argue that guidelines for compiling the portfolio are crucial (Brown & Abeywickrama 2010: 131-132). Besides clear objectives, Brown and Abeywickrama (2010: 133) state that teachers must designate time for the assessment process. First, there should be time “within the curriculum” (Brown & Abeywickrama 2010: 133), which means that pupils should have enough time to work on their portfolios. Moreover, periodic reviews and conferencing sessions must be scheduled because they fulfill formative functions (Brown & Abeywickrama 2010: 133).

#### *2.4.3.2. Self- and peer-assessment*

Self-assessment in language teaching has increased over the last decades (Oscarson 2014: 713). According to Oscarson (2014: 713f.), self-assessment has gained in popularity due to various changes in language teaching, for instance, the emergence of communicative language teaching or the move towards learner-centered teaching. According to Brown and Abeywickrama (2010: 144-145), self- and peer-assessment are based upon two crucial principles of Second Language Acquisition. These are autonomy and intrinsic motivation, two principles that reciprocally influence language learning.

From a theoretical general education point of view, self-assessment is “an integral part of effective acquisition of knowledge and skills” (Oscarson 2014: 716). The benefits of self- and peer-assessment procedures in language learning contexts have been summarized, for instance by Brown and Hudson (1998). They indicate that involving students in the assessment process can boost pupils’ motivation, encouragement and self-reliability in language learning (Brown & Hudson 1998: 665-666). Claims have been made by Nunan (1988; quoted in Ekbatani & Pierson 2000: 2) that self-assessment might be a valuable asset to classroom assessment, because it promotes self-awareness, which describes the development of learning skills about how to learn best. Moreover, research into self-

assessment shows that it “promotes reflection and helps learners to take responsibility for their own learning, it enables [learners] to see gaps in their learning and enables learners to take risks” (Ekbatani 2000: 6–7). Stern (2010: 55) claims that self-assessment procedures enable pupils to differentiate between satisfying learning outcomes and those in need of improvement. Therefore, fostering the development towards autonomously working language learners is a major concern in ESL/EFL contexts. The assessment of strengths, weaknesses and progress by the students themselves leads to “a heightened awareness of learning and of achievement levels in the various linguistic skills” (Dam and Legenhausen 1999: 98).

In 1991, the Council of Europe proposed the implementation of the European Language Portfolio (ELP), which is a tool used by learners for self-assessment (Little 2009: 3) that, in further consequence, promotes learner autonomy (Lenz 2004: 22). The main functions of the ELP are defined by Lenz (2004: 22) as being a companion to the learner in the learning process and as serving as a documentation tool. According to Little (2009), the ELP is a device where learning meets assessment. It entails three parts, which are the language passport, the language biography, which is a primarily formatively functioning assessment tool, and the dossier. In terms of assessment the centerpiece is a variety of checklists entailing can do-descriptors. The descriptors are designed in a way that learners can record and reflect on certain learning objectives and the learning process. Therefore, the ELP is a tool that facilitates the “visibility” of learning processes (Kohonen 2001: 10). It is assumed to be the basis for a new language learning culture that acknowledges what the Common European Framework of Reference (CEFR) intended to be in the first place, which is “the full integration of learning, teaching and assessment” (Little 2009: 17).

Peer-assessment can be defined as the process whereby peers evaluate or critique learning outcomes (Topping *et al.* 2000: 150; quoted in Hansen Edwards 2014: 730). In EFL contexts, the process is also known as “peer response, peer editing, or peer review” (Hansen Edwards 2014: 730) especially in assessing writing. Peer assessment is supported by interactionist theory, which emphasizes progress in language acquisition via communication, meaning negotiation between peers (Hansen Edwards 2014: 731). It is frequently meant to be formative, thus accompanying learning processes and aiming at the improvement of learning by expressing and receiving feedback.

Nowadays, self- and peer-assessment is actively proposed by various European and Austrian institutions. For instance, the OECD argues that developing self-assessment skills

is one of the major assets to alternative assessment procedures (OECD 2005: 4). The Council of Europe promotes self-assessment because it is “an effective complement to tests and teacher assessment” (2001: 191), noting that it might work best if it is conducted in correlation with descriptors that are based on objectives and standards. The ELP consists for a large part of self-assessment procedures. Austrian institutions like the *Österreichisches Zentrum für Persönlichkeitsbildung und soziales Lernen (ÖZEPS)* ‘Austrian center for personal development and social learning’ have acknowledged the potential of self- and peer-assessment and argue in favor of its implementation in Austrian schools (Stern 2010: 55).

#### **2.4.3.3. Journals**

Journals are frequently categorized as alternatives in language assessment. A journal, learner diary or learning log is a written document that contains a battery of questions which provoke students to think, reflect and write about their learning experiences critically (Stern 2010: 55). It is dedicated to make learners aware of their learning progress as well as positive and negative developments in their language proficiency (Brown & Abeywickrama 2010: 134).

Brown and Abeywickrama (2010: 136-138) provide guidelines for the smooth implementation of journals as classroom-based assessment. They argue that carefully introducing pupils to journal writing is of paramount importance. It is essential that students know what to include, therefore guidance like predesigned questions or criteria might be useful. In addition, the objectives or the purpose of the journal needs to be stated clearly. Even though journals are considered an alternative in assessment, it is important to declare whether or not the journals will be evaluated and how the evaluation might be part of a final grade. If so, the criteria must be stated concisely and must be understood. According to Stern (2010: 55), teachers can deal with the evaluation of journal work in two possible ways. On the one hand, evaluation at regular intervals is a possibility. The advantage is the discovery of the students’ attitudes and motivation. On the basis of the journal entries, the teacher might be able to alter future instruction. In this regard, a collaborative approach is promoted, that is a journal that implies interaction between teacher and pupils. On the other hand, it could be beneficial for pupils and their teacher if the journal is not handed in for regular evaluation. Students would have the possibility to write more freely. They could vent their frustration and would not exaggerate their learning processes and experiences

(Stern 2010: 55). In terms of evaluation, Law and Eckes (2007: 141) argue that it might be better to leave out correction of any sort in order to encourage pupils' engagement in journal writing.

Journals fulfill different purposes and appear in a variety of forms. In terms of language teaching contexts, a classification of journals fulfilling a variety of purposes can be found in Brown and Abeywickrama (2010: 136-137). Popular formats are language-learning logs but also journals dedicated to specific language skills or features, like grammar journals or reading logs (Brown & Abeywickrama 2010: 136-137).

As stated in Brown and Abeywickrama (2010: 138-139), journals, similar to portfolios, score very well in terms of washback and authenticity. If administered properly, "the potential in dialogue journals is off the charts" (Brown & Abeywickrama 2010: 138). Journals receive critique when practicality is considered. Administration takes great effort and much time goes into their design and evaluation. Reliability might be problematic; however, "[r]eliability can be maintained by the adherence of journal entries to stated purposes and objectives" (Brown & Abeywickrama 2010: 138).

#### *2.4.3.4. Observations and rubrics*

To some, it might be surprising that observations, which are part of every teacher's daily routine, could qualify as an alternative in language assessment. However, in order to assess pupils' language performances, teachers have to gather information about what pupils can do in a language. While several assessment tools are usable to collect information, Brown and Abeywickrama (2010: 141) argue that teachers usually know their students very well without the help of tests or other forms of examination. Law and Eckes (2007: 18) emphasize that assessment becomes effective if it is based on appropriate, representative data about student knowledge, skills and abilities from various sources. They call this balanced assessment, meaning that the pupils' work and progress should be measured and evaluated via a variety of procedures with different purposes (Law & Eckes 2007: 39). In this regard, the concept of triangulation in data collection is acknowledged in the area of language assessment. "[T]riangulation refers to the collection of data or information from three different sources or perspectives" (Huerta-Macías 2002 [1995]: 341) in order to ensure reliability. Observation is one viable form to gather information about a pupil's language learning process, because every contribution of the pupils, every question and response constitutes information for teachers about their pupils' learning process and

language skills. It is also argued that assessment based on a variety of different procedures might “counterbalance the effects of large-scale, high stakes tests” (Law & Eckes 2007: 117).

For observation to be effective, teachers should have systematized the process of observation, which means that it is important to define which types of information are important, and how it is possible to gather and record them. Brown and Abeywickrama (2010: 141) describe observation as an alternative in assessment if it aims to observe pupils’ performances while pupils are not aware of being evaluated in order to avoid any bias due to the assessment situation. Thus, observation is an assessment method which requires close integration of instruction and assessment. What can actually be observed in a language classroom has been defined by Anthony (1991; quoted in Law and Eckes 2007: 117), who proposed the theoretical framework The Quad. The framework distinguishes four areas of classroom assessment. Two domains within The Quad deal with formatively working observation; one focuses on observing the learning process, the other domain emphasizes the observation of the product. In addition to Anthony’s The Quad, Cambourne (1988; quoted in Law and Eckes 2007: 119) classifies several categories of information as an approach to structuring observation endeavors in ESL classrooms. In terms of assessment, particularly noteworthy categories are learning strategies pupils use to develop their English skills and “the control they display over language in all its forms” (1988; quoted in Law and Eckes 2007: 119).

Finally, observations usually need to be recorded. Law and Eckes remind that the collection of data must be “simple, efficient, and they must not interfere with [...] teaching” (Law & Eckes 2007: 158). Therefore, a systematized process is crucial. In order to do so, checklists, rubrics or anecdotal notes have proven successful (cf. Law & Eckes 2007: 158-179; Brown & Abewickrama 2010: 143-144).

According to Brown and Abeywickrama (2010:128), the use of rubrics is one “hot issue” in language assessment. O’Malley and Valdez Pierce (1996: 5) emphasize the helpful nature of rubrics in regard to performance assessment, which is based on observation and judgment. Rubrics assist the assessment process by providing assessment criteria. The criteria described are linked with stages of performance, which are sometimes subject to numerical scaling, that is, every performance level is usually assigned a number in ascending order (O’Malley & Valdez Pierce (1996: 5).

Aschbacher (1991; quoted in O'Malley & Valdez Pierce 1996: 5) notes that a crucial aspect of rubrics-based assessment is the availability of rubrics criteria for pupils, because that supports self-evaluation processes. Self-assessment is facilitated because rubrics provide points of reference for learning objectives as long as they are designed for pupils to use. This type of rubrics is also called instructional rubrics and need to describe criteria as concisely as possible in understandable language as well as refer to possible weaknesses in student work and appropriate strategies to overcome them (Andrade & Boulay 2003: 22).

In Austria, the BIFIE picked up on the beneficial aspects of rubrics and published rating scales and accompanying guidelines for the assessment of writing performances in EFL at CEFR levels B1 and B2. Especially important is the rubric for assessing written performances at CEFR level B2 because teachers are obligated to assess the written part of the partly standardized school leaving exam in English at CEFR level B2 based on the rubric. Of course, the scales should be made available for students to familiarize themselves with the criteria and to provide the possibility of self-evaluation (BIFIE 2014).

#### *2.4.3.5. Performance assessment*

In areas where standardized testing is the main source of information on pupils' competences, performance assessment has gained a lot of attention over recent years because it is described to measure pupils' actual language performances (Brown & Abeywickrama 2010: 126). It is also considered to be a variety of authentic assessment (cf. O'Malley and Valdez Pierce 1996: 4) which, in turn, is used as an umbrella term for several assessment tools. What they share is the perception of authenticity as a crucial feature of language assessment.

Generally speaking, performance assessment comprises constructed response tasks or personal response tasks, which are any tasks that require pupils to produce language instead of choosing provided answers (Yu 2014: 617). In addition, performance assessment involves the accomplishment of "complex and significant tasks, while bringing to bear prior knowledge, recent learning, and relevant skills to solve realistic or authentic problems" (Herman, Aschbacher & Winters 1992: 2). In the realm of language assessment, Bailey discusses performance assessment as requiring students to accomplish tasks involving "comprehending and producing language under the types of contextual constraints that would be involved in performing one's job" (Bailey 1998: 208). This, by definition, emphasizes the importance of authentic tasks and authentic stimuli.

Additional, specific characteristics have been summarized by Herman, Aschbacher and Winters (1992: 6-7). They argue that performance assessment involves the production of language in all possible forms, higher-order thinking usually induced by open-ended questions and integration of language skills and content knowledge. In addition, the process as well as the product are valued and therefore assessed (Herman, Aschbacher & Winters 1992: 7). All of those characteristics overlap with the features established for alternatives in language assessment, which is why performance based assessment is often discussed as being an alternative to traditional testing in its various forms.

## **2.5. Alternatives in (language) assessment in Austria**

Eder, Neuweg and Thonhauser (2009) analyzed assessment in Austrian schools as part of the “*Nationaler Bildungsbericht 2009 [National education monitoring report for Austria] 2009*”, which is a triennially published report on the situation of education in Austria provided by the BIFIE. They proposed development prospects in assessment for Austria. One of the prospects was the regular use of innovative, alternative forms of assessment. As part of the “*Nationaler Bildungsbericht 2015 [National education monitoring report for Austria 2015]*”, Schmidinger, Hofmann and Stern (2015: 71) dedicated a contribution to the assessment in Austria, emphasizing the importance of formatively functioning alternatives in assessment. They acknowledge that there has been no comparable research on formative assessment in Austria as it has been in the English-speaking educational community. In addition, they argue that Austrian teachers hold on to the summative function of assessment and that a systematic use of formative assessment can only be perceived rudimentarily (Schmidinger, Hofmann & Stern 2015: 72). Nevertheless, certain recent efforts aim at developing a formative assessment culture in Austria. For instance, the ÖZEPS published the pamphlet “*Förderliche Leistungsbewertung [Formative assessment]*” (Stern 2010) that emphasizes and recommends the potential of assessment in terms of facilitating learning and discusses the shortcomings of traditional tests. Stern (2010) provides a variety of practical suggestions for the implementation and administration of formative assessment routines in school, ready to use by teachers. Besides, Schmidinger, Hofmann and Stern (2015: 75) claim that progress in formative assessment use requires a newly developed LBVO which is hoped to foster self-reliant learning and the incorporation of self and peer assessment. A new LBVO is expected to be implemented in the next legislative period (Schmidinger, Hofmann and Stern 2015: 76).



What has been implemented fairly recently in the Austrian educational assessment system are processes of standardization. Two essential parts of this development are the introduction of educational standards in connection with the CEFR and the standardization of the Austrian school-leaving exam. Eder, Neuweg and Thonhauser (2009: 265-267) argued that assessment needs to be more objective and partly externally administered in order to tackle regional differences and to provide comparability in pupils' performances. According to Schmidinger, Hofmann and Stern (2015: 59), standardization processes were introduced to achieve objectivity and fairness.

It can be assumed that the implementation of educational standards and the orientation for competences has had an effect on teaching and assessment. Teaching towards competences leads to assessing competences; in the case of language learning, it is first and foremost a question of communicative competence. In order to assess communicative competence, assessment must take place in authentic contexts and provide possibilities for pupils to draw upon their experiences. Therefore, Brock (2010: 352) emphasizes that assessment cannot resemble mere reproduction of knowledge. A change in learning culture requires a change in assessment. Ritt (2008: 387) argues that alternative, formative forms in assessment are of paramount importance in order to meet standards. He sees a strong connection between instruction that is based on educational standards and competences, and formatively functioning assessment that fosters the development of competences and initializes the coordination of insights and future instruction (Ritt 2008: 387).

In conclusion, assessment development in Austria has recently been dominated by standardization. However, what started in language testing as a movement to tackle shortcomings in traditional, usually standardized testing in countries like the United States of America and has become known as alternatives in assessment are becoming the center of attention in Austria. The formative function of assessment in particular has been thoroughly researched internationally and its effectiveness has been proven in various areas, including EFL/ESL teaching. On the grounds of the lack of considerable research on the administration of formative assessment in Austria (Schmidinger, Hofmann and Stern 2015: 71) and the tendency of emphasizing summative assessment by all assessment stakeholders (Schmidinger 2013: 2), Schmidinger, Hofmann and Stern (2015) recommend the use of alternative, innovative and formative assessment; they plead for distinct Austria-related research in the area of formative assessment and changes in educational policy that will consolidate the concept as obligatory practice.

### **3. Empirical study**

This part of the diploma thesis is dedicated to the empirical study that was conducted in order to gain insights into assessment practices and teacher opinions and attitudes concerning alternatives in assessment in Austrian EFL classrooms. Before the collected data are analyzed and discussed, the research questions and objectives, research methodology, research design and procedural issues are explained concisely.

#### **3.1. Research questions and aims**

The overall aim of the study is to investigate language assessment practices in Austrian EFL classes with a distinct focus on alternatives in assessment. While Chapter 2 established the theoretical framework for the study by reviewing the current state of research on alternatives in language assessment, its characteristics in contrast to traditional language assessment and the effectiveness of formative assessment, it is unknown to a certain extent which types of language assessment are actually used in Austrian EFL classrooms and what the factors are that seem critical for the teachers' decisions to use one or the other assessment tool. Examining the reasons for certain teacher behavior is part of the teacher cognition research agenda (Borg 2009: 1). Borg describes the term teacher cognition as research into what "teachers know, believe and think" (Borg 2003: 81). He argues that teacher cognitions, like beliefs or attitudes interact with their instructional choices (Borg 2003: 81). Buehl and Beck (2015: 66) agree with this statement by acknowledging the potential effect of teacher beliefs on teacher practice by reviewing research about teacher beliefs and teacher instruction. Fives and Buehl (2012: 478) explain that teacher cognitions, in particular beliefs, fulfill three functions. First, beliefs can be construed as filters, thus influencing how the relevance of experiences is interpreted. Next, beliefs can frame problematic issues, meaning that they help to conceptualize them. Finally, they argue that beliefs are used as guides for following actions (Fives & Buehl 2012: 478). While research suggests that teacher beliefs are the main causal factor (Skott 2015: 21) for their actions in classrooms, Borg (2003: 106) emphasizes that the context teachers work in also has to be researched because, for instance, institutional or social factors influence teacher practice to a large extent. Therefore, he concludes that teacher cognition research needs to consider contextual factors in order to avoid any deficient inferences on the relation between teacher cognition and teacher practice (Borg 2003: 106).

According to Borg (2003: 105), teacher cognition studies in the area of language teaching is acknowledged across the field. However, it took a decade longer for teacher cognition research to be acknowledged as a viable research area in second and foreign language teaching compared to research in general education. Borg admits that while some areas in second and foreign language teaching receive significant attention in terms of teacher cognition research, for instance teaching grammar or reading skills, other areas have not been thoroughly researched or have received no attention at all (Borg 2009: 4). One of these scarcely researched topics is teacher cognitions and their impact on language assessment practice in second or foreign language teaching contexts (Yin 2010: 175). Yin (2010: 176-177) reviewed existing teacher cognition research that aims at understanding language assessment as part of her own research project. As a conclusion, she acknowledges the “complexity and richness of teachers’ assessment cognition” (Yin 2010: 177) and emphasizes teacher agency in terms of assessment practice (Yin 2010: 193). All in all, following Freeman’s argument that looking into “the hidden side of teaching” (Freeman 2002: 1) might be a way to improve teaching practices, teacher cognition research might be an applicable option to improve teachers’ assessment practices.

Thus, in view of the teacher cognition research agenda and its potential to explain the teachers’ instructional assessment practices, the empirical part seeks to answer the following research questions:

1. What types of assessment are used by current EFL teachers in Austria?
2. What are teachers’ attitudes towards the use of alternative assessment in contrast to traditional assessment in Austrian EFL classrooms?
3. What factors contribute to teachers’ decisions about using specific assessment procedures?

### **3.2. Research methodology**

In order to answer the research questions, two research methods were shortlisted immediately. Both quantitative and qualitative research methods are valid choices to collect the necessary data. However, because the primary purpose is to examine a general trend in teachers’ use of assessment, I decided to use quantitative research methodology primarily with the possibility of additional qualitative measures if the quantitative data showed a potential gap that needed to be closed by applying qualitative research.

While quantitative data can be obtained by different methods, I decided to use the well proven questionnaire, which is the most frequently used method in second language research (Dörnyei 2003: 1). It was chosen to collect the essential data because, as stated by Dörnyei (2007: 101), questionnaire surveys “aim at describing the characteristics of a population by examining a sample of that group”. In addition, questionnaires are able to provide three specific types of data, which are of a “factual, behavioral and attitudinal” (Dörnyei 2003: 8) nature. Calling to mind the research objectives, all three types of data are necessary to answer the research questions.

Additional considerations for choosing the research methodology were concerns of practicality and anonymity. Firstly, questionnaires are exceptionally popular because they can be designed easily and can help to collect a large amount of data in a short timespan. In other words, administering a questionnaire saves time, effort and cost on the part of the researcher (Dörnyei 2003: 9). Practicality dictates that the questionnaire is designed to be self-administered in order to be distributed online. Furthermore, self-administration ensures anonymity on the part of the participants, which is argued to be a vital characteristic to provide accurate and honest answers, especially when sensitive topics are involved (Dörnyei 2003: 24).

### **3.3. The sample**

Within the scope of this research project, it was not possible for the entire target population to have been asked to complete the questionnaire, therefore the participants are a sample of the target population consisting of 40 Austrian English teachers. Dörnyei (2007: 96) states that the sample needs to feature the same characteristics as the target population in order to draw valid conclusions. Therefore, only Austrian teachers of English as a foreign language have been asked to participate. However, one important criteria that characterized the sampling process was accessibility to participants. Thus, the sample type of this study can be considered what Dörnyei describes as convenience sampling, which is the prevailing sampling type in second language research (Dörnyei 2003: 72). That is, “practical criteria, such as [...] availability at a certain time, or accessibility” (Dörnyei 2003: 72) characterized the sampling process.

Practicality was also of concern in terms of questionnaire distribution, which was done exclusively via the internet in order to reach as many potential participants as possible in a short time. Two approaches proved to be effective; the questionnaire was sent to English

teaching colleagues and former university acquaintances via email, who, in turn, sent the questionnaire on to colleagues. Additionally, certain groups on social network websites especially created for teachers were used in order to reach the intended target group.

### **3.4. Questionnaire design**

After deciding on a method to gather suitable data, research was done looking for similar endeavors featuring questionnaires with foci on teacher attitudes and know-how and the use of different assessment procedures. However, the research was fruitless which meant that the questionnaire had to be designed from scratch without any clues from other projects.

Before the writing of items began, general decisions were made considering language, length and item formats. As has already been established, the target group of this study is English teachers in Austria. While researchers claim that it might be appropriate to ask potential participants in their mother tongue (Dörnyei & Taguchi 2010: 49), it can be assumed that English teachers should not have any issues with questionnaires written in English due to their everyday English language practice. Therefore, the introduction as well as all the items were written in English, with the exception of Austria-relevant terminology which was given in German in order to avoid misunderstandings.

In terms of layout and length, it is advised by Dörnyei that “anything that is more than 4-6 pages long and requires over half an hour to complete may be considered too much” (Dörnyei 2003: 18). Besides, personal experiences with questionnaires suggest that shorter questionnaires have better chances to attract participants. Therefore, a maximum of fifteen minutes to complete the questionnaire influenced the design process and the scope of items.

The items in the questionnaire were designed mainly as closed-ended items, which is the prevailing type of items in questionnaires (Dörnyei 2003: 35). The advantages of this type are twofold. On the one hand, closed-ended items ease the analysis process because they can be easily coded and analyzed electronically. On the other hand, participants do not have to produce subjective answers due to prepared responses, which avoids rater subjectivity, which in turn enhances reliability (Dörnyei 2003: 35). A few open-ended questions are included in the questionnaire because with some items, the range of possible responses could not be anticipated. Thus, the majority of open-ended questions can be distinguished as specific open questions, which are characterized by Dörnyei (2003: 48) as questions that ask for facts or preferences, or short-answer questions.

The questionnaire (see Appendix 3, p. 103) consists of five parts. As an introduction, the background, the researcher, and the purpose of the questionnaire are introduced. In addition it is explained at whom the questionnaire is targeted and that confidentiality and anonymity are assured.

Part One asks factual questions in order to find out about the sociodemographic characteristics of the participants and serves as a warm up. The sociodemographic data are valuable for interpretation of the questionnaire results. The scope of items comprises place of graduation, years of teaching and the place of employment.

Part Two deals with assessment use and habits as well as know-how. It aims to find out about teachers' familiarity with certain significant assessment-related concepts and terms. Considering habits, participants have to indicate what types of assessment they use in their EFL classroom. Moreover, it should be revealed what the most important intentions behind administering language assessment are. This part of the questionnaire aims at answering the question about what types of assessment are used in Austria's EFL classroom in general and which alternatives in language assessment are used in particular.

Parts 3, 4 and 5 of the questionnaire deal with differences in habits, attitudes and experiences concerning alternatives in assessment in contrast to their traditional counterparts. Therefore, it was necessary to provide concise characterizations of both concepts to ensure that every participant had the same understanding of both concepts.

Part Three reveals the participants' perceptions of teacher education and assessment. For the first time, the questions emphasize a distinct difference between traditional and alternative assessment. This part was designed to infer whether or not teacher education could be a decisive factor that influences teacher decisions on assessment procedures and habits.

Part Four is dedicated to teachers' attitudes and beliefs about assessment in general and alternative in contrast to traditional assessment in particular. First, the participants are asked whether they agree or disagree with general statements about assessment. Then, the same group of questions are posed in relation to traditional and alternative assessment. These questions are designed to shed light on the teachers' thoughts and opinions about alternative assessment.

The last part comprises open ended questions about the most useful assessment tools in EFL classroom and the possibility of stating the factors that influence the choice of the

assessment tools applied. In addition, the participants have the possibility to leave their email address in case anyone would be willing to participate in a face to face interview.

### **3.5. Procedure**

Before the received data is analyzed, procedure-related issues shall be described, namely digitalizing the questionnaire, piloting the study and preparing the results for data analysis.

#### **3.5.1. Digitalization**

One important decision had to be made during the design process about which software should be used for digitalization of the questionnaire. Nowadays, a number of different providers offer software solutions to create and distribute online surveys. Due to previous positive experiences with the web-based application Google Forms (Google Inc. 2012) and a registered Google profile, I decided to design the questionnaire with Google Forms.

#### **3.5.2. Piloting the questionnaire**

Piloting the study serves various purposes and usually avoids time-consuming work at a later stage (cf. Dörnyei 2003: 64-65). In the case of this questionnaire, it was piloted twice during the design process. The first draft was piloted among a group of colleagues who are in the final stage of their teacher education. All of them are pre-service English teachers. Therefore, they can be assumed to fit the characteristics of the intended target population. They were instructed to pay particular attention to item wording, clarity and layout. Whether or not the set time limit can be maintained was also of interest. Valuable feedback was received about wording and clarity. Time was not an issue. The questionnaire was revised and distributed once more to the same group of colleagues. The feedback was incorporated into the final version of the questionnaire.

#### **3.5.3. Analyzing the data**

As previously discussed, the questionnaire comprises both closed-ended and open-ended items, and all data collected was analyzed quantitatively. Naturally, open-ended items produce more diverse answers than closed-ended items; nevertheless, it was attempted to group the answers into categories to make quantitative analysis possible.

Google Forms provides the option to export the answers in a spreadsheet that can be imported into Microsoft Excel 15.17 (Microsoft Corporation 2015) for further processing. In Microsoft Excel, the responses were coded as digits, for instance the response *strongly agree* was coded *one* and the option *strongly disagree* was coded *five*. After the coding process, the spreadsheet was imported into IBM SPSS Statistics 24 (IBM Corporation 2016) for further statistical processing. The data analysis constitutes mainly descriptive statistics in order to provide an overview of the obtained data to interpret them in terms of the research questions. Two types of statistical tests were deployed to find correlations or differences between questionnaire variables; the chi-square test in order to find correlations between two or more categorical variables (Field 2009: 688) and the Wilcoxon-signed-rank test to compare attitudes towards two related sets of items indicated by the same participants (Field 2009: 552).

## 4. Results

The results are presented in four sections. Section 4.1. deals with sociodemographic data of the participants. Section 4.2. is dedicated to assessment use and know-how. Section 4.3. describes the results in regard to teacher education and assessment routine and finally, Section 4.4. reveals the results in terms of teacher attitudes and opinions about alternative in contrast to traditional assessment.

### 4.1. Sociodemographic data

The sociodemographic data comprise four items, which are the place the teachers graduated from, the subjects they teach, the duration they have been teaching and the school type they are working at.

Table 1 shows the answers to the questionnaire item *What university/teacher training college (PH) did you graduate from?*<sup>1</sup>. The results indicate that more than half of all

---

<sup>1</sup> Teachers in Austria are trained at two different types of academic institutions, universities on the one hand and teacher training colleges on the other. Commonly known *Pädagogische Akademien (PÄDAKs)* are teacher training colleges that are nowadays called *Pädagogische Hochschulen (PH)*. Graduates from PÄDAKs/PHs are trained for teaching at compulsory school level, like the New Secondary School (NMS), while graduates from university teacher education can work in Lower and Upper Academic Secondary Schools (AHS), Schools for Intermediate Vocational Education (BMS) as well as Colleges for Higher Vocational Education (BHS/HTL/BAKIP).



participants graduated from the University of Vienna, namely 57.5%. 42.5% of all respondents graduated from other Austrian academic institutions. Thereof, 17.5% graduated from *PÄDAKs/PHs*.

**Table 1: Answers to item no. 1  
“What university/teacher training college (PH) did you graduate from?”**

		Responses	
		N	Percent
1. What university or teacher training college (PH) did you graduate from?	University of Vienna	23	57.5%
	University of Graz	4	10.0%
	University of Innsbruck	4	10.0%
	PÄDAK Graz	3	7.5%
	Alpen-Adria-Universität Klagenfurt	2	5.0%
	PÄDAK Wien	2	5.0%
	PÄDAK Eisenstadt	1	2.5%
	PÄDAK Linz	1	2.5%
Total		40	100.0%

**Table 2: Answers to item no. 2 “What subject(s) do you teach?”**

		Responses	
		N	Percent
2. What subject(s) do you teach?	English and History	7	17.5%
	English	6	15.0%
	English and Geography	4	10.0%
	English and Spanish	4	10.0%
	English and Biology	4	10.0%
	English and German	3	7.5%
	English and PP	3	7.5%
	English and Math	2	5.0%
	English and French	2	5.0%
	English and Computer Science	1	2.5%
	English and Physical Education	2	5.0%
	English and Science	1	2.5%
	English, Music and Computer Science	1	2.5%
Total		40	100.0%

Table 2 illustrates the frequencies of certain subjects or subject combination taught by the respondents. Of course, every participant is an English teacher; however, Austria’s teacher education system requires prospective teachers to choose at least two subjects they should

be teaching later. Nevertheless, it is possible that teachers will teach only one subject depending on available vacancies. 15% of the respondents only teach English at the moment. Fifteen respondents are language teachers, which equals 37.5% of all respondents. Languages represented are English, German, French and Spanish. History is the most represented content subject; 17.5% of all respondents teach History besides English. At 10 percent each, Geography and Biology constitute above-average subject combinations with English.

Another sociodemographic question was asked with the item *How long have you been teaching?* The respondents were requested to tick one of six categories. Almost two thirds of the respondents have not been teaching longer than 6 years. While 37.5%, which represents the largest group of respondents, have been teaching up to 3 years, 27.5% have been teaching between 4 and 6 years. The third largest group consisting of 12.5% of all respondents has been teaching at least 20 years. Therefore, it can be stated that the majority of the sample population is either rather new to the job or very experienced.

**Table 3: Answers to item no. 3 “How long have you been teaching?”**

		Responses	
		N	Percent
3. How long have you been teaching?	0-3 years	15	37.5%
	4-6 years	11	27.5%
	20+ years	5	12.5%
	7-9 years	4	10.0%
	10-14 years	3	7.5%
	15-19 years	2	5.0%
Total		40	100.0%

The final question in the first part investigates the school types the respondents work at. Multiple answers are possible for each participant because multiple job locations are conceivable. The results are presented in Table 4. The column Percent shows the percentage based on all ticked answers. The column Percent based on respondents shows the percentage of respondents working in the respective school type. That is, 45% of the participants teach at BHS. The second biggest group at 35% is teachers at NMS, which is interesting because only 17.5% of the sample graduated from teacher training colleges. However, since the introduction of new secondary schools in the Austrian educational system, teachers who attended teacher training at university level are working in a peer teaching configuration at new secondary schools. 32.5% of the sample population work at

lower academic secondary schools and 30% at upper academic secondary schools. Only 5% teach at colleges for the training of nursery school teachers (BAKIP). However, none of the respondents work at pre-vocational schools or in the apprenticeship training system.

**Table 4: Answers to item no. 4 “Which sort of school are you teaching at?”**

		Responses		Percent based on respondents
		N	Percent	
4. Which sort of school are you teaching at?	BHS (HTL, HAK, HLW)	18	28.1%	45.0%
	NMS	14	21.9%	35.0%
	AHS (lower secondary)	13	20.3%	32.5%
	AHS (upper secondary)	12	18.8%	30.0%
	BMS (Handelsschule/Fachschule)	5	7.8%	12.5%
	BAKIP	2	3.1%	5.0%
Total		64	100.0%	160.0%

Table 5 shows where those teachers work, who are simultaneously teachers at NMS. The biggest overlap is found at vocational schools. Of the participating BHS teachers, 27.8% are employed at New Secondary Schools. In addition, 20% of the participating BMS teachers are also NMS teachers.

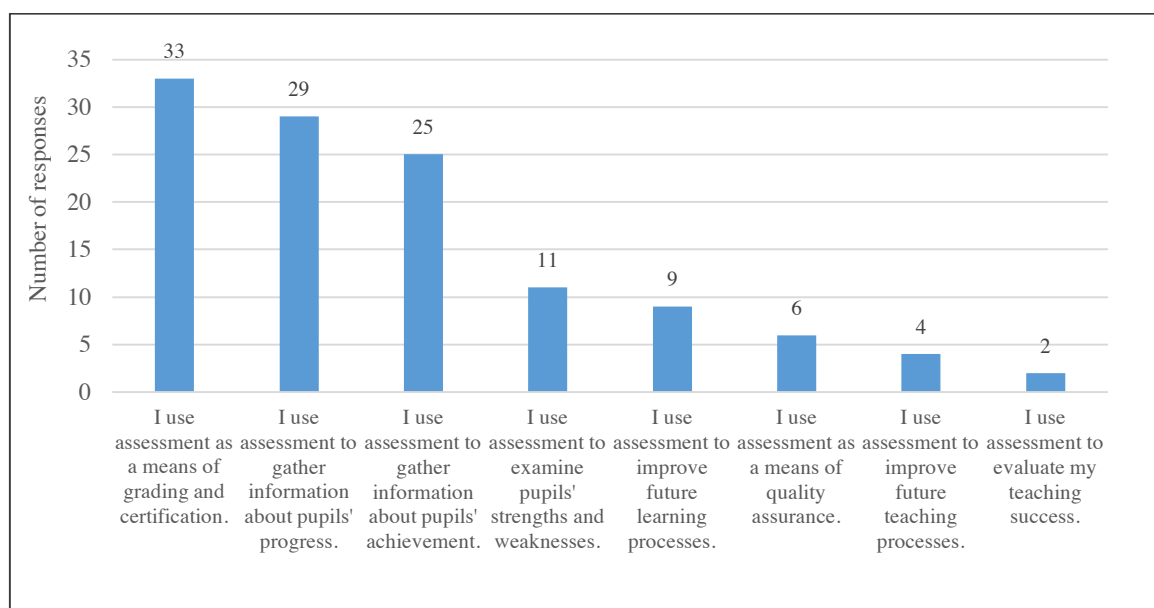
**Table 5: Teachers working at NMS and other school types**

		AHS (lower secondary)	BMS (Handelsschule/ Fachschule)	AHS (upper secondary)	BHS (HTL, HAK, HLW)	BAKIP
Not working at an NMS	Count	11	4	11	13	2
	%	84.6%	80.0%	91.7%	72.2%	100.0%
Working at an NMS	Count	2	1	1	5	0
	%	15.4%	20.0%	8.3%	27.8%	0.0%
Total	Count	13	5	12	18	2
		100%	100%	100%	100%	100%

## 4.2. EFL assessment use and know-how

Part Two of the questionnaire is designed to shed light on general assessment know-how and the assessment habits of the respondents. The first item in this part of the questionnaire asks for the three most important purposes teachers administer assessment for. Every participant was asked to choose the 3 most important purposes for language assessment, that is, multiple answers from each respondent were required. However, one respondent

picked only two purposes as being most important in his/her EFL classroom, which is why the total number of responses equals 119 rather than 120.



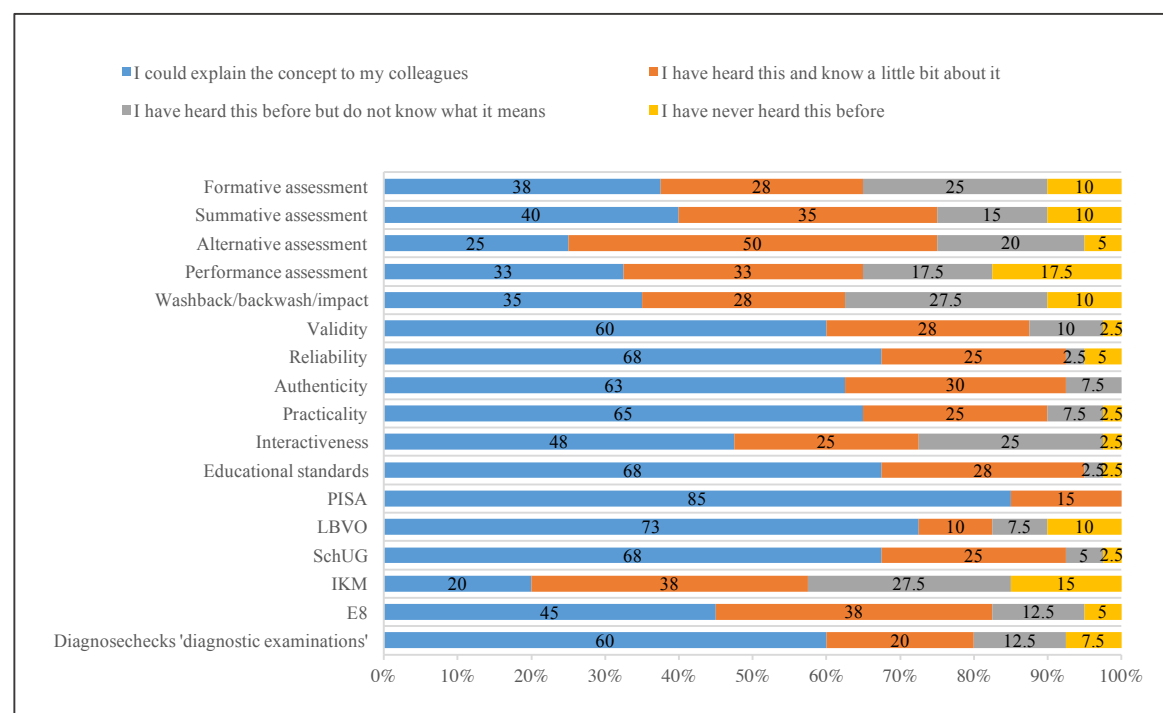
**Figure 2: Answers to item no. 5 “Please indicate the three most important purposes of language assessment in your English teaching in general.”**

The results depicted in Figure 2 show that three purposes are most important for the participants. Assessment as a means of grading and certification was picked as being a crucial purpose of assessment 33 times. Gathering information about pupils’ progress was picked 29 times and 25 participants indicated that they administer assessment to collect data about pupils’ achievement. Together those three represent more than two thirds of all picked options. In contrast, all other alternatives lag far behind. However, additional results are that 11 respondents use assessment in order to examine pupils’ strengths and weaknesses and 9 teachers try to improve pupils’ future learning processes via assessment.

In addition, it was tested whether there are correlations between indicated assessment purposes and certain groups within the sample. The focus was on teaching experience and degree. The participants were allocated into two categories for each factor; 0-6 years and 7+ years in terms of teaching experience as well as university graduates and PH/PÄDAK graduates in terms of degree. In order to find significant correlations, cross tabulations were calculated and the chi-square test was applied (see Appendix 2, p. 96). While the chi-square test did not find a significant link between assessment purposes and the teachers’ teaching experiences in years  $\chi^2(8) = 9.947$ ,  $p > 0,05$ , a significant association between assessment purposes and the teachers’ degree was found  $\chi^2(8) = 19.329$ ,  $p < 0,05$  (see Appendix 2, p. 96). Apparently, participants who earned a degree for teaching at compulsory school level use assessment for different purposes in contrast to teachers who graduated from

universities. The most striking differences are that 27.3% of the participants who graduated from universities use assessment to improve future learning processes, whereas no PH/PÄDAK graduate indicated to use assessment for this reason. On the other hand, while 42.9% of PH/PÄDAK graduates apply assessment as a means of quality assurance, only 9.1% of university graduates use assessment for the same reason. However, the small sample size might have an effect on the calculated results. Despite usual suggestions to ensure a minimum count of 5 for each cross tabulation cell, Rasinger (2008: 148) argues that a minimum count is not necessarily obligatory, which is why cell counts smaller than 5 were ignored. Nevertheless, the results should be treated with caution.

The next question investigates the respondents' familiarity with assessment related concepts. The question comprises sixteen items and the possible responses range from *I could explain the concept to my colleagues* to *I've never heard of this before*. In order to provide the opportunity to compare the familiarity with the concepts, Figure 4 visualizes all items in one figure:



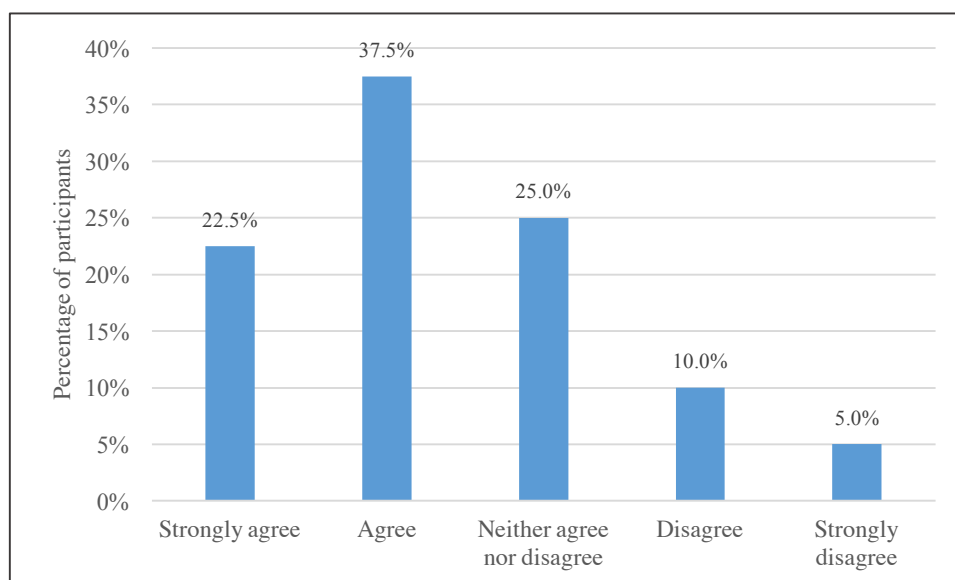
**Figure 3: Answers to items no. 6-22**  
**“Select the response which best describes how well you know the term.”**

Upon first glance Figure 3 shows that except for the items PISA and authenticity, a certain share of the sample population was unfamiliar with the item concerned. However, for most items the option *I could explain the concept to my colleagues* predominates the result analysis. All of the respondents are familiar with PISA and more than 90% of the participants know at least a little about the SchUG. This makes both concepts the best

known of all items. The test usefulness characteristics are also well known except for washback/backwash/impact and interactiveness. Moreover, concepts which are not well known or even unknown are IKM, performance assessment and formative assessment.

Furthermore, it was tested whether there is a connection between the indicated familiarity with alternative assessment and two factors, namely the teachers' teaching experience and degree. Cross tabulations were calculated and the chi-square test was applied in order to find statistically significant correlations (see Appendix 2, p. 98-101). However, there was neither a statistically significant association between familiarity with alternative assessment and groups of teachers with varying teaching experience  $\chi^2 (3) = 2.198, p > 0.05$  nor between familiarity with alternative assessment with the teachers' degree  $\chi^2 (3) = 1.212, p > 0.05$ . The same calculations were completed in terms of familiarity with formative assessment in relation to both factors. The test statistics show similar results. That is, no statistically significant association was found between familiarity with formative assessment and the teachers' experience  $\chi^2 (3) = 0.833, p > 0.05$  nor between familiarity with formative assessment and the teachers' degree  $\chi^2 (3) = 1.453, p > 0.05$ .

Item number 23 asks for the respondents' self-assessment as to whether or not they are familiar with the legal regulations in Austria considering assessment in schools. Figure 4 shows the results:



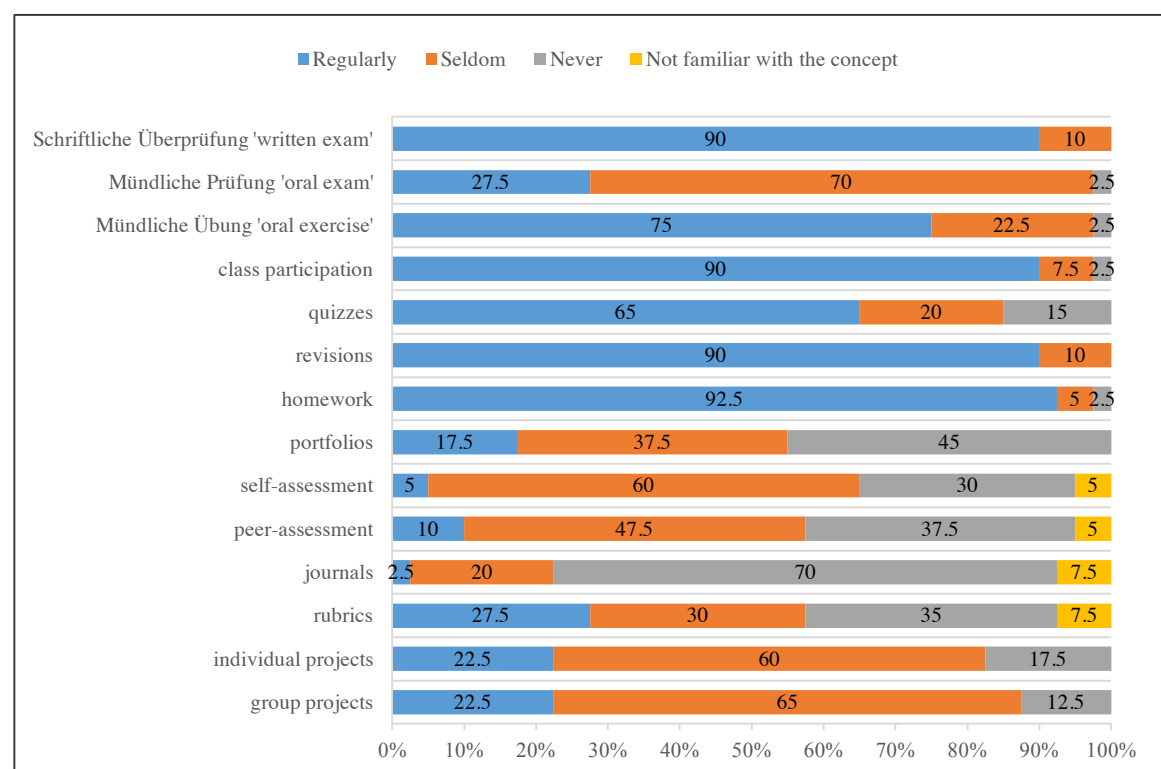
**Figure 4: Answers to item no. 23 “I’m well aware of the legal regulations as they are described in the SchUG and the LBVO concerning assessment in Austrian schools.”**

While only 22.5% of all respondents state that they strongly agree on the posed statement, 37.5% of the participants agree with it. However, that implies that almost half of all the

respondents (40%) feel that they are not well aware of the legal regulations considering educational assessment.

The remaining items in this part of the questionnaire are designed to reveal the assessment habits of all respondents. In order to do that, a series of assessment tools were listed and the respondents were asked to state how frequently the procedure is used in one specific EFL class. In case the respondents have different assessment habits in different classes, the items were provided twice. In all, thirteen respondents used the items a second time (items no. 39-52). However, the results revealed that some of the respondents did not understand the prompt entirely because they filled in the same habits twice for different school grades or answered the question a second time for the same school grade with the exact same outcome. Therefore, all answers to the second set of questions are not taken into consideration for the analysis.

Figure 5 illustrates if and how often specific assessment procedures are used by the surveyed teachers without showing differences concerning school grades:



**Figure 5: Answers to items no. 24-37 “How often do you use the listed types of assessment procedures in your EFL classroom?”**

Figure 5 shows that written examinations, class participation, revisions and homework are used by at least 90% of all respondents on a regular basis. The figure shows a remarkable difference in usage between written and oral examinations. While written examinations are used regularly by 90%, 70% seldom use oral examinations. Oral exercises are administered

regularly by 75% and quizzes by 65% of the respondents. In comparison to written exams, oral exercises, class participation, quizzes, revisions and homework, all other concepts are used regularly by considerably smaller proportions of respondents. Projects, whether they are designed as individual or group projects, are seldom used by more than half of the sample population. Portfolios, self-assessment, peer-assessment and journals are not that common. Obviously, journals are not very common in EFL classrooms: only one of the respondents uses them in the classroom regularly. However, 70% of the sample population do not use them at all and 7.5% is unfamiliar with the concept. Besides journals, only four other assessment concepts, namely rubrics and self- and peer-assessment are unknown to a very small proportion of all respondents.

The final item, number 54 considering assessment use and know-how, was the request *Please state if you use any other assessment tools and the grade you administer them in.* This item was open-ended and optional. Only four participants made use of this opportunity. Two respondents entered “presentations” and one respondent each entered “criterion-referenced assessment” and “observations in all classrooms”.

#### 4.3. Assessment and teacher education

The third part of the questionnaire aims at revealing the relation between assessment know how, assessment habits and teacher education. The participants were presented with a set of twelve statements and each respondent indicated the extent of agreement or disagreement. The final item of this part was an open-ended question.

**Table 6: Answers to items no. 55-58**  
**“Agreement/disagreement with statements concerning teacher education.”**

	Strongly Agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	N/A
55. The teacher education program at university/teacher training college prepared me for all kinds of language assessment forms.	2.5%	20.0%	20.0%	35.0%	22.5%	0.0%
56. The courses at university/teacher training college focused on traditional assessment.	17.5%	50.0%	17.5%	12.5%	2.5%	0.0%
57. The courses at university/teacher training college focused on alternative assessment.	0.0%	12.5%	17.5%	40.0%	27.5%	2.5%
58. The amount of in-service training in assessment is appropriate.	2.5%	15.0%	7.5%	45.0%	22.5%	7.5%

Table 6 illustrates the agreement and disagreement with four core statements. The majority (35% disagree and 22.5% strongly disagree) does not agree that their teacher education



prepared them for all kinds of assessment. In addition, 50% of the participants agree and 17.5% strongly agree that their courses aimed at traditional assessment. Only a small number of teachers indicate that their teacher education focused on alternative assessment; 12.5% agree with the statement in item 57. Whether the amount of in-service training in assessment is appropriate was asked in item 58. While 45% disagree, a further 22.5% strongly disagree with the respective statement.

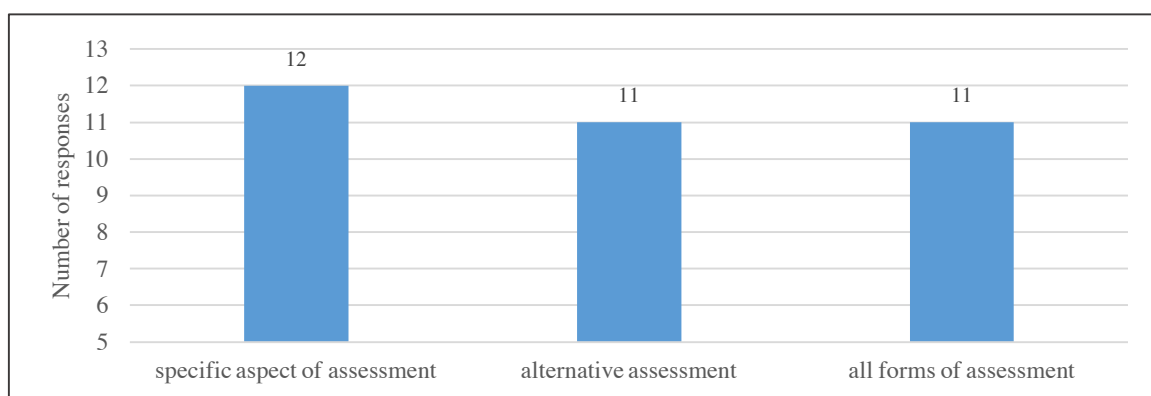
**Table 7: Answers to items no. 59-66**  
**“I feel comfortable doing the following in regard to language assessment.”**

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	N/A
59. Administering traditional forms of language assessment.	25.0%	60.0%	12.5%	2.5%	0.0%	0.0%
60. Administering alternative forms of language assessment.	12.5%	27.5%	20.0%	40.0%	0.0%	0.0%
61. Utilizing different kinds of assessment techniques.	12.5%	47.5%	30.0%	10.0%	0.0%	0.0%
62. Preparing examinations.	27.5%	52.5%	17.5%	2.5%	0.0%	0.0%
63. Scoring examinations.	25.0%	50.0%	10.0%	15.0%	0.0%	0.0%
64. Using assessment to enhance pupils' learning processes.	15.0%	40.0%	25.0%	17.5%	2.5%	0.0%
66. Using assessment to provide feedback about pupils' achievement.	30.0%	45.0%	17.5%	5.0%	2.5%	0.0%

Whether or not teachers feel comfortable with certain aspects of teaching was asked in items number 59-66, and the results are displayed in Table 7. Item 65 is not considered in the analysis due to an orthographic mistake which altered the meaning of the statement. The majority, namely 60%, agree and a further 25% strongly agree that they feel comfortable administering traditional language assessment, while only 2.5% of the participants disagree. The same statement regarding alternative forms of language assessment resulted in 40% disagreeing and 40% agreeing, but 12.5% strongly agree that administering alternative language assessment is no problem for them. Feeling comfortable using different kinds of assessment techniques was agreed by 47.5% and strongly agreed by 12.5% of the participants. Items 62 and 63 are dedicated to preparing and scoring examinations, a typically traditional form of assessment. Only small shares of the participating teachers have issues preparing and scoring examinations, that is to say 2.5% and 15% respectively. Items 64 and 66 describe alternative uses of assessment which can be described as washback-related assessment purposes. While 40% agree and 15% strongly agree that they feel prepared to use assessment in order to enhance the pupils' learning processes, a fifth (17.5% disagree and 2.5% strongly disagree) of all respondents object to

the statement. Neither agreeing nor disagreeing with the statement is indicated by 25%. Finally, 45% agreed and a further 30% strongly agree that they feel comfortable providing feedback on pupils' achievement on the basis of assessment.

The final item considering assessment in regards to teacher education asks participants to state forms of assessment they would have liked to have discussed during their teacher education program or in-service training, and is an open-ended item. Therefore, the results are quite diverse. Thirty-five respondents answered the question. One respondent's answer does not specify assessment related content but rather the method how assessment should be taught in teacher education. The applicable answers can be pooled into three categories, which can be encountered in Figure 6.



**Figure 6: Answers to item no. 67 “Assessment you would have liked to have discussed in more detail during teacher education or in-service training.”**

Twelve respondents voiced that they would have liked to have learned about a specific aspect of assessment in more detail. In this respect, only two matters came up more than once, which are *assessment to improve my teaching* and *formative assessment*. Eleven respondents indicated that they would have liked a focus on alternative assessment during their teacher training courses. Eleven respondents are of the opinion that actually all forms of assessment are neglected during teacher training even though it is such a big part of every teachers' professional life.

#### 4.4. Teachers' beliefs and attitudes

Part four of the questionnaire tries to elicit teachers' opinions, beliefs and attitudes about language assessment in general and alternative in contrast to traditional assessment in particular. The first set of items (no. 68-85) examines teachers' opinions about what assessment can do or should do. The participants were required to express their agreement

with 18 statements. The second set of items is asked in regard to traditional language assessment (items no. 86-96) and alternatives in assessment (items no. 97-107).

**Table 8: Answers to items no. 68-85**  
**“Agreement/disagreement with statements concerning language assessment.”**

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	N/A
68. Traditional assessment is superior to alternative assessment.	5.0%	25.0%	37.5%	30.0%	0.0%	2.5%
69. Assessment should mainly serve selective and allocative functions mainly.	0.0%	25.0%	27.5%	30.0%	12.5%	5.0%
70. Assessment is primarily used to determine pupils' (final) grades.	25.0%	52.5%	2.5%	12.5%	7.5%	0.0%
71. Assessment is a valid procedure to enhance instruction.	7.5%	40.0%	37.5%	7.5%	5.0%	2.5%
72. Assessment fosters pupils' learning processes.	7.5%	47.5%	27.5%	12.5%	2.5%	2.5%
73. Assessment helps to diagnose the strengths and weaknesses of instruction.	15.0%	57.5%	15.0%	10.0%	2.5%	0.0%
74. Assessment should nurture intrinsic pupil motivation.	22.5%	37.5%	25.0%	7.5%	0.0%	7.5%
75. Differences in learning styles should be accommodated in language assessment.	22.5%	40.0%	27.5%	10.0%	0.0%	0.0%
76. Self-assessment should be taken into account when assessing a pupil's language achievement.	7.5%	45.0%	20.0%	25.0%	2.5%	0.0%
77. Peer-assessment should be taken into account when assessing a pupil's language achievement.	10.0%	42.5%	22.5%	22.5%	2.5%	0.0%
78. It is important to use both formative and summative assessment in language classes.	30.0%	45.0%	20.0%	5.0%	0.0%	0.0%
79. In addition to grades, feedback on pupil performance is essential in language assessment.	57.5%	40.0%	2.5%	0.0%	0.0%	0.0%
80. Contextualization and authenticity are important in language assessment.	55.0%	35.0%	10.0%	0.0%	0.0%	0.0%
81. Assessment requires tasks involving higher level thinking and problem-solving skills.	32.5%	47.5%	20.0%	0.0%	0.0%	0.0%
82. I would prefer assessment that can be rated quickly by machines rather than scored by humans.	22.5%	27.5%	20.0%	10.0%	15.0%	5.0%
83. Continuous assessment is crucial in order to measure the pupils' progress.	25.0%	52.5%	15.0%	5.0%	2.5%	0.0%
84. Traditional assessment tools can be used for formative purposes.	2.5%	40.0%	40.0%	2.5%	2.5%	12.5%
85. Assessment can fulfill all purposes (selective, allocative, diagnostic, quality assurance) simultaneously.	7.5%	25.0%	25.0%	27.5%	7.5%	7.5%

The answers to the first set of questions are displayed in Table 8. It shows that most items received more positive (*Strongly agree* and *agree*) than negative responses (*Disagree* and *strongly disagree*). Only the items 68, 69 and 85 obtained similar percentages of agreement and disagreement. It is also remarkable that the calculated median of all *Neither agree nor disagree* answers is 21.25%, which indicates that about a fifth of all respondents could not decide whether to agree or disagree.

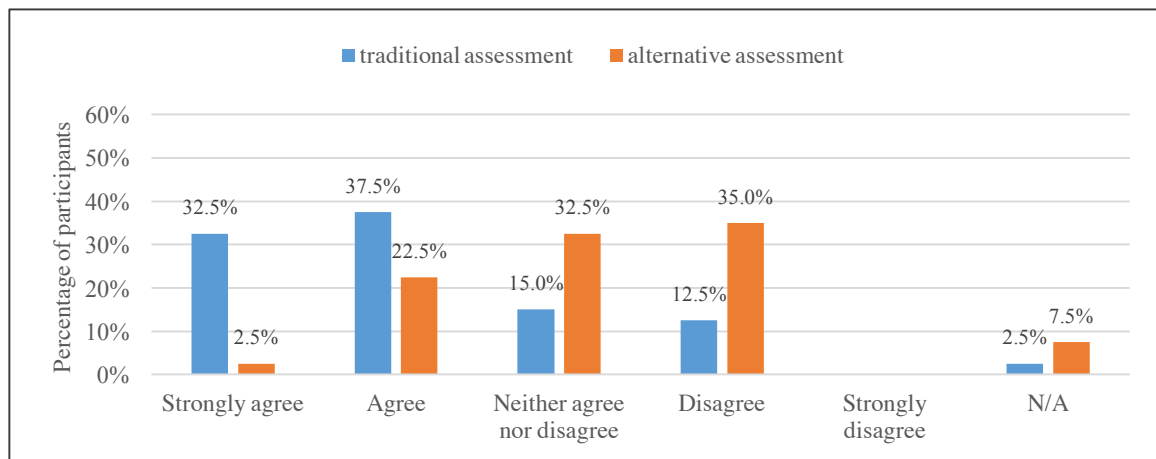
A vast majority of positive responses are indicated at items 70, 71, 72, 73, 78-81, 83 and 84. Items 79-81 stand out because they did not receive any negative responses at all. The participants expressed their opinion about the importance of feedback when they answered item number 79. The conflated positive responses (57.5% strongly agree and 40% agree) amount to 97.5%. Contextualization and authenticity is viewed as similarly important; 55% strongly and a further 35% agree that both concepts are essential in language assessment. Item 81 required the participants to reveal their attitude towards whether higher level thinking and problem-solving skills should be addressed by language assessment. The positive responses amount to 80% (32.5% strongly agree and 47.5% agree).

Less clear but still mainly positive responses were obtained from the answers to items number 70, 73, 74, 75, 78, 83 and 84. While 25% strongly agree and 52.5% agree that assessment is primarily used to determine pupils' grades, the majority of all surveyed teachers expressed their agreement on whether assessment helps to find out about strengths and weaknesses of instruction, whether intrinsic motivation is an issue and whether differences in learning styles should be accommodated. Related to those answers is the participants' opinion on the statement in item 78. The positive responses amount to 75%.

Items 68, 69, 76, 77 and 85 received less positive expressions of agreement. The largest amount of negative responses (30% disagree and 12.5% strongly disagree) was obtained in response to assessment serving mainly selective and allocative functions. Some surveyed teachers are also skeptical whether or not self- and peer-assessment should be a part of language assessment. Both items dealing with self- and peer-assessment received negative responses from about a quarter of all participants.

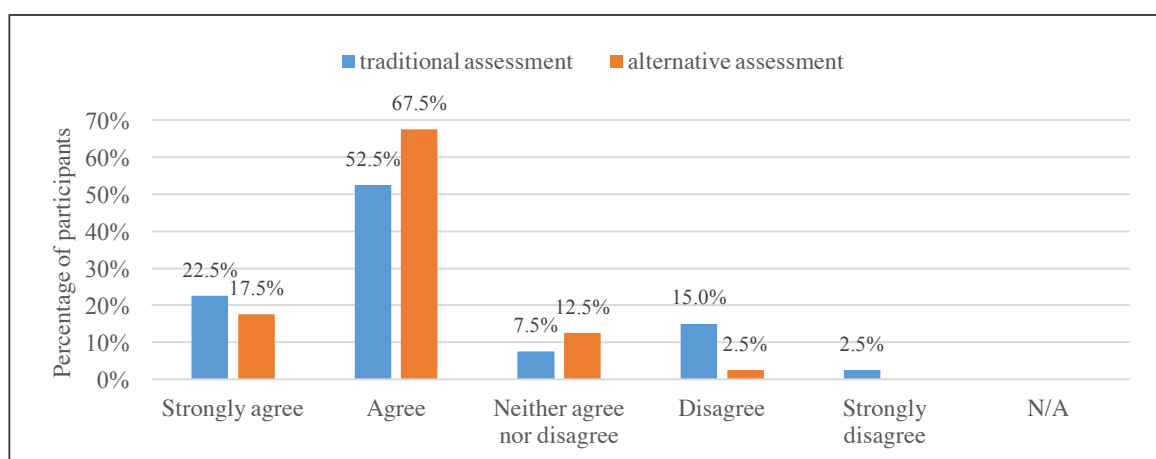
The results for the second set of items, which was asked in relation to traditional assessment and alternative assessment, are presented in individual bar charts for each item-pair to illustrate the teachers' attitudes towards both concepts. That is, the participants were requested to indicate their agreement with statements that were related to traditional assessment first and to alternative assessment afterwards. For instance, participants had to

express their agreement with the statement *Traditional forms of assessment should be applied because they are obligatory to determine pupils' grades* in item no. 86 and *Alternative forms of assessment should be applied because they are obligatory to determine pupils' grades* in item no. 97.



**Figure 7: Answers to items no. 86/97 “... they are obligatory to determine pupils' grades.”**

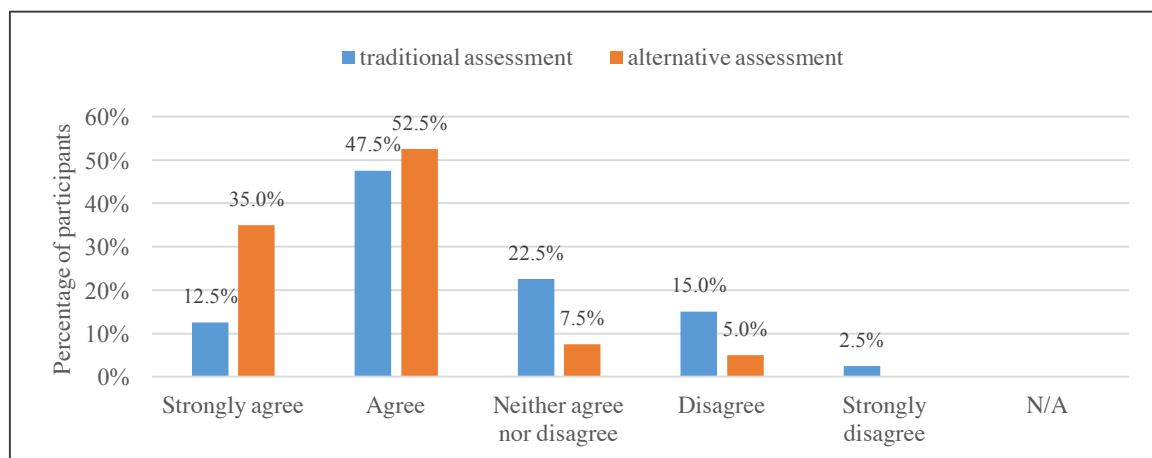
The results for items 86 and 97 reveal that a greater proportion of participants associate traditional assessment with determining pupils' grades in contrast to alternative assessment. While 32.5% strongly agree and 37.5% agree that traditional assessment should be used because it is necessary for grading purposes, large proportions of participants are either undecided (32.5%) or disagree (35%) with the statement if alternative assessment is evaluated in terms of grading.



**Figure 8: Answers to items no. 87/98 “... they provide information about pupils' achievement.”**

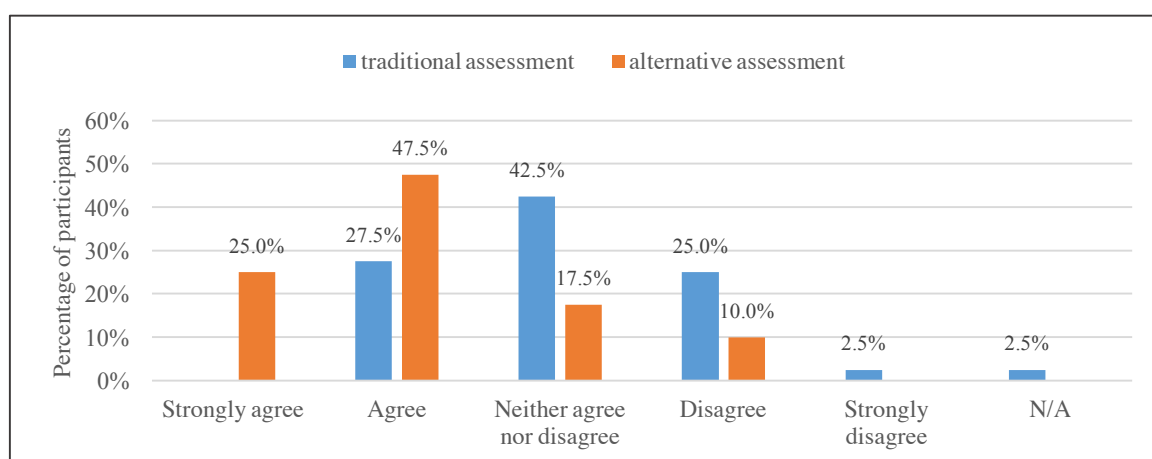
Figure 8 illustrates an entirely different picture in comparison to Figure 7. However, it explains why the Wilcoxon-signed-rank-test calculated insignificant differences between the concepts of traditional and alternative assessment when providing information about pupils' achievement is in question (see p. 72 and Appendix 2, p. 102). Thus statistically

speaking, both traditional and alternative assessment are believed to provide information about pupils' achievement equally well.



**Figure 9: Answers to items no. 88/99 “... they provide information about pupils' strengths and weaknesses.”**

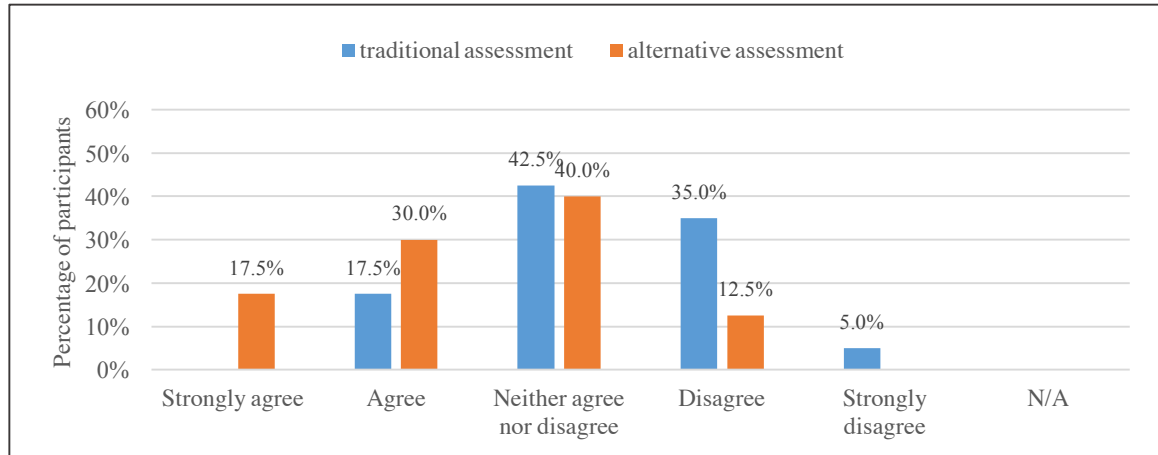
Looking at Figure 9, it becomes evident that alternative assessment is believed to be more appropriate when figuring out students' strengths and weaknesses is of interest. While 35% strongly agree and 52.5% agree that alternative assessment should be used, only 12.5% strongly agree and 47.5% agree with the same statement in relation to traditional assessment. In addition, disagreement with the statement considering traditional assessment was expressed by 15% of the participants, while only 5% think that pupils' strengths and weaknesses cannot be discovered via alternative forms of assessment.



**Figure 10: Answers to items no. 89/100 “... they are essential in fostering learning processes.”**

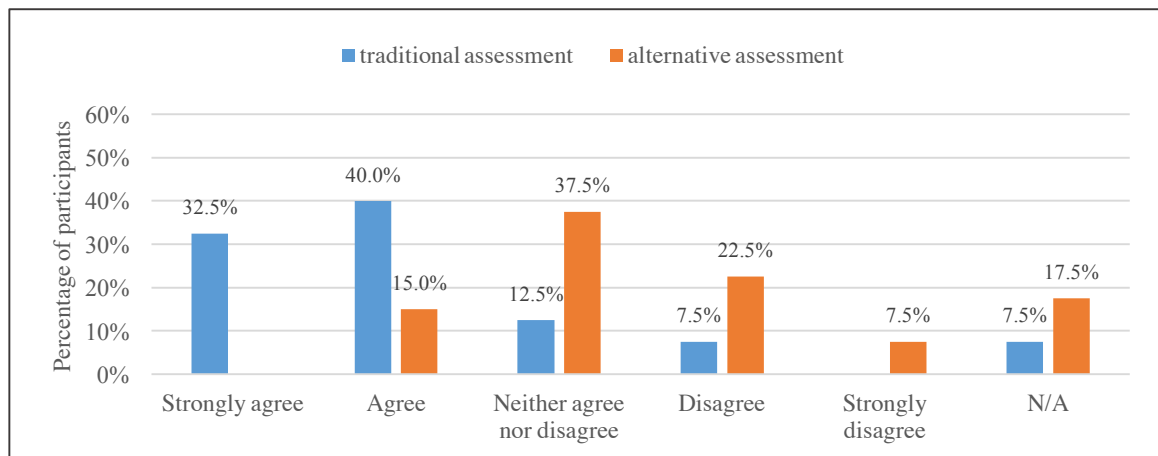
Considering the ability to foster learning processes, agreement clearly shifts towards alternative assessment forms, as can be seen in Figure 10. 25% strongly agree and 47.5% agree that alternatives in language assessment should be used in order to enhance learning processes. The data obtained from item 100 are not informative. Similar proportions of teachers agree (27.5%) and disagree (25%) with traditional assessment being essential to

foster pupils' learning processes. However, most importantly considering traditional assessment, 42.5% neither agree nor disagree with the statement.



**Figure 11: Answers to items no. 90/101 “... they provide information about the quality of teaching.”**

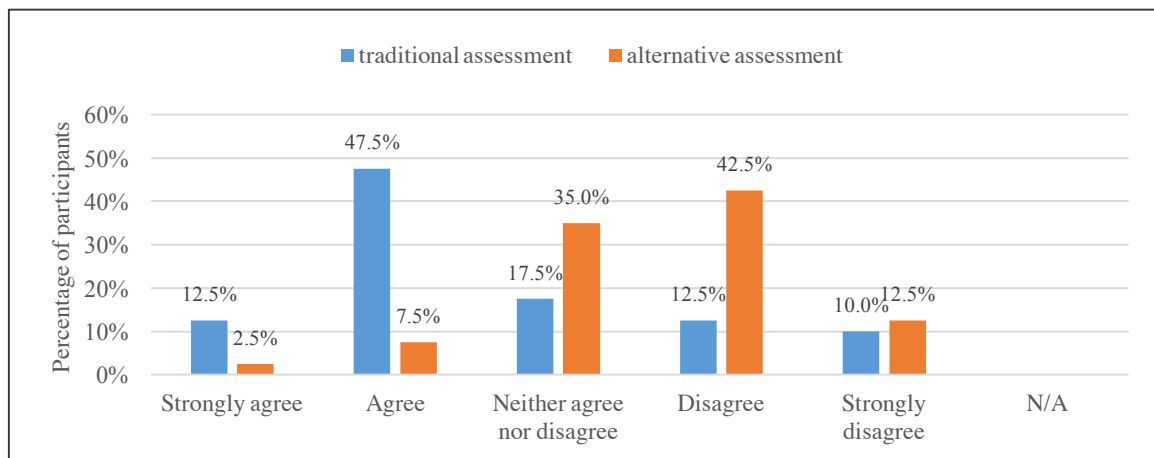
The most significant result depicted in Figure 11 is that almost half of the respondents are undecided on whether they agree or disagree with either option to evaluate teaching quality. However, comparing both options, a larger proportion of surveyed teachers expressed positive responses (17.5% and 30%) when it comes to alternative assessment while more participants expressed negative responses (35% and 5%) when traditional assessment was related to the statement.



**Figure 12: Answers to items no. 91/102 “... they are required by the LBVO and the SchUG.”**

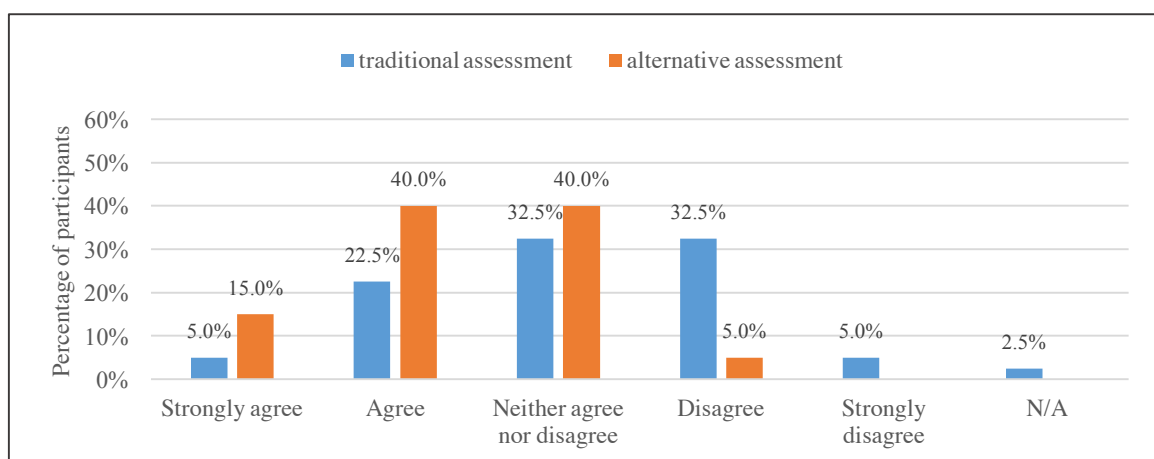
According to the data obtained by items 91 and 102, one major reason for administering traditional assessment is that it is required by the LBVO and the SchUG. This is supported by the fact that 32.5% strongly agree and 40% agree with the statement that traditional forms should be applied due to requirements stated in the two legally binding documents. Whether or not legal considerations are a reason to use alternative forms of assessment is not that obvious as almost 40 percent neither agree nor disagree. However, more teachers

do not associate alternative assessment with legal requirements compared to those who do (15% agree vs. 22.5% disagree and 7.5% strongly disagree).



**Figure 13: Answers to items no. 92/103 “... they are practical (easy to design, easy to score, ...).”**

When the teachers were asked to give their opinion on practicality-matters, 47.5% agreed that traditional forms should be applied when practicality is of importance compared to 7.5% who think the same of alternative assessment. On the contrary, 42.5% disagree that practicality is a reason for administering alternative assessment forms. As far as *Neither agree nor disagree* responses are concerned, 35% are indecisive whether or not alternative assessment should be applied because they are practical.

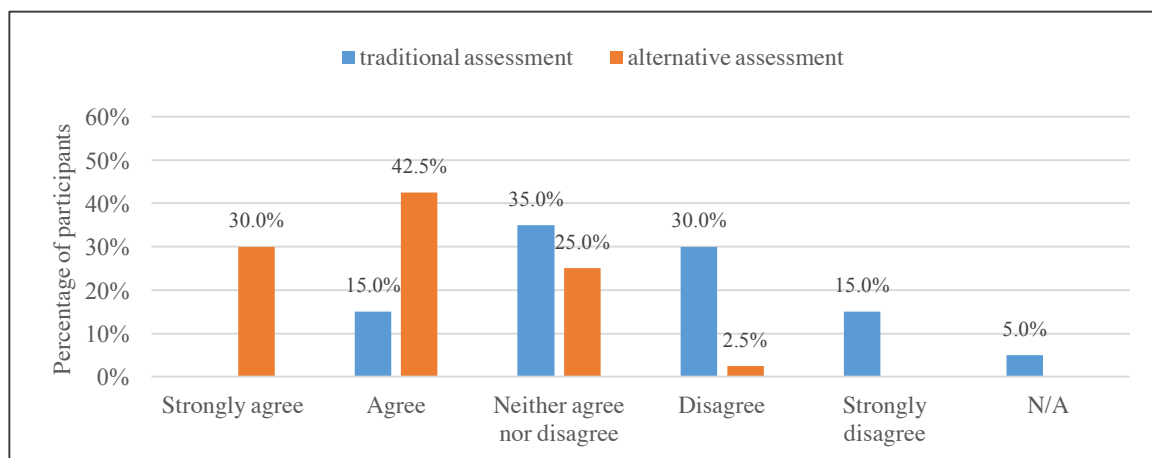


**Figure 14: Answers to items no. 93/104 “... they provide information for improving future instruction.”**

Whether or not both assessment approaches shall be used in order to improve future teaching is illustrated in Figure 14. While 15% strongly agree and 40% agree that alternative forms of assessment might provide information to enhance future instruction, a further 40% percent appear to be irresolute. In comparison, only 5% strongly agree and 22.5% agree that traditional assessment can be used for the same purpose and even more

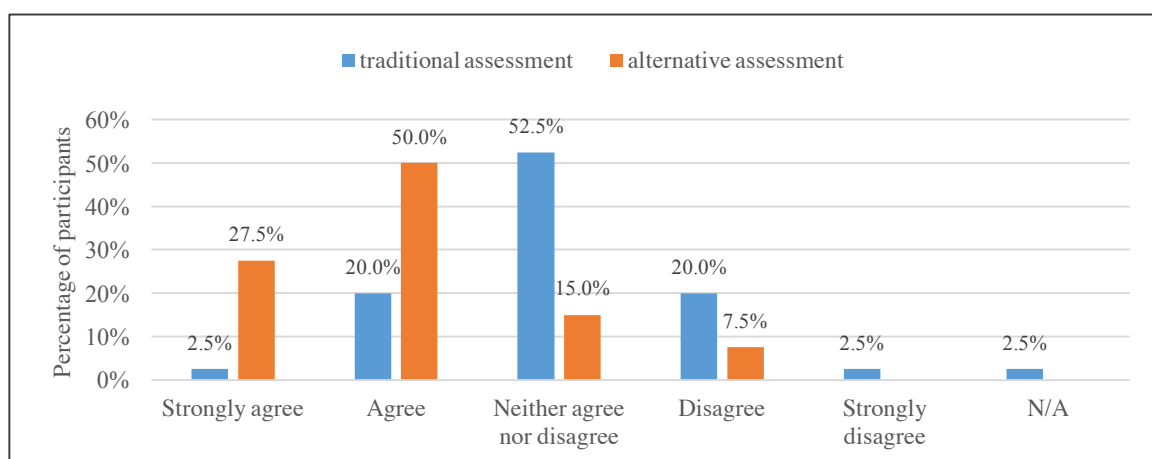


participants, 32.5% who disagree and 5% who strongly disagree, are convinced that traditional assessment can not be used to do so.



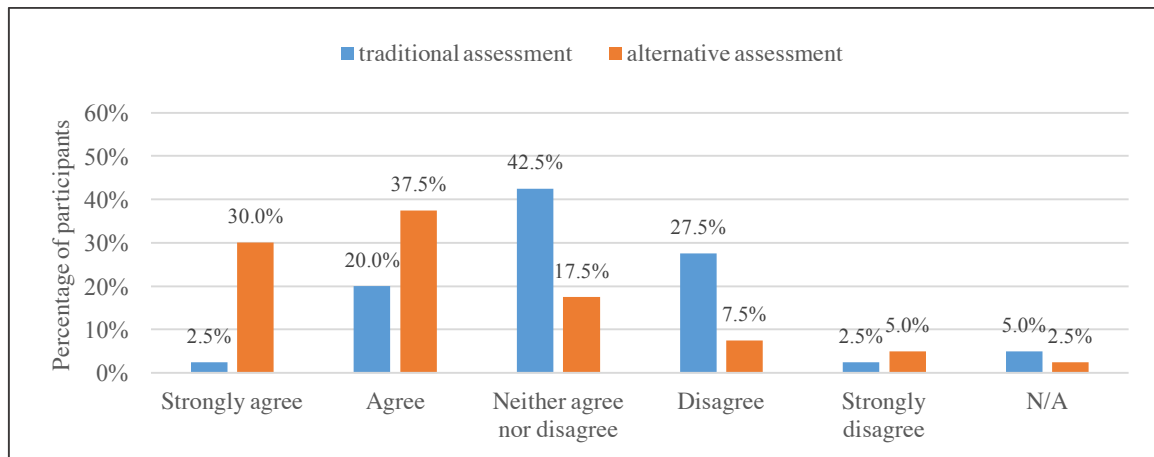
**Figure 15: Answers to items no. 94/105 “... they motivate pupils to learn English.”**

Items number 94 and 105 examined if the participants think that either form of language assessment rather tends to support motivation for learning English. As can be seen in Figure 15, more teachers expect alternative assessment to trigger language learning motivation as 30% strongly agree and 42.5% agree that alternative assessment motivates pupils while only 15% of the teachers think that traditional assessment also motivates pupils.



**Figure 16: Answers to items no. 95/106 “... they provide the opportunity to assess the pupils' performance in real-world tasks and exercises.”**

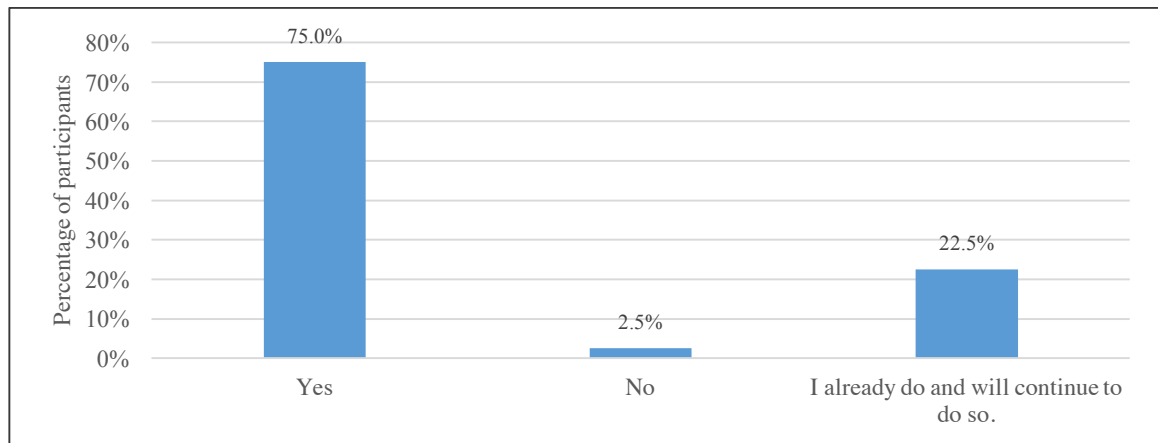
Looking at the data obtained from items 95 and 106, it can be inferred that teachers expect alternative assessment being more suited to incorporate real world tasks set in real world situations. Considering alternative assessment, conflated positive responses amount to almost 80%. Whether teachers believe that traditional assessment supports the application of authentic tasks cannot be inferred from the results as the same proportion of teachers who agree (2.5% and 20%) also disagree (20% and 2%) with the statement. In addition, more than half of all surveyed teachers neither agree nor disagree (52.5%) to item 105.



**Figure 17: Answers to items no. 96/107 “... pupils benefit from the assessment process.”**

The last items that compare teachers’ attitudes towards traditional and alternative assessment in terms of a specific characteristic deal with the question as to whether pupils benefit from either assessment process. According to the surveyed teachers, the majority (30% strongly agree and 37.5% agree) is convinced that alternatives can be beneficial, while a proportion of participants (2.5% strongly agree and 20% agree) believe that pupils profit from traditional assessment.

The Wilcoxon-signed-rank-test was applied in order to determine whether the presented participants’ beliefs towards statements related to traditional assessment differ from the participants’ beliefs towards the same statements in relation to alternative assessment. It is the appropriate test for non-parametric data based on two related sets of data provided by the same participants (Field 2009: 552). The Wilcoxon signed-rank test compares each participant’s answer to both items. The differences are calculated and ranked for all participants. If the p-value is smaller than 0.05, there are statistically significant differences between the participants’ agreement on the statement related to traditional assessment on the one hand and alternative assessment on the other hand. The test statistic (see Appendix 2, p. 102) shows statistically significant differences for all item pairs except for item pair 87/98. A p-value of 0.222 indicates that there is no statistically significant difference between the participants’ beliefs towards both concepts of assessment in relation to their capabilities to provide information about pupils’ achievements.



**Figure 18: Answers to item no. 108 “Would you consider using alternatives in assessment in your future EFL classrooms?”**

The final item of this questionnaire section asked the teachers if they would consider alternatives in language assessment as a possibility for their future assessment routine. Only 2.5% of all participants do not consider alternatives as a valid choice for future language assessment in school. All others have a positive attitude towards alternative assessment tools. Three quarters consider alternatives in language assessment for the future and 22.5% already administer alternatives and will continue to do so.

#### **4.5. Final comments**

The final part of the questionnaire comprises three open-ended questions. The analysis in this subsection of questionnaire results is strictly in terms of frequency of answers while the interpretation follows in Chapter 5 as part of the discussion of the findings.

Item 109 reads *Which assessment tools do you find most useful in your English language classes and why?* Of the forty participants, thirty-six took the chance to answer the question (see Appendix 4, p. 111). The first step in analyzing the multifaceted answers is the attempt to group the answers in order to receive a general overview: 17 answers can be associated with traditional assessment forms, 11 answers can be categorized as alternative forms, 4 answers could not be grouped as either traditional or alternative and 4 answers are not specific to the item but rather general statements about assessment. In general, most responses deal with single assessment tools but only a few answers indicated why the said tool is most useful.

Item 110 asks for factors teachers consider when they decide to use a certain assessment tool (see Appendix 4, p. 112). Six respondents did not answer the question and four answers could not be interpreted as an answer to the question. The answers are even more diverse

for this item than for item 109. Nevertheless, some of the responses occur more often than others. For instance, issues considering practicality like time constraints were mentioned very often. The characteristics of pupils were definitely a concern, for instance age group/maturity of pupils. Another frequently voiced subject for the choice of particular assessment tools was legal regulations like the LBVO. In addition, an answer given by a number of respondents was that teacher education influences assessment use.

Responses to item 111 reveal whether the teachers use traditional assessment formatively (see Appendix 4, p. 112). The question was optional, which resulted in fewer answers compared to items 109 and 110. Sixteen responses were recorded. Generally speaking, the answers can be categorized into three groups. The biggest group comprises responses in which teachers explain that they use traditional forms formatively and how they manage to do so. A small group of three teachers point out that they do not use traditional forms in a formative way and, lastly, 3 respondents explicitly admit that they are not familiar with the differences between traditional and formative assessment, which means they could not answer the question.

## **5. Discussion**

Chapter 5 is dedicated to discussing the results of the questionnaire presented in Chapter 4. In doing so the results will be interpreted in order to provide answers for the research questions:

Q1.What types of assessment are used by current EFL teachers in Austria?

Q2.What are teachers' attitudes towards the use of alternative assessment in contrast to traditional assessment in Austrian EFL classrooms?

Q3.What factors contribute to teachers' decisions about assessment procedures?

Chapter 5 is divided into three parts, each one dedicated to one research question. Section 5.1. sheds light on the assessment approaches and the assessment tools which are being used by Austria's EFL teachers, emphasizing the distinction between traditional and alternative forms of assessment (Q1). Section 5.2. will interpret the teachers' responses in terms of beliefs about and attitudes towards alternative and traditional assessment (Q2). Finally, Section 5.3. determines whether certain factors influence the teachers' decisions about utilizing assessment tools (Q3).

### **5.1. EFL assessment use & know-how**

Generally speaking, this section is divided into two intertwined parts. First, conclusions shall be drawn in terms of assessment purposes, because the reason for using assessment in the first place influences the selection of assessment types, which is why a connection might be established between assessment purposes and the use of either traditional or alternative assessment methods. Second, it shall be inferred which tools are used out of conviction.

Section 2.2 already explained the various purposes of educational assessment. In Section 4.2 the frequency of answers concerning purposes for EFL assessment was analyzed. At 82.5%, using assessment as a means of grading and certification was the option that was ticked by almost all of the respondents. In Austria, grading and certification can be described as the final stage of evaluating achievement. It is a process that is a specific, legal requirement which needs to be satisfied by all Austrian teachers. Closely related to the purpose of grading and certification is the use of assessment in order to evaluate achievement, which was indicated as very important by approximately two thirds of the participating teachers. In particular the evaluation of achievement, which is considered as fulfilling the summative purposes of assessment, is associated with traditional language assessment by Hurta-Macías (2002: 339). The actual second most frequently voiced purpose for administering assessment is gathering information about pupils' progress. Evaluating progress might be interpretable as relating to both traditional and alternative assessment. While measuring pupils' progress could be used as a paraphrase for measuring pupils' achievement towards learning objectives, the evaluation of progress is also a crucial part of formative and therefore alternative assessment. As indicated by Wiliam and Thompson (2008, quoted in Turner and Purpura 2016: 256), determining where pupils are in their learning process is a vital part of formative assessment, because it lays the groundwork for determining objectives and deciding how to reach those objectives.

Two more purposes for language assessment were chosen by approximately 20% of the participants. First, eleven teachers picked examination of pupils' strengths and weaknesses as one of the three most important functions of assessment in their EFL classes. Investigating pupils' strengths and weaknesses is strongly associated with formative uses of assessment. "Being responsive to learner needs" (Dickins and Gardner 2000: 217) is described as a key component of formative assessment. However, Black and Wiliam (1998: 140) emphasized that the information needs to be used to alter learning processes and/or teaching processes to be really considered as formative assessment. Nevertheless, the

evaluation of strengths and weaknesses could also be a means of gathering information that is strictly used to evaluate achievement for grading purposes. However, assessment to improve future learning was also selected as one of the three most significant assessment functions by 22.5 percent of all respondents.

The attempt to find correlations between assessment purposes and certain categories of participants was partly successful. The idea was to examine whether teaching experience and/or the type of degree has an impact on the participants' assessment practice. According to the test statistics, there is a significant association between the type of degree and the reasons assessment is used. The data show that participants that graduated from PÄDAK/PHs use assessment for different purposes than University graduates. What is most striking about the results is that almost a third of all University graduates indicated that they use assessment in order to foster future learning processes which can be associated with alternative forms in assessment or formative assessment. None of the PÄDAK/PH graduates chose the same purpose. What could be inferred from the different approaches to assessment is that different teacher education programs might focus on different assessment concepts.

The second part of this section discusses how often certain assessment tools are used by EFL teachers. Items 24-37 revealed how often teachers administer certain assessment tools. The given answer options are either largely associated with traditional or alternative forms of language assessment. The analysis of the data obtained revealed that assessment forms that are perceived as typically traditional are administered more often than prime examples for alternative EFL assessment tools. Moreover, the data show that those assessment procedures that are explicitly mentioned in the LBVO are represented more frequently than others. For instance, the most commonly used assessment methods used in descending order are homework, *Schriftliche Überprüfungen* 'written exams', revisions and class participation. This is not very surprising, because homework and written exams are obligatory by law. The LBVO might also be the reason for the less regular use of oral exams, because oral exams are not mandatory by the LBVO. However, every student is entitled to one oral exam per term. The more interesting investigation is how teachers make use of typical alternative forms of assessment. Looking at Figure 5 (see p. 63), a clear distinct difference becomes apparent between typically categorized concepts of alternative assessment and traditional assessment. The procedures that are associated with alternative language assessment are used very infrequently in comparison to traditional assessment procedures. While a small number uses them regularly, considerable numbers of

respondents seldom work with alternative assessment. For instance, portfolios are used regularly by 17.5% of the teachers. Peer-assessment, self-assessment and journals are applied regularly by even fewer teachers, although self- and peer-assessment is used seldom by 47.5% and 60% of the participants respectively. Journals seem to be the least favorite assessment procedure, as 70% of the teachers report that they never use them. Rubrics, which have been characterized as an alternative language assessment tool (cf. Brown & Hudson 1998; Brown & Abeywickrama 2010), are used by more than half of all teachers (11 teachers use them regularly and 12 seldom use them), although further investigation might be necessary to determine if those are used merely for grading purposes or if they are used formatively, for example for feedback purposes.

Another issue worth considering is unfamiliarity with certain assessment tools. Three teachers were unfamiliar with rubrics and journals and two teachers were unfamiliar with peer-assessment and self-assessment. All four concepts are considered alternative forms of assessment (cf. Brown & Hudson 1998; Brown & Abeywickrama 2010), often described as working formatively (cf. Stern 2010). One questionnaire item asked specifically for the participants' familiarity with the concept of alternative assessment. The attempt to associate familiarity with alternative assessment with the teachers' teaching experience or their type of degree was not successful. Statistically speaking, the test results show that familiarity with alternative assessment is independent of the participants' experience as teachers and their type of degree. That implies that neither a rather recent degree ensures familiarity with alternative assessment despite its rather recent popularity nor does a teaching degree from an Austrian university or an Austrian teacher training college. The same factors were taken into account in order to look for correlations between the familiarity with the concept of formative assessment and the teachers' teaching experience or their type of degree. The results, however, imply that familiarity with formative assessment cannot be associated with a recent degree or longer teaching experience than 6 years, nor with university teaching degrees or PH/PÄDAK degrees.

Item 108 asked for the most useful assessment tools in the teachers' EFL classes. Twenty-eight answers were recorded. As mentioned in Section 4.5, the majority of answers can be associated with traditional assessment. While some participants indicated "traditional assessment" or simply "traditional", others mentioned particular assessment tools and tried to explain why it is the most useful tool for their EFL classes. For instance, one of the respondents wrote down "SAs", which is an abbreviation for *Schularbeiten*, "because they provide a lot of information about pupils' achievement" or "Regular revisions – clearly

shows [sic] learning curve of students”. Tests in various forms are very popular in general. The teachers mentioned “tests”, “written exams”, “quizzes” and “check ups” several times. On the other hand, 12 teachers indicated that alternative assessment tools are most useful in their classes, mostly with more detailed explanations for their use. For instance, one respondent indicated “peer assessment” as most useful because “talking about strength [sic] and weaknesses is involved, students become aware of weaknesses and acquire a skill of analyzing their own work”. Feedback in various forms was also mentioned, for instance, “peer feedback, because it means a lot to students and boosts their motivation”.

It can be summarized that while alternatives in language assessment are used to some extent by almost all of the teachers, the prevailing tools used in Austrian EFL classes are forms of traditional assessment.

## **5.2. Teachers’ attitudes towards alternative assessment**

Besides the actual use of assessment tools, this study attempts to shed light upon teachers’ beliefs and attitudes concerning assessment. Of particular concern are the teachers’ attitudes towards alternatives in language assessment in contrast to traditional language assessment. The questionnaire comprises various items that tried to elicit the teachers’ beliefs and attitudes. Three sets of items are of particular interest. Section 4.4. illustrated the questionnaire results which shall now be interpreted. The first set of items were chosen to deduce what teachers think assessment is able to do and should do. The second and third set of items asked if teachers assign specific characteristics of language assessment to traditional assessment methods or alternatives in language assessment. The discussion refers to items from all sets simultaneously if they deal with related characteristics.

As already established in Section 5.1, grading as the final result of assessment processes is a major factor for teachers. The vast majority agrees that assessment is primarily used by teachers to determine grades, which should not be too surprising because grading is a mandatory feature of the Austrian educational system. In order to grade pupils, 70% of the teachers think that traditional assessment should be foremost be deployed to do so, while in comparison only 25% believe that alternatives can be used for grading purposes. In this regard, the results of items 91 and 102 should also be considered. 72.5% of the teachers indicated that traditional assessment should be used because it is mandatory by law. At the same time, being mandatory is not a decisive factor for administering alternatives in language assessment. So it can be inferred that more profound deliberations lead to the use



of alternative assessment, while one major reason for using traditional forms like examinations is the statutory requirements.

What is agreed upon by the majority of teachers is that, in general, assessment should feature various characteristics that have been associated with alternatives in language assessment in the literature so far. It can be inferred from the questionnaire results that teachers are aware of the benefits of formative assessment. First and foremost, 75% of the participants indicated that language assessment should feature formative as well as summative features. Whether or not this can be achieved by administering solely traditional forms of assessment, alternative forms of assessment or both could not be established from the questionnaire though. However, the sparse responses to item 110 reveal that about a third of the teachers manage to use traditional forms of assessment formatively. The item that asked whether assessment can fulfill several purposes simultaneously received 32.5% positive responses and 34.5% negative responses, which means that almost the same proportion of teachers believes in and reject the possibility of assessment fulfilling several purposes at the same time.

Moreover, assessment seems to be expected to foster learning processes, as 47.5% of the teachers agreed and 7.5% strongly agreed. Enhancing learning has been established as a characteristic of alternative assessment, especially formative assessment (Rea-Dickins & Gardner 2000: 217) which is also the view of the ARG, who first and foremost acknowledged the power of assessment to enhance learning (Wiliam 2011:10). The surveyed teachers assign the quality of fostering learning mainly to alternative assessment as can be seen in Figure 10. While 27.5% think traditional assessment might be able to improve pupils' learning, 72.5% believe that alternative forms should be used to do so. In close connection to the fostering of learning processes, more than 70% indicated that assessment should diagnose strengths and weaknesses which can be assumed to be crucial for enhancing learning in the first place. Similarly, when teachers were asked which forms of assessment lend themselves to this purpose, more than 85% think that alternatives in assessment are the forms to stick to. Nevertheless, 60% of the teachers indicated that traditional assessment procedures can also be a valid option to investigate strengths and weaknesses of students.

One of the premises of alternative language assessment is that the information gathered about pupils needs to be used proactively in order to satisfy their needs (Black & Wiliam 1998: 140). One way to do this is by adapting instruction. While agreement of teachers in

this regard is not as straight forward, 47.5% indicated that it is a viable option to alter instruction based on assessment results. In addition, more teachers believe that assessment is an option to investigate the strengths and weaknesses of their instruction. When asked whether traditional or alternative assessment should be used to do so, the data showed that teachers believe in alternative language assessment being the better choice for gathering information to improve future teaching. One possible reason for the teachers' attitudes and beliefs might be that alternative assessment is more open to the altering of instruction due to its process orientation. While traditional assessment is characterized by assessing products, for instance evaluating whether or not certain objectives have been met, alternatives in assessment emphasize the learning process, depicting an ongoing evaluation of learning. Therefore, it might be arguable that the improvement of instruction can be done while it is underway, thus answering pupils' needs on the spot instead of retrospectively. Nevertheless, for both options, more than 30% of the teachers are undecided if either assessment form is the more suitable option.

In relation to whether or not assessment might be usable to gather information to alter instruction meaningfully, it was also asked if assessment can be used to evaluate the quality of teaching. Stern (2010: 32) argues that the results of assessing pupils' achievement can be perceived as an indicator of teaching quality. The results of the questionnaire revealed that only 15% of the teachers indicated that assessment as a means of quality assurance is one of the main purposes for assessment in the first place. However, the surveyed teachers seem to have quite strong opinions about which form of assessment can provide the information to do so. By comparison, it becomes evident that alternative assessment might be the better choice if assessment should elicit information to evaluate the quality of teaching, as 30% of the teachers agreed and 17.5% strongly agreed that alternative assessment is a suitable option whereas only 17.5% are in favor of traditional assessment.

According to the teachers, feedback is another characteristic component of effective assessment. Except for 2.5%, all of the teachers believe that it is necessary to be provided in addition to grades. While feedback might be a supplement to traditional assessment forms like tests, it is a core feature of formative assessment (Black & Wiliam 2012: 18). In relation to already discussed items, feedback that functions formatively is described as descriptive feedback or formative feedback because it aims to explicitly improve learning positively (Shute 2008: 154) by providing information about future steps in the pupils' learning processes (Wiliam 2010: 21).

As Gardner states, learning a second language is not a necessity for most people, hence he infers that motivation to do so must play a crucial part in the process (Gardner 2007:10). Lightbown and Spada argue that there is a body of research that empirically establishes a connection between motivation and the “willingness to keep learning” (2006: 63). In addition, they argue that it is to some extent up to the teacher to boost the motivation of pupils for language learning. While the effects of motivation on acquiring a second language are considered a complex matter (Lightbown & Spada 2006: 63), the vast majority of the participating teachers indicated that assessment should foster intrinsic motivation for language learning. Intrinsic motivation is closely linked to driving language learning (Harlen 2012: 173) and it is the type of motivation all teachers would prefer to be displayed by their pupils. The results of the questionnaire suggest that the teachers value motivation based on interest instead of motivation caused by grades, prizes or similar incentives. When it comes to the comparison of traditional and alternative assessment in relation to motivation for learning English, 72.5% believe that alternative assessment should be used because they can motivate pupils to learn English. On the contrary, almost half of the teachers indicated that traditional assessment is not capable of motivating pupils to learn English. One possible reason for arguing for alternatives in language assessment could be that they are associated with the application of authentic material and tasks. Relevant content, as Lightbown and Spada (2006: 64) state is crucial to make a “positive contribution” to the motivation of pupils to learn a language. Contextualization and authenticity are crucial for teachers, as 55% strongly agreed and 35% agreed on their importance in language teaching. The teachers were asked which kind of assessment might be more suitable for integrating real world exercises and tasks. The data reveal that almost 80% of the teachers think that alternative assessment should be used to assess language performance in real world situations, while more than half of the teachers are not sure if traditional assessments could be used. Besides authenticity, Lightbown and Spada (2006: 64) argue that a supportive atmosphere is similarly important. Feedback, which is considered to be support in various stages of learning and a key component of formative assessment, might thus also be a crucial factor that turns teacher attitudes towards alternative forms in language assessment.

About half of the surveyed teachers think that self- and peer-assessment should be taken into account when pupils’ achievements are assessed. One reason for these attitudes might be an ongoing shift towards learner-centered teaching and assessment. Whether or not self- and peer-assessment can be implemented in traditional forms of assessment was not part of

this study; however, both concepts are considered a primary feature of alternatives in language assessment in general and formatively functioning assessment for learning in particular (Broadfoot *et al.* 2002). The ARG promotes the development of reflective and self-managing learning which is a main concern of self- and peer-assessment as part of the assessment for learning concept. Similarly, the CEFR was explicitly designed to enable self-directed learning, which promotes the application of self-assessment procedures (Council of Europe 2001: 6). Another conceivable reason for the teachers' beliefs is based on the introduction of educational standards and the concomitant orientation towards competences that has been implemented in the Austrian educational system. Besides subject-specific competences, cross-curricular competences are promoted, for instance as part of all Austrian curricula. Specified therein are *Didaktische Grundsätze* 'didactic principles' and *Allgemeine Bildungsziele* 'general educational objectives' among which so-called self-competence can be found which builds self-evaluation and self-reliance (Eder & Hofmann 2012: 72).

Practicality seems to be of major concern for teachers when it comes to assessment. The results obtained for the open questions revealed that practicality issues like time constraints are essential for the choice of assessment forms. Half of the teachers would even prefer assessment that can be scored automatically, for instance with the help of machines or computers. A quarter of all teachers would not consider automatically scored assessment a viable option. Nevertheless, practicality apparently dictates that teachers use traditional assessment procedures. Close to 50% of all teachers agree and 12.5% strongly agree that traditional assessment forms are the rational choice when practicality is concerned. The attitudes towards alternatives in this regard show quite a different picture because more than 50 percent of all teachers do not think administering alternative assessment is practical. A possibility for teachers thinking this way might be simply familiarity with certain tools. Administering new forms of assessment, that is, procedures which have never been used by the teachers, takes time and effort on the part of the teachers. Assessment needs to be prepared in advance but also takes up time from the lessons in order to make the pupils familiar with those tools. Traditional forms like tests, quizzes, revisions or homework are well known amongst teachers and pupils, thus presenting a considerable advantage in practicality over lesser-known alternative forms of assessment.

Finally, the last question asked whether teachers think that pupils might benefit from traditional or alternative assessment. Figure 17 (see p. 72) shows that 67.5% of the teachers think that alternatives should be used because pupils might benefit from the assessment

process. Even so 22.5% think that students can also benefit from traditional approaches. However, even more teachers (30%) think that pupils will not benefit from traditional assessment. The reason for those beliefs seems to be rather obvious. While traditional assessment focuses on the evaluation of achievement, usually at the end of learning processes when certain learning objectives should have been achieved, alternatives, especially formative assessment, aim to support and improve pupils' language learning. So the purposes of both approaches imply that pupils will benefit from the administering of alternatives in language assessment. However, it needs to be considered that there are possibilities to use traditional forms of assessment formatively, meaning in alternative ways, because 25% of the surveyed teachers indicate that they somehow use assessment this way. While realization varies, a group of teachers tries to incorporate feedback into traditional forms of assessment. One teacher wrote that when tests are administered they use lessons to give detailed feedback to the students analyzing strengths and weaknesses and suggesting ways to improve. Others try to work with homework assignments and their corrections by using them to further self-reliance on the part of the pupils.

Summing up, teachers definitely acknowledge various assessment purposes and associate language assessment with characteristics that are associated with alternative forms. In addition, it can be concluded that the majority of teachers believe that alternative procedures can fulfill those purposes best, at least when they are contrasted with traditional assessment.

### **5.3. Factors contributing to teachers' decisions about assessment procedures**

In order to answer the third research question, the most important item from the questionnaire is item 109 because it explicitly asks for factors that contribute to the teachers' decisions about using specific forms of assessment. Frequently given answers were practicality and the relevant regulations. Thus, it can be inferred that traditional assessment forms like tests are used because of Austrian law. In combination with practicality issues, which are associated with alternatives in assessment for the most part, the preference towards administering traditional assessment forms is not surprising.

Another important factor that teachers raised was teacher education, which is why the results of items 54 to 66 are discussed here. Generally speaking, teachers' assessment routines seem to be highly influenced by teacher education experiences. The part of the questionnaire dealing with teacher education revealed that a proportion of 67.5% of the

teachers experienced a focus on traditional language assessment during their teacher training. Close to 60% of the teachers do not feel prepared for handling various kinds of language assessment. Only 12.5% experienced a focus on alternative methods during their assessment related courses. Furthermore, in-service training in assessment is not available according to 67.5 percent of the surveyed teachers.

Nevertheless, when the teachers were asked whether they feel comfortable doing certain tasks in regard to language assessment, it seems that most of them managed to teach themselves the different approaches to assessment and how to use them appropriately, but it is still the traditional forms of language assessment the outright majority (85%) of teachers feel most comfortable with. In relation to traditional language assessment, teachers indicated that preparing (80%) and scoring (75%) examinations do not present any problems. The item that asked whether or not teachers feel comfortable using alternatives in language assessment received 40% positive responses and 40% negative responses. Moreover, teachers are no strangers to washback related assessment features. The majority of teachers feel confident to use assessment for enhancing pupils' learning processes, altering instruction appropriately and for feedback purposes. Overall, 60% feel comfortable utilizing different kinds of assessment procedures.

In all, the most important decisive factors for teachers in the survey favor traditional forms of assessment. Regulations as well as practicality issues affect the choices of applied assessment approaches sustainably. Moreover, an emphasis of traditional assessment during teacher training also takes an essential role.

## **6. Conclusion**

The main purpose for carrying out this study was to investigate language assessment habits in Austrian EFL classes. The focus was on so-called alternatives in language assessment in contrast to traditional language assessment. Alternatives are promoted by researchers and educational institutions likewise and were classified as a "hot topic" (2010: 17) in language assessment by Brown and Abeywickrama. At the same time, the ÖZEPS published Stern's pamphlet "*Förderliche Leistungsbewertung [Formative assessment]*" (2010) in Austria. Institutions like the OECD promote alternative assessment for example for formative purposes (OECD 2005: 2). Generally speaking, alternatives in assessment have been part of the discourse of assessment for quite some time, which is why the questionnaire aimed

to reveal whether these forms of assessment have found their way into Austrian EFL classes and what attitudes to these assessment forms exist among EFL teachers.

The overall impression that can be received from the questionnaire results is that alternatives in language assessment are used to some extent but traditional assessment forms prevail in Austrian EFL classes. The majority of the teachers who participated in the survey think that traditional forms of assessment are most useful in their language classes. Written exams, homework assignments and revisions are very important for EFL teachers and therefore used on a regular basis. Typical alternative forms are not used frequently, for instance more than three quarters never use journals and some teachers are not even familiar with the concept. Some alternatives like portfolios are used more often; however, they are used much more infrequently in comparison to traditional language assessment procedures.

The reasons for this were found rudimentarily. Teachers argue that regulations influence their choice of assessment, which can be found first and foremost in the Austrian LBVO. Practicality issues are of concern and to some extent whether or not the teachers are trained to use certain assessment approaches appropriately or, more generally, know about the concepts at all. Teacher education apparently focuses on traditional forms of assessment and a considerable number of teachers would have liked to deal with alternative assessment forms in more detail when they were teacher trainees. In addition, while everyone might agree that assessment is a crucial, if not one of the most crucial aspects of a teacher's daily routine, basically all of the surveyed teachers would have welcomed more or specific courses about language assessment, not necessarily solely about alternative forms but language assessment in general. Personally, I believe that future teachers should be familiarized with all forms of assessment during their teacher education. That way, an adjustment in assessment habits will probably be simpler and an easily conceivable anxiety of tapping into new assessment grounds can be avoided in the first place.

Parts of the questionnaire focused on teachers' attitudes towards assessment, what they perceive to be the purpose of assessment, their experiences about what it can do and if they believe that some assessment purposes might be better fulfilled by applying a specific assessment approach. Overall, it can be concluded that teachers believe more strongly in the capabilities of alternatives in language assessment compared to traditional forms. Even when assessment purposes like the evaluation of language achievement are challenged, teachers believe that alternatives suit those needs equally well. This is somehow problematic because, naturally, assessment that poses advantages should be the one to use.

The majority of teachers believe that pupils would benefit from alternatives but less so from traditional forms of assessment. If that is the case, the question remains why traditional assessment is applied more frequently than its counterpart. After all, the underlying premise in teacher cognition studies is that instructional practice is highly influenced by the teachers' thoughts, beliefs and attitudes. In the case of this diploma thesis, the positive attitudes and beliefs concerning alternatives in language assessment would suggest that teachers would use them more frequently. It appears, however, that environmental circumstances compared with certain doubts concerning single aspects of alternative assessment favor traditional forms.

**Word count: 31,383**



## 7. Bibliography

- Andrade, Heidi G.; Bouley, Beth A. 2003. "Role of rubric-referenced self-assessment in learning to write". *The Journal of Educational Research* 97(1), 21-30.
- Andrews, Stephen. 1994. "The washback effect of examinations: its impact upon curriculum innovation in English language teaching". *Curriculum Forum* 1, 44–58.
- Anthony, Robert J.; Johnson, Terry D.; Mickelson, Norma L.; Preece, Alison. 1991. *Evaluating literacy: A perspective for change*. Portsmouth: Heinemann.
- Bachman, Lyle. 1990. *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachmann, Lyle F.; Palmer, Adrian S. 1996. *Language testing in practice*. Oxford: Oxford University Press.
- Bailey, Kathleen M. 1996. "Working for washback: a review of the washback concept in language testing". *Language Testing* 13(3): 257–279.
- Bennett, Randy E. 2009. *A critical look at the meaning and basis of formative assessment*. Princeton, NJ: Educational Testing Service.
- Bloom, Benjamin. S. 1969. "Some theoretical issues relating to educational evaluation". In Tyler, Ralph W. (ed.). *Educational evaluation: New roles, new means*. Chicago, IL: University of Chicago Press, 26-50.
- BIFIE. 2013. *Standardisierte kompetenzorientierte Reifeprüfung. Reife- und Diplomprüfung [Standardized competence-based school leaving exam]*. Wien: BIFIE. [https://www.srdp.at/fileadmin/user\\_upload/downloads/SRDP/SRDP-Grundlagen-Entwicklung-Implementierung-2013-11-04.pdf](https://www.srdp.at/fileadmin/user_upload/downloads/SRDP/SRDP-Grundlagen-Entwicklung-Implementierung-2013-11-04.pdf) (20 Mar. 2017)
- BIFIE. 2016a. *Bildungsstandards [Educational standards]*. <https://www.bifie.at/bildungsstandards> (30 Aug. 2016).
- BIFIE. 2016b. *PISA-Studie für 15-/16-Jährige [PISA survey for 15-/16-year-olds]*. <https://www.bifie.at/pisa> (30 Aug. 2016).
- Birenbaum, Menucha. 1996. "Assessment 2000: towards a pluralistic approach to assessment". In Birenbaum, Menucha; Dochy, Filip J. R. C. (eds.). *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston, MA: Kluwer, 3-30.
- BMB. 2017a. *Schulrecht [School laws]*. <https://www.bmb.gv.at/schulen/recht/index.html> (11.04.2017)
- BMB. 2017b. *Standardisierte Reife- und Diplomprüfung [Standardized school leaving exam]*. <https://www.srdp.at/> (04 Jan. 2016).
- BMBWK. 2002. *White Book. Quality development and quality assurance in the Austrian system of education*. Vienna: BMBWK.
- Borg, Simon. 2003. "Teacher cognition in language teaching: a review of research on what language teachers think, know, believe, and do". *Language Teaching* 36(2), 81-109.
- Borg, Simon. 2009. *Introducing language teacher cognition*. <http://www.education.leeds.ac.uk/research/files/145.pdf> (30 Feb. 2017).
- Broadfoot, Patricia M.; Daugherty, Richard; Gardner, John; Gipps, Caroline; Harlen, Wynne; James, Mary; Stobart, Gordon. 1999. *Assessment for learning: beyond the black box*. Cambridge: University of Cambridge School of Education.
- Broadfoot, Patricia M.; Daugherty, Richard; Gardner, John; Gipps, Caroline; Harlen, Wynne; James, Mary; Stobart, Gordon. 2002. *Assessment for learning: 10 Principles*. Cambridge: University of Cambridge School of Education.

- Brock, Rainer. 2010. "Die Bildungsstandards E8: kompetenzorientierter Unterricht in Englisch [Educational standards E8: competence-based instruction in English]". *Erziehung und Unterricht* 3(4), 348-354.
- Brown, Douglas H.; Abeywickrama, Priyanvada. 2010. *Language assessment: principles and classroom practices*. (2<sup>nd</sup> edition). New York, NY: Pearson Longman.
- Brown, James D.; Hudson, Tom. 1998. "The alternatives in language assessment". *TESOL Quarterly* 32(4), 653-675.
- Buehl, Michelle M.; Beck, Jori S. 2015. "The relationship between teachers' beliefs and teachers' practices". In: Fives, Helenrose; Gill, Michele G. (eds.). *International handbook of research on teachers' beliefs*. New York, NY: Routledge, 66-84.
- Buschmann, Ingrid; Thonhauser, Josef. (2000). "Im Namen, aber nicht im Sinne des Gesetzes: zur Praxis der Leistungsbeurteilung [In the name of but not within the meaning of the law: assessment practice]". *Erziehung und Unterricht*, 150(1-2), 173-191.
- Cambourne, Brian. 1988. *The whole story: natural learning and the acquisition of literacy in the classroom*. Auckland: Ashton Scholastic.
- CERI. 2005. "Assessment for Learning. Formative Assessment". <http://www.oecd.org/site/educeri21st/40600533.pdf> (5 Jan. 2017)
- Cheng, Liying; Rogers, Todd W.; Hu, Huiqin. 2004. "ESL/EFL instructors' classroom assessment practices: purpose, methods, and procedures". *Language Testing* 21(3), 360-389.
- Chudowsky, Naomi; Behuniak, Peter. 1998. "Using focus groups to examine the consequential aspect of validity". *Educational Measurement: Issues and Practice* (17)4, 28-38.
- Chun, Christian W. 2006. "Commentary: An analysis of a language test for employment: the authenticity of the PhonePass test". *Language Assessment Quarterly* 3(3), 295-306.
- Council of Europe. 2001. *Common European Framework of Reference for languages: learning, teaching, assessment*. <http://www.coe.int/lang-CEFR> (28 Apr. 2017).
- Council of Europe. 2002. "Legislative acts and other instruments". [http://www.coe.int/t/dg4/linguistic/UE-CECR-2002\\_EN.pdf](http://www.coe.int/t/dg4/linguistic/UE-CECR-2002_EN.pdf) (26.12.2016).
- Council of Europe. 2014. *Education and languages, language policy*. [http://www.coe.int/t/dg4/linguistic/cadre1\\_en.asp](http://www.coe.int/t/dg4/linguistic/cadre1_en.asp) (28.8.2016).
- Council of Europe. 2016. "European Language Portfolio (ELP)" <http://www.coe.int/en/web/portfolio> (26.12.2016)
- Crooks, Terence J. 1988. "The impact of classroom evaluation practices on students". *Review of Educational Research*, 438-481.
- Dalton-Puffer, Christiane; Faistauer, Renate; Vetter, Eva. 2011. "A country in focus: research on language teaching and learning in Austria (2004-2009)". *Language Teaching* 44(2), 181-211.
- Dochy, Filip. 2001. "A new assessment era: different needs, new challenges". *Research dialogue in learning and instruction* 2, 11-20.
- Dörnyei, Zoltán. 2001. *Strategies in the language classroom*. Cambridge: Cambridge University Press.
- Dörnyei, Zoltán. 2003. *Questionnaires in second language research: Construction, administration, and processing*. Mahwah, NJ: Lawrence Erlbaum.
- Dörnyei, Zoltán. 2007. *Research methods in applied linguistics: Quantitative, qualitative and mixed methodologies*. Oxford: Oxford University Press.
- Eder, Ferdinand; Neuweg, Georg H.; Thonhauser, Josef. 2009. "Leistungsfeststellung und Leistungsbeurteilung [Performance assessment and performance evaluation]". In Specht, Werner (ed.). *Nationaler Bildungsbericht Österreich 2009, Band 2:*

- Fokussierte Analysen bildungspolitischer Schwerpunktthemen [National education monitoring report for Austria 2009, Volume 2: focused analysis of key areas of education policy]*. Graz: Leykam, 247-267.
- Eller, Barbara; Schratz, Michael. 2009. "Country case study Austria". In Vlaardingerbroek, Barend; Taylor, Neil (eds.). *Secondary school external examination systems: reliability, robustness and resilience*. New York, NY: Cambria Press, 225-240.
- Ekbatani, Glayol; Pierson, Herbert. (eds.) 2000. *Learner-directed assessment in ESL*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Ekbatani, Glayol. 2000. "Moving toward learner-directed assessment". In Ekbatani, Glayol; Pierson, Herbert. (eds.), *Learner-directed assessment in ESL*, 1-11. Mahwah, NJ: Lawrence Erlbaum Associates.
- Fend, Helmut. 2006. *Neue Theorie der Schule: Einführung in das Verstehen von Bildungssystemen [New theories of school: introduction to an understanding of educational systems]*. Wiesbaden: Verlag für Sozialwissenschaften.
- Feigl, Susanne. 2007. *Informationsblätter zum Schulrecht Teil 3: Leistungsfeststellung und Leistungsbeurteilung [Fact sheets on school laws part 3: performance assessment and performance evaluation]*. Wien: BMUKK.
- Field, Andy. 2009. *Discovering statistics using SPSS*. (3<sup>rd</sup> edition). London: SAGE.
- Fives, Helenrose; Buehl, Michelle M. 2012. "Spring cleaning for the "messy" construct of teachers' beliefs: What are they? Which have been examined? What can they tell us?". In: Harris, Karen R.; Graham, Steve; Urdan, Tim (eds.). *APA educational psychology handbook, volume 2: individual differences and cultural and contextual factors*. Washington, DC: APA, 471-499.
- Gallagher, C. J. 2003. "Reconciling a tradition of testing with a new learning paradigm". *Educational Psychology Review* 15(1), 83-99.
- Genesee, Fred; Upshur, John A. 1996. *Classroom-based evaluation in second language education*. Cambridge: Cambridge University Press.
- Goodman, Jodi S.; Wood, Robert E.; Hendrickx, Margaretha. 2004. "Feedback Specificity, Exploration, and Learning". *Journal of Applied Psychology* 49(2), 248-262.
- Google Inc. 2012. *Google Forms*. [Computer Program]. <https://docs.google.com/forms/> (20 Dec. 2016)
- Green, Anthony. 2007. *IELTS washback in context: preparation for academic writing in higher education*. Cambridge: Cambridge University Press.
- Hansen Edwards, Jette G. 2014. "Peer assessment in the classroom". In Kunnan, Anthony J. (ed.). *The companion to language assessment. Vol. 2*. Chichester: Wiley Blackwell, 730-750.
- Hamp-Lyons, Liz; Condon, William. 2000. *Assessing the portfolio: principles for practice, theory, and research*. Cresskill, NJ: Hampton Press.
- Hattie, John; Timperley, Helen. 2007. "The power of feedback". *Review of Educational Research* 77(1), 81-112.
- Heritage, Margaret. 2007. "Formative assessment: what do teachers have to know and do?". *Phi Delta Kappan* 89(2), 140-145.
- Herman, Joan L.; Aschbacher, Pamela L.; Winters, Lynn. 1992. *A practical guide to alternative assessment*. Alexandria, VI: ASCD.
- Holec, Henri. 1979: *Autonomy and foreign language learning*. Strasbourg: Council of Europe.
- Hoyle, Eric; John, Peter D. 1995. *Professional knowledge and professional practice*. London: Cassell.

- Hu, Guangwei. 2005. "Using peer review with Chinese ESL student writers". *Language Teaching Research* 9(3), 321-342.
- Huerta-Macías, Ana. 2002 [1995]. "Alternative assessment: responses to commonly asked questions". *TESOL Journal*, 5, 8-11.
- Hughes, Arthur. 2003. *Testing for language teachers*. (2nd edition). Cambridge: Cambridge University Press.
- IBM Corporation. 2016. *IBM SPSS Statistics*. (Version 24). [Computer Program]
- Jouqhlin, Gordon. 2009. "Introducing: refocusing assessment". In Jouqhlin, Gordon (ed.). *Assessment, learning and judgment in higher education*. Wollongong: Springer, 1-12.
- Kahl, Stuart. 2005. "Where in the world are formative tests? Right under your nose!". *Education Week* 25, 26 Oct., 11.
- Kirk, Jerome; Miller, Marc L. 1986. *Reliability and validity in qualitative research*. Beverly Hills, CA: SAGE.
- Kohonen, Viljo. 2001. "Exploring the educational possibilities of the "Dossier": suggestions for developing the pedagogical function of the European Language Portfolio". In Kohonen, Viljo; Westhoff, Gerard. *Enhancing the pedagogical aspects of the European Language Portfolio (ELP)*. Strasbourg: Council of Europe.
- Kulhavy, Raymond W. 1977. "Feedback in written instruction". *Review of Educational Research* 47, 211-232.
- Lantolf, James P.; Poehner, Matthew E. 2004. "Dynamic assessment: bringing the past into the future". *Journal of Applied Linguistics* 1(1), 49-74.
- Lantolf, James P.; Poehner, Matthew E. 2014. *Sociocultural theory and the pedagogical imperative in L2 education: Vygotskian praxis and the research/practice divide*. London: Routledge.
- LBVO, *Leistungsbeurteilungsverordnung [Austrian act for assessment]*. 2016. <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009375> (11 May 2017).
- Leahy, Siobhan; Lyon, Christine; Thompson, Marnie; Wiliam, Dylan. 2005. "Classroom assessment: minute-by-minute and day-by-day". *Educational Leadership* 63(3), 18-24.
- Lenz, Peter. 2004. "The European Language Portfolio". In Morrow, Keith (ed.). *Insights from the Common European Framework*. Oxford: Oxford University Press, 22-31.
- Leung, Constant; Mohan, Bernard. 2004. "Teacher formative assessment and talk in classroom contexts: assessment as discourse and assessment of discourse". *Language Testing* 21(3), 336-359.
- Lewkowicz, Jo A. 2000. "Authenticity in language testing: some outstanding questions". *Language Testing* 17(1), 43-64.
- Little D. 2009. *The European language portfolio: where pedagogy and assessment meet*. <http://www.coe.int/portfolio> (15.9.2016).
- Messick, Samuel J. 1989. "Validity". In Linn, R.L. (ed.). *Educational measurement* (3rd ed.). New York, NY: Macmillan, 13-103.
- Microsoft Corporation. 2015. *Microsoft Excel*. (Version 15.17). [Computer Program].
- Morgan, Carol. 2006. "Appropriate language assessment in content and language integrated learning". *Language Learning Journal* 33, 59-67.
- Morrow, Keith. 1979. "Communicative language testing: revolution of evolution?". In: Brumfit, C.K., Johnson, K. (eds.), *The communicative approach to language teaching*. Oxford: Oxford University Press, 143-159.
- Moser, Wolfgang. 2007. "Bildungsstandards für Fremdsprachen [Educational standards for foreign languages]". *Erziehung und Unterricht* 7-8, 613-621.

- Mousavi, Seyyed A. 2009. *An encyclopedic dictionary of language testing* (4<sup>th</sup> edition). Tehran: Rahnama Publications.
- Neuweg, Georg H. 2014. *Schulische Leistungsbeurteilung [Educational assessment]* (5<sup>th</sup> edition). Linz: Trauner.
- Norris, John M.; Brown, James D.; Hudson, Thom; Yoshioka, Jim. 1998. *Designing second language performance assessments*. Honolulu: University of Hawai'i Press.
- Nuffield Foundation. 2016. Assessment Reform Group.  
<http://www.nuffieldfoundation.org/assessment-reform-group> (26.12.2016).
- Nunan, David. 1988. "Syllabus design". In Candlin, Christopher N.; Widdowson, Henry G. (eds.). *Language Teaching: a scheme for teacher education*. Oxford: Oxford University Press.
- Oller, John W. 1979. *Language tests at school: a pragmatic approach*. London: Longman.
- O' Malley, Michael J.; Valdez Pierce, Lorraine. 1996. *Authentic assessment for English language learners: practical approaches for teachers*. Harlow: Longman.
- Oscarson, Mats. 2014. "Self-assessment in the classroom". In Kunnan, Anthony J. (ed.). *The companion to language assessment*. Vol. 2. Chichester: Wiley Blackwell, 712-729.
- Paulson, Leon F.; Paulson, Pearl R.; Meyer, Carol A. 1991. "What makes a portfolio a portfolio?". *Educational Leadership* 48(5), 60-63.
- Pellegrino, James W.; Chudowsky, Naomi; Glaser, Robert. (eds.). 2001. *Knowing what students know: the science and design of educational assessment*. Washington, DC: National Academy Press.
- Poehner, Matthew E. 2008. *Dynamic assessment: a Vygotskian approach to understanding and promoting second language development*. Berlin: Springer.
- Ramaprasad, Arkalgud. 1983. "On the definition of feedback". *Systems Research and Behavioral Science* 28(1), 4-13.
- Rasinger, Sebastian. 2008. *Quantitative research in linguistics: an introduction*. London: Continuum.
- Rea-Dickins, Pauline; Gardner, Sheena. 2000. "Snares and silver bullets: disentangling the construct of formative assessment". *Language Testing* 17(2), 215-243.
- Rea-Dickins, Pauline. 2001. "Mirror, mirror on the wall: identifying processes of classroom assessment". *Language Testing* 18(4), 429-462.
- Rea-Dickins, Pauline. 2004. "Editorial: understanding teachers as agents of assessment". *Language Testing* 21(3), 249-258.
- Rea-Dickins, Pauline. 2006. "Currents and eddies in the discourse of assessment: a learning-focused interpretation". *International Journal of Applied Linguistics* 16(2), 163-188.
- Sadler, Royce D. 1989. "Formative assessment and the design of instructional systems". *Instructional Science* 18(2), 119-144.
- Saville, Nick; Khalifa, Hanan. 2016. "The impact of language assessment". In: Tsagari, Dina and Banerjee, Jayanti (eds.). *Handbook of second language assessment*. Boston, MA: De Gruyter Mouton, 77-94.
- Schmidinger, Elfriede; Hofmann, Franz; Stern, Thomas. 2015. "Leistungsbeurteilung unter Berücksichtigung ihrer formative Funktion [Assessment in consideration of its formative function]". Bruneforth, Michael; Eder, Ferdinand; Krainer, Konrad; Schreiner, Claudia; Seel, Andrea; Spiel, Christiane (eds.). *Nationaler Bildungsbericht Österreich 2015, Band 2: Fokussierte Analysen bildungspolitischer Schwerpunktthemen [National education monitoring report*

- for Austria 2015, Volume 2: focused analysis of key areas of education policy]. Graz: Leykam, 59-94.
- SchUG, Schulunterrichtsgesetz [Austrian school education act]. 2016. <https://www.ris.bka.gv.at/GeltendeFassung.wxe?Abfrage=Bundesnormen&Gesetzesnummer=10009600> (11 May 2017)
- Scriven, Michael. 1967. "The methodology of evaluation". In Tyler, Ralph W.; Gagné, Robert M.; Scriven, Michael. (eds.). *Perspectives of curriculum evaluation. Volume 1*. Chicago, IL: Rand McNally, 39–83.
- Scott, Tony. 2004. "Teaching the ideology of assessment". *Radical teacher* 71(4), 30-37.
- Shohamy, Elana. 1993. *The power of tests: The impact of language tests on teaching and learning: NFLC occasional paper*. Washington, DC: National Foreign Language Center.
- Shohamy, Elana. 2001. *The power of tests: a critical perspective on the uses of language tests*. London: Pearson.
- Shute, Valerie J. 2008. "Focus on formative feedback". *Review of Educational Research* 78(1), 153-189.
- Stern, Thomas. 2010. *Förderliche Leistungsbewertung [Formative assessment]*. Wien: ÖZEPS.
- Thonhauser, Josef. 2000. "Neue Lehrpläne: Versuch einer Beurteilung unter erziehungswissenschaftlichem Aspekt [New curricula: an assessment attempt in terms of educational science]". *Erziehung und Unterricht* 150(7-8), 841-851.
- Tsagari, Dina; Banerjee, Jayanti. (eds.) 2017. *Handbook of second language assessment*. Boston, MA: De Gruyter Mouton.
- Tunstall, Pat; Gipps, Caroline. 1996. "Teacher feedback to young children in formative assessment: a typology". *British Educational Research Journal* 22(4), 389–404.
- Turner, Carolyn E.; Purpura, James E. 2016. Learning-oriented assessment in second and foreign language classrooms. In Tsagari, Dina; Banerjee, Jayanti (eds.). *Handbook of second language assessment*. Boston, MA: De Gruyter Mouton, 255-273.
- Vygotsky, Lev S. 1997. "The historical meaning of the crisis in psychology: a methodological investigation". In Rieber, Robert W.; Wollock, Jeffrey (eds.). *The collected works of L.S. Vygotsky. Volume 3: problems of the theory and history of psychology*. New York, NY: Plenum, 233-344.
- Vygotsky, Lev S. 1978. *Mind in society: the development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Weeden, Paul; Winter, Janand. 1999. *The learn project: learners' expectations of assessment for learning national: report for the qualifications and curriculum authority*, September 1999. Bristol: University of Bristol, Graduate School of Education.
- Weigle, Sara C. 2002. *Assessing writing*. Cambridge: Cambridge University Press.
- Wiater, Werner. 2009. "Zur Definition und Abgrenzung von Aufgaben und Funktionen der Schule [About the definition and differentiation of duties and functions of school]". In Blömeke, Sigrid; Bohl, Thorsten; Haag, Ludwig; Lang-Wojtasik, Gregor. (eds.). *Handbuch Schule. Theorie – Organisation – Entwicklung [Handbook school: theory – organization – development]*. Bad Heilbronn, 65-71.
- Wiliam, Dylan; Thompson, Marnie. 2008. "Integrating assessment with instruction: what will it take to make it work?". In Dwyer, Carol A. (ed.). *The future of assessment: shaping teaching and learning*. Mahwah, NJ: Routledge, 53-82
- Wiliam, Dylan. 2011. "What is assessment for learning?". *Studies in Educational Evaluation* 37(1), 3-14.
- Wiliam, Dylan. 2010. "An integrative summary of the research literature and implications for a new theory of formative assessment". In Andrade, Heidi. L.; Cizek, Gregory

- J. (eds.). *Handbook of formative assessment*. New York, NY: Taylor & Francis, 18-40.
- Xi, Xiaoming; Davis, Larry. 2016. "Quality factors in language assessment". In Tsagari, Dina; Banerjee, Jayanti (eds.). *Handbook of second language assessment*. Boston, MA: De Gruyter Mouton, 61-76.
- Yin, Muchun. 2014. "Portfolio assessment in the classroom". In Kunnan, Anthony J. (ed.). *The companion to language assessment. Vol. 2*. Chichester: Wiley Blackwell, 659-676.
- Yu, Guoxing. 2014. "Performance assessment in the classroom". In Kunnan, Anthony J. (ed.). *The companion to language assessment. Vol. 2*. Chichester: Wiley Blackwell, 617-630.

## **Appendix 1: Abstracts**

### **Abstract**

This diploma thesis examines assessment in Austrian EFL classes and focuses on juxtaposing traditional language assessment and the critically discussed alternatives in language assessment. The distinction between traditional and alternative language assessment is based upon characteristics that are prevalent in the literature. The empirical study follows a quantitative approach, investigating what types of assessment are administered by English teachers in Austrian secondary schools and to what purpose these procedures are applied. In addition, the study provides insights into teachers' attitudes and beliefs concerning alternatives in language assessment and crucial factors for the teachers' decisions about utilizing certain assessment tools. The analysis and interpretation of the obtained data established that the majority of the participating Austrian EFL teachers prefer traditional assessment methods, and alternatives in language assessment are infrequently used. Nevertheless, teachers seem to be convinced of the potential of alternative assessment, although decisive factors for using either approach favor traditional over alternative language assessment. Those factors are primarily issues of practicality and legal regulations.



## **Zusammenfassung**

Diese Diplomarbeit untersucht die Leistungsbeurteilungsgewohnheiten österreichischer Englisch-LehrerInnen. Besondere Berücksichtigung finden dabei alternative im Gegensatz zu traditionellen Leistungsbeurteilungsmethoden. Die Abgrenzung beider Leistungsbeurteilungskonzepte erfolgt dabei durch Merkmale, wie sie in der entsprechenden Forschungsliteratur zu finden sind. Der empirische Teil verfolgt einen quantitativen Ansatz. Mittels eines Online-Fragebogens sollen drei Forschungsfragen beantwortet werden. Erstens soll geklärt werden, welche Leistungsbeurteilungsmethoden von österreichischen Englisch-LehrerInnen der Sekundarstufe verwendet werden. Außerdem werden die Einstellungen der LehrerInnen gegenüber alternativer Leistungsbeurteilung erforscht. Schließlich zielt der Fragebogen darauf ab, die Faktoren für die Entscheidung über den Einsatz spezifischer Leistungsbeurteilungsmethoden aufzudecken. Aus der Analyse der Daten geht hervor, dass die Mehrzahl der Englisch-LehrerInnen nach wie vor traditionelle Leistungsbeurteilung wie z.B. Tests bevorzugen. Alternative Leistungsbewertung wird eher spärlich verwendet. Gründe für die seltene Nutzung alternativer Formen sind in erster Linie praktische Sachverhalte sowie gesetzliche Vorgaben. Dennoch zeigt die Analyse der Daten, dass obwohl meist traditionellen Formen der Vorzug gegeben wird, die LehrerInnen oft alternativer Leistungsbeurteilung mehr zutrauen als deren traditionellen Pendants.

## Appendix 2: Test statistics

		University		Teacher training college	
		N	Percent	N	Percent
Assessment Purposes	I use assessment to gather information about pupils' achievement.	21	63.6%	4	57.1%
	I use assessment to gather information about pupils' progress.	24	72.7%	5	71.4%
	I use assessment as a means of grading and certification.	28	84.8%	5	71.4%
	I use assessment to examine pupils' strengths and weaknesses.	9	27.3%	2	28.6%
	I use assessment to evaluate my teaching success.	0	0.0%	2	28.6%
	I use assessment as a means of quality assurance.	3	9.1%	3	42.9%
	I use assessment to improve future teaching processes.	4	12.1%	0	0.0%
	I use assessment to improve future learning processes.	9	27.3%	0	0.0%
Total		33	100%	7	100%

### Pearson Chi-Square Tests

		Degree (University/PH)
Assessment purposes	Chi-square	19.329
	df	8
	Sig.	.013

		0-6 years experience		7+ years experience	
		N	Percent	N	Percent
Assessment purposes	I use assessment to gather information about pupils' achievement.	17	65.4%	8	57.1%
	I use assessment to gather information about pupils' progress.	19	73.1%	10	71.4%
	I use assessment as a means of grading and certification.	22	84.6%	11	78.6%
	I use assessment to examine pupils' strengths and weaknesses.	7	26.9%	4	28.6%
	I use assessment to evaluate my teaching success.	0	0.0%	2	14.3%
	I use assessment as a means of quality assurance.	2	7.7%	4	28.6%
	I use assessment to improve future teaching processes.	4	15.4%	0	0.0%
	I use assessment to improve future learning processes.	6	23.1%	3	21.4%
Total		25	100%	15	100%

#### Pearson Chi-Square Tests

		years of experience
Assessment purposes	Chi-square	9.947
	df	8
	Sig.	.269

			Experience in years		Total
			0-6 years	7+ years	
Familiarity with the concept alternative assessment	I could explain the concept to my colleagues.	N	4	6	10
		Percent	28.6%	23.1%	25.0%
	I've heard this and know a little bit about it.	N	6	14	20
		Percent	42.9%	53.8%	50.0%
	I've heard this before but don't know what it means.	N	4	4	8
		Percent	28.6%	15.4%	20.0%
	I've never heard this before.	N	0	2	2
		Percent	0.0%	7.7%	5.0%
Total		N	14	26	40
		Percent	100.0%	100.0%	100.0%

#### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	2.198	3	.532
Likelihood Ratio	2.811	3	.422
Linear-by-Linear Association	.081	1	.776
N of Valid Cases	40		

			Degree		Total
			University	Teacher training college	
Familiarity with the concept alternative assessment	I could explain the concept to my colleagues.	N	9	1	10
		Percent	27.3%	14.3%	25.0%
	I've heard this and know a little bit about it.	N	16	4	20
		Percent	48.5%	57.1%	50.0%
	I've heard this before but don't know what it means.	N	6	2	8
		Percent	18.2%	28.6%	20.0%
	I've never heard this before.	N	2	0	2
		Percent	6.1%	0.0%	5.0%
Total	N	33	7	40	
	Percent	100.0%	100.0%	100.0%	

#### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	1,212	3	,750
Likelihood Ratio	1,583	3	,663
Linear-by-Linear Association	,110	1	,740
N of Valid Cases	40		

			Experience in years		Total
			7+ years	0-6 years	
Familiarity with the concept formative assessment	I could explain the concept to my colleagues.	N	10	5	15
		Percent	38,5%	35.7%	37.5%
	I've heard this and know a little bit about it.	N	6	5	11
		Percent	23.1%	35.7%	27.5%
	I've heard this before but don't know what it means.	N	7	3	10
		Percent	26.9%	21.4%	25.0%
	I've never heard this before.	N	3	1	4
		Percent	11.5%	7.1%	10.0%
Total		Count	26	14	40
		Percent	100.0%	100.0%	100.0%

#### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	.833	3	.842
Likelihood Ratio	.826	3	.843
Linear-by-Linear Association	.116	1	.734
N of Valid Cases	40		

			Degree		Total
			University	Teacher training college	
formative assessment	I could explain the concept to my colleagues.	N	13	2	15
		Percent	39.4%	28.6%	37.5%
	I've heard this and know a little bit about it.	N	8	3	11
		Percent	24.2%	42.9%	27.5%
	I've heard this before but don't know what it means.	N	9	1	10
		Percent	27.3%	14.3%	25.0%
	I've never heard this before.	N	3	1	4
		Percent	9.1%	14.3%	10.0%
Total	N	33	7	40	
	Percent	100.0%	100.0%	100.0%	

#### Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	1.453	3	.693
Likelihood Ratio	1.427	3	.699
Linear-by-Linear Association	.037	1	.847
N of Valid Cases	40		

Test Statistics											
	Items 86/97: ... they are obligatory to determine pupils' grades.	Items 87/98: ... they provide information about pupils' achievement.	Items 88/99: ... they provide information about pupils' strengths and weaknesses.	Items 89/100: ... they are essential in fostering learning processes.	Items 90/101: ... they provide information about the quality of teaching.	Items 91/102: ... they are required by the LBVO and the SchUG.	Items 92/103: ... they are practical (easy to design, easy to score, ...)	Items 93/104: ... they provide information for improving future instruction.	Items 94/105: ... they motivate pupils to learn English.	Items 95/106: ... they provide the opportunity to assess the pupils' performance in real world tasks and exercises.	Items 96/107: ... pupils benefit from the assessment process.
Z	-2,727 <sup>b</sup>	-1,221 <sup>c</sup>	-2,863 <sup>c</sup>	-2,996 <sup>c</sup>	-4,170 <sup>c</sup>	-2,909 <sup>b</sup>	-3,842 <sup>b</sup>	-3,253 <sup>c</sup>	-4,241 <sup>c</sup>	-4,007 <sup>c</sup>	-3,505 <sup>c</sup>
Asymp. Sig. (2- tailed)	0,006	0,222	0,004	0,003	0	0,004	0	0,001	0	0	0



## Appendix 3: Questionnaire

### Alternatives in assessment in Austrian EFL classes

Dear participants,

First of all, thank you for taking interest in this survey.

My name is Michael Schneller and I am a student at the University of Vienna, studying English and Geography/Economics. As part of my diploma thesis, I am undertaking an empirical study. With the help of this questionnaire, I wish to investigate the following questions:

1. What types of assessment are used by current EFL teachers?
  2. What are teachers' attitudes towards the use of alternative assessment in Austrian EFL classrooms?
  3. What factors contribute to teachers' decisions about assessment procedures?
- Hence the questions featured in this survey are targeted at those who are currently teaching English in Austrian classrooms.

I am now kindly asking you to help me complete my research project by filling in this online survey. This questionnaire has been designed to ask for your personal behavior and know-how as well as your personal opinion, attitudes and beliefs, so there are no right or wrong answers to the questions. It should not take any longer than 15 minutes to complete the survey. The results of this questionnaire will be solely used for research purposes and will at all times be treated confidentially and anonymously.

Thank you very much for your participation!  
Michael Schneller

*\*Required*

### Warm up – details about your teaching career

**1. What university/teacher training college (PH) did you graduate from? \***

---

**2. What subject(s) do you teach? \***

---

**3. How long have you been teaching? \*** *Mark only one oval.*

- ☐ up to 3 years
- ☐ 4-6 years
- ☐ 7-9 years
- ☐ 10-14 years
- ☐ 15-19 years
- ☐ 20+ years

**4. Which sort of school are you teaching at? \*** *Tick all that apply.*

- ☐ NMS
- ☐ AHS (lower secondary)
- ☐ Polytechnische Schule, Berufsschule
- ☐ BMS (Fachschule/Handelsschule)
- ☐ AHS (upper secondary)
- ☐ BHS (HTL, HAK, HLW)
- ☐ BAKIP
- ☐ Other:

## EFL assessment know how

**5. Please indicate the three most important purposes of language assessment in your English teaching in general. \***

- ☐ I use assessment to gather information about pupils' achievement.
- ☐ I use assessment to gather information about pupils' progress.
- ☐ I use assessment as a means of grading and certification.
- ☐ I use assessment to examine pupils' strengths and weaknesses.
- ☐ I use assessment to evaluate my teaching success.
- ☐ I use assessment as a means of quality assurance.
- ☐ I use assessment to improve future teaching processes.
- ☐ I use assessment to improve future learning processes.
- ☐ Other:

**The list below includes some terms related to language assessment. Some terms might be familiar and others might be new to you. For each term, select the response which best describes how well you know the term. \***

*Mark only one oval per row.*

	I've never heard this before.	I've heard this before but don't know what it means.	I've heard this and know a little bit about it.	I could explain the concept to my colleagues.
6. formative assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. summative assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. alternative assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
9. performance assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. washback/backwash/impact	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. validity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. reliability	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
13. authenticity	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
14. practicality	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
15. interactiveness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
16. educational standards	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
17. PISA	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
18. LBVO	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
19. SchUG	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Informelle Kompetenzmessung (IKM)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
21. E8 educational standards	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
22. Diagnosechecks	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**23. I'm well aware of the legal regulations as they are described in the SchUG and the LBVO concerning assessment in Austrian schools. \***

*Mark only one oval.*

Strongly agree	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	Strongly disagree
----------------	--------------------------	--------------------------	--------------------------	--------------------------	--------------------------	-------------------

## EFL assessment habits

The next question asks for your assessment habits. While answering the question, please think of one typical school grade and indicate which one you meant subsequently. If your assessment habits vary in different school grades, please feel free to provide answers in regard to more than one school grade. The questions are provided twice.

**How often do you use the listed types of assessment procedures in your EFL classroom? \*** *Mark only one oval per row.*

	Regularly	Seldom	Never	Not familiar with the concept
24. written exam (schriftliche Überprüfung)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
25. oral exam (mündliche Prüfung)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
26. oral exercises (Mündliche Übungen)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
27. class participation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
28. quizzes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
29. revisions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
30. homework	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
31. portfolio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
32. self-assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
33. peer-assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
34. journals (used for reflections on learning processes)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
35. rubrics (assessment grids)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
36. individual projects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
37. group projects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**38. Please indicate which grade the preceding answers refer to.** *Mark only one oval.*

- |                                    |                                     |
|------------------------------------|-------------------------------------|
| <input type="checkbox"/> 5th grade | <input type="checkbox"/> 10th grade |
| <input type="checkbox"/> 6th grade | <input type="checkbox"/> 11th grade |
| <input type="checkbox"/> 7th grade | <input type="checkbox"/> 12th grade |
| <input type="checkbox"/> 8th grade | <input type="checkbox"/> 13th grade |
| <input type="checkbox"/> 9th grade |                                     |

**How often do you use the listed types of assessment procedures in your EFL classroom?** *Mark only one oval per row.*

	Regularly	Seldom	Never	Not familiar with the concept
39. written exam (schriftliche Überprüfung)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
40. oral exam (mündliche Prüfung)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
41. oral exercises (Mündliche Übungen)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
42. class participation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
43. quizzes	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
44. revisions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
45. homework	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
46. portfolio	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
47. self-assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
48. peer-assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

49. journals (used for reflections on learning processes)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
50. rubrics (assessment grids)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
51. individual projects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
52. group projects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**53. Please indicate which grade the preceding answers refer to.** *Mark only one oval.*

- |                                    |                                     |
|------------------------------------|-------------------------------------|
| <input type="checkbox"/> 5th grade | <input type="checkbox"/> 10th grade |
| <input type="checkbox"/> 6th grade | <input type="checkbox"/> 11th grade |
| <input type="checkbox"/> 7th grade | <input type="checkbox"/> 12th grade |
| <input type="checkbox"/> 8th grade | <input type="checkbox"/> 13th grade |
| <input type="checkbox"/> 9th grade |                                     |

**54. Please state if you use any other assessment tools and the grade you administer them in.**

---

## Traditional and alternative assessment forms

The questions in the following sections deal with your attitude towards and your beliefs about both traditional and alternative assessment forms. Due to the possibility of different perceptions of both concepts, I want to provide the brief ensuing definitions in order to ensure everybody has the same understanding of the terms. Please read through both characterizations and keep them in mind while answering the following questions.

Traditional assessment, first and foremost, comprises all tools and processes which serve summative purposes, that is, those that measure and evaluate achievement. It is considered to be product-oriented and administered under tight time constraints. Other characteristics of traditional assessment forms are decontextualization and standardization of tasks.

Alternative assessment, on the other hand, comprises tools and processes which serve formative purposes by enhancing learning processes and discovering pupils' strengths and weaknesses. Thus, it is product- and process-oriented. Additionally, the focus is on authentic so-called real world tasks which require higher-order thinking skills. Open disclosure of assessment criteria is a requirement.

## Assessment and teacher education

**Please indicate how much you agree or disagree with each of the statements listed below.** \* *Mark only one oval per row.*

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Not applicable
55. The teacher education program at university/teacher training college prepared me for all kinds of language assessment forms.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
56. The courses at university/teacher training college focused on traditional assessment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
57. The courses at university/teacher training college focused on alternative assessment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
58. The amount of in-service training in assessment is appropriate.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

**I feel comfortable doing the following in regard to language assessment. \*** *Mark only one oval per row.*

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly Disagree	Not applicable
59. Administering traditional forms of language assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
60. Administering alternative forms of language assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
61. Utilizing different kinds of assessment techniques.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
62. Preparing examinations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
63. Scoring examinations.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
64. Using assessment to enhance pupils' learning processes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
65. Using assessment to alter my teaching instructions.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
66. Using assessment to provide feedback about pupils' achievement.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**67. Which forms of assessment would you have liked to have discussed in more detail during your teacher education program or in-service training? \***

---

## EFL assessment – beliefs and attitudes

**Please indicate how much you agree or disagree with each of the statements listed below. \*** *Mark only one oval per row.*

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Not applicable
68. Traditional assessment is superior to alternative assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
69. Assessment should mainly serve selective and allocative functions mainly.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
70. Assessment is primarily used to determine pupils' (final) grades.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
71. Assessment is a valid procedure to enhance instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
72. Assessment fosters pupils' learning processes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
73. Assessment helps to diagnose the strengths and weaknesses of instruction.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
74. Assessment should nurture intrinsic pupil motivation.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
75. Differences in learning styles should be accommodated in language assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
76. Self-assessment should be taken into account when assessing a pupil's language achievement.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
77. Peer-assessment should be taken into account when assessing a pupil's language achievement.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
78. It is important to use both formative and summative assessment in language classes.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
79. In addition to grades, feedback on pupil performance is essential in language assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
80. Contextualization and authenticity are important in language assessment.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
81. Assessment requires tasks involving higher level thinking and problem-solving skills.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
82. I would prefer assessment that can be rated quickly by machines rather than scored by humans.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

83. Continuous assessment is crucial in order to measure the pupils' progress.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
84. Traditional assessment tools can be used for formative purposes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
85. Assessment can fulfill all purposes (selective, allocative, diagnostic, quality assurance) simultaneously.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### In your opinion, traditional forms of assessment should be applied because ...

\* Mark only one oval per row.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Not applicable
86. they are obligatory to determine pupils' grades.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
87. they provide information about pupils' achievement.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
88. they provide information about pupils' strengths and weaknesses.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
89. they are essential in fostering learning processes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
90. they provide information about the quality of teaching.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
91. they are required by the LBVO and the SchUG.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
92. they are practical (easy to design, easy to score, ...)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
93. they provide information for improving future instruction.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
94. they motivate pupils to learn English.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
95. they provide the opportunity to assess the pupils' performance in real world tasks and exercises.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
96. pupils benefit from the assessment process.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### In your opinion, alternative forms of assessment should be applied because ...

\* Mark only one oval per row.

	Strongly agree	Agree	Neither agree nor disagree	Disagree	Strongly disagree	Not applicable
97. they are obligatory to determine pupils' grades.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
98. they provide information about pupils' achievement.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
99. they provide information about pupils' strengths and weaknesses.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
100. they are essential in fostering learning processes.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
101. they provide information about the quality of teaching.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
102. they are required by the LBVO and the SchUG.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
103. they are practical (easy to design, easy to score, ...)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
104. they provide information for improving future instruction.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
105. they motivate pupils to learn English.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
106. they provide the opportunity to assess the pupils' performance in real world tasks and exercises.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
107. pupils benefit from the assessment process.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

### 108. When you think of your future career as an English teacher, would you consider using alternatives in assessment in your EFL classroom? \*

Mark only one oval.

- ☐ Yes
- ☐ No
- ☐ I already do and will continue to do so.

## **Final comments on alternative assessment**

**109. Which assessment tools do you find most useful in your English language classes and why? \***

---

**110. Which decisive factors, if any, influence the choice of your applied assessment tools? \***

---

**111. Do you use typical traditional assessment procedures in a formative way? If so, please explain how.**

---

**112. Would you be willing to answer a few questions about your assessment practices face-to-face? Do you have experience with traditional assessment and/or alternative forms in assessment and want to share your thoughts and opinions? Please leave your email address and I'll contact you soon.**

## Appendix 4: Answers to the open ended questions

	<b>Answers to item no. 54 “Please state if you use any other assessment tools and the grade you administer them in.”</b>
1	I use observations in order to gain information about the achievements of my students in all classrooms.
2	No
3	Criterion-referenced assessment, all grades
4	presentations in 10th grade
5	none
6	presentations

	<b>Answers to item no. 66 “Assessment you would have liked to have discussed in more detail during teacher training or in-service training.”</b>
1	Alternative forms of assessment that can do it all, like ones that can be used for grading but can also be used for giving the students good feedback where they need to improve and how they can improve.
2	Types of assessment to alter my teaching
3	Portfolios
4	Alternative assessment in general
5	Criterion-referenced assessment
6	alternative assessment, oral assessment
7	More real life examples, hands on approach
8	all
9	ASSESSING INDIVIDUAL AND GROUP PROJECTS
10	alternative forms of assesement
11	Ussing assessment to alter my teaching
12	alternative forms
13	Assessment of oral performances
14	Various forms of alternative assessment
15	Alternative!
16	Giving no marks at all
17	All forms of assessment.
18	all kinds. there were not enough courses on assessment in general.
19	alternative ways of assessment other than schularbeiten
20	Basically all kinds of assessment. But I think I've never heard of alternative assessment during my teacher training. As far as I remember was the focus in one course on scoring and statistics...
21	Legal regulations. Preparing examinatin (written and oral)
22	both
23	in general more information on how to enhance pupil's learning processes and what might be options to adjust and consequentially render my personal teaching more beneficial
24	all of them
25	written exams
26	all of them
27	Assessment motivating student's learning without necessarilygrading it
28	various forms of formative assessment
29	alternative forms of language assessment
30	Formative Assessment procedures
31	using assessment to provide feedback
32	all of them



33	alternative
34	All of them.
35	All of them basically. When I think that assessment is such a big part of every teachers life, the training at university is everything but not enough. Alternative assessment is basically neglected at all.
36	all kinds of assessment

	<b>Answers to item no. 109 “Which assessment tools do you find most useful in your English language classes and why?”</b>
1	Most useful in order to decide upon pupils' grades are Schularbeiten and quizzes and check ups in order to have a variety of notes on participation. At the moment I don't have the time to use assessment, which aims at enhancing learning of the pupils.
2	Oral revisions, vocabulary check ups
3	alternative
4	Regular revisions - clearly shows learning curve of students
5	Criterion-referenced assessment, rating scales = alternative tools
6	Feedback tools
7	Lms
8	Assessing is my least favorite part of teaching
9	peer assessment because talking about strength and weaknesses is involved, students become aware of weakness and acquire a skill of analyzing their own work
10	peer feedback, because it means a lot to students and boosts their motivation.
11	a mixture of traditional and alternative forms but i hate testing anyway; language learning works without testing too, there a too many Tests in Austrian schools
12	activity, more rhymes rhythmicals
13	traditional assessment because it is easy to prepare
14	Revisions, oral performances, participation
15	Classroom activity of the pupils
16	Monitoring individual & pair work because the T can concentrate on the observation role
17	Portfolio
18	SAs because they provide a lot of information about pupils' achievement.
19	tests, revisions and presentations
20	Progress check, check-up, online tools/cyber homework (helbing languages)
21	tests, revisions, quizzes or check ups
22	Homework
23	Portfolio
24	face-to-face feedback
25	tests and quizzes
26	written exam in combination with oral practice
27	tests, quizzes, check-ups
28	Oral assessment peer assessment group assessment
29	CEFR and assessment grids in general because of the descriptors
30	all of them are useful
31	Vocabulary revisions because they are easy to prepare
32	quizzes, oral revisions
33	tests, quizzes, revisions
34	traditional
35	Being a novice teacher, I currently struggle with all kinds of assessment in order to do everything by-the-book. Plus, while I'd really like to try out different forms of assessment (incl. alternative forms), I don't have the time to prepare everything that is needed at the moment.
36	I think that every form has its use. So, it depends on a lot of things.

	<b>Answers to item no. 110 “Which decisive factors, if any, influence the choice of your applied assessment tools?”</b>
1	One factor is definitely time. Being a new teacher, I have my hands full with preparation and post-processing of my instruction. Therefore I try to use practical, time-saving tools I know will help to decide upon the grades of the students.
2	Students' level of proficiency
3	Topic, time, students' motivation
4	Age group, Level of responsibility /maturity displayed by the learner group
5	Process-oriented tools
6	the purpose of assessing: checking understanding or grading or feedback for students, etc.
7	Practical, doable, time issues
8	purpose (final grade, students achievements)
9	SCHOOL SYSTEM LBVO AND TIME
10	the learning target, students should always be encouraged to higher-level thinking,
11	creative answer opportunities
12	Students and classroom atmosphere, motivation of students
13	Time!
14	Havent thought about it
15	Age of the pupils, time constraints, practicality, the relevant content/competences
16	Time constraints and the maturity of the students
17	classroom level
18	My teacher training practice which only prepared me for certain kinds of assessment formats but also what I managed to teach myself and what I've learned from my colleagues and from my "Betreuungslehrer" during the UP. And of course, first and foremost, the law... what is obligatory by law so to speak.
19	Regular basis
20	I need to assess the students using Schularbeiten, which I don't find very valueable for the learning process of the students.
21	authenticity, practicality
22	my teacher training of course, the age of my pupils and how mature they are
23	real life situations
24	my teacher training, student preferences, time constraints
25	School requirements students needs
26	assessment criteria used in the new matura
27	legal requirements, time, timing , practicability
28	Practicality and usefulness
29	time, validity
30	teacher education, age of pupils, practicality issues
31	studied and applied
32	decisions made by the Fachgruppe; instructions given to the teachers by the headmaster/headmistress
33	Time!, legal regulations of course and to some extent the teacher training I enjoyed
34	eg. time, age of kids, content/grammar knowledge or skill, familiarity with assessment procedure.

	<b>Answers to item no. 111 “Do you use typical traditional assessment procedures in a formative way? If so, please explain how.”</b>
1	Especially with Schularbeiten, I try to provide feedback to students about the things they did well and the things they need to work on in particular, although it depends on the available time how detailed the feedback can be.
2	Schularbeit as part of all the information collected
3	NO
4	yes, for slow learners

5	Cambridge exam type questions
6	To be honest, I don't know?!
7	I don't know.
8	I honestly don't know very much about the difference between traditional and formative assessment but what I read in the definition, I try to find out about the pupils' skills and knowledge by doing tests, revisions.
9	Drafts for hw
10	unfortunately, not! But I will give it a thought.
11	I mark mistakes in homeworks and then let students find out origins, reasons and correction of the mistakes or let them rate their own perception of the mistake (sloppy, serious, funny,...) I also use traditional forms of assessment in peer assessment (students grade each others' work)
12	giving feed back to the students
13	No, I've never done that.
14	I'm really not that familiar with the difference between traditional and formative assessment, so, I can't.
15	To some extent I'd say, yes. E.g. I always try to give detailed feedback in addition to test grades. In terms of writing tasks, I try to use the rubrics from the BIFIE to give detailed feedback to the pupils. I definitely try to provide as detailed feedback as possible.
16	yes, I try to incorporate feedback into traditional assessment processes.