Teachers'
Assessment
Literacy
Enhancement

Needs analysis report

Karin Vogt (with help from Ines Sperling and Moritz Brüstle)

The TALE consortium wishes to thank all learners and teachers who participated in the questionnaire survey and helped us gain insights into and increased our understanding of issues related to language testing and assessment.

Project partners:

University of Cyprus, Department of English Language Studies / Department of Social and Political Sciences - Cyprus

University of Education Heidelberg, English Department – Germany

Hellenic Open University, Faculty of Humanities – Greece

University of Debrecen, Department of English Linguistics – Hungary

Oslo Metropolitan University, Department of Primary and Secondary Teacher Education –Norway

University of Bedfordshire, Centre for Research in English Language Learning and Assessment – UK

Contents

Executive Summary

Foreign language teachers use 30 to 50% of their time on assessment-related activities. Foreign language teaching and assessment, moreover, should be seen as one unit and not separate entities. To ensure professional assessment practices in the foreign language classroom that efficiently aim at enhancing learning needs expertise and professionality on the side of the teachers. The attitudes, knowledge and skills related to efficient assessment practices have recently been coined Language Assessment Literacy. Vogt & Tsagari (2014: 377) offer the following definition of LAL: 'the ability to design, develop and critically evaluate tests and other assessment procedures, as well as the ability to monitor, evaluate, grade and score assessments on the basis of theoretical knowledge'. There has been ample research activity in the field of LAL, e.g. theoretical conceptualisations, needs analyses for teachers etc. However, the needs of teachers and learners have not been established yet at the same time with a view to designing an accessible digital training material for teachers as an important group of stakeholders.

The purpose of the present study was to identify currently used assessment practices in English as a Foreign Language (EFL) classrooms across Europe from the perspective of two important groups of stakeholders in the foreign language assessment process, namely EFL teachers and learners. Secondary school learners in different European educational contexts were for the first time targeted with regard to assessment practices in order to establish what assessment practices help them develop their language proficiency in the foreign language. Moreover, EFL teachers' confidence levels with a range of assessment formats were identified as well as their perceived training needs in this field.

To this end, in a quantitative study, a questionnaire survey was administered 852 EFL teachers and 1788 learners in state schools in Cyprus, Germany, Greece and Hungary. The design of the questionnaire was aligned to the standards for teacher competence in educational assessment suggested by the American Federation of Teachers (1990) and Joint Committee on Standards for Education (2015), supplemented with EFL-related aspects such as linguistic skills or the link to the Common European Framework of Reference for Languages. Teacher and learner questionnaires were designed to contain the same questions where feasible in order to compare the perceptions of the different stakeholders. The results will be briefly summarized below.

With regard to **assessment practices**, the linguistic skills and subskills (vocabulary, grammar) are all evenly represented in teachers' reported assessment practices. Learners report that their teachers assess their writing, speaking, grammar and vocabulary most often. Feedback on assessment results tend to be given by way of marks and / or brief comments. Active class

participation, tests with closed answers and extended writing range among the most frequently used assessment methods EFL teachers in the sample use, with some variance across countries that can be attributed to the different assessment cultures in the respective educational contexts. Learners report that teachers use tests with closed answers, active class participation and extended writing most often. They also report that these very assessment methods help them learn English; however, it has to be noted that they can only judge the methods that they are used in their classrooms. Discrepancies between teacher and learner questionnaires regard assessment practices such as translation, active class participation and oral presentations, the two latter with a statistically significant difference.

Concerning **teacher confidence in EFL assessment**, it can be said that generally no teacher reported to be very confident in any of the given areas. Teachers said to be most confident in assessing reading and writing skills and explaining results to stakeholders such as parents and learners. Other more recent methods or aspects of assessment such as assessing learners with specific learning difficulties (SPLDs), using self-assessment, peer assessment or portfolio assessment and identifying the relevance of the CEFR in assessment showed teachers to be less confident in. The frequency of use of assessment methods seems to relate to teachers' reported confidence levels. A significant correlation could be shown between teachers who received assessment training in the past and their confidence levels in foreign language assessment overall. In accordance with reported confidence levels, EFL teachers in the sample voiced the **need for training** in the areas they feel least confident in, namely assessing students with SPLDs, using student portfolios (51%), using self-assessment (41%), identifying the relevance of the CEFR (40%) and using peer assessment (40%).

Regarding **online training resources**, the majority of teachers (74%) stated that they have no experience in online training. Nevertheless, they are very open towards online resources, particularly embedded videos, online peer discussions and trying out and evaluating materials.

The conclusion to be drawn from the survey results is that teachers and learners seem to overlap in their perceptions of assessment practices in the EFL classroom despite some discrepancies. Teachers report not to be very confident in any of the areas of assessment suggested to them in the questionnaire, and identified rather recent and innovative assessment methods as areas that they would need training in. These areas will be part of the training resource for foreign language teachers. Moreover, they are open towards digital training formats despite the fact that many of them have never used them before, and identified contents and formats that would be helpful for

them to explore in an digital training resource. The TALE Online Course has therefore considered all these elements.

1. Introduction

English language teachers (ELTs) in Europe and beyond face growing challenges in the area of Language Testing and Assessment (LTA). The growth in use of accountability systems and the influence of external frameworks in educational policy-making, such as the CEFR (Council of Europe 2001), have increased both the amount of LTA required of teachers and the importance placed on it. ELTs are now expected to design, adapt and score a wide variety of language tests that are relevant to their own particular teaching context, carry out innovative assessment procedures, provide useful feedback to learners based on results of such assessments and align their LTA procedures with language curricula or educational policies in ways that meet national or European language assessment standards.

Studies such as Tsagari (2009), Cheng, Andrews & Yu (2011), Vogt & Tsagari (2014) or Tsagari & Vogt (2017) have shown that LTA can have a strong impact on the quality of the learning outcomes. Yet, ELTs cannot deliver professional results if they are not sufficiently trained in the area of LTA. Unfortunately, in many educational systems across Europe, ELTs experience a lack of what has been termed "language assessment literacy" (Hasselgreen, Carlsen & Helness 2004, Tsagari & Csépes 2011, Inbar-Lourie 2013, Hill 2017). Given the current state of affairs, there is an urgent need to develop an efficient, relevant, accessible and sustainable LTA training infrastructure for ELTs which can eventually be beneficial to other language teachers as well. The contents and types of delivery of this assessment infrastructure addressed primarily at pre-service and in-service ELTs are not determined by academic syllabi as is the case currently at many universities. Instead, their real training needs are to be taken into consideration and put into practice. Teachers are one vital group of stakeholders in foreign language assessment processes, but not the only ones. Learners are often considered as subjects in the assessment process whose performance is assessed without taking into account that the results of assessment should be used to feed back into teaching and, more importantly, into learning. More often than not, learners' views and perceptions of and towards assessment are not taken into consideration either (Tsagari 2013). One of the aims of the needs

analysis, therefore, is to give learners a voice and enquire about their views on assessment and its role in learning foreign languages, as language assessment literacy (LAL) in learners.

The project consortium has undertaken to identify LAL training needs through a comprehensive literature review, a European survey of ELTs and EFL learners in the different educational contexts of the partner countries. In a first step, the different educational contexts of the partner countries will be detailed in order to clarify the research context. The design of the needs analysis study will be explained and data collection and data analysis methods as well as the procedure will be outlined. Subsequently the results will be reported for the individual countries involved, using a variety of descriptive and inductive statistics procedures. Finally, the results will be discussed in the light of implications for the project, before a conclusion is drawn.

2. Language assessment literacy

2.1 Language Assessment Literacy (LAL): definition and conceptualization

The traditional perception of literacy as including solely 'reading and writing skills and practices' has been widened with the addition of new types of literacies such as 'media' literacy, 'computer' literacy, 'science' literacy, 'academic' literacy and other types of literacies in recent academia (Taylor, 2013; Harding and Kremmel, 2016). Therefore the addition of 'assessment' literacy to the growing body of literacies 'to be acquired in contemporary life, together with language assessment literacy as a potentially subordinate or overlapping category' was fairly anticipated (Taylor, 2013: p.405).

The American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) in 1990, produced one of the first attempts to define the standards of what areas teachers should be competent at, in terms of assessment literacy. This came to be known as the 'Standards for Teacher Competence in Educational Assessment of Students'. These included selecting assessments, developing assessments for the classroom, administering and scoring tests, using scores to assist instructional decisions, communicating results to stakeholders, and finally being aware of inappropriate and unethical uses of tests. The Joint Committee on Standards for Education Evaluation has recently published updated standards for classroom assessment in Pre-K12 contexts (JCSEE, 2015). The Standards for Teacher

Competence in Educational Assessment of Students were used as a theoretical background for the questionnaire employed for the needs analysis (cf. section 4).

The term 'assessment literacy' was first coined by Stiggins (1991) who describes the term as 'the ability to understand, analyze and apply information on student performance to improve instruction' (in Falsgraf, 2005, p.6). Other definitions of LAL explicitly focus on teachers as one of the the most important stakeholder groups. Lam (2014, p. 4) sees LAL as "teachers' understanding and mastery of assessment concepts, measurement knowledge, test construction skills, principles about test impact, and assessment procedures which can influence significant educational decisions (evaluation of student learning) within a wider sociocultural context". Likewise, Tsagari and Vogt (2014: 377) offer the following definition that preliminarily targets teachers but potentially encompasses other stakeholders: 'the ability to design, develop and critically evaluate tests and other assessment procedures, as well as the ability to monitor, evaluate, grade and score assessments on the basis of theoretical knowledge'.

For Taylor (2013), assessment literacy is the skill-based 'know-how' but 'with a multilayered set of competences such as the awareness of and ability to draw on a deep knowledge-base of assessment for critical reflection on one's own assessment practice and the practices of others' (Harding and Kremmel, 2016), thereby explicitly considering various groups that are involved in the assessment process. Pill & Harding (2013, p.382), for example, define LAL as the ability "to understand, evaluate and create language tests and analyze test data". Fulcher (2012) offers a rather detailed definition of LAL:

> 'The knowledge, skills and abilities required to design, develop, maintain or evaluate, large-scale standardized and/or classroom based tests, familiarity with test processes, and awareness of principles and concepts that guide and underpin practice, including ethics and codes of practice. The ability to place knowledge, skills, processes, principles and concepts within wider historical, social, political and philosophical frameworks in order to understand why practices have arisen as they have, and to evaluate the role and impact of testing on society, institutions, and individuals'. (p. 125).

O'Loughlin (2013, p. 363) also takes a broader view of LAL with a definition of LAL to entail 'the acquisition of a range of skills related to test production, test score interpretation and use and test evaluation in conjunction with the development of a critical understanding about the roles and functions of assessment within society' (2013: p. 363). Kremmel and Harding (2016) identify several stakeholders involved in assessment procedures that require satisfactory levels of LAL.

The commonalities in the various definitions of LAL are monitoring educational progress, design, develop, maintain, or evaluate of large-scale tests for stakeholders, teachers, administrators, and more recently the general public.

Within the field, various attempts have been made to enlist the key concepts of what requirements LAL should entail for teachers, stakeholders and as mentioned above, for the general public. Brindley (2001) for example, considers five components that are entailed in the nature of LAL; however, he only targets teachers as an important group of stakeholders. Two components are necessary while the other three are optional. He contends that teachers should be conversant in 1) the social context of assessment, 2) the definition and description of proficiency, 3) constructing and evaluating language tests, 4) the role of assessment in the language curriculum and 5) putting assessment into practice.

The diversity of levels of expertise among stakeholders inevitably leads to the emergence of different levels of literacy. The range of needs that will emerge across the levels will reveal the 'types of knowledge most useful for stakeholders' (Pill and Harding 2013, p.383). Due to the inextricable affiliation of teachers and assessment, the consensus is that all five components should be mandatory for teachers (Brindley, 2001; Inbar-Lourie, 2008; Harding and Kremmel, 2016).

Davies (2008) considers the five components proposed by Brindley (2001) and proposed a three-core system that includes skills, knowledge and principles to be the areas of competencies in LAL. In line with Davies (2008), Inbar-Lourie (2008, p.390) deems three areas that need to be addressed in LAL; the 'what' that reflects the 'knowledge', the 'why' which entails the 'principles', and the 'how' that is refers to the 'skills'. A more synthesized view of what LAL should involve was proposed by Taylor (2009) who stresses that 'training for assessment literacy entails an appropriate balance of technical know-how, practical skills, theoretical knowledge, and understanding of principles, but all firmly placed within a sound understanding of the role and function of assessment within education and society' (p 27).

The aforementioned approaches sought to reveal the key components of LAL. There have been, however, various attempts to describe the aspects of LAL through a grading pattern (Bybee, 1997; Kaiser & Willander, 2005; Pill and Harding, 2013). For example Pill and Harding (2013, p.383) produced the following grading system from illiteracy to multidimensional language assessment literacy:

A continuum model of language assessment literacy (Pill & Harding, 2013: 383)

Illiteracy:	Ignorance of language assessment concepts and methods
Nominal literacy:	Understanding that a specific term relates to assessment, but may indicate a misconception
Functional literacy:	Sound understanding of basic terms and concepts
Procedural and conceptual literacy:	Understanding central concepts of the field, and using knowledge in practice
Multidimensional literacy:	Knowledge extending beyond ordinary concepts including philosophical, historical and social dimensions of assessment

Table 1: Pill & Harding, A continuum model of language assessment literacy

These grading patterns have not remained unchallenged. For example, the grading pattern proposed by Pill and Harding, (2013) has been criticized due the fact that the level of language assessment skills different stakeholders should have, was unclear. Also, the scale itself depends mainly on theoretical assumptions rather than practice (Harding and Kremmel, 2016).

In a first attempt to provide a conceptual framework of LAL, Lam (2014) has proposed a twofold model with the establishment of a theoretical component of LAL and a rationale for developing LAL with (pre-service) teachers, his primary target group. Lam (2014, p. 4) sees a clear need to develop LAL in order to "untangle the two seemingly incompatible cultures, namely testing and learning cultures".

2.2 Research on LAL

The growing interest in assessment literacy seems to be due partly to the central role that assessment plays in student learning (Black & Wiliam, 1998), and to the strong evidence that teachers are key agents in educational assessment (Leung, 2014). The demand and therefore the necessity particularly for teachers' assessment expertise has also been triggered by the attention devoted to students' performance on standardized achievement tests and the introduction of formative approach to classroom-based assessment in many educational contexts (see Vogt and Tsagari, 2014). Evidence has been gathered concerning their efficacy in assessment, the knowledge and skills that teachers need to be considered assessment literate, their training requirements, as

well as contextualized understanding of assessment literacy (e.g., DeLuca & Klinger, 2010; Gottheiner & Siegel, 2012; Plake, Impara & Fager, 1993).

2.2.1 Teachers' assessment practices

The extent to which teachers are prepared to face these challenges has been treated with skepticism (Pellegrino et al, 2016). Even though theoretical arguments have repeatedly accentuated the need to prepare teachers and stakeholders for their assessment responsibilities, only a minority of teachers seem to be prepared 'to face the challenges of classroom assessment because they have not been given the opportunity to learn to do so' (Stiggins, 2002, p.762).

In his discussion of the professionalization of language testing in the 21st century, Bachman (2000, p. 19-20) points out that 'the majority of practitioners who develop and use language tests, both in language classrooms and as part of applied linguistics research, still do so with little or no professional training'. This view is also maintained by Alderson (2005, p. 4), who stresses that teachers' assessment competencies are limited: 'Tests made by teachers are often of poor quality, and the insight they could offer into achievement, progress, strengths and weaknesses is usually very limited indeed'. Coniam (2009, p. 227) presented findings from previous studies, which showed that tests made by teachers tended to be of poor quality, 'were too difficult or too easy; … measured content that had not been taught in class or not specified in the syllabus… [and] did not show what students had actually achieved'. In empirical research undertaken, Tsagari (2016) found that her Greek and Cypriot primary school EFL teachers employed mainly a summative orientation towards evaluating their students' performance. In terms of content, teacher-made tests include a limited range of language skills, usually restricted to the assessment of vocabulary and grammar and lacked creativity with regard to the task types. Furthermore, the criteria teachers use to select testing materials or provide feedback on test results is inconsistent. Teachers also seemed to have unclear ideas about the purposes and implementation of formative assessment, mainly due to lack of professional training in language assessment.

Coniam's study (2009) explored the effects of a basic training course in assessment (provided as part of a graduate course in English Language Teaching) on the quality of teacher-produced tests. Although teachers were able to improve their material by editing, reviewing and redrafting, even after passing through the recommended stages of test construction, the resulting tests generally failed to satisfy basic quality criteria. In addition to insufficient training, Coniam's (2009) informants

reported that practical constraints had a negative impact on their work: lack of time, resources and institutional support, sharing the work with colleagues etc., restricted the attention they could give to developing effective assessments and to improving their own knowledge and skills (also in Tsagari, 2012).

2.2.2 Teachers' conceptions of assessment

To inform the gap regarding the sufficiency of teachers' knowledge and skills in assessment matters, many researchers have set out to examine teachers' conceptions of assessment. For example, Hidri (2015) investigated secondary and university teachers' assessment conceptions in Tunisia using a four-factor teachers' conceptions of assessment (TCoA) inventory (Brown, 2006). Results denoted teachers' misconceptions about assessment but stressed that teachers still conceive of assessment in a positive way (see also Brown & Michaelides, 2011; Brown, 2011; Kitiashivili, 2014; Gebril & Brown, 2013). In his study, Gebril (2016) also investigated the assessment conceptions among English teachers in Egypt based on a questionnaire. The results of the study showed a tendency for different assessment practices among in-service teachers (preferring a combination of formative and summative assessment practices) and pre-service teachers (favored a relatively summative paradigm). The study also showed that both groups had negative perceptions about the effectiveness of teacher training programs for language assessment development and highlighted that teacher educators should focus on understanding and also changing the attitudes towards assessment in teacher training to insure better assessment practices in schools.

Other researchers have taken a more qualitative approach to teachers' understanding of language testing concepts. Sahinkarakas and Buyukkarci (2011) describe a case study carried out with two experienced ELT teachers and how these teachers' personal theories relating to assessment practices, particularly formative assessment practices, have changed over a 15-week period using the Repertory Grid Technique, a derivative of Kelly's Personal Construct Theory (Kelly, 1995). The results indicate that eliciting the personal constructs of teachers about assessment practices helps them become aware of their own and others' personal views on assessment, which, in turn, might result in effective formative assessment practices.

Another line of research into teacher assessment literacy has set off to describe the appropriate content and standards of knowledge and skills needed by teachers in order to be considered assessment literate through the use of tests of AL.

2.2.3 Tests of assessment literacy

Attempts have been made to create instruments that would be used to measure teachers' assessment literacy following the initiation of the 'Standards for Teacher Competence in Educational Assessment of Students' (DeLuca, LaPointe-McEwan, & Luhanga, 2015).

One of the first attempts was an instrument initiated by Plake and Impara (1992) who produced the "Teacher Assessment Literacy Questionnaire (TALQ)" which consisted of 35 items. The TALQ was administered to a sample of 555 in-service teachers around the US. The results revealed that even though teachers spent up to 50 percent of their time on assessment-related activities, teachers' lack of preparation in issues related to assessment of students' performance was evidenced by the average score of 23 out of 35 items that were marked as correct (Plake, Impara, & Fager, 1993).

Targeting the same area, Campbell, Murphy, and Holt (2002) used an altered version of the questionnaire, the Assessment Literacy Inventory (ALI), to measure the assessment literacy level of 220 undergraduate pre-service teachers who had completed a course in tests and measurement. The results showed an average score of 21 out of 35 items correct, thus indicating a fairly low level of proficiency in assessment literacy. Using the same instrument as Plake et al. (1993), O' Sullivan and Johnson (1993) designed a course, with the aim of integrating the standards into a measurement course with graduate-level students. The teachers were evaluated with a pretest (M= 24.2) and a post-test (M= 27.3). The results of the study revealed a minor improvement of the teachers in assessment literacy.

Another survey that was conducted by Mertler (2003) also used the TALQ as the basis for the development of an instrument that would 'measure and compare pre-and in-service teachers' assessment literacy' (p.2). The modified version of TALQ was renamed as Classroom Assessment Literacy Inventory (*CALI*) and was used in Mertler's (2003) survey, which was one of the first attempts to compare results from pre- and in-service teachers. The average score that was obtained was 22 out of 35 items that were marked as correct. The results were similar to Plake et al's (1993) survey with an average of 23 correct items. One of the most significant finding of the study was that in-service teachers outperformed pre-service teachers on nearly every subscale of the study. The data suggested the urgency for the re-evaluation of the assessment training courses that are integrated to the educational program of pre-service teachers. Mertler (2003) also stresses the

importance of the elevation of the standards of assessment literacy among the community of all educational stakeholders.

Other researchers have taken up Plake and Impara's (1993) approach and applied it to language testing. Newfields (2006), for example, constructed a test of language teacher assessment literacy designed for self-diagnosis which covered four broad topic areas: *Terminology*, *Procedures*, *Test Interpretation* and *Assessment Ethics*. However, the majority of the questions focus on the use and interpretation of statistics. Kaftandijeva (2008) further refined Newfields' test to create a 27-item True/False test for online administration. Unfortunately, the reduction in length comes at the cost of an even more restricted coverage of topics. Although testing the assessment literacy of teachers appears a promising theme, no results are available showing how well teachers perform on either the Newfields (2006) or Kaftandijeva (2008) measures.

To sum up, although the strong and weak areas found in these studies varied with different samples, the consensus was that teacher assessment knowledge was generally inadequate relative to standards and expectations.

In addition to the use of tests of assessment literacy as a means of collecting more direct evidence of the extent to which language teachers are aware of key testing concepts, another, less developed strand of research was proposed, that of surveys of language testing courses.

2.2.4 Surveys of language testing courses – testing instructors

Recurrent complaints have been expressed by language testing specialists as to the quality of teacher training programmes in language assessment. There is a perception that even programmes aimed at more advanced students and experienced teachers fail to provide sufficient input on assessment. Taylor (2009: 23), for example, complained that graduate programmes for language teachers 'typically devote little time or attention to assessment theory and practice, perhaps just a short (often optional) module'. Research efforts have been made both to ascertain what is generally covered in teacher training courses and to evaluate how well teachers are able to apply what they have learned.

With the questionnaire aiming to gather information about the instructor of the course, the course itself and the students' attitudes, Bailey and Brown conducted two surveys, one in 1996 and the other in 2008 (Bailey and Brown 1996, Brown and Bailey 2008). The results of both surveys combined showed similarities in responses to items in both the new and the old version with an

acceptance that throughout their data, more advanced language testing courses are been provided in a number of graduate programs around the world. This showed indications of a stable knowledge base that is evolving and expanding, rather than shifting radically. Courses consistently featured such core topic areas as validity theory, reliability, measurement error, statistics for test and item analysis, the critique and analysis of test content, and item writing skills. Bailey and Brown point out the fact that 'further research might focus on the international distribution of language testing courses and how such courses differ in different types of institutions, from country to country' (2008, p. 373).

Another study that used the questionnaires designed by Bailey and Brown, was conducted by Jin (2010). The aim of the study was to explore the teaching content, the material and the teaching methodology of language testing and assessment (LTA) courses conducted in the tertiary-level education programs for foreign teachers. The major areas of comparison of the study was drawn between LTA courses provided in graduate and undergraduate levels and the second comparison was between LTA courses 'provided in normal universities, that is, universities that train future teachers, and non-normal universities' (Jin 2010, p. 557). The results found similar evidence of the enduring appeal for tutors of established course constituents. Test validity and reliability, principles and practice of item writing, and the construction of (multiple-choice) test questions again featured among the most commonly taught topics. Showed that despite expectations when judged in terms of the educational background and professional experience of the instructors 'the majority of the instructors had a master's or doctoral degree and the title of 'associate professor' or 'professor'. When the strengths of the subgroups were compared, it is worth noting that the normal universities, trainers of future teachers, was relatively weaker than the non-normal universities, for example in tertiary institutions that covered the essentials theoretical and practical aspects of language testing were as normal universities did not adequately provide the essentials (p. 566). Jin emphasised the need to improve the quality of LTA courses, and provide foreign language teachers with 'the core competences' (Inbar-Lourie, 2008, p. 396) in assessment.

Bailey and Brown surveys were a starting point to Jeong's 2013 survey, who used a mix method approach to investigate the instructors' backgrounds, e.g. the topics they covered, and their students' apparent attitudes toward those courses (Jeong, 2013, p. 347). The difference between Bailey and Brown (1996, 2008) and Jeong (2013) was that the latter separated the instructors into two groups: 66 language testers (LTs) - individuals or professionals that focus in the areas of language testing a total of participants and 74 non-language testers (non-LTs) – with an interest in other areas of language teaching (e.g. second language acquisition) though they have the knowledge in language

assessment matters (e.g. developed standardized test or worked with a testing agency). The interview results confirmed that non-LTs are less confident in teaching technical assessment skills and have a tendency to focus more on classroom assessment issues. However, the final result that Jeong's survey provided was that 'there was little difference between the two groups' (p. 350). Jeong stresses the importance of possessing a common understanding of assessment literacy among stakeholders within the testing community, but also among non-LTs who teach language assessment courses to maintain course quality and to better meet student teachers' needs' (pp. 356-357). Jeong's overall conclusion is to protect their field in assessment, and to share the knowledge to those who are part of the language assessment culture. 'It is important for LTs to preserve their specialty, but also it is essential to share the knowledge and make it accessible to those who are part of the language assessment culture. It is the role of the LT community to make the field approachable to others'. (p. 357)

Lam (2014) scrutinized the teacher education programmes of Hong Kong universities involved in teacher education and conducted focus group interviews with 40 students and 9 instructors. He found that university-based language assessment training does not adequately support pre-service teachers' LAL. Instructors and students both reported that the social dimension of LAL, e.g. ethics and fairness, were not included in most assessment courses. Training was reported to be too academic, leading to a wide theory-practice gap that was perceived as unsatisfactory.

When language teachers have received in-service training related to LTA, they are likely to develop LAL and the learning outcomes of their learners can be enhanced, as the study by Amirian, Pourfarhad and Nafchi (2016) shows. In their particular case, they administrated an in-service training course for a group of teachers instructing IELTS candidates and measured the effect of their literacy on the writing skill developments of their learners. Teachers who participated in the training course were more aware of the genre principles of writing. Their learners achieved higher scores than their fellow learners because their teachers were using a more effective approach to teaching their learners the demands of the test.

2.2.5 Surveys with stakeholders, e.g. teachers and others

In addition to the above lines of research, another, less developed strand of research has proposed the use of surveys of assessment literacy as a means of collecting more direct evidence of the extent to which language teachers and other stakeholders are aware of key testing concepts.

In 2004, Hasselgreen et al. launched a survey to identify the training and educational needs of three types of stakeholders: language teachers, language teacher trainers and experts from 37 European countries and 50 representatives from non-European countries. The survey revealed that all stakeholder groups were in need of formal education and training in assessment matters due to lack of training. Teacher themselves felt the need for further training. The authors conclude with the observation that these needs are not being catered for by European teacher development programs.

Fulcher (2012) conducted an international online survey. The study was aimed at language teachers' assessment training needs. The date that was collected from June to September 2009 from a total of 278 that responded 85% held a higher degree (e.g. MAs or PhDs). The results revealed that from specific feedback and the different language testing textbooks, language teachers are aware of assessment needs that are not currently supplied in existing materials designed to improve assessment literacy. Fulcher 2012 concluded that language teachers want a comprehensible and realistic textbook that will have appropriate activities for the language teachers that are also testers. The outcome of Fulcher research was used to plan and produce a textbook (Fulcher, 2010) and to develop a website for all users of assessment issues (http://languagetesting.info).

The survey by Pill and Harding (2013) regarding assessment issues was conducted in Australia and looked into the English proficiency for overseas-trained doctors as part of their registration processes. The aim was to investigate the types of misconceptions about language testing in the communication, especially policy makers and other parties related in matters of language proficiency assessment. The survey was based on feedback that was interpreted through level descriptors and the test specifications. The results from Pill and Harding demonstrated 'the lack of understanding of both language and testing issues and the lack of familiarity with the tools used and with their intention, can lead to meaningful misconceptions.'(p.304).

O'Loughlin's (2013) study investigated the assessment literacy needs of test score users in relation to the IELTS test in Australian higher education. That included the purpose and content of test, the meaning of the test scores, the appropriateness of cut-off levels, the test's validity, reliability, predictive power and comparability with other accepted forms of evidence of English proficiency. The survey included 84 participants from two universities, 43 from university A and 41 from university B responded and completed the survey. Reduction from 84 was done by the question whether they used the IELTS to which a total of 50 responded from university A 23 and 27 from university B. The main focus however was the combination of data collection that directly used the test in their work, this lead to the final result of 15 whom were 2 admissions officers, 6 marketing

officers, 2 academic, 3 language and 2 other staff members. The second part of the survey was interviews with three research questions. Results showed indications that identified other usage for the test, which were communicating with other university staff and overseas agents, and making scholarship decisions. O'Loughlin (2013) concluded that 'participants mostly needed information about IELTS for advising prospective students about English language entry requirements and making admissions decisions' (p. 378). Language testers and teachers who need to acquire a much broader knowledge base 'grounded in theory and epistemological beliefs, and connected to other bodies of knowledge in education, linguistics and applied linguistics' (Inbar-Lourie, 2008, p. 396).

Another study conducted by Kvasova & Kavytska (2014) aimed at identifying strengths and weaknesses of Ukrainian university foreign language teachers in assessment as well as their training needs. The first part of the survey was a replication of the European Survey of Language Testing and Assessment Needs (Hasselgreen et al, 2004) The second part of the survey, which was the authors' own design, revealed that FL teachers were quite competent in assessment-related tasks such as using ready-made tests, administering tests appropriately and providing feedback to students. The authors concluded with the assumption that 'it is to be expected that overall assessment literacy has not yet reached an appropriately high level' (p. 175).

A large scale survey conducted by Vogt & Tsagari (2014) aimed at measuring EFL teachers' assessment literacy as well as identifying their training needs in this field. The study replicated the questionnaire study by Hasselgreen et al. (2004) but has taken a step forward by conducting follow-up interviews with the participants and thus collecting both quantitative and qualitative data. The participants who responded to the questionnaire were FL teachers from seven European countries that had completed their pre-service teacher training. Two of these European countries (Greece and Germany) were selected for the qualitative data collection by way of interviews. The results from both the interviews and the questionnaires showed that teachers had received little or no training in the field of language testing and assessment (LTA). The study provides insights into the training needs of FL teachers from different regional backgrounds and makes suggestions for the development of an in-service teacher training 'format' that involves cooperation with teachers that could cater for their needs (Vogt & Tsagari 2014, p. 392; Tsagari & Vogt, 2017)

The following section describes the different educational contexts as a background to the needs analysis study laid out in this paper. The section also contains a description of the overall study context.

2.3 Language assessment in the educational contexts of the partner countries

Assessment is one of the cornerstones in education. Traditionally, assessment procedures are used to generate results for the individual students which in turn help take decisions that may determine their further educational and professional careers. Although this general description holds true for most educational sectors such as primary and secondary as well as tertiary education, there are substantial differences in assessment practices throughout the participating countries. Various factors determine individual approaches to assessment, and the function of assessment in the respective country. While there are certainly individual varieties in assessment practices among teachers based on professional experience and personal preferences, and requirements arising from the specific characteristics of the respective subject, there is an inevitable influence from various external factors which have an immense impact on educational assessment. Some of these factors result immediately from the context of language learning such as curricula at the individual educational level and appropriate assessment methods, while other factors can be found at more distanced levels such as educational systems and governmental policies.

2.3.1 Curriculum and aims of foreign language learning

The first aspect which requires consideration are the various curricula. In Cyprus, Greece and Hungary national curricula are in place. In Germany, each of the sixteen federal states is granted educational sovereignty with an individual curriculum based on National Educational Standards (Bildungsstandards) defined by the Conference of Ministers of Education (KMK). This results in a rather heterogeneous picture for foreign language learning (FLL) in German primary education: English is either taught from grade one or three, or plays overall a marginal role. Moreover, there are no National Educational Standards for FLL in place.

The most important aspect which needs to be considered concerning curricula in the context of assessment is the aim of FLL per educational level and per country: FLL at primary level in Cyprus for instance has a clear focus on the development and integration of the four language skills speaking, listening, reading and writing with an emphasis on the development of intercultural awareness and positive attitudes towards foreign languages and cultures in general. This provides the basis for secondary education which requires students to acquire competencies for communication and interaction in authentic communication-statements while gaining knowledge at

various components of language such as grammar, vocabulary, or syntax. In Hungary, FLL at secondary level qualifies for various paths in the students' professional careers, and is closely oriented at the language levels defined in the CEFR. In Germany, the development of oral competencies is the main goal in FLL for primary level in reference to the general requirements of modern pedagogy of FLL. Equivalent to Cyprus, FLL at secondary level in Germany has now turned from a rather linguistic emphasis towards a communicative approach based on the CEFR aiming at communicative and intercultural competencies (KMK 2003, p. 6). In all countries, FLL grades play a key role for school- leaving certificates which influence placement at universities and vocational training.

The curricular specifics and aims of FLL are reflected in assessment policies which vary among the participating countries. In Greece, the new "Integrated Foreign Languages Curriculum" (IFCL) integrates the updated levels of proficiency described in the CEFR.

2.3.2 Assessment policy and stakeholders

Educational assessment has become a key aspect in education due to a growing importance of evaluation practices at various levels. This is reflected in general assessment policies for each country which, however, have their emphasis on very different aspects. This reflects national traditions, aims and attitudes towards assessment.

Cyprus makes a clear distinction between the aims of assessment in primary and secondary education. Assessment in primary education reflects FLL goals: it aims at the achievement of the objectives for development of the language, and their integration, at the development of learning strategies and the cultivation of intercultural awareness (Figure 1). In addition to that, assessment is expected to focus on behaviour and social skills of the children as part of their development into democratic citizens.
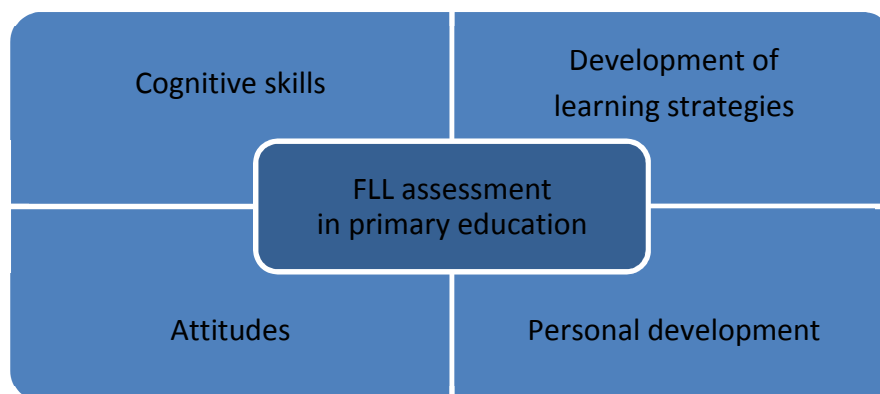
Figure 1: *Overview of assessment areas in Cyprus primary education*

Assessment guidelines require teachers to protect and maintain the positive atmosphere that is created in the classroom, and pupils' positive attitudes towards language learning and learning in general. The assessment approach is proposed to include a combination of structured and alternative ways of assessment activities. Among alternative assessment methods, portfolio plays a key role emphasizing a democratic evaluation process through the focus on the individual student's performance rather than the peers' performance. Secondary level assessment requires both summative and formative assessment procedures. This way, teaching activities in terms of planning and improvement can be guided through assessment. Formative assessment is to be carried out in a systematic way, while written and oral tests reflect summative assessment procedures. In terms of inclusive learning, provisions are to be made for learners with special abilities. Reporting to stakeholders plays an important role in the assessment guidelines for Cyprus.

Educational assessment in Greece is based on two cornerstones: *monitoring* and *evaluating*. This underlines the combination of both formative and summative aspects in assessment by defining educational assessment as (1) *monitoring the learning process* and (2) *measuring the degree to which the educational aims have been achieved*. In reference to the *Integrated Foreign Languages Curriculum* (IFLC), FLL assessment is explicitly viewed as a dynamic, holistic and collaborative process. The Greek Assessment guidelines define explicitly the individual stakeholders such as the learner, peers, parents and teachers which contribute to the educational process and provide feedback (Figure 2). The emphasis here is on systematic peer-assessment.
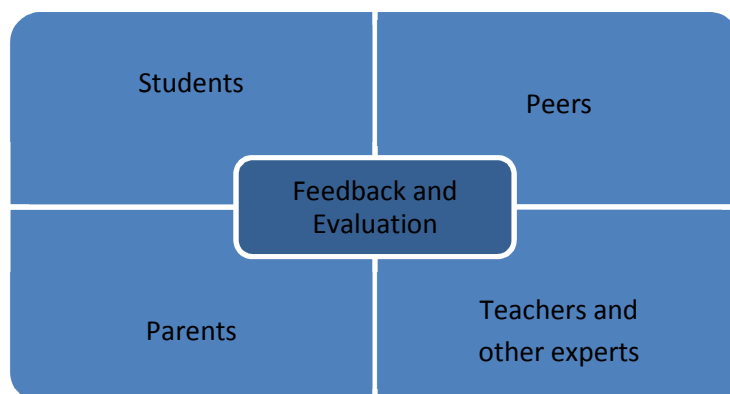
Figure 2: Overview of stakeholders providing feedback in educational assessment in Greece

Similar to Cyprus, the focus in assessment and feedback is on quantitative rather than on an exclusively qualitative description of the outcomes of assessment. In primary education, alternative assessment is the preferred approach to educational assessment while progress tests are encouraged to be taken from grade four or five. Suggested alternative assessment methods include a wide range of internationally acknowledged and approved methods such as self-assessment, systematic observation and learning portfolios. Beyond classroom-based assessment, testing plays a major role from the beginning of secondary education for grading and placement purposes (Figure 3).
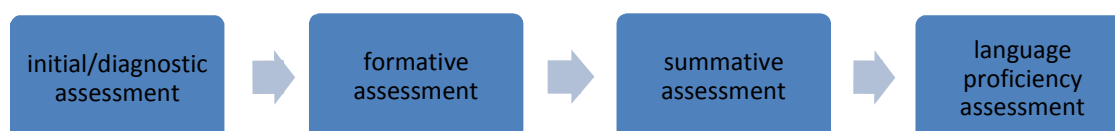
Figure 3: Overview of suggested types of assessment in Greece

Although Germany has a federal system of government, various assessment practices have been introduced on national and international levels in addition to classroom-based assessment. Resulting from the mediocre outcomes of the first PISA test, a monitoring system has been put in place in 2003/2004. In turn, assessment is now carried out at different levels.

In line with the FLL curriculum, assessment in primary education has a strong focus on formative assessment, e.g. by way of observation sheets or portfolio assessment, while standardised tests or FL certificates play no role. The change towards an outcomes-oriented approach in education elicited new forms of assessment, and a tendency towards more large-scale and standardized assessment. Despite these enhancements, there is a persistent dominance of classroom-based

assessment ranging from traditional vocabulary tests to design of high-stakes tests such as school leaving examinations. Classroom-based assessment is also used for the generation of final marks by means of teacher-designed written tests and other pen and paper formats. In reference to the CEFR, the dominant focus on writing skills in FLL has now been complemented by the integration of listening skills in written tests and speaking as a part of classroom instruction in the foreign language. In contrast to countries like Cyprus or Greece, however, further specifications of assessment methods and techniques, aims of assessment, or stakeholders in the assessment process for classroom-based assessment have not been published on a national level. With the introduction of a monitoring strategy in education, a range of standardized assessment procedures has been implemented. Continued participation in international standardized tests (PISA, TIMMS, etc.) has become a cornerstone in monitoring and feeds into pedagogical and didactic research projects. Outcome monitoring via standardized tests is carried out by the IQB (Institut für Qualitätsentwicklung), an institution on national level. Furthermore, central final school-leaving examinations have been introduced in various federal states.

As stated before, assessment procedures on different levels have been implemented which, in turn, have specific functions in educational monitoring. While international large-scale tests like DESI for English (www.dipf.de/desi) evaluate the outcomes of the German educational system in its entirety, standardised tests such as VERA (compared tests) provide feedback on (1) school-level and (2) for improvements of teaching processes. Tests like VERA are taken one or two years before the actual end of the respective end of educational level (year three and year seven) to serve as the foundation of the evaluation and adaptation of learning and teaching for the remaining time at school.
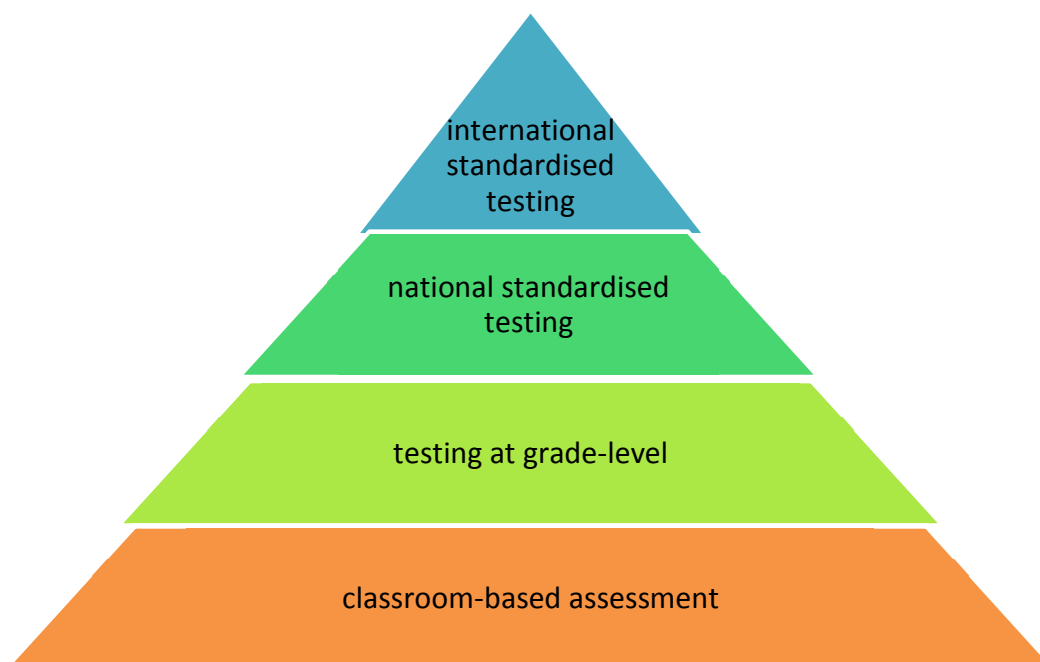
Figure 4: Educational assessment levels in Germany

The introduction of inclusive education has brought forth many changes and adaptations in teaching practice. Although teachers experience many practical challenges arising from heterogeneous learner groups and the concept of inclusion, assessment in an inclusive setting currently poses immense challenges (Liebers & Seifert, 2012) in Germany.

Hungary has a strong examination-dominant culture. Test requirements and achievement levels are based on the CEFR. Test results are the foundation for admission in higher education and tertiary education. In the public education sector, target attainments in foreign languages are also clearly defined and measured in the school-leaving exams covering all the four language skills. However, language certification is a flourishing business, which strongly affects the classroom assessment practices of the teachers. Exam preparation and exam coaching are top priority of most teachers.

2.3.3 Testing

The growing importance of educational assessment has brought forth a large variety of test formats and test design levels. Throughout the countries, FLL assessment guidelines reflect the philosophy of modern pedagogy of foreign languages by highlighting the significance of assessment being based on

the students' capacity for communication and interaction in authentic communication-statements. This means, assessment must not just focus on the individual and isolated ability in the various components of language such as grammar, vocabulary, syntax etc. Moreover, tasks are required to be appropriate and relevant to the students' age group and experience (Greece).

The individual test types are being designed by respective experts depending on the test level. German FLL teachers have a wide range of responsibilities regarding test design involving classroom-based assessment, tests written by all pupils of their grade, final grades and school-leaving examination.In contrast to tests designed at school level, test items for standardised tests at national (VERA, DESI) and international level (PISA, TIMMS) comply with academic standards. These tests are designed collaboratively by test experts, trained item writers and teaching experts and the tests are piloted. Test results, however, do not count into students' grades which is a main difference of standardized testing in Germany in comparison to the other countries. Due to the tradition of rather independent assessment procedures with responsibility on the individual teacher, teachers often perceive national and international testing as additional burden adding to their workload. Another important aspect of standardised testing in German education is that national and international tests, with the exception of VERA, are taken by a sample cohort of students, not by all students of the respective group. Standardised tests in Germany serve a specific function: compared tests such as VERA provide feedback on school level and are used for the evaluation and improvement of teaching processes. Results from international standardised test such as PISA or TIMMS evaluate the success of the national educational system in it is entirety.

Unlike Cyprus, in Hungary, a state accredited language certificate (CEFR B2 or C1 level) is in place which provides bonus points for admission into higher education. In tertiary education, the B2 level state accredited language certificate is part of the study requirements, but from 2020 onwards, it will be a prerequisite for entrance to higher education.

The Greek State Certificate of Language Proficiency (KGP) Examinations are a further example of standardised tests designed and taken on a national level (Hellenic Republic, Ministry of National Education and Religious Affairs).

2.3.4 Pre-Service Teacher Training and Career-long Professional Learning

With the growing importance and relevance of educational assessment the question arises, into how far aspects of educational assessment and its multifaceted approaches are being considered in pre-service teacher education and in in-service teacher training. Again, there are differences to be found throughout the countries.

Hungary has acknowledged the need of teachers' assessment skills by the introduction of new assessment trends in teacher education in 2006(15/2006 [IV.3] Government Decree). The requirements of designing assessment tools, fostering self-assessment skills and the interpretation and use of test results are among the competences which need to be acquired by teacher trainees. Knowledge about the CEFR and academic standards for test design are equally considered in the teacher training curriculum. A further reform in teacher education in 2013 requires teacher trainees to acquire additional competencies in areas such as assessment to be conceived and carried out in motivating ways, the use of various forms of assessment, and the consideration of individual needs of the learners (8/2013 [I.30] Government Decree).

Teacher education in Germany is the responsibility of the 16 sovereign federal states. The acquisition of assessment skills during teacher education, thus, depends on the curricula and guidelines of the individual federal states.

In many contexts, formative assessment has gained immense significance and teachers are required to implement formative assessment techniques in their teaching practices. A study on formative assessment with UK teachers carried out by Black et al. (Black et al., 2004), however, showed that positive attitudes towards formative assessment and theoretical knowledge are insufficient for the successful execution of formative assessment procedures. In their study, Black et al. (ibid.) provided expert support to teachers to help them change their teaching practice. Furthermore, this support allowed a change in student behavior which was the foundation to successful implementation of formative assessment. As a main outcome, Black et al. learned that in most cases teachers were not fully knowledgeable about the concept of formative assessment. Teachers' false perceptions of formative assessment led them to inappropriate application of formative assessment techniques. This, in turn, signifies the importance of high-quality pre-service teacher training which provides future teachers with the necessary concepts and aims of assessment and allows them to acquire competence and skills in this area. In-service teachers gain a high level of knowledge and expertise during their professional career throughout all areas of teaching. The

changes which assessment and the philosophy behind it have undergone, however, require in-service teacher to be trained at areas of assessment policy, methods and techniques and aims of assessment. Hence, the need of career-long professional learning (CLPL) as highlighted by Livingston and Hutchinson (2017) gains strong significance and needs to be implemented in all areas of in-service teacher training.

2.3.5 Preliminary conclusion

Assessment as part of teaching is practiced in all areas of education. It is, however, influenced by a wide range of factors which vary among the countries. The following figure presents contextual factors in educational assessment.
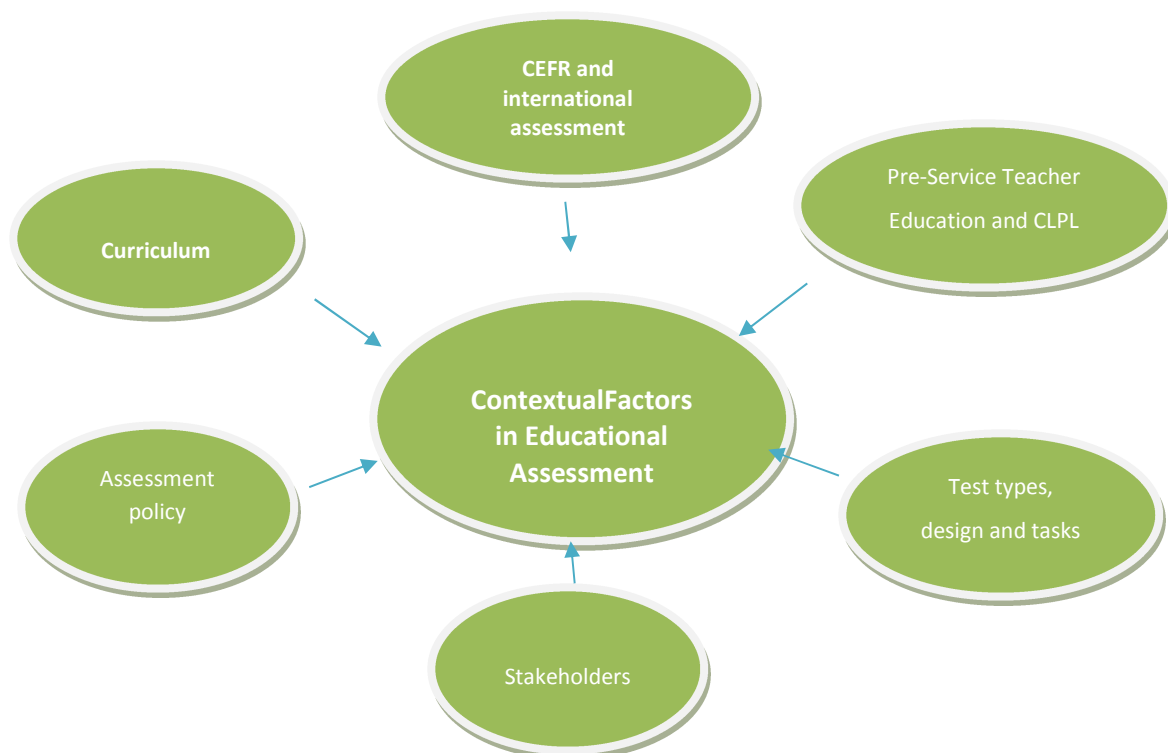
Figure 5: Overview of contextual factors with an impact on educational assessment

As the previous paragraphs showed, none of these factors functions independently in assessment. On the contrary, all factors are interconnected. The CEFR for instance feeds into the curriculum, and, in turn, into assessment requirements. The curriculum, on the other hand, is based on national educational guidelines which also determine stakeholders in assessment.

A main result of the explications in this paper is that the presented influencing factors impact assessment practices to a different extent in the individual countries. The evaluation of assessment practices, thus, needs to consider all factors and their extent to which they feed into educational assessment to be able to provide knowledge and explanation for regional and national differences. The present overview of contextual factors serves as foundation for analysis, comparison, and evaluation of current assessment practices and may yield new insights and best practices for the benefit of each participating country. One focus area in the domain of educational assessment should be teacher education and career-long professional learning which provide opportunities for enhancements in successful educational assessment.

The research questions the study is based on can be formulated as follows:

1. *What assessment practices in the EFL classroom do teachers and learners report?*
2. *Which assessment practices do they consider help learners to learn English?*
3. *What level of confidence concerning LAL do teachers have?*
4. *In what areas do they wish for training?*

3. LAL Training Needs Analysis

3.1. Study design and research context

The study is quantitative in nature in order to reach as many teachers and learners as possible and in order to be in a position to make generalizations of certain statements. The purpose of the study is a differentiated picture of assessment practices of teachers and learners across educational sectors and contexts. The study can be characterized as a questionnaire survey. For Bailey and Nunan (2009, p. 125), the purpose of questionnaires is seen as "a snapshot of conditions, attitudes, and / or events". It is precisely that snapshot that we intended to obtain from the two different groups of informants. Despite the single data collection method, the research design incorporates triangulation, namely the triangulation of perspectives (Elsner & Viebrock, 2014) of teachers and learners. The aim of the triangulation is to consider the different perspectives of stakeholders engaged in and affected by LAL, in this case teachers who form a typical target group and important informants of assessment practices and training needs in LAL and learners as a quantitatively important but often neglected stakeholder group.

The educational contexts of the study were Cyprus, Germany, Greece and Hungary, members of the project consortium. Questionnaires were targeted at pre-service and in-service teachers in the primary and secondary sectors as well as learners at secondary schools. It has been deemed necessary that the informants approached spend as little of their time as possible, since many of the teachers were also approached for the piloting of the online course. Informants took part in the survey on a voluntary basis and so care was taken that the teachers who took part were not overtaxed in terms of the amount of time they devoted to the project. Regarding the target group of learners, primary school learners were excluded because they could not be expected to answer questions on the metalevel required in the survey questionnaires. The sample can be described as a convenience sample (Cohen, Manion, and Morrison, 2011), with the researchers activating their respective networks of schools and teachers.

The study consists of two questionnaire surveys, one for EFL teachers and one for EFL learners. The questionnaires were aligned as much as possible as to make the statements of teachers and learners comparable. The teachers were EFL teachers mainly in primary schools and secondary state schools, some were also teaching in the adult education sector but not at university so as to obtain a sample that is as homogenous as possible. Teachers were teaching English as a foreign language in Cyprus, Germany, Greece and Hungary. 852 teachers and 1788 learners took part in the study. The learners attended primary and secondary schools, some of them came from the adult education sector but not the tertiary sector. They also were located in Cyprus, Germany, Greece and Hungary. The sampling represents a non-probability sampling, more precisely a convenience sampling with members of the consortium approaching members of their networks, the university networks of teaching practice schools etc. This can be seen in a critical light because the teachers who were approached were likely to be motivated and inclined to use innovative methods and / or open for in-service teacher education in general. In addition, in some subsamples, e.g. the German sample, pre-service teachers were over-represented, which might impact on the data because the target group of pre-service teacher only has limited teaching experience. The table below shows the distribution of the informants.

Countries	Teachers' questionnaire	Students' questionnaire
Cyprus	404	909
Hungary	230	300
Germany	127	285
Greece	91	294
Total	*852*	*1788*

Table 2: Distribution of the informants

3.2 Data collection and analysis

Two different questionnaires were distributed to 852 English as a Foreign Language teachers and 1788 EFL learners. They were asked about assessment practices in the EFL classroom, assessment-related feedback mechanisms as well as training needs (for teachers) and assessment practices that enhance their learning (for learners). The items of the questionnaire were aligned in order to enable a comparison of the results from the different perspectives.

In terms of questionnaire design, the standards for teacher competence in educational assessment suggested by the American Federation of Teachers, et al. (1990) and Joint Committee on Standards for Education (2015) were used as a theoretical basis and underlying construct. The AFT standards were then translated into questionnaire items in the teacher questionnaire in the section on teachers' confidence levels. Other aspects such as the link to the CEFR and more specific skills-related competences were added. Table 3 shows an overview of the standards translated into questionnaire items for the teacher questionnaire.

AFT Standards	Questionnaire item
Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.	1,2,4,5,6,7,8,9,10,11,12
Teachers should be skilled in developing assessment methods appropriate for instructional decisions.	3,4,5,6,7,8,9,10,11,12
Teachers should be skilled in administering, scoring and interpreting the results of both externally-produced and teacher-produced assessment methods.	1,2,4,5,6,7,8,9,10,11,12
Teachers should be skilled in using assessment results when making decisions about individual students/learners, planning teaching, developing curriculum, and school improvement.	14,15
Teachers should be skilled in developing valid pupil grading procedures	14

which use pupil assessment.	
Teachers should be skilled in communicating assessment results to students/learners, parents, other laymen and other educators.	18,19
Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information.	20

Table 3: Can-do statements translated from AFT Standards (1990)

The teacher questionnaire comprises four parts. The first part includes questions about general biographical information, such as age, gender, qualification, years of teaching experience, language(s) taught, age range of their learners etc. In this section teachers were also asked if they have ever received training in testing and assessment. The second part of the questionnaire involves assessment practices of teachers in the EFL classroom. Teachers are asked which skills areas they usually assess, and they are asked about feedback mechanisms that they employ in the EFL classroom. Informants are also invited to give information on the various types or methods of assessment they use in their classrooms and how often these types or methods are used. The second part of the questionnaire is parallel to one of the parts in the learner questionnaire. The third part of the teacher questionnaire focuses on teachers' assessment profiles and training needs. The first question is about their confidence levels with competences laid down in the AFT standards. These competences are presented as 'can do' descriptors with a four-point Likert scale ranging from "very confident" to "not confident". Other aspects such as the relevance of the CEFR for assessment and descriptors specifying e.g. alternatives in assessment or assessing specific skills were added.

While the teacher questionnaire was administered in English, the learner questionnaire was translated into the language of schooling in the respective countries. The questionnaires were piloted with a small number of teachers and learners, resulting in small changes in the questionnaires, mostly concerning layout and minor formulations.

Regarding data analysis, teacher and learner questionnaires were analysed using descriptive and inductive statistics. Descriptive statistics included percentages and means besides crosstabulations. Inductive statistics were also used, mainly concerned ANOVA tests for correlations between e.g. confidence levels of teachers in language testing and assessment and training received in this area. Other tests, e.g. Mann-Whitney U Tests to compare the means for results on assessment types for teachers and learners were carried out as well.

4 Results

To identify currently used assessment practices in the EFL classroom and a potential format of an online learning course on assessment, two separate questionnaire surveys were conducted, one for EFL teachers and another one for EFL learners. As previously mentioned, the questionnaire for the EFL teachers consists of four parts (I. General information, II. Assessment practices, III. Assessment profiles and training needs, IV Use of technology). The EFL learners' questionnaire has been designed to be conducted parallel to part II of the teachers' questionnaire and thus make the triangulation of the results possible. It consists of only three parts (I. General information, II. Assessment practices, III. Assessment needs and wants).

In this section, the structure of the EFL teachers' questionnaire will be used as an outline to first illuminate the overall results, followed by the results of each sub-sample.

4.1 Overall results – all countries

4.1.1 Teacher questionnaires

I.   General information

In terms of gender distribution, the sample represents the current state of the teacher profession. Therefore, the predominant part of the sample being female EFL teachers (87%) was to be expected (cf. figure 6).
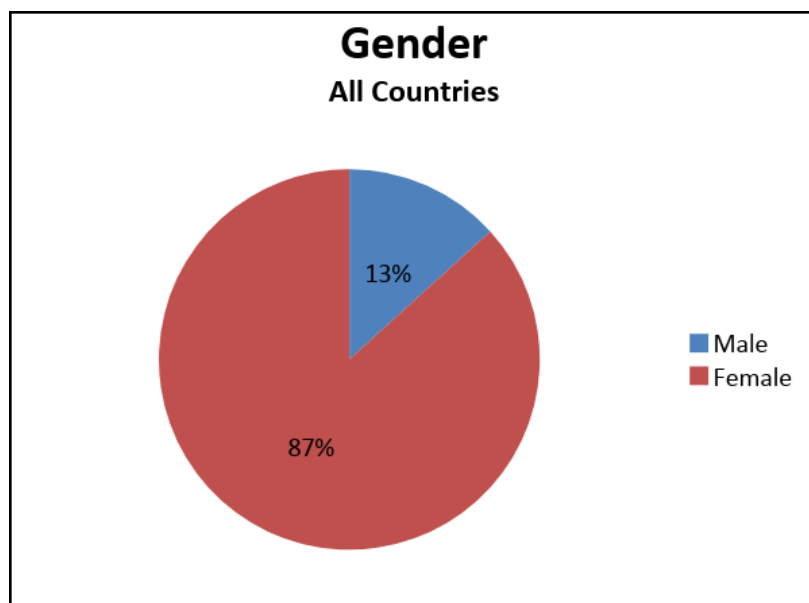
Figure 6: Gender distribution overall data

The age of the EFL teachers was much more equally distributed, as figure 7 shows, just as their teaching experience ranges from pre-service teachers to 15+ years. 44% of the teachers in the sample even have a substantial amount of teaching experience (figure 8).
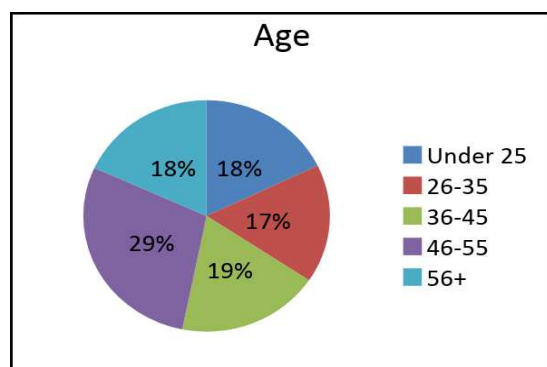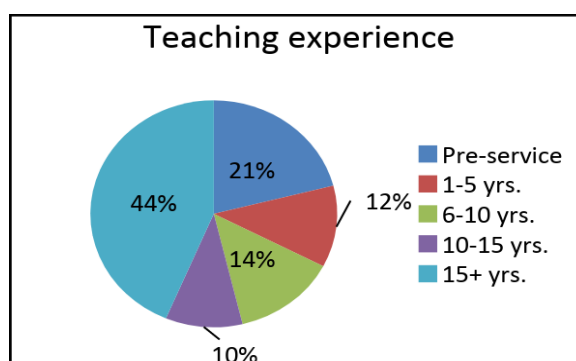

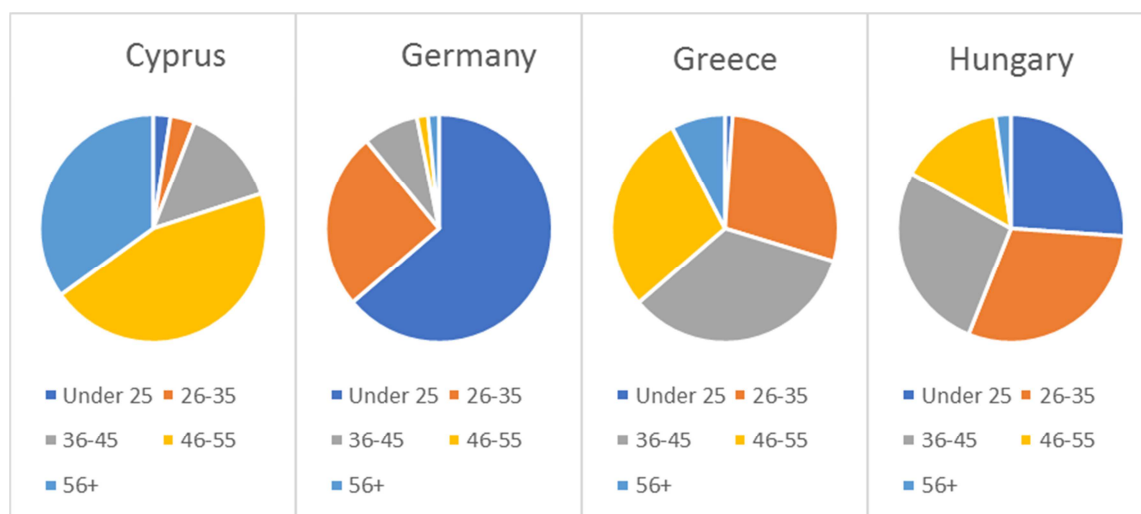*Figure 7: Age distribution*

Figure 8: Teaching experience

Figure 9: Teachers' age sub-samples

The county the informants were from was another factor that needs to be highlighted. Almost half of the EFL teachers (47 %) and more than half of the learners (51 %) were teaching in Cyprus (cf. figure 10) must be considered when looking at any further results. The remaining halves of the sample were quite evenly distributed between Greece, Hungary, and Germany. The same is true for the learners who participated in the study.
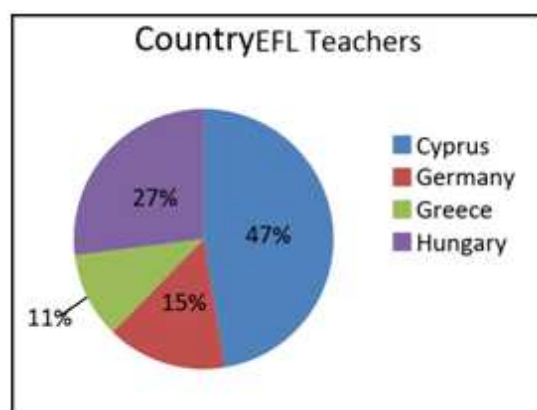
Figure 10: Country teachers
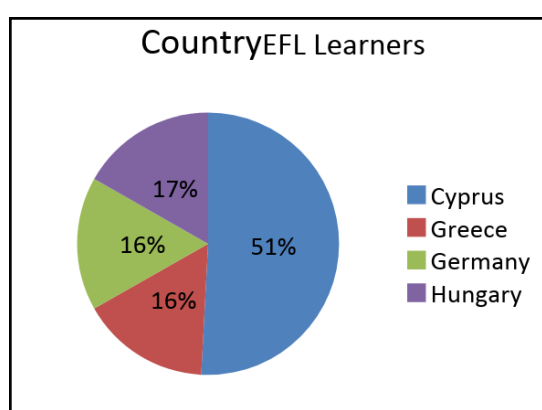
Figure 11: Country learners

This study was aimed exclusively at teachers and learners of English as a foreign language. Therefore, it is not surprising that 98% of the teachers stated English as the main language they teach in school.

Teachers were finally asked if they had previously received any testing and assessment training which 63 % of them answered with yes (cf. figure 12). This result becomes particularly interesting when analysing the currently used assessment practices and the teachers' confidence when using them in class. The following parts will indicate these relations and elaborate fair assumptions that can be made.
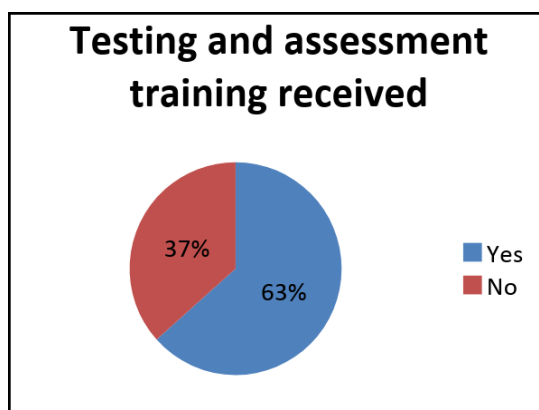

Figure 12: Testing and assessment training received

II.   Assessment practices

The second part of the teachers' questionnaire comprised three items to precisely determine which assessment methods are currently used in the EFL classroom. These results need to be considered in the context of the sample composition which were described in the previous part. The first item of the second part in the teachers' questionnaire was aimed at the skills which are assessed in the EFL classroom. The teachers were asked to state what skills they usually assess and was able to add other skills.

Which of these skills/areas do you assess?		
	N	*%*
Writing	795	93,31
Vocabulary	783	91,90
Grammar	768	90,14
Reading	755	88,62
Speaking	738	86,62
Listening	706	82,86
Other	59	6,92

Table 4: Skills assessed in the EFL classroom

Table 4 shows that the six given skills are very evenly distributed. More conventional skills, such as writing and reading are assessed more often, although just slightly. This reflects the state of foreign language assessment, as other results will confirm, because pen and paper tests that test written skills seem to be the most dominant assessment tool of teachers in the EFL classroom. Reasons for that could be previous training and ready-made material available for assessment and their accessibility.

In terms of feedback given by teachers on their learners' assessment results, marks and brief comments are by far the most common ways of feedback (cf. table 5).

10. What feedback do you give on your learners' assessment results?		
	N	%
Mark (e.g. letter, percentage)	741	86,97
Brief comments	691	81,10
Detailed comments	507	59,51
Comments / hints on how to improve their learning	591	69,37
Other	40	4,69

Table 5: Feedback given on learners' assessment results

Detailed comments and hints on how to improve their learning, elements that are vital in effective feedback procedures, seem to be less often used by the EFL teachers in the sample. As mentioned above though, there are several possible reasons for these results and one cannot be sure if the previously received testing and assessment training is connected. Also, feedback in the EFL classroom often is restricted to corrective feedback (Sheen & Ellis, 2011) and has not been identified as an instrument for formative assessment by teachers yet.

The final item to collect current assessment practices was designed to determine how often certain assessment methods are used. On a scale of 0 (never) to 3 (very frequently), teachers answered the question "How often does your teacher ask you to [assessment method]?" A compilation of all answers given by the EFL teachers can be seen in figure 13 below.
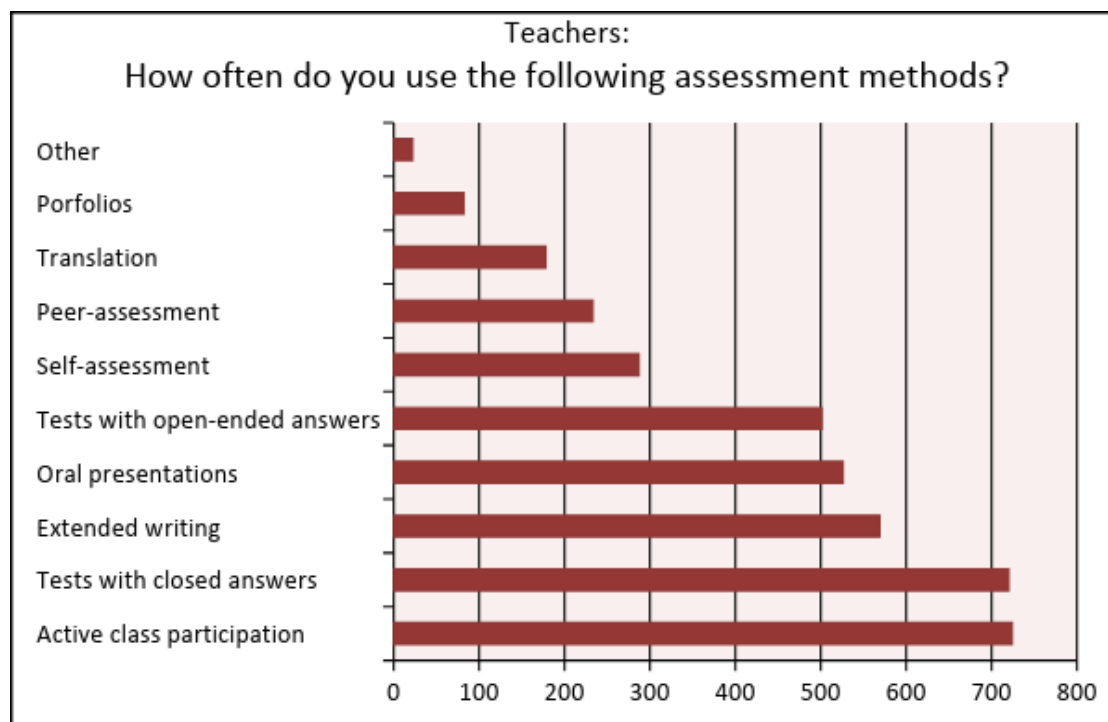
Figure 13: Frequency of assessment methods teachers

Teachers in our sample seemed to rely most on active class participation and tests with closed answers. Extended writing, oral presentations and tests with open-ended answers also played a role. More innovative assessment methods such as portfolios or peer-assessment are used less often according to the informants. Comparing the teachers' and learners' answers reveals new insights, e.g. when comparing the teachers' answers to the results of the learners' questionnaire (Qu. 6) "How often does your teacher ask you to […]?". The learners' questionnaire also included an item with 10 questions, each with a four-point Likert scale (3 = very frequently, 2= frequently, 1 = sometimes, 0 = never) so this comparison is possible. The compiled results of both stakeholders can be seen in table 6.

Questions	*Cyprus*	*Germany*	*Greece*	*Hungary*	*Total*
Frequency of assessment types (Qu. 11)	**Teachers**				
01. Oral presentations	1.85*	1.48	1.93	1.72	1.76
02. Tests with open-ended answers	1.85	0.96	1.79	1.57	1.63
03. Portfolio assessment	0.62	0.56	0.70	0.43	0.56
04. Peer assessment	1.09	0.95	1.19	1.04	1.06
05. Tests with closed answers (e.g. gaps, multiple choice, matching exercises)	2.24	1.67	2.32	2.29	2.18

06. Self-assessment	1.23	0.93	1.53	1.29	1.23
07. Extended writing, e.g. letters, essays	1.96	1.18	1.87	1.74	1.77
08. Active class participation	2.45	2.12	2.43	2.16	2.32
09. Translation (L1/L2)	0.64	0.70	1.48	1.12	0.85
Frequency of assessment types (Qu. 6)	**Learners**				
01. Oral presentations	0.87*	1.21	1.52	0.95	1.05
02. Tests with open-ended answers	1.67	1.43	1.73	1.64	1.64
03. Portfolio assessment	0.47	0.83	1.40	0.23	0.64
04. Peer assessment	1.01	1.26	1.16	1.01	1.07
05. Tests with closed answers (e.g. gaps, multiple choice, matching exercises)	2.23	2.17	2.38	1.82	2.17
06. Self-assessment	1.26	1.12	1.68	0.86	1.24
07. Extended writing, e.g. letters, essays	1.66	1.90	1.68	1.84	1.73
08. Active class participation	2.20	1.55	2.31	1.48	1.99
09. Translation (L1/L2)	1.61	1.58	1.72	1.96	1.68

*Based on a Likert scale from 0 - 3

Table 6: Teacher and student responses to frequency of assessment methods by country

It can be observed that the most frequently used methods according to the teachers are active class participation (mean rating 2.32), tests with closed-ended answers (2.18) and extended writing (1.77). The same teachers reported that portfolio assessment (0.56), translation (0.85) and peer assessment (1.06) are used least frequently in the EFL classroom. The learners, on the other hand, reported that tests with closed answers (2.17), active class participation (1.99), and extended writing (1.73) are used most frequently while portfolio assessment (0.64), oral presentations (1.05), and peer assessment (1.07) are the least frequently used methods. It can be seen that the data overlaps to some degree, but there are some methods that seem to be perceived differently by learners and teachers.

One particularly striking item regarding such discrepancies is *translation* that has values of 1.68 in the learner questionnaire and 0.85 in the teacher questionnaire. This means that the teachers report to not use this type of assessment very frequently while the learners state it is used rather often. The reason for this discrepancy could lie in different perceptions that the two groups of informants have about assessment practices in the classroom. However, the present data only allow speculations as to possible reasons. Other discrepancies that can be seen in table 5 concern *active class participation* and *oral presentations*. Interestingly, both items showed an asymptotic significance in a Mann-Whitney U Test with a significance level of .05 (cf. appendix 2).

III. Assessment profiles and training needs

The third part of the teachers' questionnaire included two items that were designed to determine the teachers' confidence levels when using different assessment methods (Qu. 12: "How confident do you feel about the following areas?") and if they would like to receive any training in each respective area. This was an item including 20 questions (i.e. areas of assessment in the EFL classroom), each with a four-point Likert scale (0 = not confident, 1 = somewhat confident, 2 = confident 3 = very confident). In the same table, teachers were prompted to tick an additional box ("I'd like training in this") for each of the 20 areas. Figure 14 displays all answers given by the teachers from all participating countries.
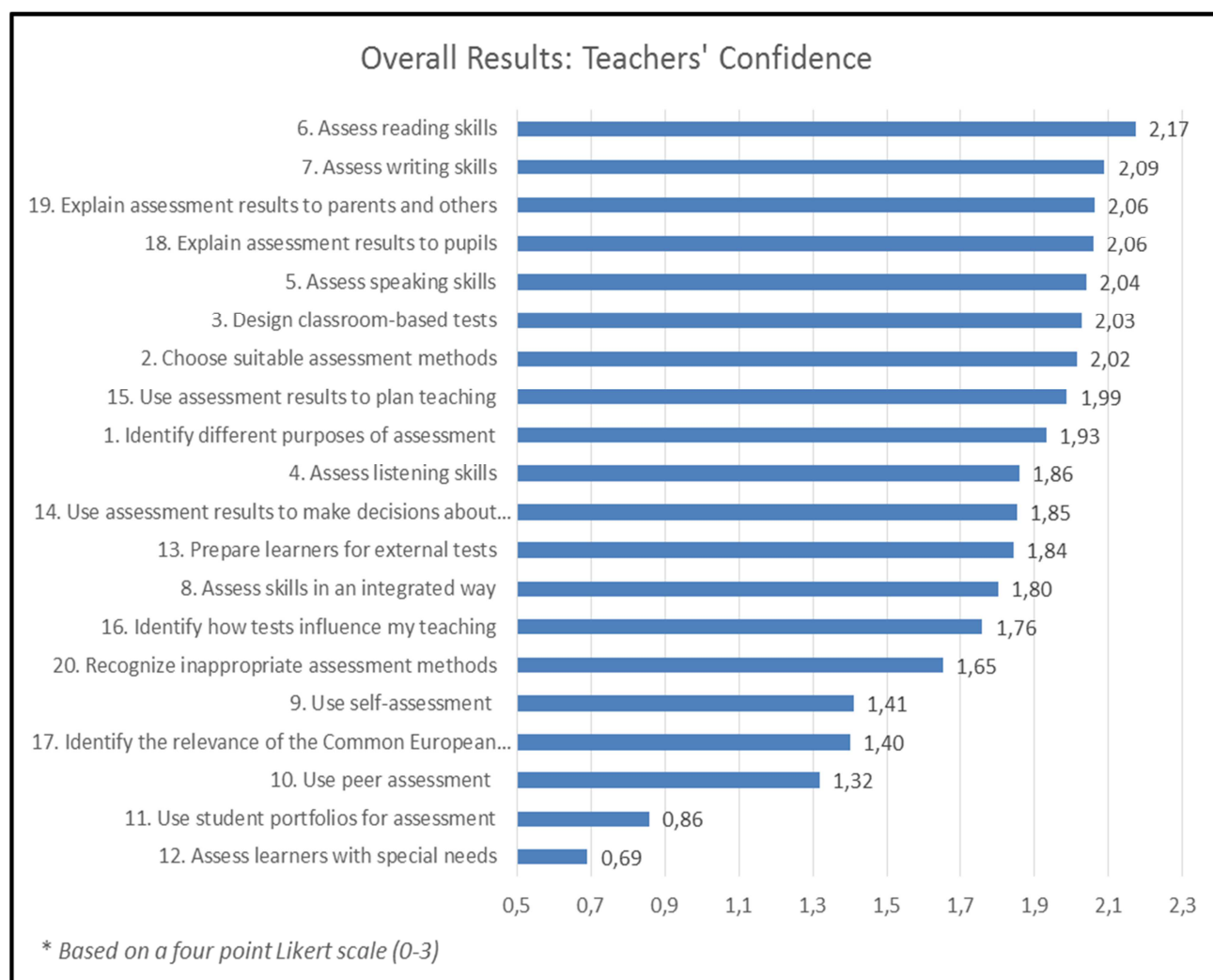
Figure 14: Teacher responses to Q12: 'Please indicate how confident you feel about the following areas'

Once again it can be seen that teachers seem to be confident in areas that are associated with assessing the four skills (reading, writing, listening, speaking). Assessing learners with special needs, using portfolio assessment, peer assessment and self-assessment as well as identifying the relevance of the Common European Framework of Reference for Languages are areas where teachers in the sample report to be the least confident. All these forms of assessment are relatively new in teachers' practices. The frequency of use seems to be partially represented in the confidence levels of the EFL teachers.

It is particularly striking that almost no teacher reports feeling very confident in any of the 20 areas (3 on the Likert scale). Figure 14 shows that the means range between .69 and 2.17. This reinforces the hypothesis that there is a demand for comprehensible assessment training for EFL teachers and that said training can boost their overall confidence. To detect this correlation between previously received training and the EFL teachers' confidence an ANOVA analysis was carried out. It showed that those teachers who received assessment training in the past stated to be significantly more confident (cf. appendix 1).

Item 13 focused on the training needs particularly by asking the teachers to mark each area they would like to be trained in.

Qu. 13. Tick the areas you need training in [In order of preference]	N	%
Assessing students with special learning needs	566	66.43
Using student portfolios	433	50.82
Using self-assessment	346	40.61
Identifying the relevance of CEFR	340	39.91
Using peer-assessment	339	39.79
Recognizing inappropriate methods	299	35.09
Choosing suitable assessment methods	282	33.1
Assessing listening skills	279	32.75
Assessing skills in an integrated way	257	30.16
Assessing speaking skills	240	28.17
Identifying different assessment purposes	239	28.05
Designing classroom-based tests	237	27.82
Preparing learners for external tests	210	24.65
Assessing writing skills	209	24.53
Identifying how tests influence teaching	198	23.24
Using assessment to plan teaching	174	20.42
Using assessment to make decisions about individual learners	159	18.66
Assessing reading skills	157	18.43
Explaining results to parents & others	150	17.61
Explaining results to pupils	123	14.44

Table 7: Teachers' perceived training needs

Once again, it can be seen that the areas teachers would like to receive training in (i.e. assessing students with special needs, using student portfolios, using self-assessment) for the most part overlap with those areas teachers do not use frequently and feel less comfortable in using. The demand across the board confirms that a well-designed training material is needed. Even the areas which were previously regarded as used more frequently and reported by teachers as the more confident ones (e.g. assessing skills) demand training according to the EFL teachers' answers (assessing listening (32.75 %), speaking (30.16 %), writing (24.53 %), and reading (18.42 %)).

IV. Use of technology

The last part of the teachers' questionnaire included only two items: Qu. 16: "Have you ever participated in any kind of online learning course?", and Qu. 17: "If yes above, please clarify." However, this part was significant for the design of the assessment and feedback training material the study was aiming at. In this context, also Qu. 14: "The format I prefer for a training event offered in an online learning environment on language assessment is …" and Qu. 15: "In a training course about assessment in an online learning environment, I would find the following [activity] useful …". These four items can be looked at collectively and were designed to unveil the basic format as well as activities of a future training event desired by the participants. Figure 15 shows the percentage of informants that regarded the respective format 'very useful'. It is remarkable that only 53 (1.5%) of the 3408 answers collected to Qu. 14 equalled the answer 'not useful at all'.
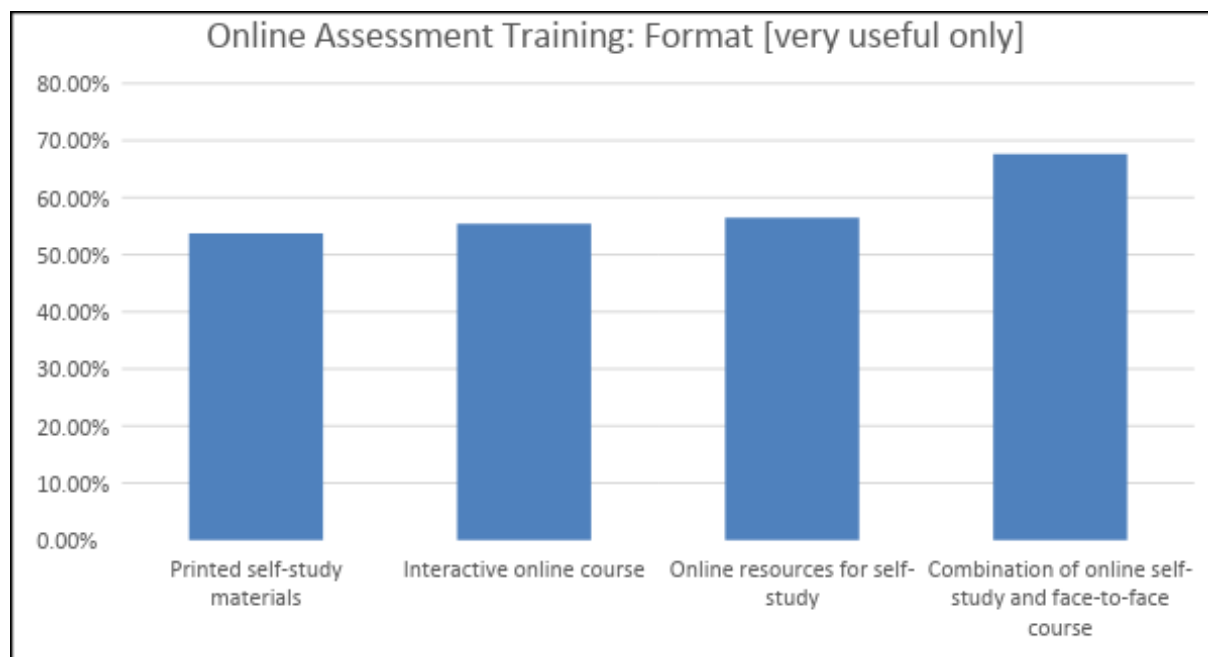
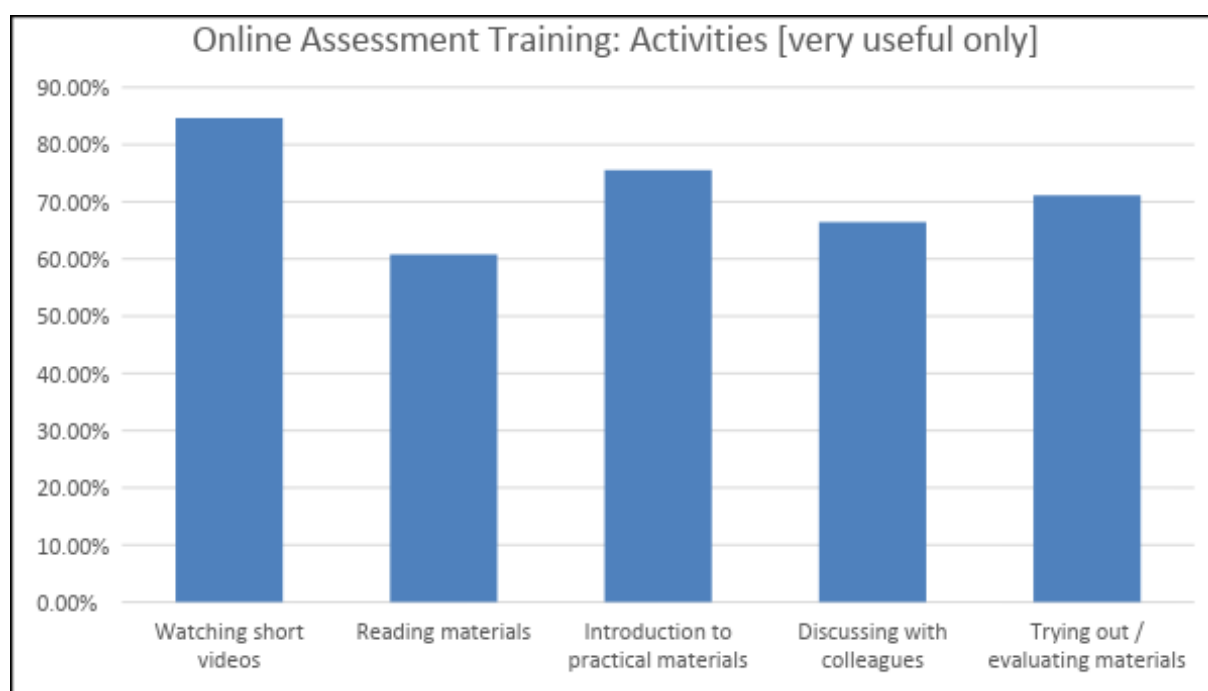*Figure 15: Online assessment training format*

Figure 16: Online assessment training: Activities

Figure 16 comprises the results of question 15 regarding potential activities in an online assessment training. Watching short videos, a hands-on introduction to materials, the possibility to try out assessment materials or to generate them as well as the opportunity to discuss issues with

colleagues was deemed very useful by at least 60% of the teachers asked. Generally, this allows the assumption that an online training event in the future would be regarded useful and thus also be used by EFL teachers. This is particularly interesting in connection to question 16: "Have you ever participated in *any* kind of online learning course?" 74% of the informants replied in the negative but they seem still be very open towards online training resources.
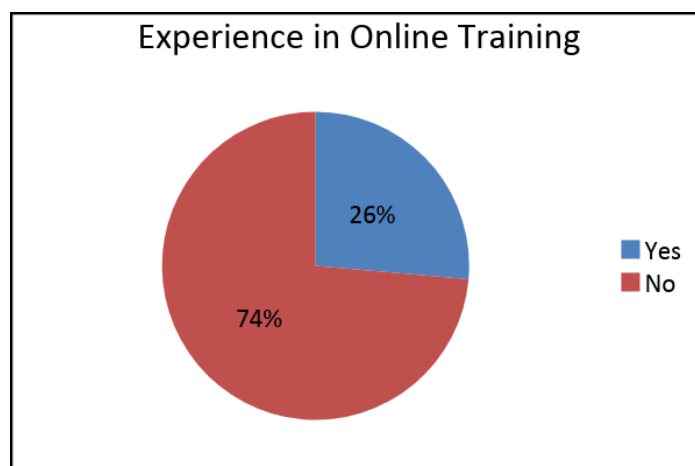
Figure 17: Experience in online training

In the context of the study, this number is important, yet not discouraging. The demand for comprehensive assessment training is definitely there, as question 13 clearly confirms. An online format was also accepted and expected to be useful by the teachers (cf. Figures 15 & 16). Thus, it can be said that the majority of EFL teachers have not yet experienced any online or blended-learning events, however, they would be interested in an online event in the future.

4.1.2 Learner questionnaires

I. General information

Parallel to the teachers' questionnaire, 1788 learners from the same countries were also prompted to participate in the study. This sample was much more evenly distributed in terms of gender as we are looking at an even split of 50% female and 50% male students. The age of the learners ranges from 10 to 20 years although almost half of the sample indicated to be 13-15 years old (cf. figure 18).
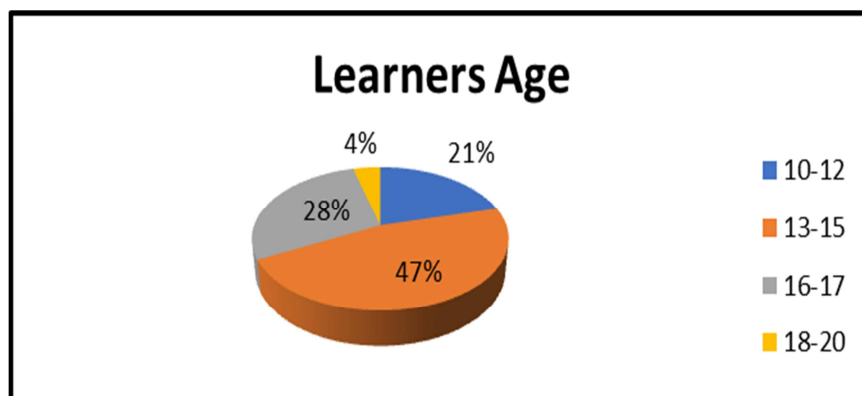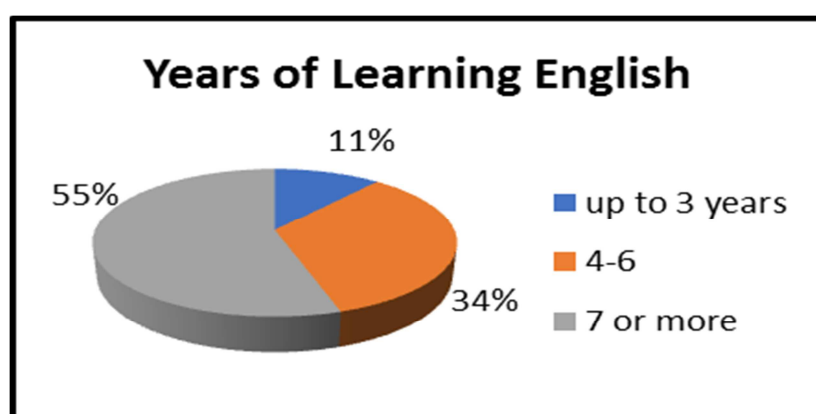
Figure 18: Learners age all countries

Figure 19: Years of learning English all countries

In order to make results from the teacher and learner questionnaires comparable, the learner questionnaires were administered in the same European countries as the teacher questionnaires. Thus, the Cypriot sample is comparably big with 51% of Cypriot learners, which again should be considered when interpreting the data. The remaining countries were relatively evenly distributed (cf. Figure 20).
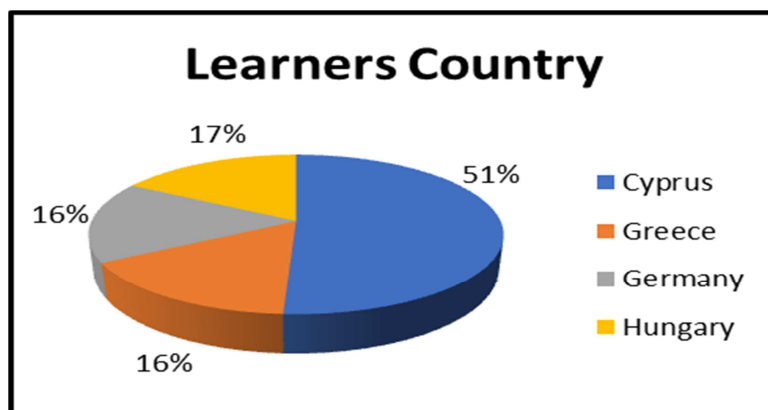
Figure 20: Learners country

II. Assessment practices

This second part of the learners' questionnaire was designed to look into the areas assessed by the teachers according to their learners. In order to do so, two items were designed. The first item (5. "My teacher(s) assess(es) my English … (Choose as many answers as appropriate)") contained the linguistic skills and subskills, parallel to the teacher questionnaire. However, the students were also given the opportunity to add areas themselves. Figure 21 shows a comparison of the skills / areas indicated by the teachers and learners.
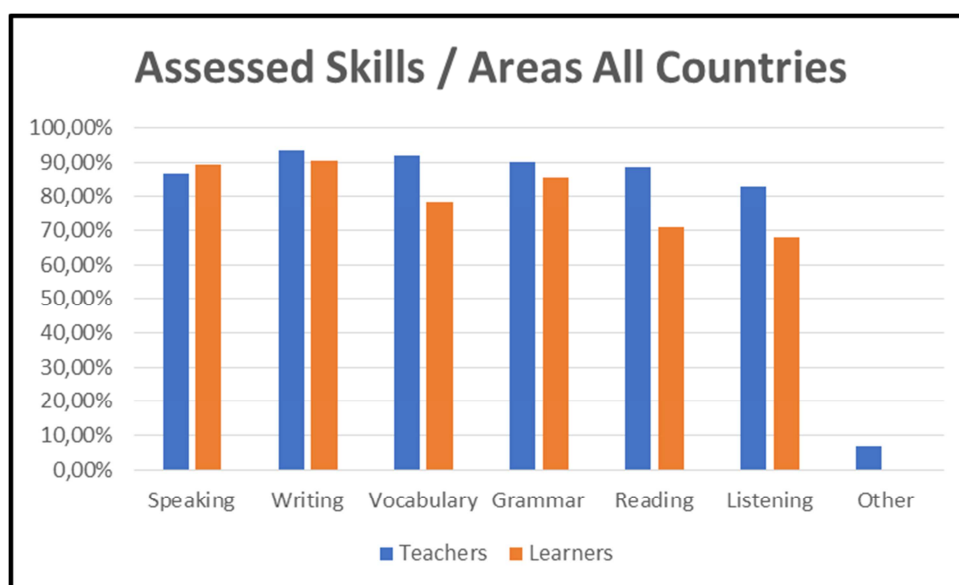
Figure 21: Assessed skills / areas all countries

It can be observed that overall the data overlaps for the most part. However, there are some minor discrepancies in vocabulary, reading and listening. All of them indicate that fewer learners have identified the respective area to be assessed. This could be due to fact that some assessment methods go unnoticed by the learners. The second item in this part was designed to determine which assessment methods are used by the teachers (6. "How often does your teachers ask you to…"). Ten predetermined areas were given, each could be answered through a four-point Likert scale (3 = very often, 2= often, 1 = sometimes, 0 = never). Figure 22 shows a comparison of the teachers' and the learners' results. The results again overlap to a certain degree but there are two very interesting and significant discrepancies.
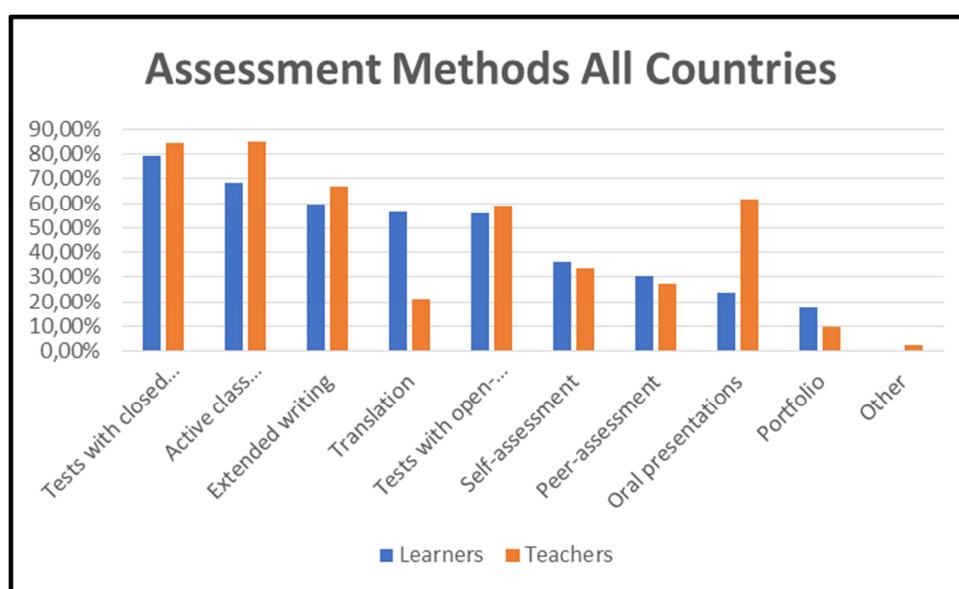
Figure 22: Assessment methods all countries

III. Assessment needs and wants

The first item included in the third part of the learners' questionnaire was designed to determine what methods are regarded useful when learning English (7. "Does it help you learn English when you…?"). Ten predetermined areas were given, each could be answered through a four-point Likert scale (3 = very often, 2= often, 1 = sometimes, 0 = never). This data is particularly interesting compared to item 11 of the teachers' questionnaire ("How often do you use the following methods to assess your learners' English?"). Such a comparison can be seen in Figure 23. Some of the

discrepancies are very striking and the method *translate sentences or texts* stands out once more. The collected data suggest that this method, although regarded useful very often by a substantial number of students, is not used very frequently in the classroom. Parallel trends can be noticed with other methods such as *assess your own work*. It can only be speculated why such discrepancies show however, it seems likely that some methods are not identified in the same way by the learners and teachers.
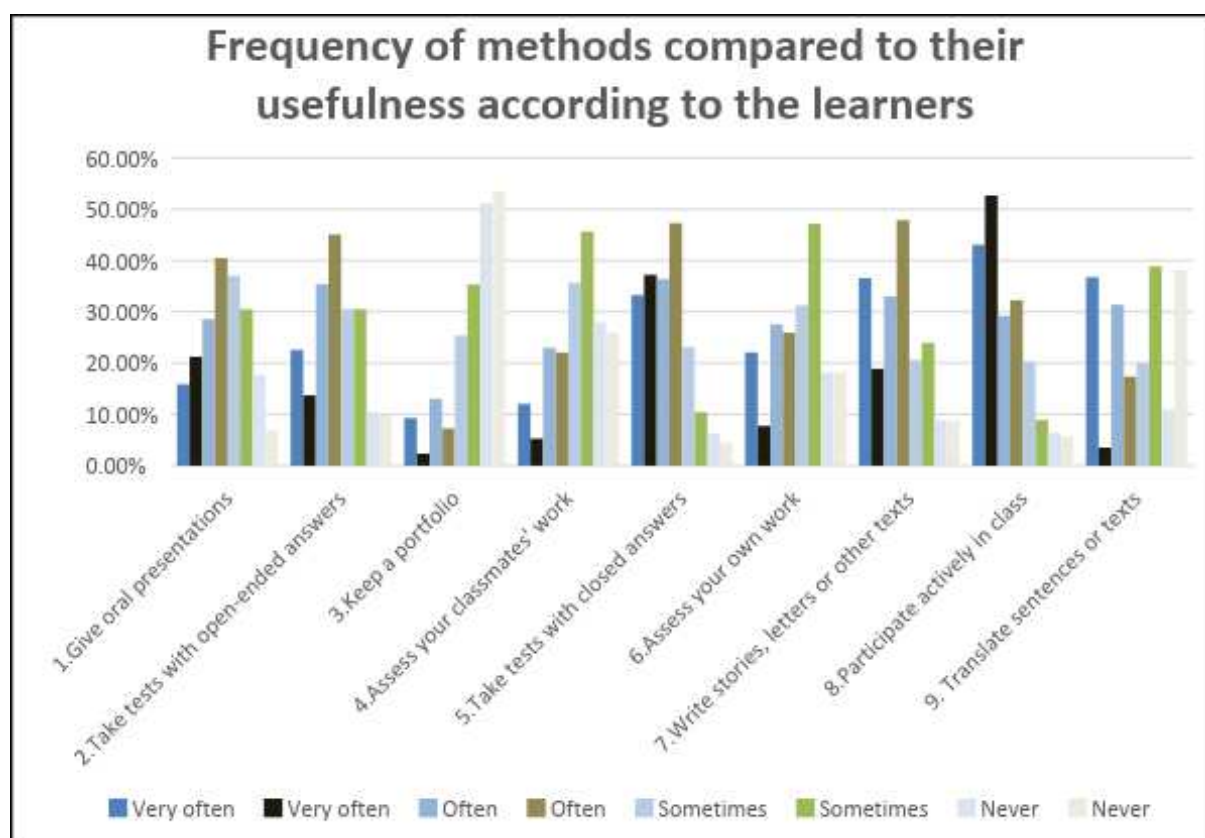

Figure 23: Frequency of methods compared to their usefulness according to the learners

The second item of the third part of the learners' questionnaire was designed to determine what kind of feedback the learners get (8. "What feedback do you get on your assessment results?"). Once again these results can be compared to the indications made by the teachers, which was done in figure 24.
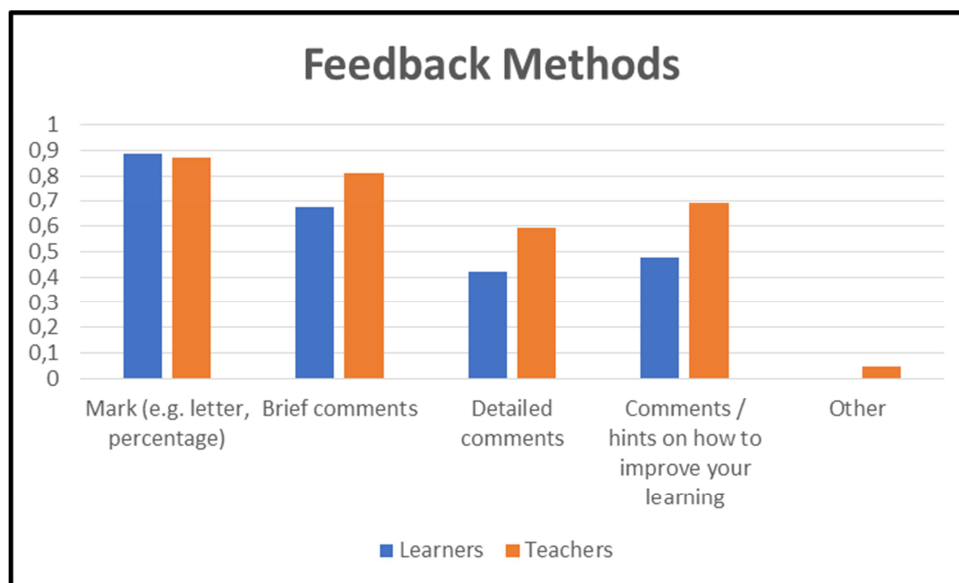
Figure 24: Feedback methods

4.2 Results – Cyprus

I. General information

As previously indicated in figures 10 and 11, 47 % of the teachers' and 51 % of the learners' sample were collected in Cyprus. Thus, the following presentation of results needs to be considered as particularly impactful on the overall results, displayed above.

The participating Cypriot EFL teachers were at least 46 years old and, therefore, more experienced in comparison to the other sub-samples. Figures 25 and 26 support these claims and furthermore show that the percentage of less experienced (10 %), young (3 %) teachers is strikingly small.
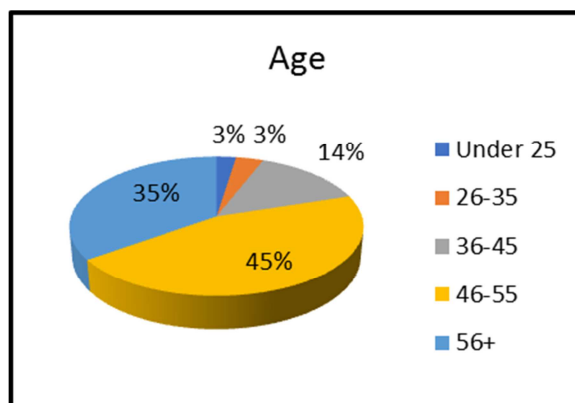
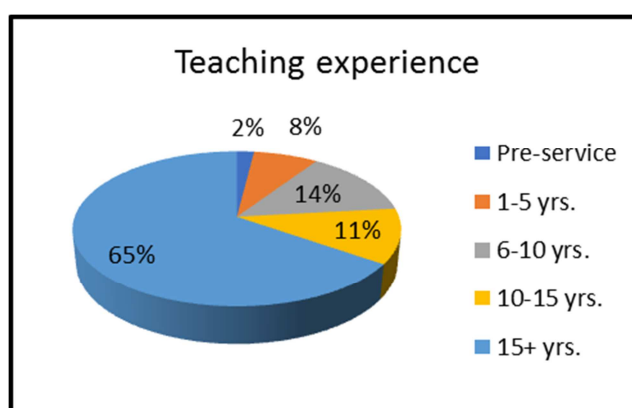*Figure 25: Age of Cypriot EFL teachers*

Figure 26: Teaching experience of Cypriot EFL Teachers

This is due to teachers only acquiring a position at state schools at a relatively late stage in their careers.

In congruence with the overall sample, the majority of Cypriot teachers (96 %) stated English as the language they teach. The same applies for the age of their learners, which was slightly less evenly distributed than the overall sample (cf. figure 27).
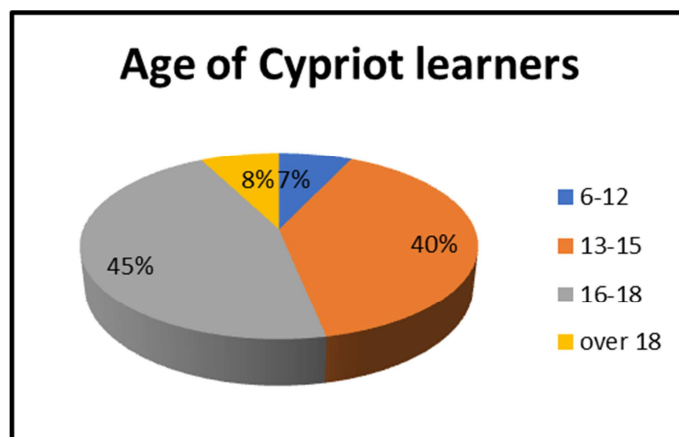
Figure 27: Age of Cypriot learners

II. Assessment practices

The second part of the questionnaire included three items which were to determine the areas that are currently assessed, what type of feedback is given, and the frequency of the different assessment types in the EFL classroom. Figure 28 illustrates a comparison of the overall answers to Qu. 11: "How often do you use the following methods to assess your learners' English?" and the answers collected in Cyprus. Furthermore, the learners' answers to Qu. 6: "How often does your teachers ask you to…" were compiled into the same graph (cf. figure 28). Both items, in the teachers' and the learners' questionnaire, were equipped with a four-point Likert scale (0 = never, 1 = sometimes, 2 = frequently, 3 = very frequently). Figure 28 shows all answers that equaled 2 or 3 on the Likert scale (frequently and very frequently).
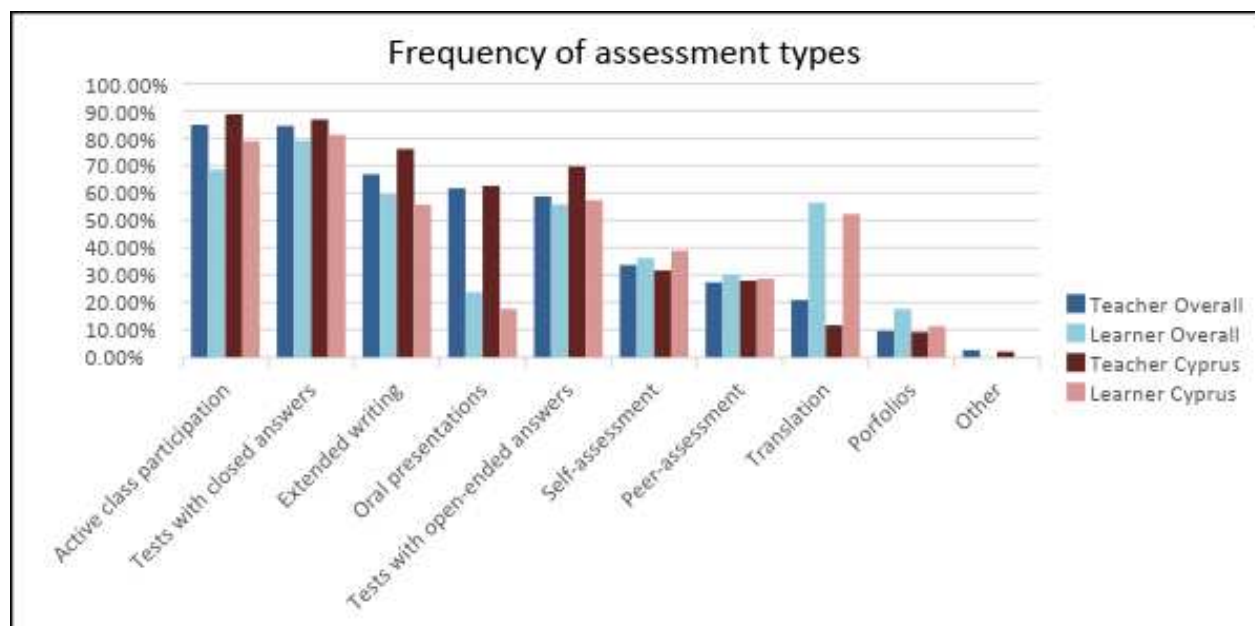
Figure 28: Frequency of assessment types: Overall results compared to Cyprus

The results are in line with the overall findings on the use of different assessment types that were presented above. However, there are some aspects that differ and should therefore be mentioned.

The overall results showed that teachers reported certain assessment types in the EFL classroom (active class participation, tests with closed answers, extended writing, oral presentations, tests with open-ended answers) to be used more frequently than their learners did. For the remaining four types (self-assessment, peer-assessment, translation, portfolio) the opposite tendency can be noted in the overall data as well as in the Cypriot data. Teachers reportedly used these assessment methods types to be used less frequently than their learners did. The Cypriot sample confirms this trend with an even wider gap between teachers' and learners' perception in several assessment methods.

One striking example is the assessment method *extended writing* which shows a noticeable discrepancy. Cypriot teachers rated extended writing as an important and frequently used part of their assessment practices in the EFL classroom since 76% stated it to be used at least frequently (cf. figure 29).
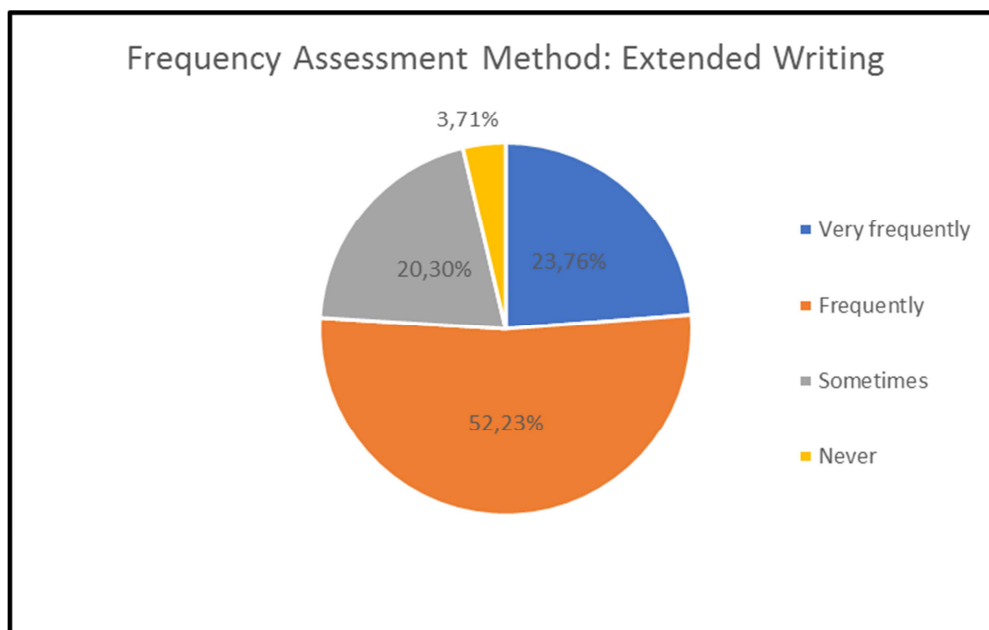
Figure 29: Frequency of extended writing as assessment method in Cyprus

Only 55.75 % of the EFL learners from Cyprus, however, rated this assessment method to be used at least frequent. Several reasons can be stated for this discrepancy. Since these items are based on the participants' ability to accurately identify and remember incidents in which a certain assessment type was used, certain inaccuracies are expected. Still, a discrepancy as significant as this suggests that there might be a reason other than false memory or lack of identification.

The frequency of all assessment types in the EFL classroom is particularly interesting in connection to question ("Does it help you learn English when you …") of the learners' questionnaire. This item included ten questions, each with a four-point Likert scale (very often=3, often = 2, sometimes = 1, never = 0) to determine how often learners evaluate the respective assessment methods to be helpful when learning English. Figure 30 shows a comparison between question 7 from the learners' and question 11 from the teachers' questionnaire ("How often do you use the following methods to assess your learners' English?"). As mentioned above, these items were designed to be parallel and therefore compared in the analysis. Therefore, question 11 from the teachers' questionnaire also included ten questions, each with a four-point Likert scale (very frequently = 3, frequently = 2, sometimes = 1, never=0).
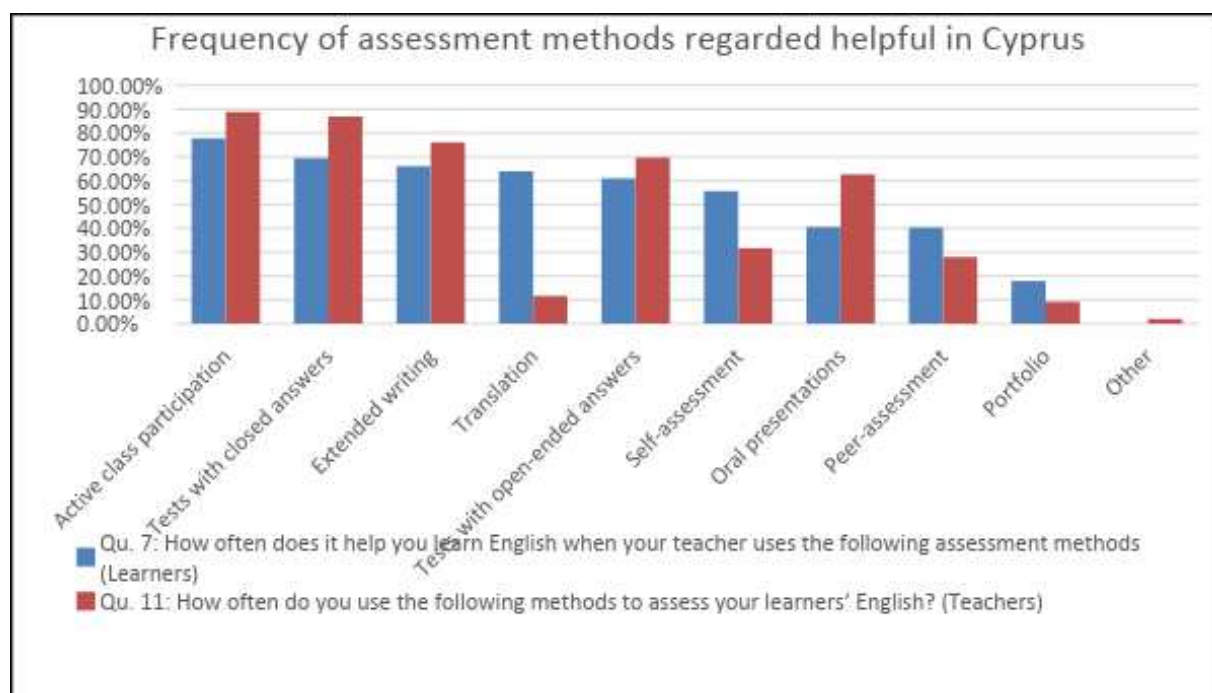

Figure 30: Frequency of assessment methods regarded helpful in Cyprus

This comparison confirms a trend that was identified in the overall results already. This time the results can be directly connected to the students' learning experience, since they ascertain how often a certain assessment method helps them to learn English. More than 50 % of the Cypriot EFL learners for example rated *translation* and *self-assessment* as being helpful in the learning process at least frequently (cf. figure 30). However, these methods are not used very frequently at all by their Cypriot EFL teachers. Once again, especially *translation* as an assessment method produces a considerable

discrepancy with only 11.63 % teachers stating it to be used frequently or very frequently, compared to the 64 % of Cypriot learners identifying this method to be useful often or very often. Translation can include rather common assessment methods such as vocabulary tests in the EFL classroom. However, teachers report that they do not use these ways of assessing very often. Once again, these results could be due to an erroneous or differing identification of the learning situation in the EFL classroom. Thus, stakeholders' perceptions of the assessment method used could be different. However, the data available cannot confirm such claims with the questionnaire data available. Possible follow-up interviews with the same learners and teachers could be a way to elaborate this discrepancy.

III. Assessment profiles and training needs

In this part of the questionnaire, the teachers' confidence levels about 20 predetermined areas connected to assessment were ascertained (Qu. 12: "Please indicate how confident you feel about the following areas"). In addition to that, the teachers were also able to indicate if they would like to receive training in each respective area. The results of the third part of the Cypriot teachers' questionnaire are basically in line with the overall results (cf. figure 31). This could be expected as the Cypriot teachers made up 47 % of the overall sample.
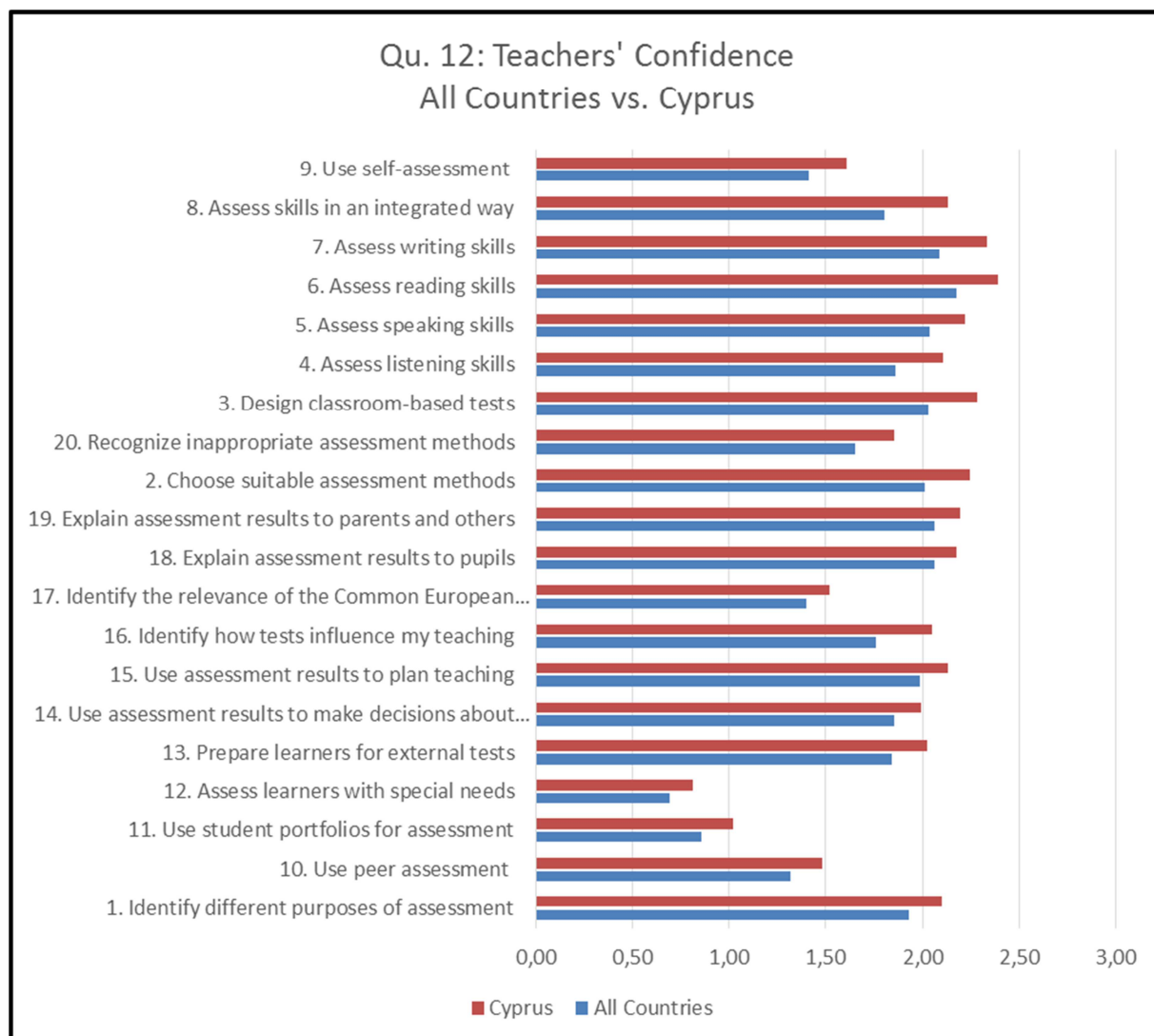
Figure 31: Teachers' confidence: All countries compared to Cyprus

It should be mentioned that on average the Cypriot EFL teachers were slightly more confident in every single point than teachers in the overall sample. As mentioned above, Cypriot teachers represented the oldest and thus most experienced sub-sample (79.95% being 46 years or older, 65% having 15+ years of teaching experience). It can be assumed that these two factors contribute to the fact that Cypriot teachers report to feel more confident in all areas. When looking at the areas where the discrepancy in confidence between Cyprus and all countries is the highest, again the more conventional assessment types can be recognised (8. Assess skills in an integrated way (.33), 3. Design classroom-based tests (.26), 7. Assess writing skills (.24)). This supports the hypothesis that more experienced teachers tend to use conventional assessment types rather than

more innovative ones. This could be due to a lack of proper assessment training in the past. In fact, it can be assumed that teachers with more than 15 years teaching experience have not received any assessment courses during their studies as it was rarely implemented in the teachers' training in the past.

A comparison of the percentage of Cypriot EFL teachers who indicated they would like to receive training in a respective area with the overall data confirms the assumption made above (cf. figure 32). It can be seen that generally, fewer teachers in Cyprus reported a need for training in any of the 20 areas. Once again, this might be connected to the EFL teachers' experience and confidence in the classroom.
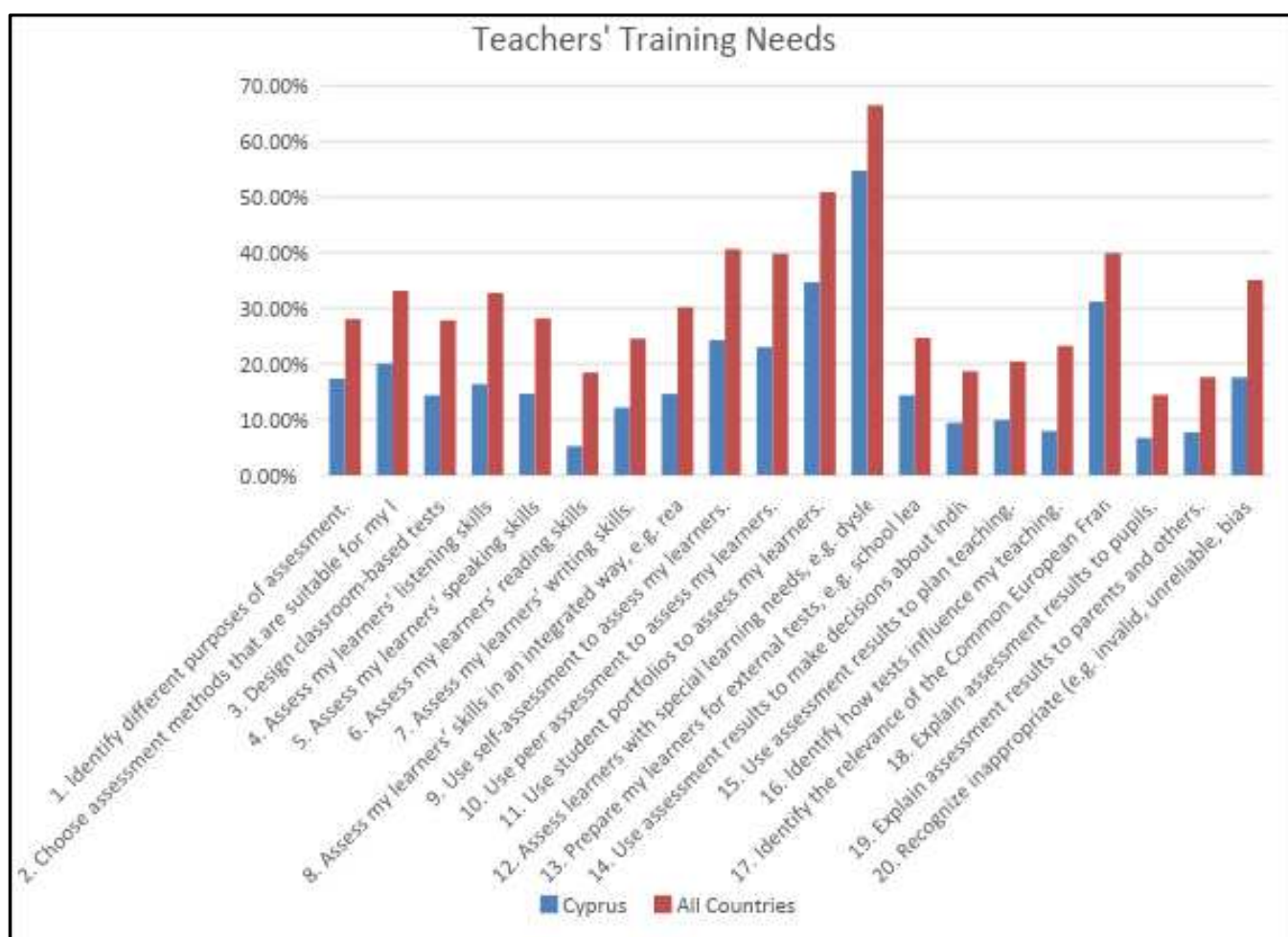

Figure 32: Teachers' training needs: Overall results compared to Cyprus

IV. Use of technology

In terms of the format of a future training event in an online learning environment on language assessment, five predetermined options were given: 1. Printed self-study materials, 2. Interactive online course, 3. Online resources for self-study, 4. Combination of online self-study and face-to-face course, 5. Other. The aim was to determine the perceived usefulness of each option, by once again providing a four-step Likert scale (0 = not useful at all, 1 = less useful, 2 = somewhat useful, 3 = very useful). The Cypriot EFL teachers demanded roughly the same formats of a potential online training event as the overall sample. Once again, this might be due to the Cypriot sub-sample making almost half of the overall sample. Still, it is interesting that the Cypriot EFL teachers seem to demand formats containing self-study more than the other two (cf. Figure 33).
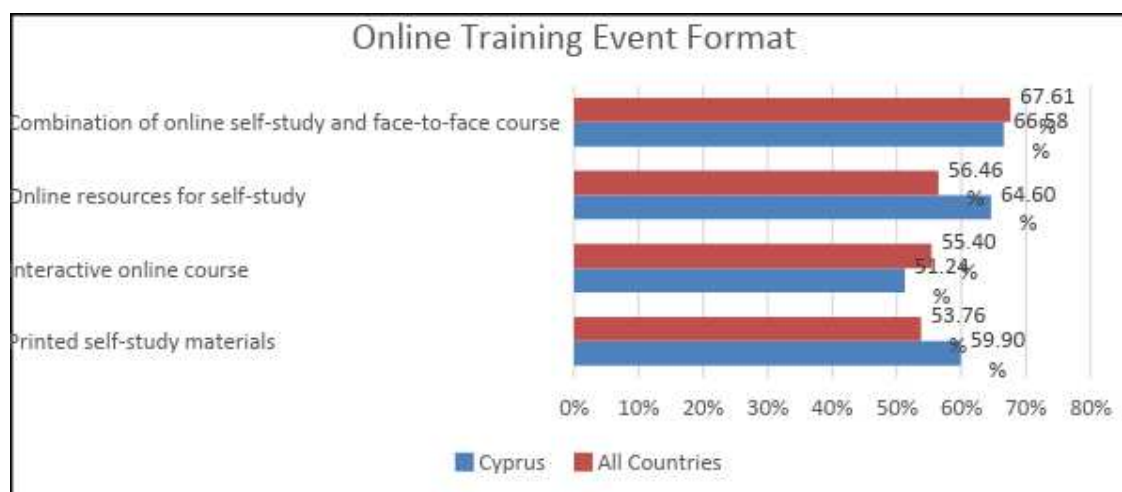

Figure 33: Online training event formats Cyprus

Also, as already pointed out in the overall results, it is interesting that no format was rated 'not useful at all' by the teachers. Across all four items, 'not useful at all' was only ticked 28 times which is equivalent to less than 2 % of all given answers as figure, as figure 34 shows.
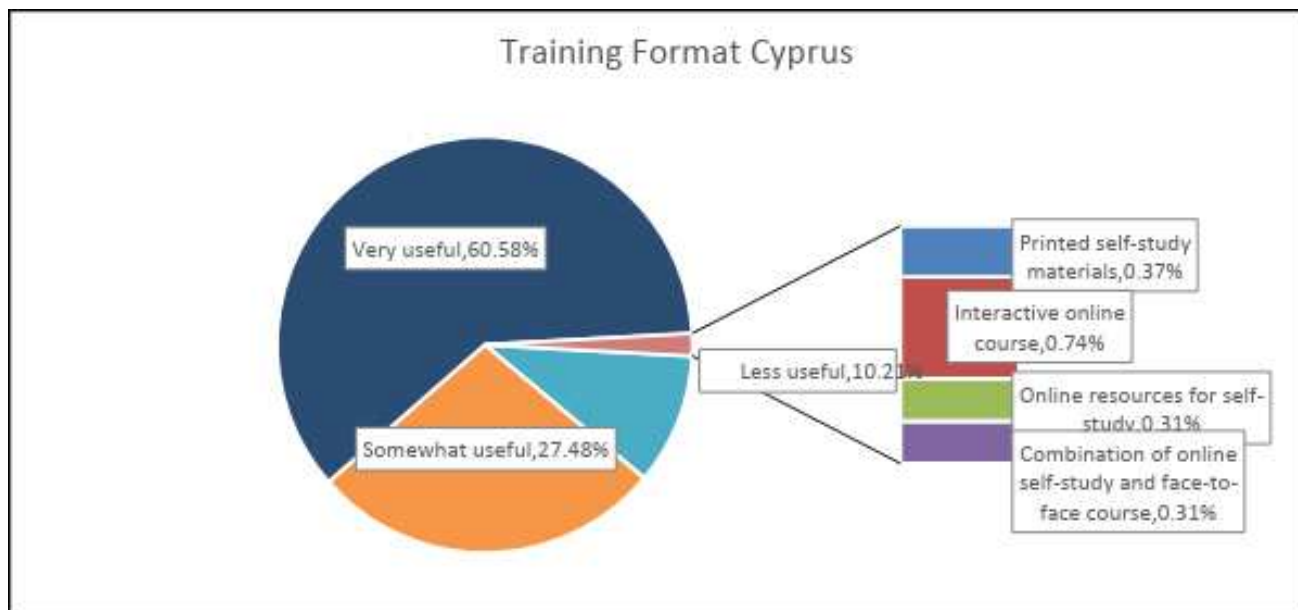
Figure 34: Online event format Cyprus

This confirms that an online assessment training event would be appreciated by almost all Cypriot EFL teachers in the sample. Question 15, which was designed to determine which activities would be regarded useful by the EFL teachers, yielded approximately the same result. Figure 35 shows how many Cypriot EFL teachers regarded which online activity as being 'very useful' compared to the overall sample, showing many overlaps with the overall data. However, when comparing to the overall sample, approximately 15 % more Cypriot EFL teachers rated 'reading materials' to be 'very useful' (cf. figure 35). This is in line with the finding that a course format that enables teachers to engage in a self-study format would be appreciated.
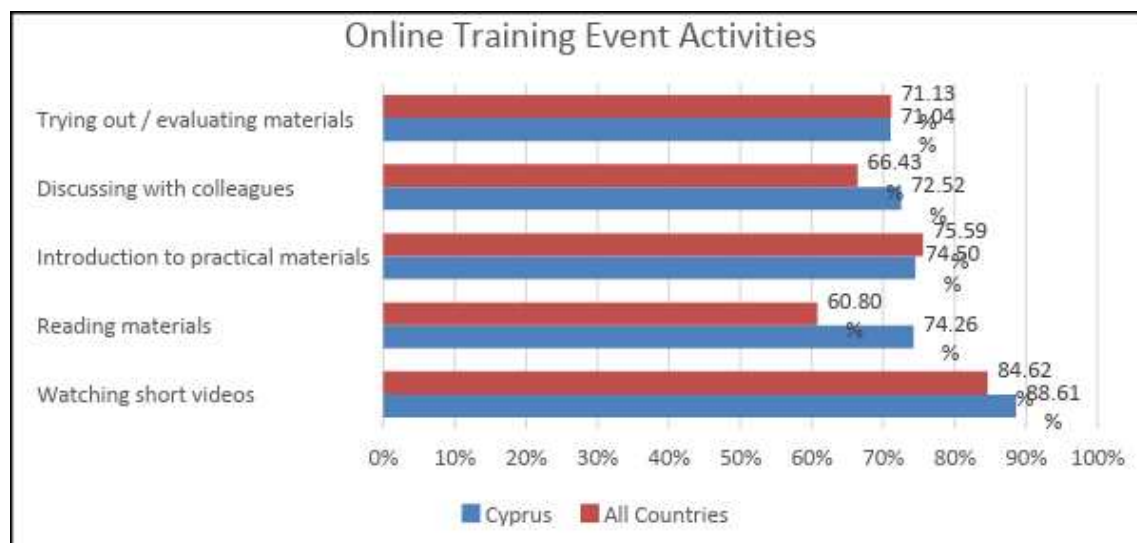
Figure 35: Online training event activities Cyprus

The claim made above, that the relative high age of the Cypriot sub-sample is responsible for several effects that were presented in this report is once again confirmed in the second last Item of the questionnaire (Qu. 16: "Have you ever participated in any kind of online learning course?"). Figure 36 shows the expected result that 81 % of the Cypriot EFL teachers have not yet participated in an online training course.
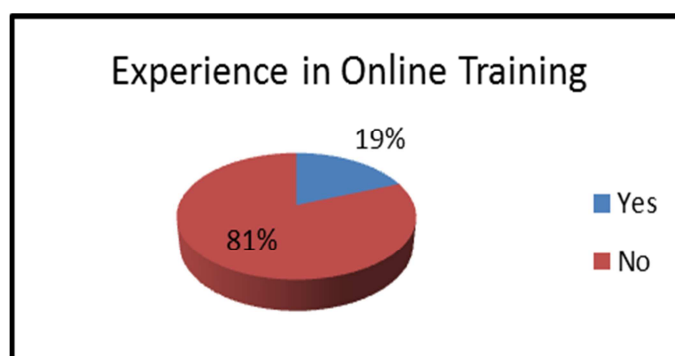

Figure 36: Teachers' experience in online training Cyprus

4.3 Results – Germany

I. General information

With only 127 EFL teachers and 285 learners, the results collected in Germany represent the second smallest part of the overall sample (13.50 % of all participants). Therefore, the results presented in this part are certainly less influential than the previous ones (Cyprus) but still show the effect of contextual factors on assessment practices and perceptions.

The first noticeable aspect is the teachers' age range which shows entirely different attributes in the German sub-sample. Figure 37 shows that almost 89 % of all German EFL teachers who participated were 35 years or younger. This results in a teacher population almost opposite to the Cypriot sample (94 % being 36 or older). The overall results show an evenly distributed age range of all EFL teachers, suggesting that the Cypriot and the German sub-samples even out the entire sample, as figure 37 illustrates.
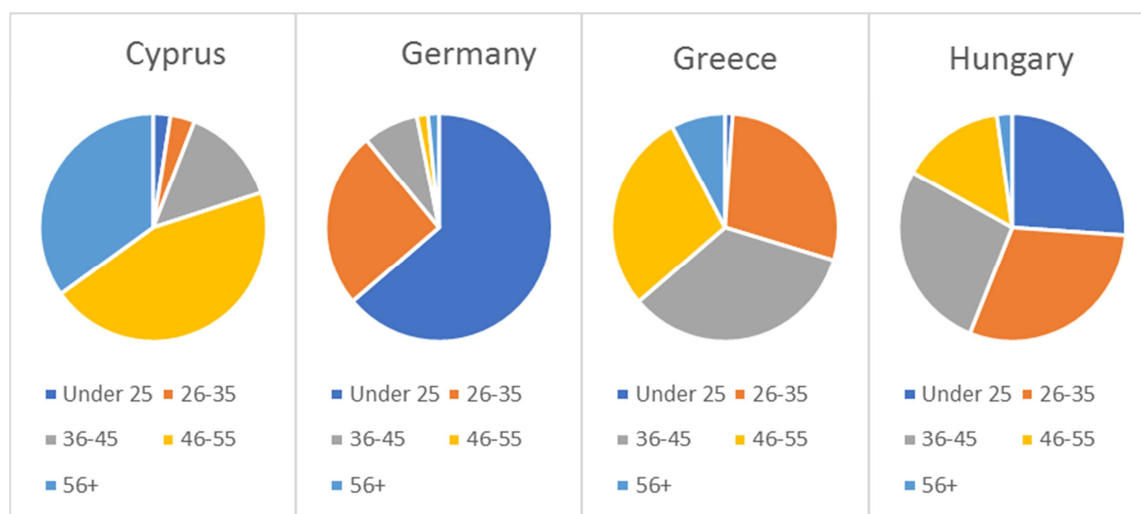
Figure 37: EFL teachers age distribution

In the German sub-sample, 74% of the teachers were pre-service teachers, as figure 38 confirms. This fact needs to be explained further though, as teacher training in Germany differs from the other countries involved in the study. In figure 38, the pre-service teachers are certainly the least experienced, but it does not mean they do not have any teaching experience at all. In Germany, teachers must successfully complete a two-year period in a school after graduating to be officially

allowed to teach. In this sub-sample, many 'pre-service' teachers are currently in this two-year phase after graduation.

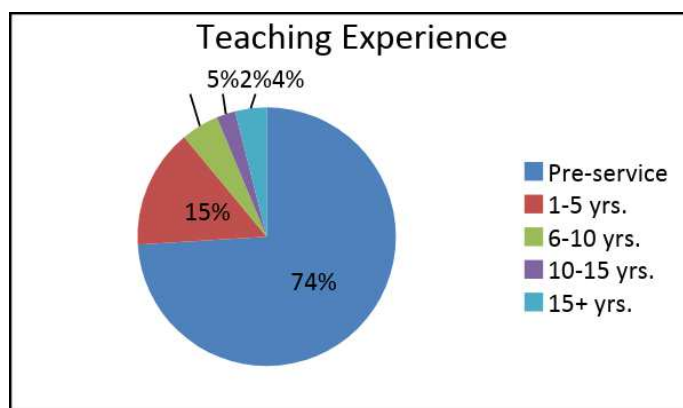*Figure 38: Teaching experience of German EFL teachers*

It is important not to misinterpret the German sub-sample in this context. Still, figures 37 and 38 once again confirm the assumption that teachers' age and experience seem to be related.

It can also be noted that zero percent of the German teachers stated a BA degree as their highest teaching qualification while 98 % stated secondary school or state exam. This is because teacher students for primary and secondary schools in the part of Germany the subsample comes from were not able to acquire a BA degree until 2015. Instead, the state exam was the only degree a teacher student could acquire, which accounts for the result for BA degree. This was changed in 2015 and students of education for primary and secondary schools now graduate with a BA degree after six and a MA degree after eight semesters, thus substituting the former state exam.

Although the German sub-sample was the youngest group of teachers on average, almost 60% stated that they had never received any testing and assessment training (Qu. 8 teachers' questionnaire). Thus, the previous assumption that older teachers lack assessment skills because it was not implemented in the past cannot be applied. There must be other reasons for the different experiences EFL teachers had with testing and assessment training in their respective country. In Germany, it can be assumed that the lack of training is due to the marginalised role that language testing and assessment still plays in pre-service teacher education, as previous research into the assessment literacy of foreign language teachers confirms (Vogt, 2011).

II. Assessment practices

As previously mentioned, the second part of the teachers' questionnaire was designed to be parallel to the learners' questionnaire. Thus, the perceptions of both stakeholders on assessment practices (assessed skills, type of feedback, and frequency of assessment methods) can be compared. In terms of skills that are assessed in Germany (Qu. 9: "Which of these skills/areas do you assess?"), the emphasis mostly lies on the four skills such as speaking, writing, and listening (cf. figure 39).
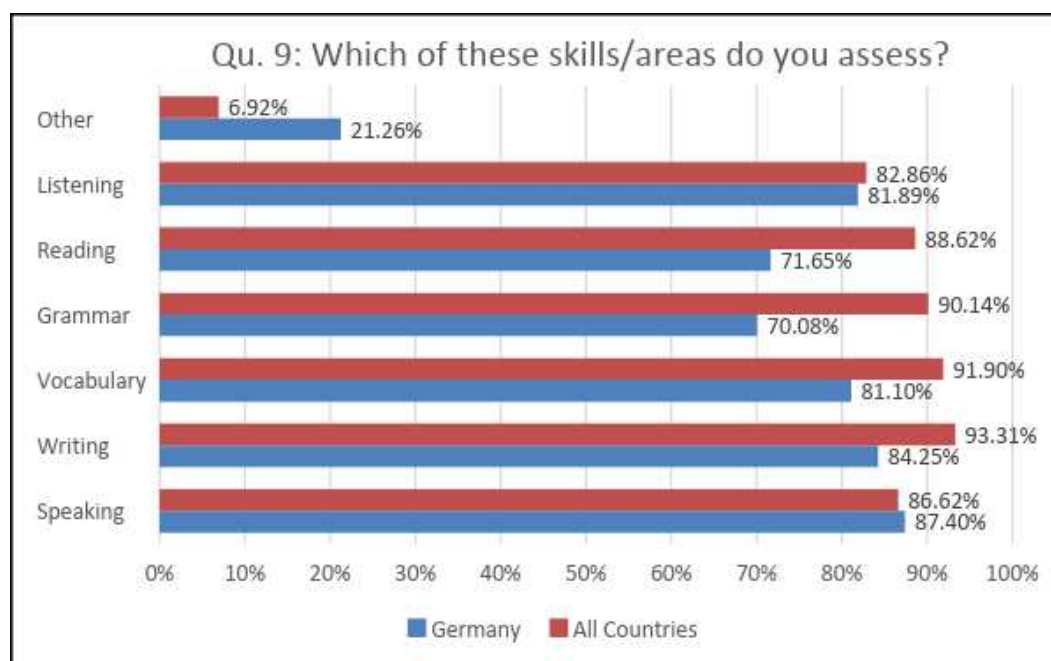

Figure 39: Skills assessed by German EFL teachers

Surprisingly, the skill *reading* does not follow this trend in the German sub-sample as almost 17 % less EFL teachers stated it to be assessed at all, compared to the overall results. Still, communicative skills in the EFL classroom seem to be important to the German teachers in the sample. This can be assumed because the skill *mediation,* which was not a given choice in the questionnaire, was still mentioned by many teachers who ticked 'other'. The fact that three times as many German EFL teachers decided to choose 'other', suggests that the six predetermined skills and subskills did not match completely the assessment practices used in class. Particularly *Intercultural competence, presentation skills,* and *mediation* were added to the skills/areas by the German EFL teachers. In addition to that, it can be observed that the most significant discrepancy occurs in the context of the area *grammar*. In Germany, 70% compared to 90% in the overall sample stated that

they assess grammar in their classrooms. This result might be due to the fact that competence-oriented language teaching has been enforced by curricula throughout Germany, placing less importance on subskills like grammar and vocabulary and in some places forbidding to assess grammar separately.
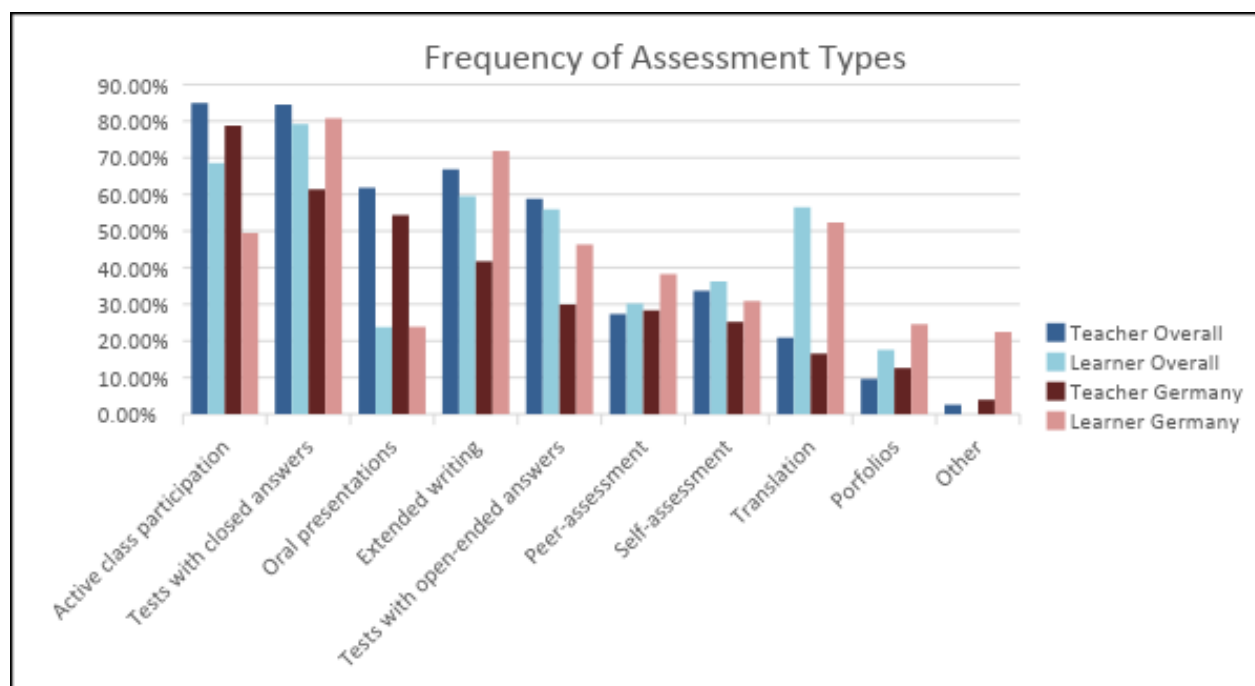

Figure 40: Frequency of assessment types: Overall results compared to Germany

As pointed out in the overall results, some discrepancies between the teachers' and learners' perception is striking. Among these are the assessment methods '*tests with closed answers*', '*extended writing*', and '*tests with open ended answers*'. As figure 40 shows, German EFL teachers report using all three of these methods considerably less often than their colleagues from the other countries. In addition to that, the EFL learners from Germany perceived those three types of assessment to be used in the classroom strikingly more often than their teachers stated. Again, this data is based on the participants' recollection and correct identification of certain situations in which they are convinced a certain assessment type was used. This could be one reason for a discrepancy of perceptions.

Nevertheless, a difference as remarkable as in the German sub-sample suggests that there must be additional reasons which cannot be ascertained on the basis of the present data alone. Figure 41 shows a comparison of how many EFL teachers in Germany rated the respective methods

to be used 'frequently' or 'very frequently' and how many EFL learners rated the same methods to be helpful often or very often when learning English. This comparison stresses the trend identified above and shows that there is not only a discrepancy in the perception of how often certain methods are used but also toward their helpfulness in the EFL classroom. Ideally, the results in figure 41 would be expected to overlap. However, there are some discrepancies in the context of the methods 'translation', 'extended writing' and 'tests with open-ended answers' in particular.
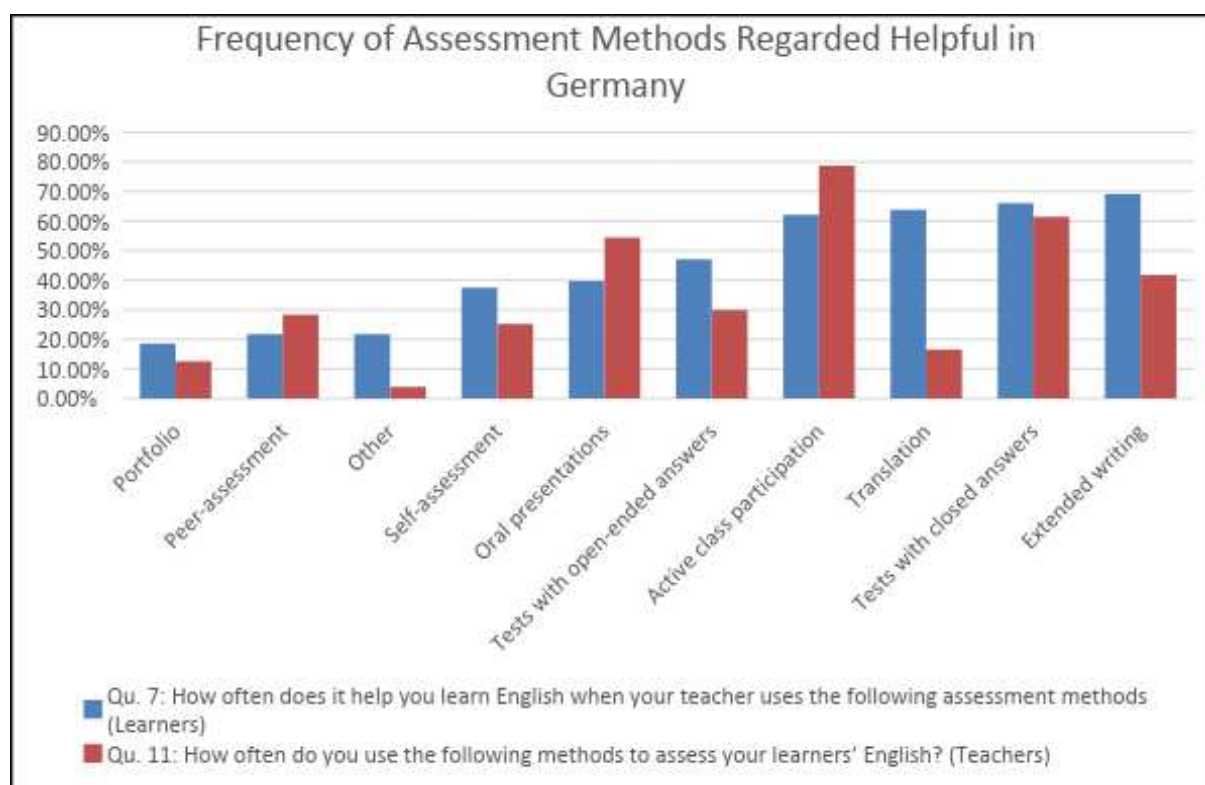
Figure 41. Frequency of assessment methods regarded helpful in Germany

III. Assessment profiles and training needs

This part of the teachers' questionnaire explored the EFL teachers' confidence levels and training needs regarding 20 predetermined areas. To ascertain the confidence of German EFL teachers, for each area a four-point Likert scale was given (0 = not confident, 1 = somewhat confident, 2 = confident, 3 = very confident). Figure 42 shows how many teachers chose confident or very confident for each respective area.
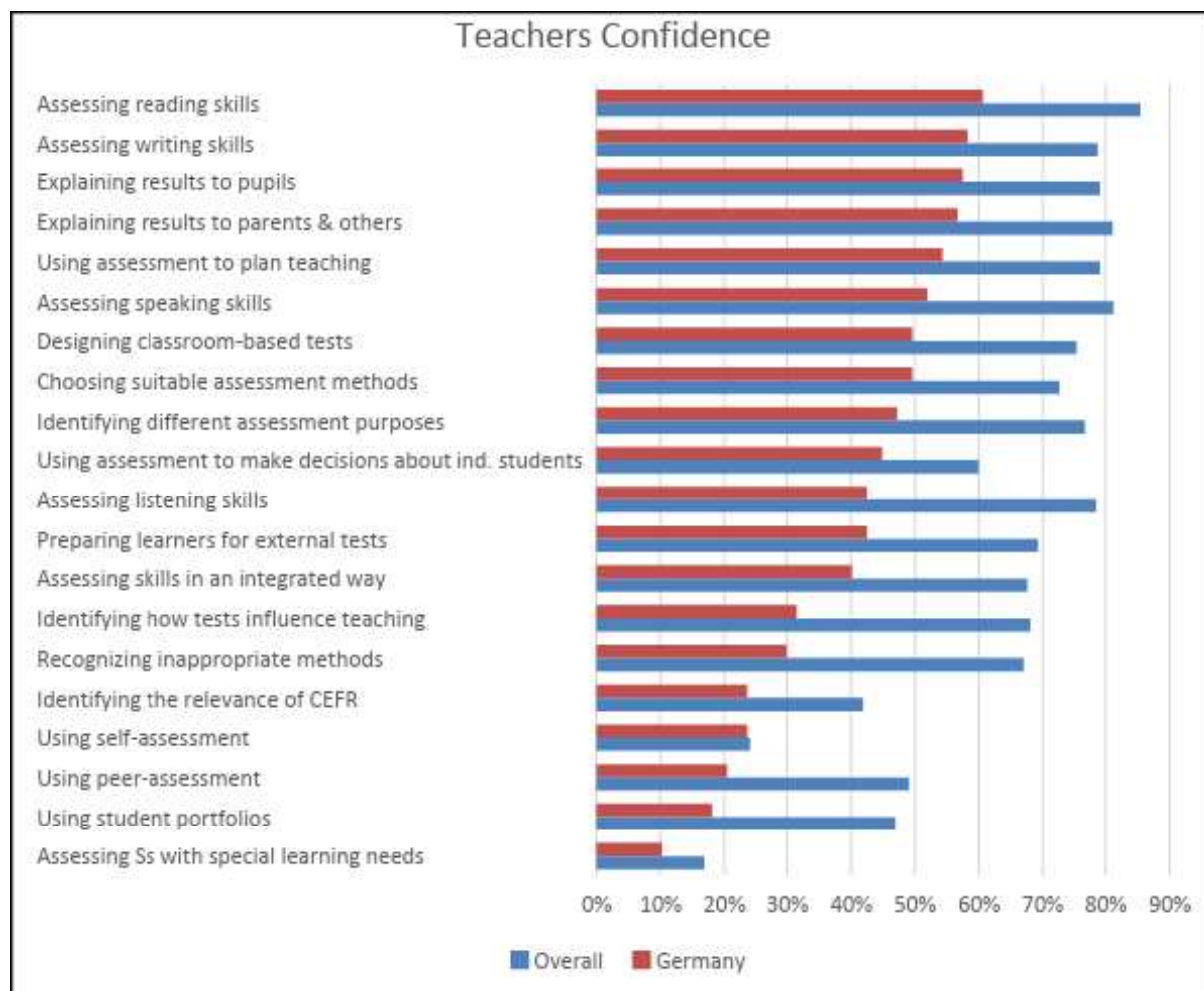
Figure 42: Teachers' confidence (very confident + confident): Overall results compared to Germany

The comparison of the German sub-sample and the overall results show very convincingly that overall fewer German EFL teachers stated to be '*very confident*' or '*confident*'. Figure 43 shows that most German EFL teachers stated to be 'somewhat confident'. However, there is still a significant number of teachers who reported they were 'not confident' in several areas (cf. figure 43).
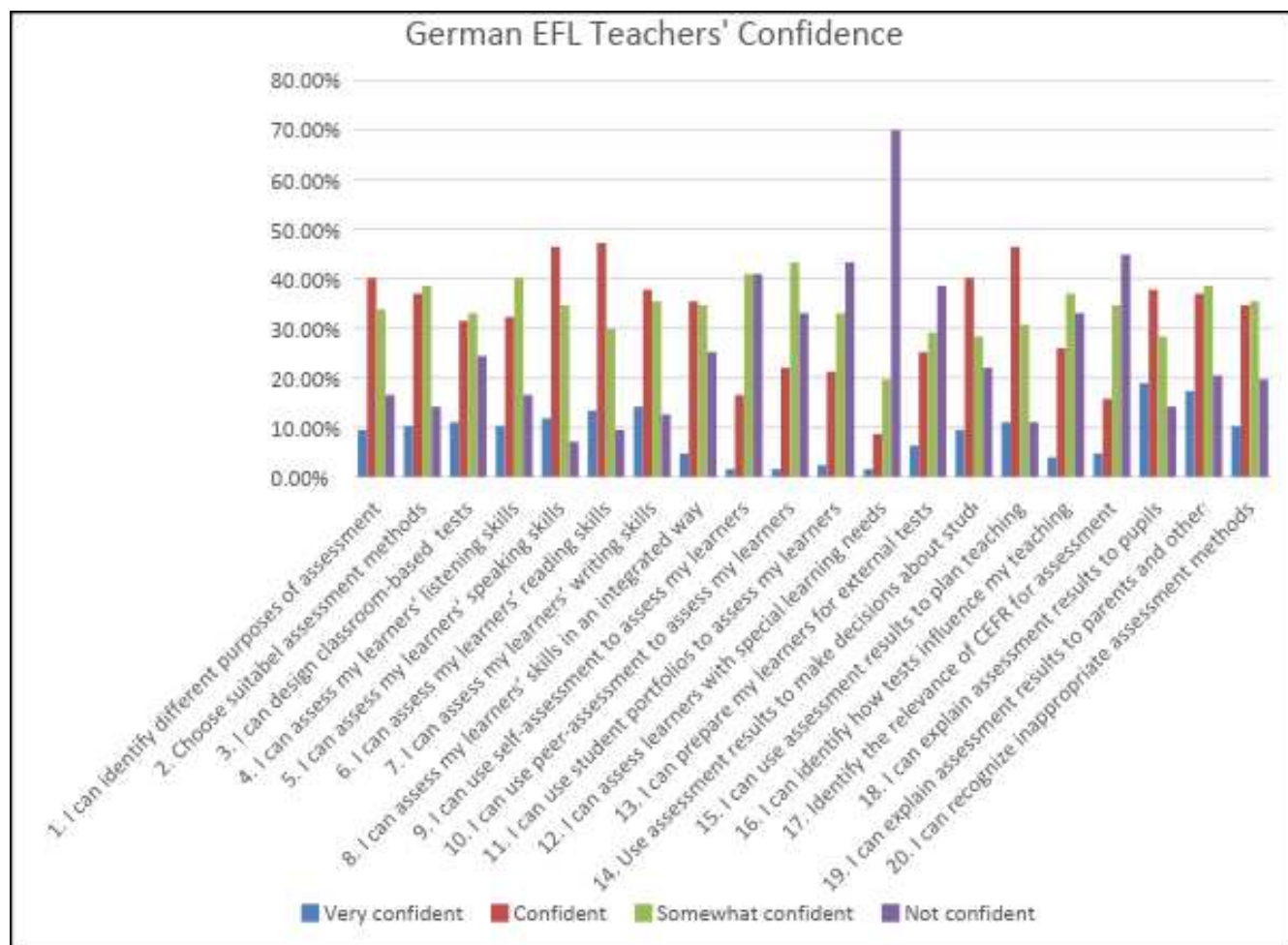
Figure 43: German EFL teachers' confidence

The confidence level of German teachers is, compared to other sub-samples in this study, relatively low. There might be several reasons for this about which we can only speculate. The first reason might be related to the age and level of experience of the informants. The majority of teachers in the German sample is at the beginning of their teaching careers and thus not very experienced. Furthermore, the result might also partly be due to contextual factors since language assessment tends to be neglected in pre-service teacher training (Vogt, 2011) and in-service teacher training is not mandatory in Germany. Teachers are not forced to take in-service teacher training courses about assessment, which might account for lower levels of language assessment literacy in the sub-sample. Previous research supports this assumption (ibid). Lastly, it might also be down to cultural differences in communication patterns since teachers might only be tempted to report their confidence in an area if they are really confident. As said before, the reasons can only be speculated on.

Question 13 (I'd like training in this) showed that generally, more German EFL teachers would like to receive training in all the 20 areas (cf. figure 44). Even in comparison to the overall results, the German sub-sample strikes for having an overall high demand for assessment training. Again, the reasons for this result can only be speculated on but experience-related factors, contextual and / or cultural factors could account for this result.
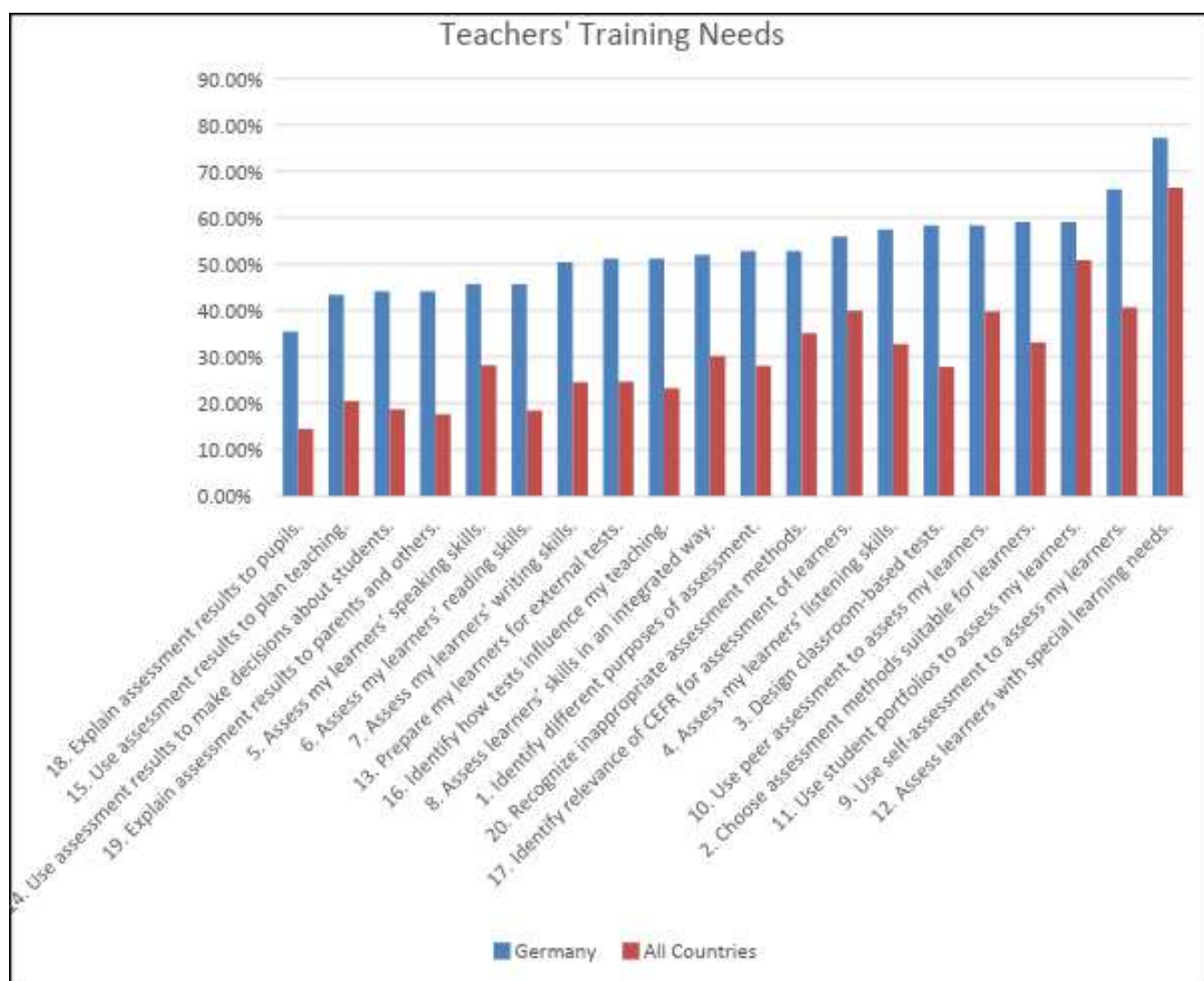
Figure 33: Teachers' training needs: Overall results compared to Germany

IV. Use of technology

To establish the teachers' preferences regarding the format of an online learning programme on language assessment, two items (Qu. 14: "Usefulness of different formats" and Qu. 15: "Usefulness of different activities") were included in the teachers' questionnaire. Both items included several predetermined options with a four-point Likert scale (0 = not useful at all, 1 = less useful, 2 = somewhat useful, 3 = very useful). The results show that all given options for a possible format were regarded useful (cf. figure 45).
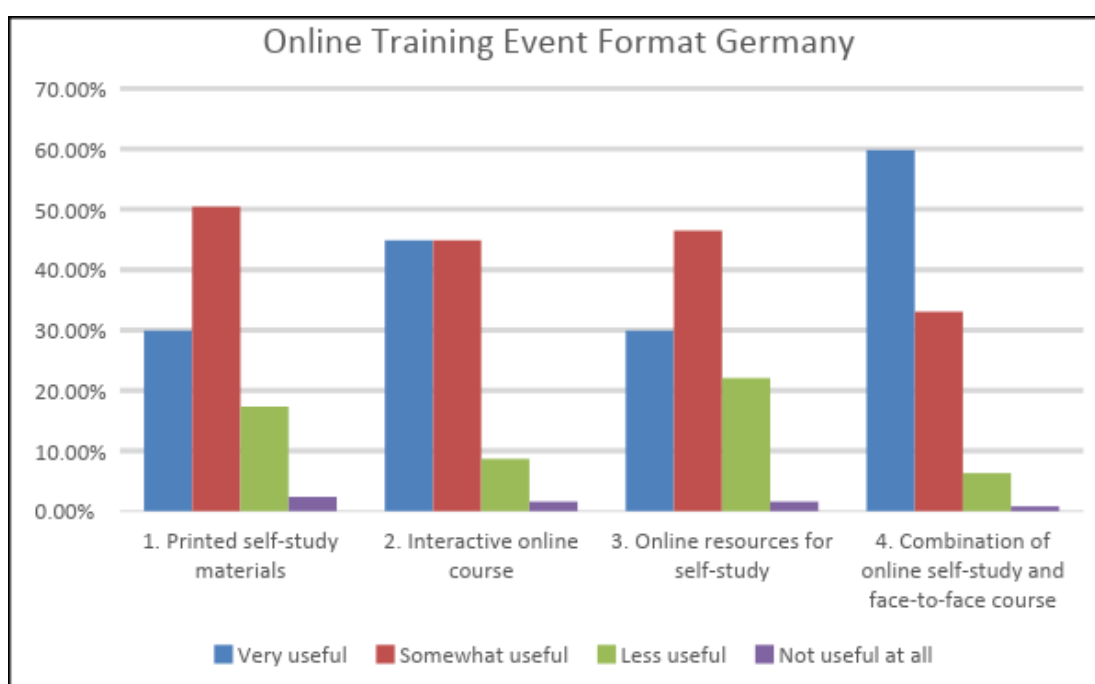
Figure 45: Online training event format Germany

Only 6.30 % of all answers given by EFL teachers asked in Germany stated any of the four options *'not useful at all'.* This result is in line with the overall sample where a total of 6.22 % of the answers equalled '*not useful at all*'. This confirms the initial hypothesis that there is a great demand across Europe for language assessment training that is delivered digitally.

The same trend can be identified regarding the demanded training methods and content. Only 3.15% of all 127 participating German EFL teachers regarded any of the given choices as '*not useful at all'* (cf. figure 46).
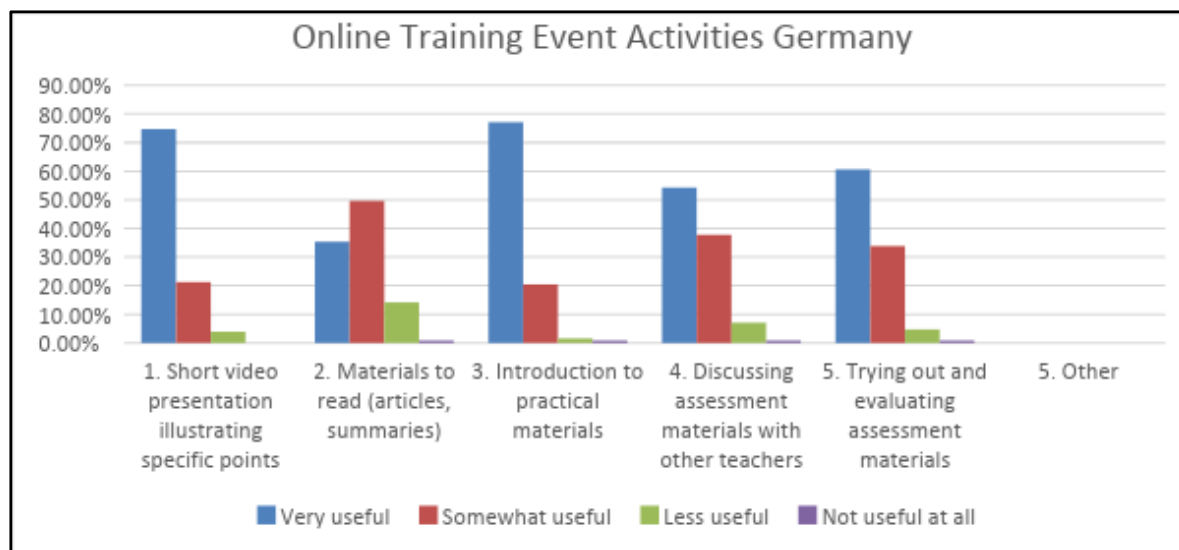
Figure 46: Online training event activities Germany

Finally, the German EFL teachers were asked if they have ever participated in any kind of online learning course (Qu. 16: "Have you ever participated in any kind of online learning course?"). The result was rather surprising, since the young age of the sub-sample indicated toward a more recent training that could have included online parts. However, more than half of the German sub-sample stated that they never had any kind of online training experience (cf. figure 47).
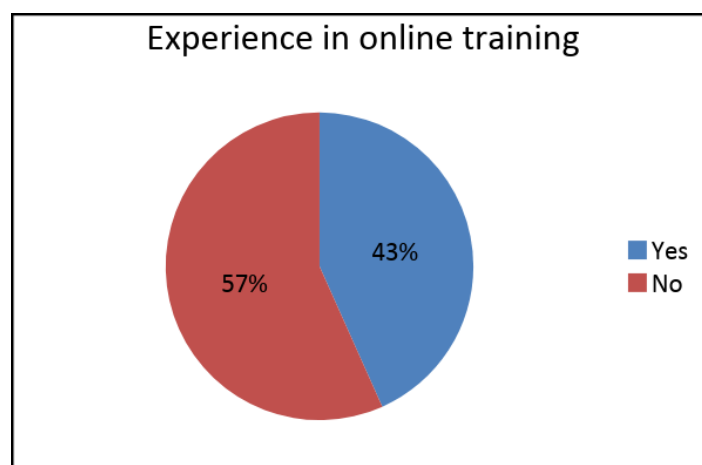
Figure 47: Teachers' experience in online training Germany

4.4 Results - Greece

I. General information

The Greek sub-sample makes the smallest part (14.58%) of the overall collected data. A total of 91 EFL teachers and 294 learners participated in the study. Once again, the gender distribution of the teacher sample corresponds to the overall data (82% female / 18 % male) as it reflects the current state of the profession.

Opposite to the previously presented German and Cypriot sub-samples, the Greek EFL teachers' age was well distributed (cf. figure 48).
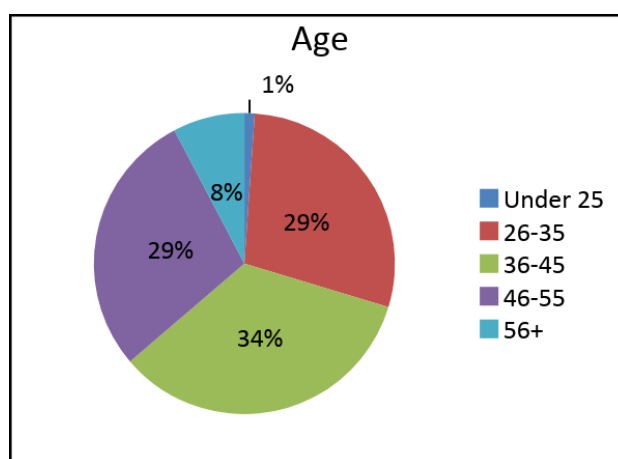
Figure 48: Age of Greek EFL teachers
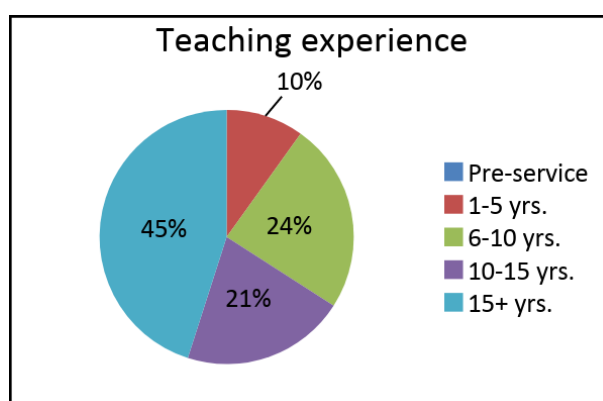
Figure 49: Teaching experience of Greek EFL teachers

This obviously reflects on the teaching experience as identified in all sub-samples thus far (cf. figure 49). It is interesting that there were no pre-service teachers at all and generally only a rather small amount of less experienced teachers (10 %). The same effect was identified in the Cypriot sub-sample, which is due to the same teacher training systems in Cyprus and Greece.

The predominantly taught language was once again English (94 % of Greek teachers stated English). Similar to the German sub-sample, only 44 % of the Greek EFL teachers stated that they had received any testing and assessment training (Qu. 8). This is almost 20 % less of what the overall result showed as 63 % of all EFL teachers confirmed that they had received some training. In contrast, Cyprus and Hungary had ratios that were comparable to the overall result.

As in the other sub-samples presented above, the gender of the EFL learners was very evenly distributed (46 % female / 54 % male). As to the learners' age, the Greek sample was rather young by average (cf. figure 50).
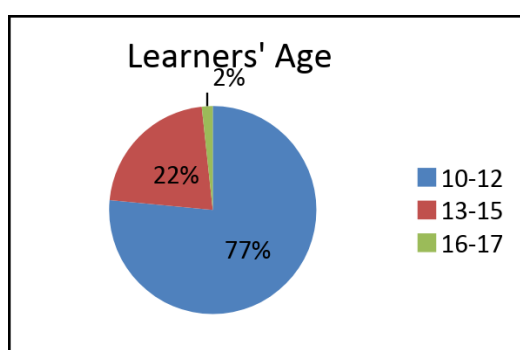
Figure 50: Age of Greek EFL learners

II. Assessment practices

As previously mentioned, the second part of the teachers' questionnaire was parallel to the learners' questionnaire. Thus, the perceptions of both stakeholders on assessment practices (assessed skills, type of feedback, frequency of assessment methods) can be compared. The areas assessed by Greek EFL teachers largely overlap with the results from all countries (cf. figure 51). However, some discrepancies can be found when looking at the skills '*listening'* and '*speaking'*.
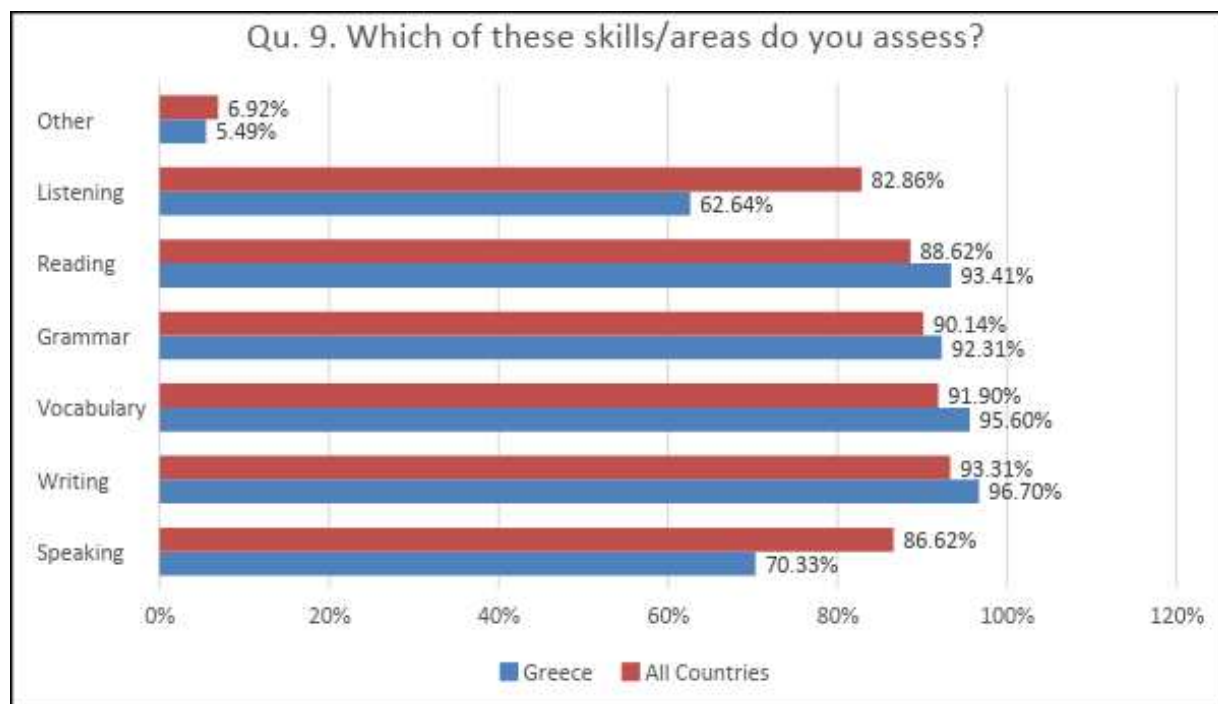
Figure 51: Skills assessed by Greek EFL teachers

Fewer EFL teachers chose '*other*' compared to the German sub-sample. Therefore, it seems that there was almost no demand for any skills/areas to be assessed in the EFL classroom beyond the five predetermined choices. In terms of feedback given by Greek EFL teachers, it can be observed that their choice generally overlaps with the results from all countries.
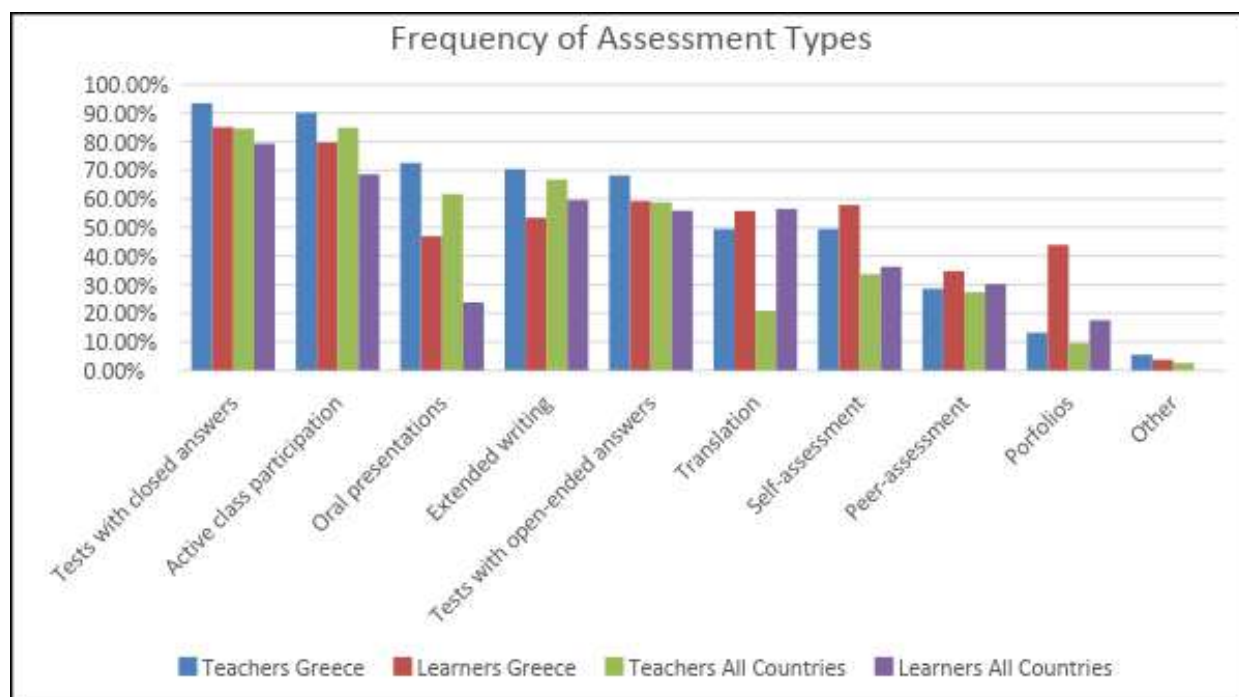
Figure 52: Frequency of assessment types: Overall results compared to Greece

Figure 52 shows that Qu. 11: "How often do you use the following methods to assess your learners' English?" produced results that were mainly in line with the overall sample. However, there are some crucial differences that need to be pointed out. Striking discrepancies in teachers' and learners' perception of the usage of different methods were presented in the overall results and have been confirmed by every sub-sample so far. Although there have been discrepancies between the data collected from Greece EFL teachers and learners show more similar answers, especially concerning o*ral presentation* and *translation*.

In addition to the frequency, the learners were also asked how helpful they find each of these assessment methods when learning English (Qu. 7: "Does it help you learn English when you…"). This item included all methods named above (cf. figure 53), each with a four-point Likert scale (0 = never, 1 = sometimes, 2 = often, 3 = very often). In figure 53, all methods used 'frequently' or 'very frequently' according to the Greek EFL teachers were compared to the learners' opinion on their helpfulness ('often' and 'very often').
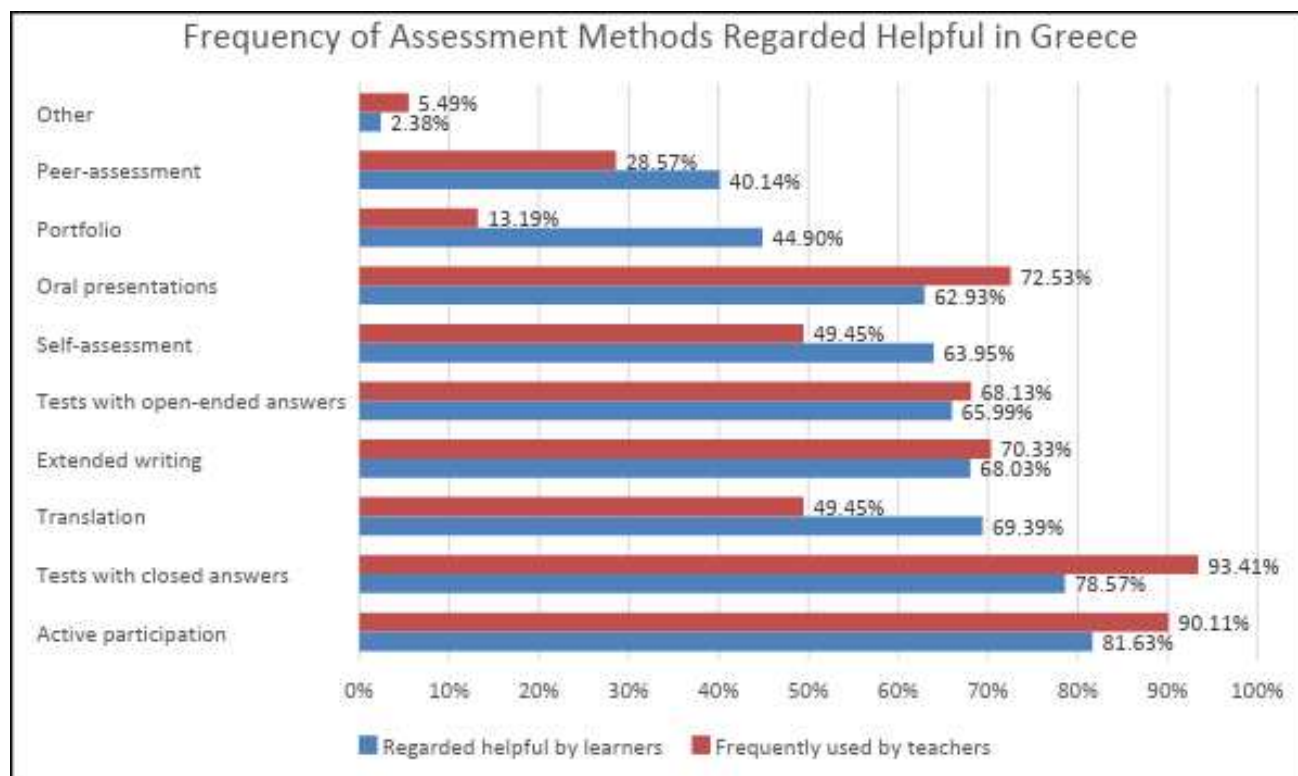
Figure 53: Frequency of assessment methods regarded helpful in Greece

As in other sub-samples, there seem to be discrepancies between the teacher and learner perceptions, particular in translation. It can be assumed that differences of EFL learners' and teachers' results might be based on whether a translation-related classroom activity is perceived as assessment or not.

III. Assessment profiles and training needs

In this part of the questionnaire, the teachers' confidence about 20 predetermined areas connected to assessment was recorded (Qu. 12: "Please indicate how confident you feel about the following areas"). In addition to that, the teachers were also able to indicate if they would like to receive training in each respective area (Qu. 13: "Wherever you feel you need training, mark the last column on the right with a tick"). Item 12 was designed with a four-point Likert scale (0 = Not confident, 1 = Somewhat confident, 2 = Confident, 3 = Very Confident). Figure 54 shows the number of Greek EFL teachers who indicated to be '*confident*' or '*very confident*' compared to the overall results.
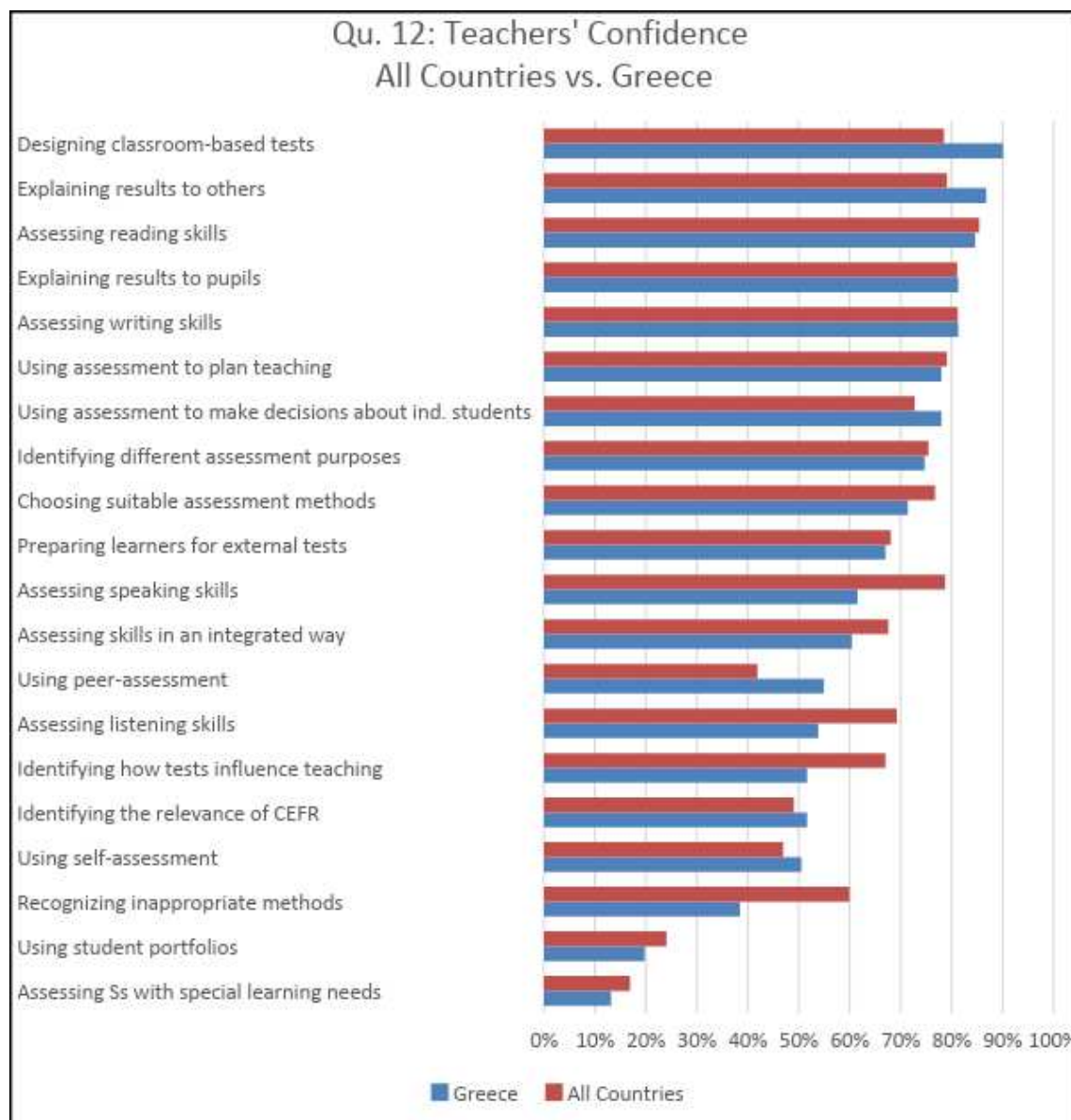
Figure 54: Teachers' confidence: All countries compared to Greece

It can be seen that the Greek results are much more congruent with the overall sample than previous sub-samples were. This means that more Greek EFL teachers stated to be 'confident' or 'very confident' and the overall confidence seems to be higher than in Germany for example. This result can be attributed to several factors, namely their amount of teaching experience or a context-related factor. As figure 54 shows, there are still some areas in which the Greek sub-sample differs from the overall data though. The areas in which fewer Greek EFL teachers stated to be '*confident'* or '*very confident*' (i.e. 'recognizing inappropriate methods', 'assessing speaking skills', 'assessing listening skills', 'identifying how tests influence teaching') stand out here.
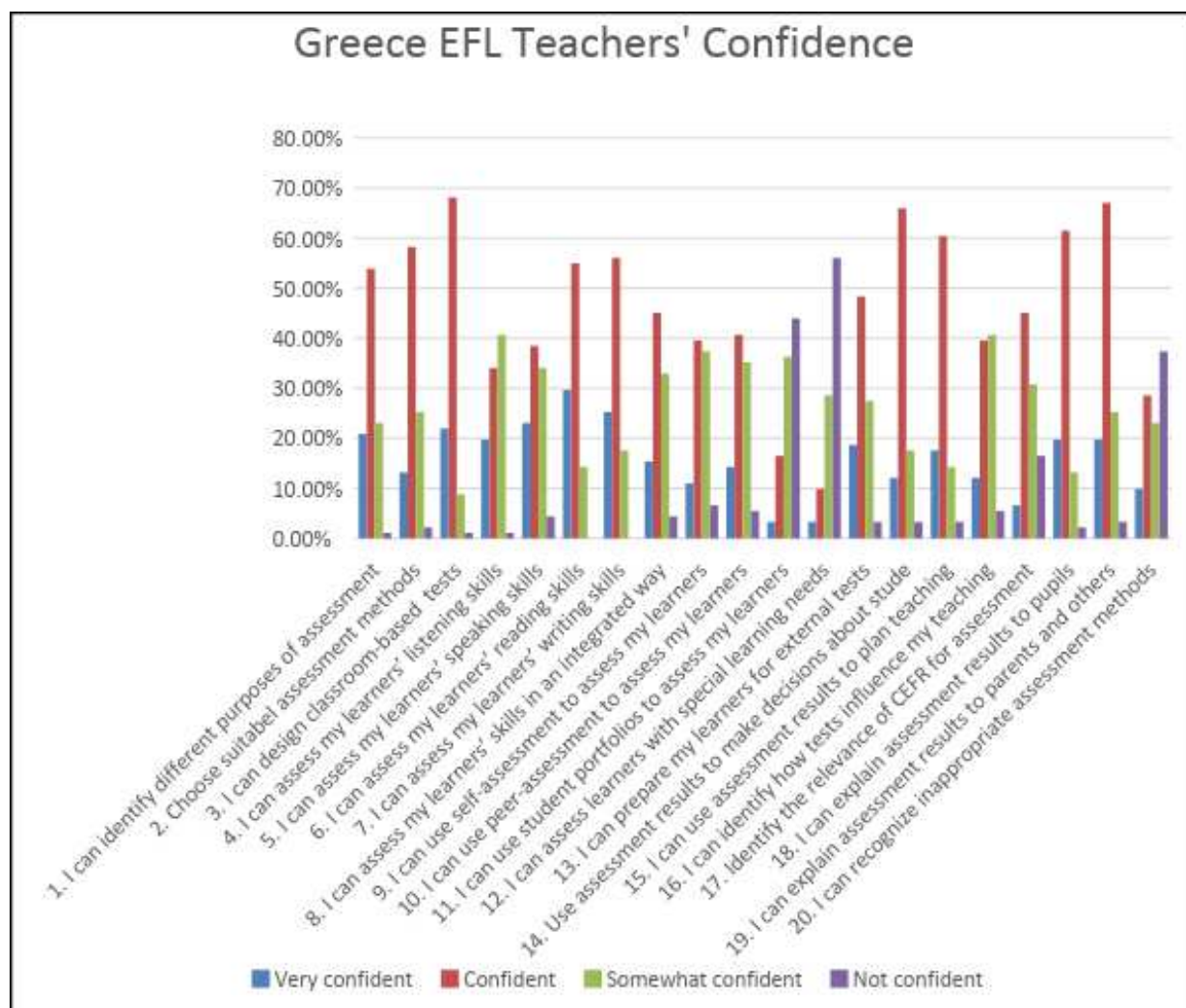
Figure 55: Greece EFL Teachers' Confidence

Figure 55 shows the exact results for each individual area and generally confirms that more Greek teachers seem to feel 'confident', for example compared to the previously presented sub-sample (Germany). Only rarely do Greek teachers state that they do not feel confident in an assessment area (exceptions: portfolios, learners with SPLDs, recognizing inappropriate assessment methods). The three areas in which they did show negative results, were expected as the overall sample indicated very low confidence levels in these particular areas as well.

Figure 56: EFL teachers' training needs Greece

Figure 56 shows the percentage of Greek EFL teachers who specified their training needs in the respective areas. The results conform to the findings from Qu. 12 above. Thus, areas that they expressed the least confidence in (portfolio assessment, assessing learners with SPLDs, recognizing inappropriate assessment methods) were in highest demand for a future training.

IV. Use of technology

In terms of the format of a future training event in an online learning environment on language assessment, five predetermined options were given: 1. Printed self-study materials, 2. Interactive online course, 3. Online resources for self-study, 4. Combination of online self-study and face-to-face course, 5. Other. The aim was to determine the usefulness of each option by once again providing a four-step Likert scale (0 = not useful at all, 1 = less useful, 2 = somewhat useful, 3 = very useful). As in previous sub-samples, it is remarkable that the option 'not useful at all' was only chosen twice in total, which is equivalent to less than 1% of all answers given by EFL teachers in Greece. In fact, less than 5% of Greek EFL teachers regarded the given options 'not useful at all' or 'less useful'. Figure 57

illustrates this result and shows that every given option was regarded 'very useful' by at least 68% of all participating Greek teachers, suggesting that the teachers asked were very open to all kinds of online training formats.
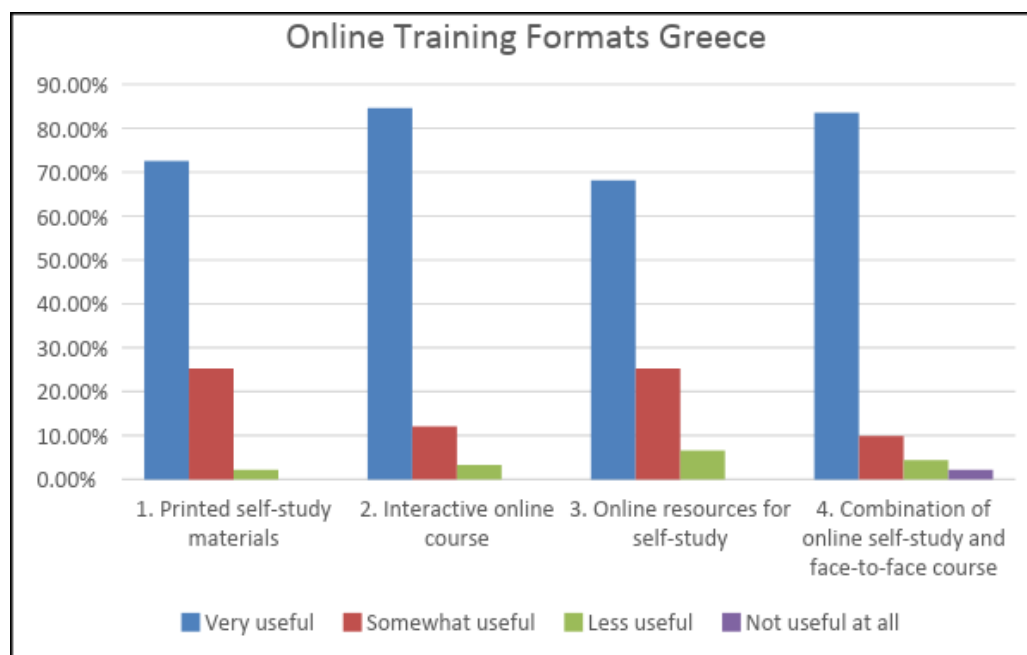
Figure 57: Online training formats Greece

Question 15, which was designed to determine which activity would be regarded useful by the EFL teachers, yielded similar results (cf. figure 58). The item had the same four-point Likert scale (0 = not useful at all, 1 = less useful, 2 = somewhat useful, 3 = very useful).
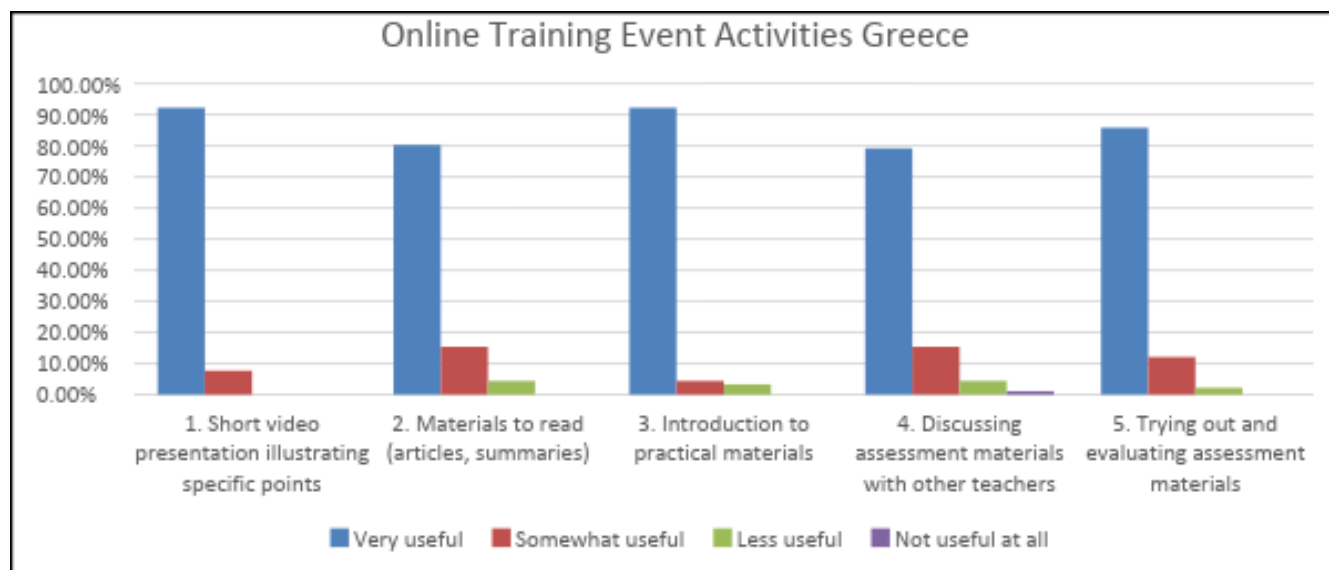
Figure 58: Online training event activities Greece

As indicated above, this result was even more significant. Thus, every given option was regarded 'very useful' by at least 79 % of the Greek EFL teachers. This is one of the most obvious results to Qu. 15 in comparison to the other sub-samples. Once again, this supports the study's claim that there is a significant demand for an online training course on assessment. This can be said even though more than half of the Greek teachers have never participated in any kind of online learning course (cf. figure 59).
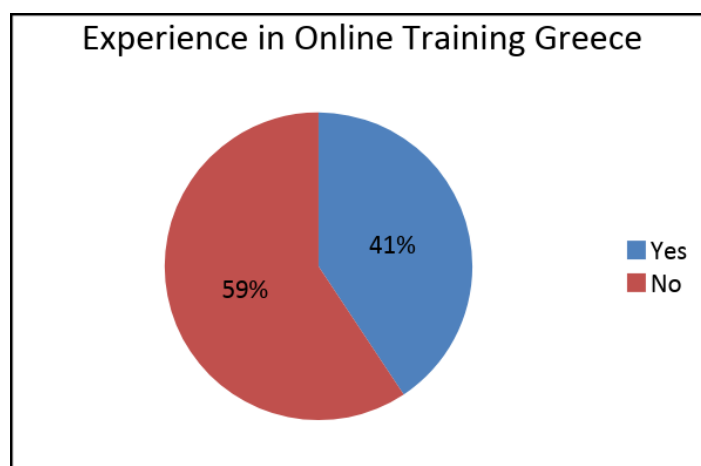
Figure 59: Teachers' Experience in Online Training Greece

4.5 Results – Hungary

I. General information

The Hungarian results comprise the second biggest sub-sample in the study with 230 EFL teachers and 300 learners. Age and teaching experience of the Hungarian teachers were both surprisingly evenly distributed (cf. figures 60 & 61).
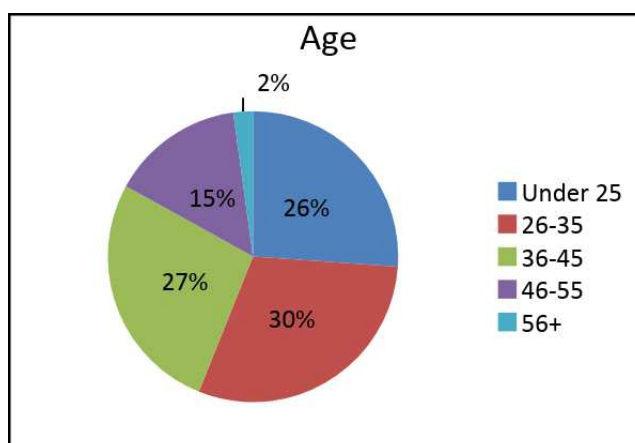
Figure 60: Age of Hungarian EFL Teachers
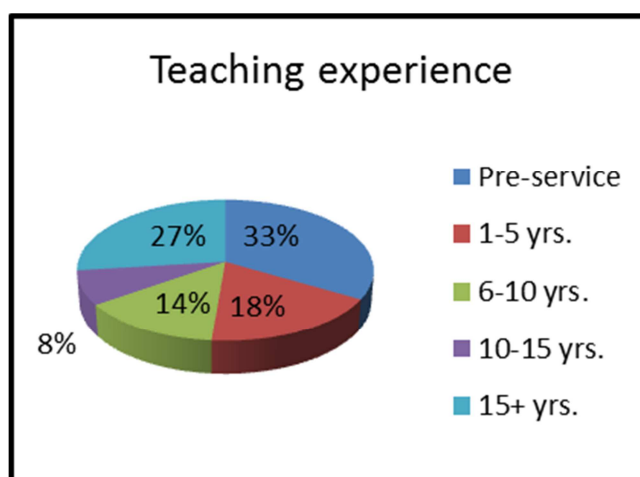
Figure 61: Teaching Experience of Hungarian EFL Teachers

In contrast to the other sub-samples, Hungarian EFL teachers stated not only English as the language they teach (Qu. 5: "What foreign languages do you / will you teach?"). 14.35 % mentioned German as a foreign language as well and another 7 % chose 'other'.

Another noticeable result can be seen for Qu. 6: "What is the highest qualification you hold in the language you teach?" 66 % of all Hungarian teachers stated that they have a MA degree, which is almost twice as many as in the overall sample (37 %). This is due to the fact that teachers in Hungary need an MA degree to be able to teach at state schools.

II. Assessment practices

The second part of the teachers' questionnaire was parallel to the learners' questionnaire. Thus, the perceptions of both stakeholders on assessment practices (assessed skills, type of feedback, frequency of assessment methods) can be compared. Regarding the areas that are assessed by Hungarian EFL teachers, the data is very much in line with the overall results (cf. figure 62).
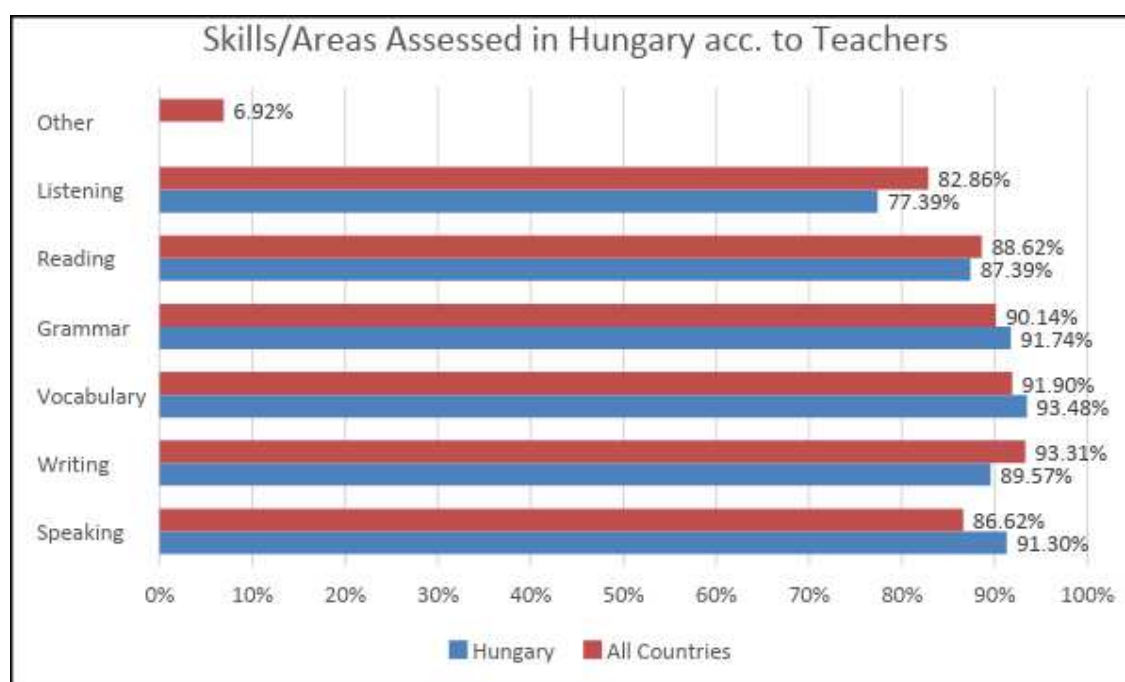
Figure 62: Skills/Areas assessed in Hungary according to teachers

The only difference that can be noticed is that no Hungarian teacher chose 'other'. Especially in contrast to the German sub-sample, where over 21 % of the participating teachers chose 'other',

which indicated that the given skills/areas seemed not to be in accordance with their assessment practices. In Hungary the opposite seems to be true.

The results of question 10: "What feedback do you give on your learners' assessment results?" are rather congruent with the overall results. It can be noticed that both 'detailed comments' and 'comments' seem to be a less preferred form of feedback in Hungary while marks (e.g. letter, percentage) are slightly more preferred in comparison to the overall results. The results of question 11: "How often do you use the following methods to assess your learners' English?" confirm the assumptions made above (cf. figure 63).
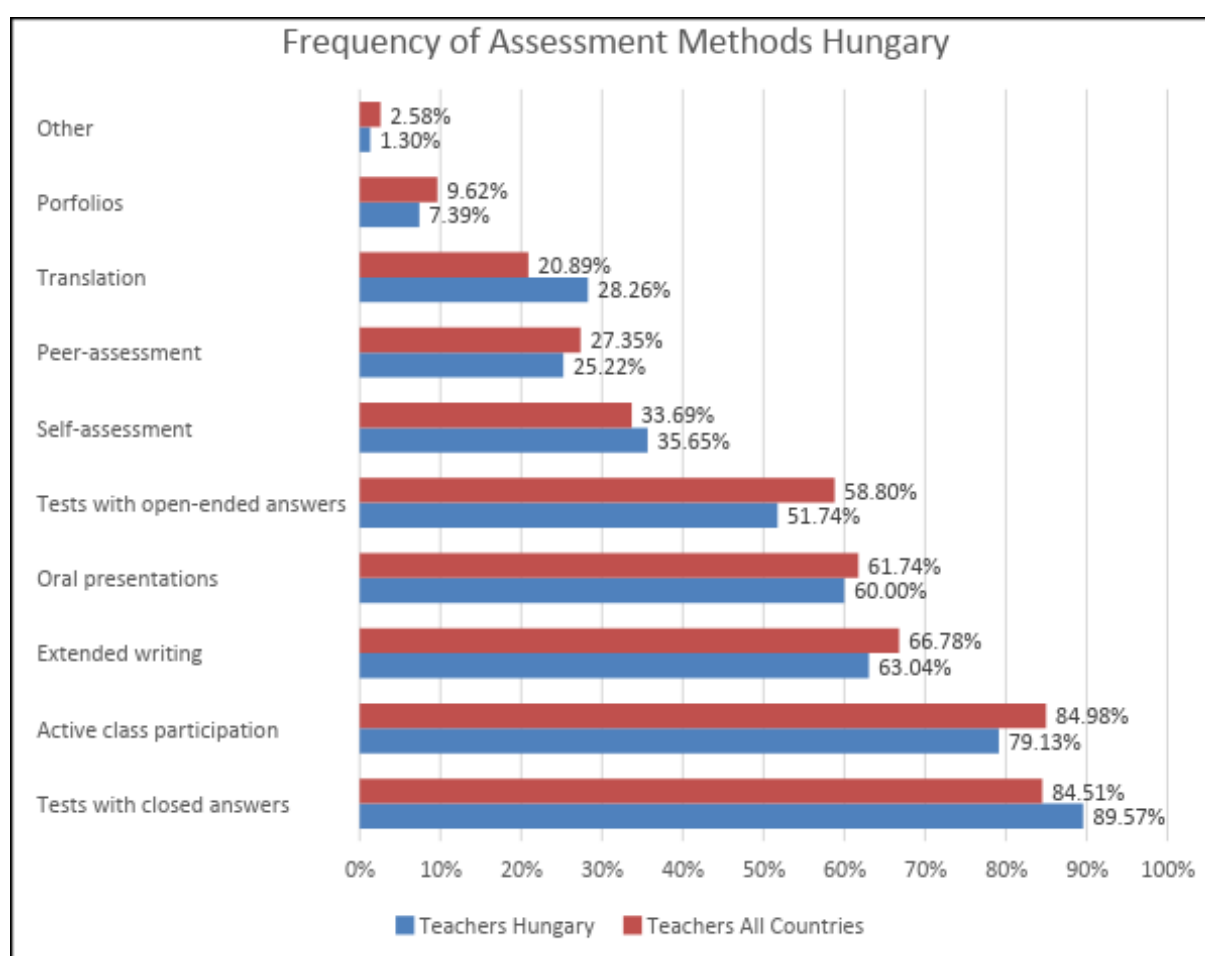
Figure 63: Frequency of assessment methods: Overall results compared to Hungary (teachers)

Figure 63 shows a comparison with the overall data of how many Hungarian EFL teachers have chosen 'very frequently' or 'frequently' with regard to the respective method. 'Tests with open-ended answers' and 'Active class participation' are the only two methods which are used considerably less in Hungary.

In figure 64, the answers 'very frequently' and 'frequently' were combined and the datasets of Hungarian EFL teachers and learners are compared to the overall data.
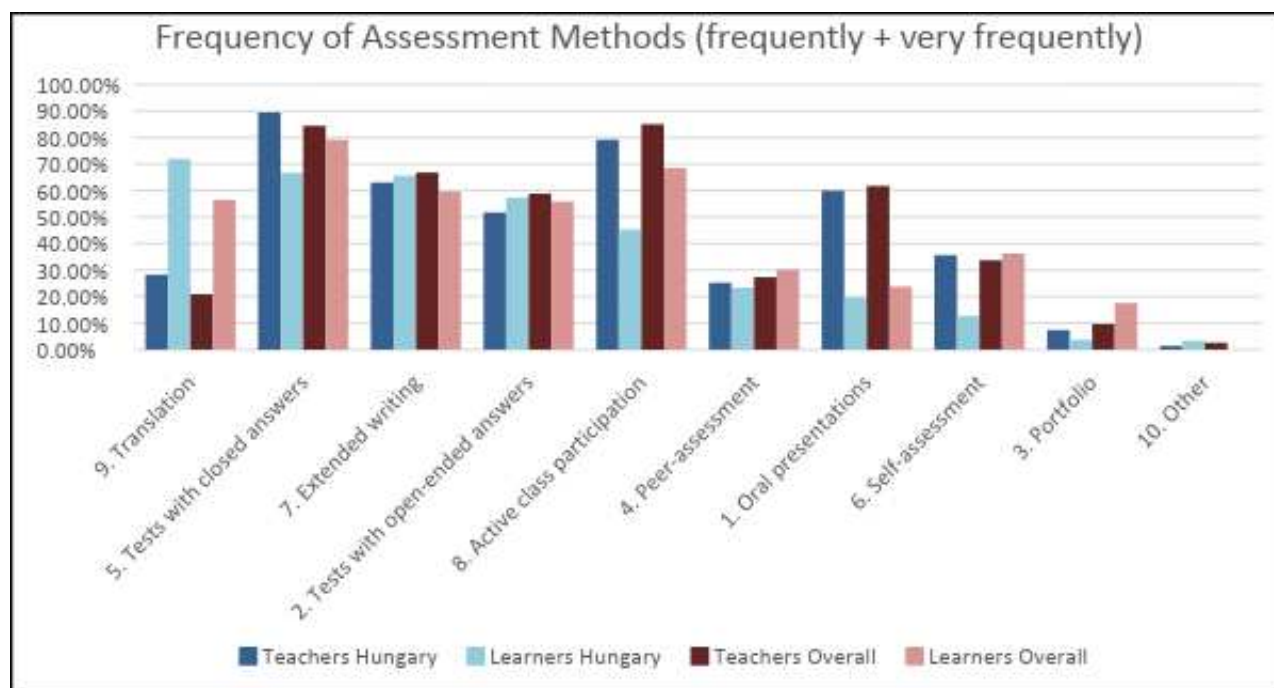
Figure 64: Frequency of assessment methods: Overall results compared to Hungary (teachers + learners)

The Hungarian results generally show the same tendencies as the overall data. A noticeable difference in learners' and teachers' perception, however, can be seen with 5. 'tests with closed answers'. While almost 90 % of the Hungarian EFL teachers use this method 'frequently' or 'very frequently', only 66 % of the learners perceived the same method to be used 'frequently' or 'very frequently' for assessment purposes. This is a discrepancy that was not seen in any other sub-sample or in the overall data. The same kind of disagreement of learners' and teachers' perceptions in Hungary applies to 'self-assessment' as an assessment method. These discrepancies can be identified in figure 65 as well. This is a comparison of the EFL teachers' perception of how frequently a method is used to the EFL learners' perception of how often the respective method is to learn English. It can be seen that 'tests with closed answers' and 'active class participation' are both used frequently by EFL teachers although their learners regard them to be less useful compared to others. This effect goes both ways, as the method 'translation' shows. Almost 80% of the EFL learners stated that this method is 'often' or even 'very often' useful to learn English which implies that it should be used in

the EFL classroom as much as possible. However, only 28% of the Hungarian Teachers use this method 'frequently' or 'very frequently'. As indicated several times above, this result might be attributed to the participants' perception of past events and their recollection. Thus, it could be assumed that some of the discrepancy is due to different ideas of translation as an assessment method.
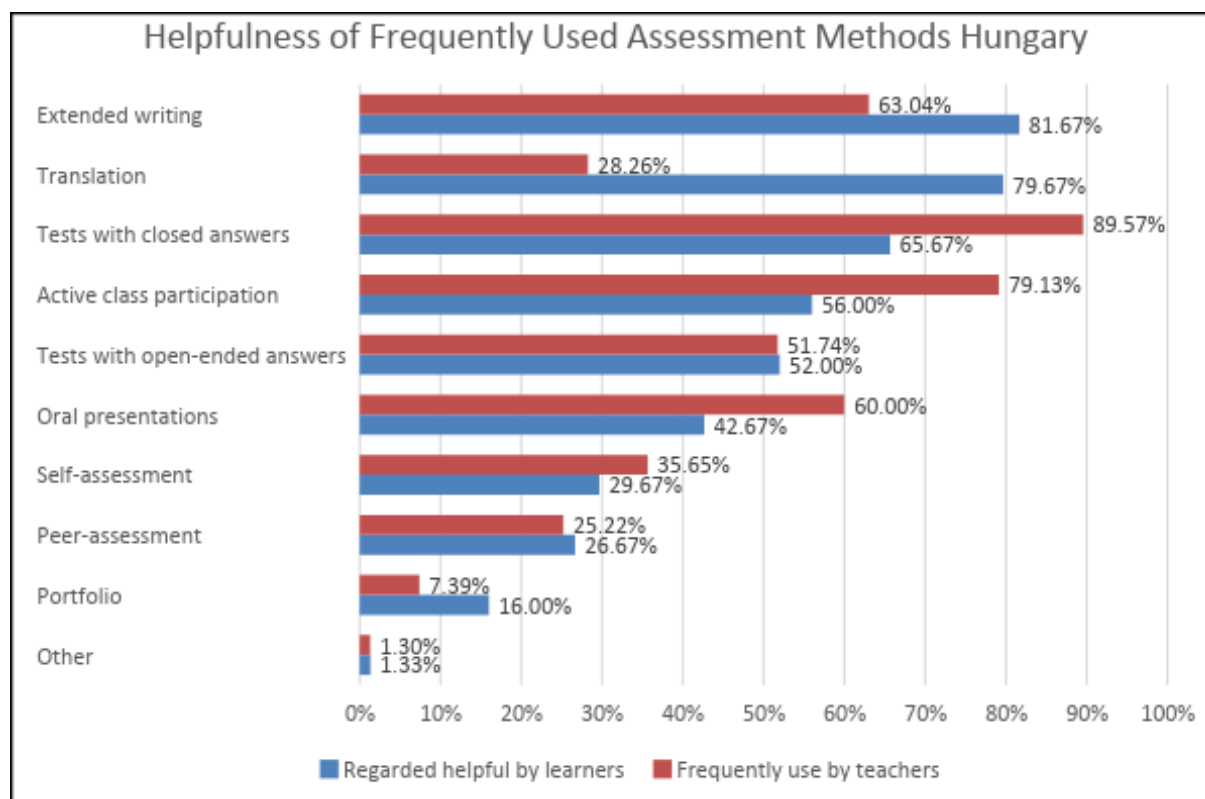
Figure 65: Frequently used assessment methods regarded helpful by learners in Hungary

III. Assessment profiles and training needs

In this part of the questionnaire, the teachers' confidence about 20 predetermined areas connected to assessment was recorded (Qu. 12: "Please indicate how confident you feel about the following areas"). In addition to that, the teachers were also able to indicate if they would like to receive training in each respective area (Qu. 13: "Wherever you feel you need training, mark the last column

on the right with a tick"). Item 12 was designed with a four-point Likert scale (0 = Not confident, 1 = Somewhat confident, 2 = Confident, 3 = Very Confident). Figure 66 shows all answers rated two or higher on the Likert-scale (confident or very confident) of all Hungarian EFL teachers compared to the overall data.
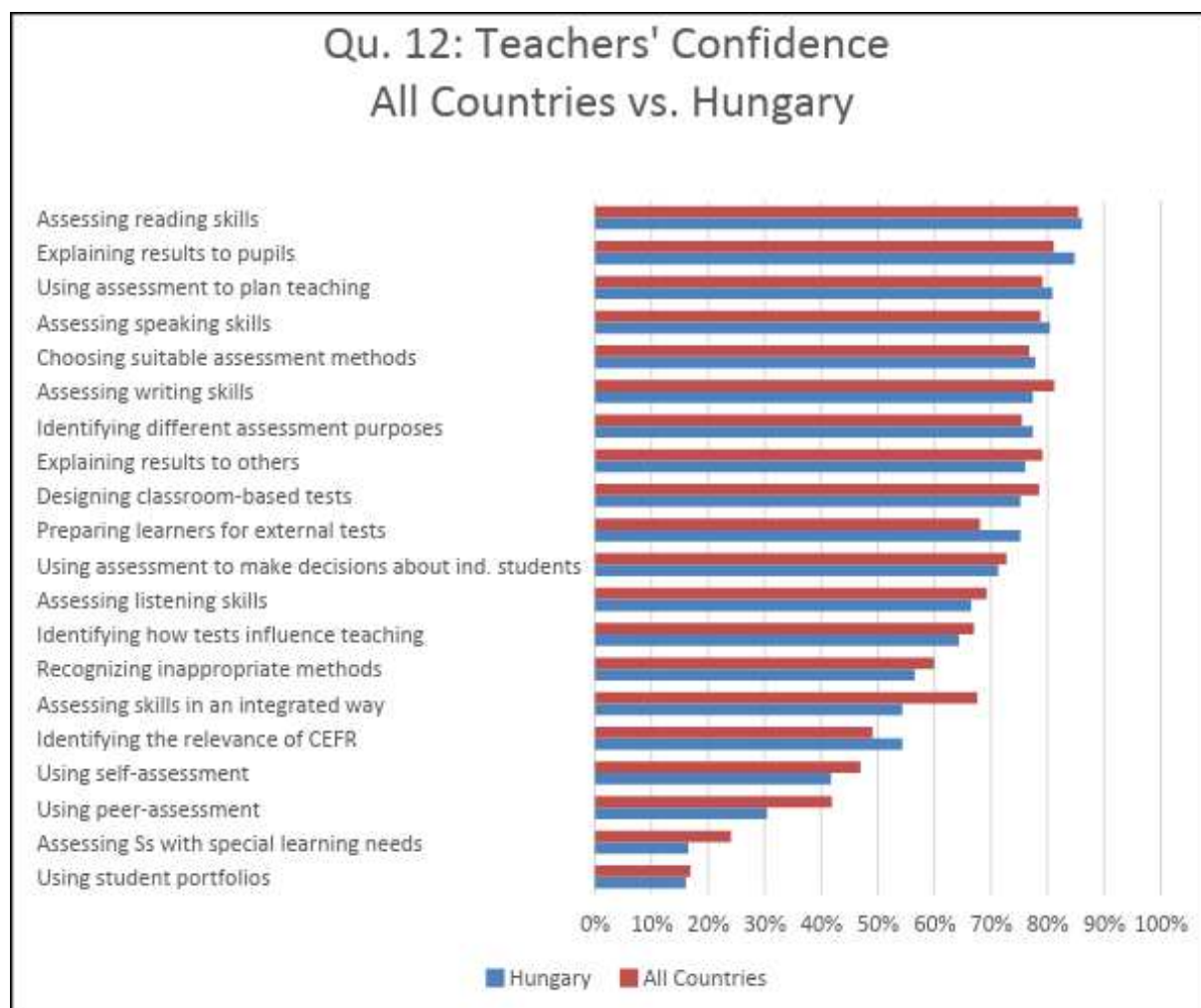
Figure 66: Teachers' confidence: All countries compared to Hungary

Once again, it can be noticed that the areas that Hungarian teachers mention are roughly the same as the teachers in the overall sample put forward. There are only few areas in which the Hungarian data was producing a visible discrepancy. The two most obvious differences can be found in the areas '*assessing skills in an integrated way*' and '*using peer-assessment*'. Both show that fewer of the Hungarian EFL teachers feel confident or very confident in these specific areas.

To give context to the claims made above, all given answers in Hungary (0-3 on the Likert-Scale) area displayed in figure 67.
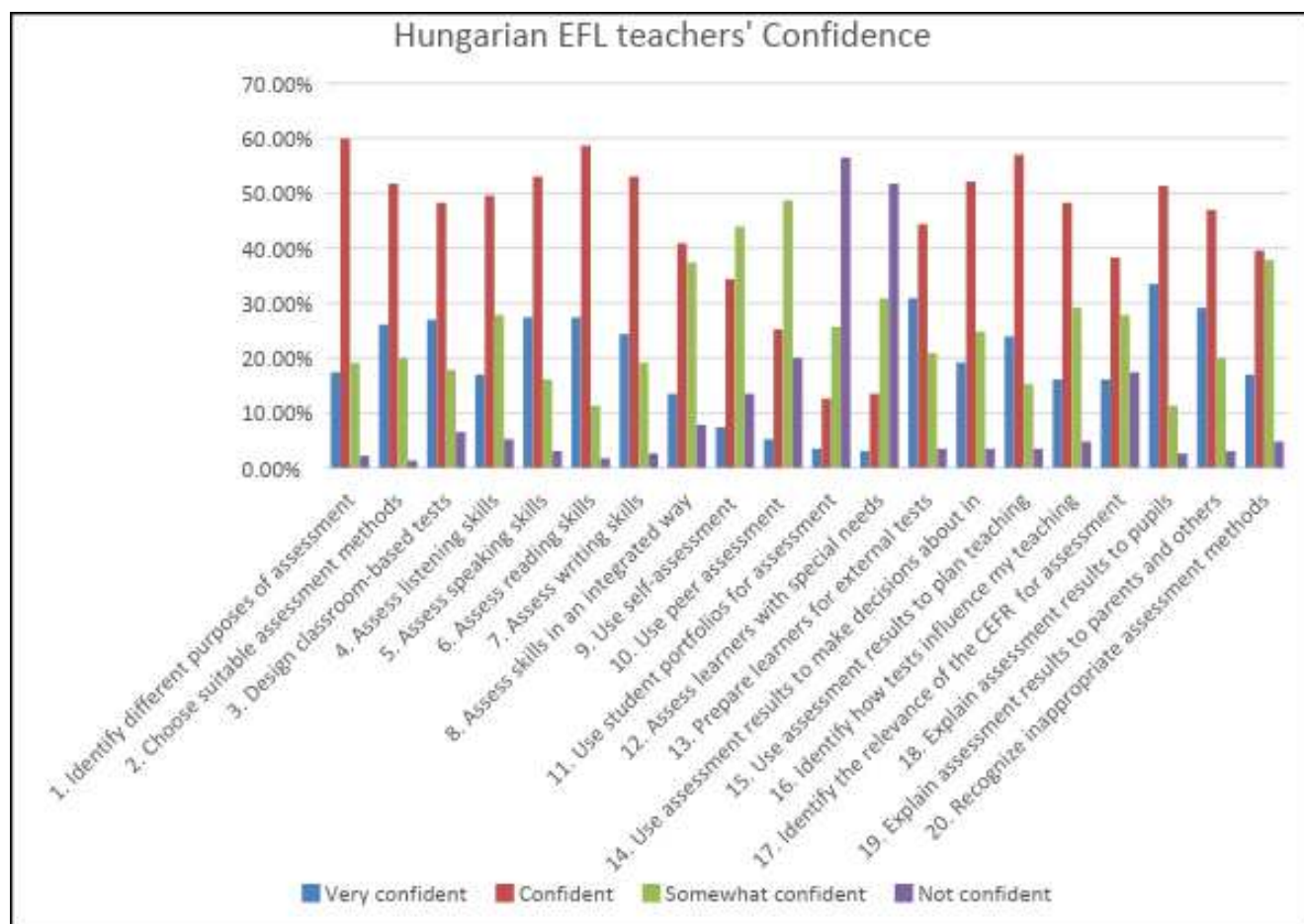
Figure 67: Hungarian EFL teachers' confidence

It can be seen that in most of the areas, the majority of the Hungarian EFL teachers felt at least 'confident', as indicated by the orange bars. However, there is a decline of confidence in integrated skills assessment, self-assessment, peer assessment, portfolio assessment and assessing learners with specific learning difficulties. Parallel to the remaining sub-samples, and thus to the overall data, the areas in which informants considered themselves more confident in areas like assessing skills, preparing learners for tests, explaining the assessment results to learners / parents / others. These are likely to have been included in teacher training for a longer time now. The areas teachers report not to be confident in represent more recent assessment methods or purposes. As already indicated, these are congruent with the overall data, indicating that there are low perceived LAL levels in these areas.

To conclude the third part of the questionnaire, the EFL teachers were prompted to indicate whether they would like to receive training in the respective area. Figure 68 shows how many of the Hungarian EFL teachers would like to receive training to each of the 20 predetermined areas.
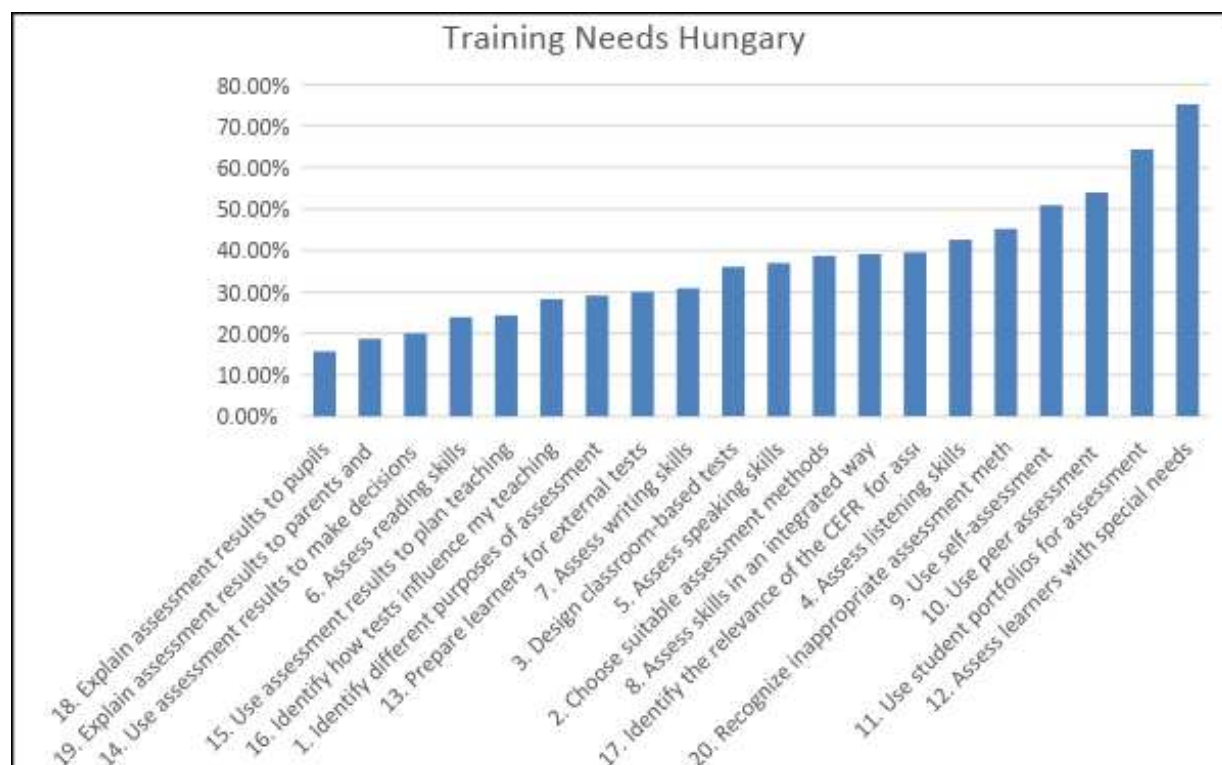
Figure 68: Training needs Hungary

It can be observed that the training needs correspond to the areas that teachers reported not to be confident in. Thus, Hungarian teachers' training needs in LAL are typical for the sample, thus confirming that there is a need for training in these areas across Europe.

IV. Use of technology

In terms of the format of a future training event in an online learning environment on language assessment, five predetermined options were given: 1. Printed self-study materials, 2. Interactive online course, 3. Online resources for self-study, 4. Combination of online self-study and face-to-face course, 5. Other. The aim was to determine the usefulness of each option, by once again providing a four-step Likert scale (0 = not useful at all, 1 = less useful, 2 = somewhat useful, 3 = very useful). As every sub-sample thus far confirmed, there seems to be not only a demand for online training courses (cf. figure 69) but also a general appreciation for possible format. This can be supposed as

only 15 (6.5%) of the 230 participating Hungarian teachers regarded any of the given options as 'not useful at all' (cf. figure 69).
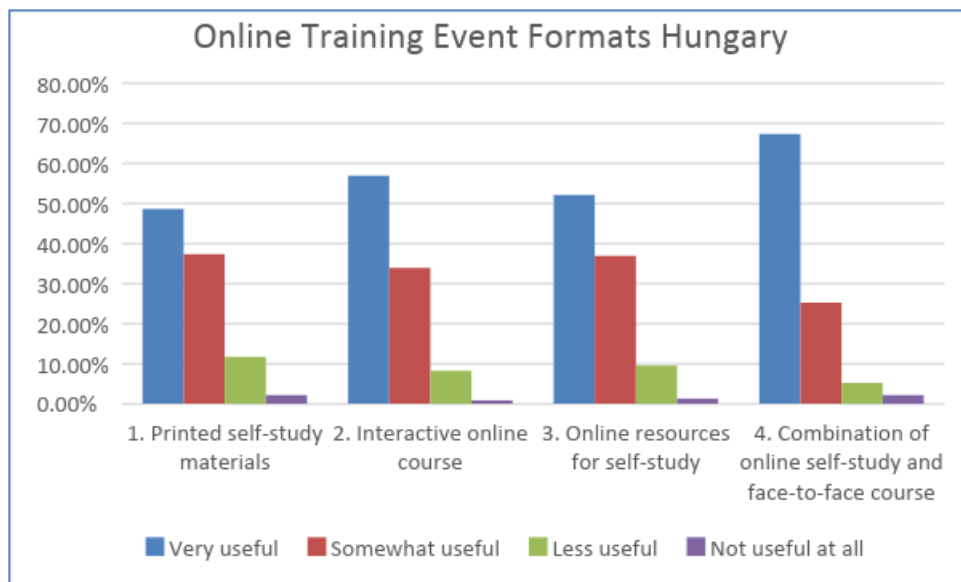
Figure 69: Online training event formats Hungary

In addition to potential training formats, the teachers were also asked to rate the usefulness of six predetermined online activities which could be included in a training course about assessment in an online learning environment (1. Short video presentation illustrating specific points, 2. Materials to read, 3. Introduction to practical materials, 4. Discussing assessment materials with other teachers, 5. Trying out and evaluating assessment materials, 6. Other). Again, a four-point Likert scale was given to indicate the usefulness of each activity (0 = not useful at all, 1 = less useful, 2 = somewhat useful, 3 = very useful). Figure 70 shows a summary of the results to this question.
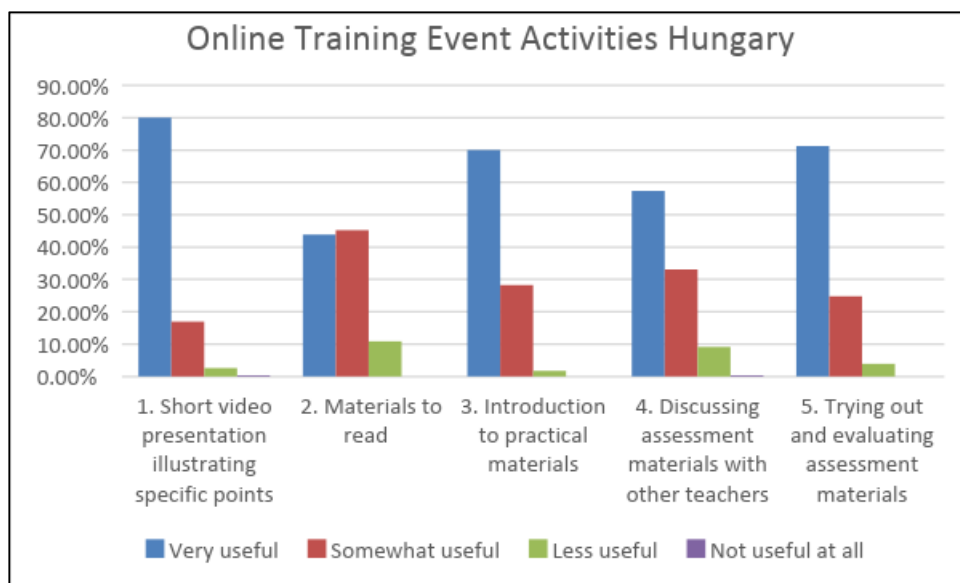
Figure 70: Online training event activities Hungary

The trend suggested that only few teachers regard the formats and activities of an online course on assessment 'not useful at all', on the contrary, most teachers asked are in favour of a digital course format. Less than 1% of all given answers in Hungary indicated any of the activities to be 'not useful at all'. The only activity that is worth mentioning is the second one 'materials to read' which in relation to the remaining ones was seen less useful but still not useless.

Finally, the teachers were asked whether they had ever participated in any kind of online training course, which 75.6% of the Hungarian EFL teachers answered with 'no'. Once again it can be seen that although teachers are not very experienced with online courses, they would appreciate assessment training in an online learning environment in the future.

5. Discussion

The purpose of the present study was to identify currently used assessment practices in English as a Foreign Language (EFL) classrooms across Europe from the perspective of EFL teachers and their learners. Moreover, confidence levels of EFL teachers pertaining to areas of LAL and their corresponding training needs were investigated.

With regard to assessment practices that EFL teachers reported, linguistic skills were all evenly represented in teachers' assessment practices, with learners reporting slightly different skills, namely writing, speaking, grammar and vocabulary. Discrepancies in results of the teacher and the

learner questionnaires concerned perceptions of assessment methods most frequently used in the EFL classroom. Also, a slight variance in the focus of assessment methods used was detected, reflecting the different assessment cultures in the respective educational contexts, a factor that has an impact on teachers' assessment practices (e.g. Fulmer, Lee & Tan, 2015). Another central finding was that feedback in the sense of process-oriented formative assessment tool does not seem to play an important role with teachers. The types of feedback given in the class are mostly marks and brief comments. Data from both teacher and learner questionnaire confirm the lack of importance of detailed comments and hints or feedforward. This seems to be a missed chance to make the most of feedback in the (formative) assessment context (cf. Vogt et al., in preparation).

Learners generally report that the assessment methods used in the foreign language classroom help them learn English. This result needs to be interpreted with caution for two reasons. One, regional variance might play a role. In Germany, for example, vocabulary tests are a very common form of classroom-based test and involve translation. Hence, many learners found translation (in vocabulary tests) not only frequently used but also reported that it helped them learn English. Two, learners cannot report assessment methods to be conducive to their EFL learning if they do not know them. Portfolio assessment, for example, scored low in terms of helpfulness of learning overall with learners but a glance at the low figures for the use of portfolio assessment in the classroom rectifies the picture. Also, checking teacher and learner results against each other has revealed that learners across educational contexts do not seem to recognize certain assessment methods as such. A more in-depth qualitative analysis into this phenomenon would shed more light on the relationship between perceptions and actual practices in the classroom.

In terms of EFL teachers' confidence levels in different areas of LAL, it is striking that no teacher in this relatively large sample that includes various educational contexts reports to be very confident in any LAL areas in the questionnaire. This finding suggests that levels of LAL are not sufficient, as perceived by teachers in our sample. Teachers report to be confident in bread-and-butter areas like written skills. Consequently, teachers express little demand in these areas, again with a slight variance in the different educational contexts (e.g. Greece expressing a demand for training even in well-developed areas). However, teachers report considerably less confidence in areas that pertain to alternatives in assessment (self-assessment, peer assessment, portfolio assessment) that have only recently been embraced in the EFL classroom. Other innovations and developments that have e.g. had an impact on the curriculum that teachers are affected by, such as the CEFR or increasing numbers of learners with specific learning difficulties (SLPDs) equally

represent areas which teachers seem to be less confident in. Consequently, the demand for training in these areas is highest across the board, corroborating previous research e.g. by Kvasova & Kavytska (2014), Vogt & Tsagari (2014).

Although the majority of teachers (74%) in the sample claimed to have no prior experience in online training, teachers were very open towards digital training formats, confirming that they found online training resources in LAL useful or very useful. When asked about their preference for particular formats or modes of delivery, they mentioned videos, online peer discussions and trying out and evaluating materials.

There are some limitations of the study that need to be discussed. As to sampling, there is a strong representation of Cypriot teachers and learners, which account for almost half the sample. It is to be expected that the results of this large sub-group will have influenced overall results. Moreover, the age and experience distributions across the different subsets are not even. In the Cypriot and Greek subsamples, an important percentage of teachers asked reported an age of 46+ years, which again leads us to interpret our data with caution.

By contrast, an overrepresentation of teachers with little experience, might have distorted or affected the results, particularly in the sub-samples. In this connection, it is important to make a distinction within the German sample between pre-service teachers who are still undergoing university training as undergraduate teacher students and teachers who are in their second phase of teacher training, which takes place at school. They work at school and teach regularly, albeit with a reduced teaching load. While these teachers have little teaching experience they would, in other educational contexts, still be counted as in-service teachers. As the questionnaire was anonymous, there is no way of knowing which of the participants in the sub-sample can be counted as "real" pre-service teachers and which would really be practicing teachers. However, despite the dangers that slightly skewed samples like this represent, the overall data have been found to be relatively homogenous.

In the framework of the project, it was only possible to do a questionnaire survey as part of a quantitative study. Questionnaires are a relatively cheap and easy means of surveying large populations (Dörnyei & Czisér, 2012), which the questionnaire was chosen as a data collection instrument. The aim was to survey as many informants as possible in various educational contexts across Europe so as to generate a general picture of EFL teachers' training needs. While the size of the sample allows us to tentatively generalize the results, they do not allow deeper insights into the data. Therefore it would have been desirable to complement the quantitative survey data with

qualitative data e.g. from classroom observations in order to shed some light on the discrepancies in teacher and learner perspectives and to garner empirical data on actual assessment practices in EFL classrooms. Moreover, interviews with matching datasets of teachers and learners would have been insightful to be able to triangulate data, methods and perspectives. Having said that, it was possible to triangulate the perspectives of two important stakeholder groups on the basis of the questionnaire data, and these yielded numerous interesting insights which can, on the basis of the large sample, be tentatively generalized. Also, the present study represents the largest LAL survey in the literature to date that includes EFL teachers **and** learners.

6. Conclusion and implications for the project

In the first phase of the TALE project, the questionnaires administered to 852 EFL teachers and 1788 EFL learners in Cyprus, Germany, Greece and Hungary yielded insights into assessment practices in the EFL classroom from the perspectives of two important groups of stakeholders, learners' perceptions of conduciveness of assessment practices for their EFL learning, EFL teachers' perceived confidence levels in foreign language assessment and teachers' perceived training needs in various areas of language assessment.

The results of the needs analysis were taken into account for the course design. The teacher questionnaire results informed the selection of topics to be dealt in the training resource. The topics would both be relevant to teachers at all career stages and comply with their current training needs.

The groups that the training materials are to target are teachers at all stages of their careers. Hence a flexible, modular system is needed to be appropriated by both pre-service and in-service teachers as well as teacher trainers.

The needs analysis results concerning the delivery mode of the training materials suggest a preference for an online course which should be flexible enough to accommodate and be adapted to different contexts, including use as a stand-alone, self-access online course, a blended learning course or a face-to-face course. The focus, however, is to be on the online component to make the training accessible to a large number of participants.

The modular nature of the course should allow participants to work on parts of the course or study the complete course contents depending on time resources available as well as their training

needs. The aim is to customize the training materials as much as possible to individual needs and wants of the participants.

The course should be relevant for teachers from as many types of schools as possible. As assessment methods in primary schools differ considerably from those used in secondary school contexts, however, a focus will be put on the assessment of secondary EFL learners.

The training materials are supposed to be designed in such a way that teachers can directly apply the procedures and principles of the course contents to their respective teaching environment. Materials have to be hands-on for the training to be worthwhile for busy teachers. The topics and materials were designed in line with the EALTA Guidelines of Good Practice (EALTA, n.d.).

The training materials are supposed to be made available to as many teachers as possible, and well beyond the partner countries. The training materials will therefore be freely accessible on the project website, with the simplest registration procedure possible in order to attract a large number of participants.

7. References

Alderson, J. C. (2005). *Diagnosing Foreign Language Proficiency. The Interface between Learning and Assessment*. London, UK: Continuum.

American Federation of Teachers. National Council on Measurement in Education. and National Education Association (1990). Standards for Teacher Competence in Educational Assessment of Students. *Educational Measurement: Issues and Practice 9*(4), 30-32.

Amirian, S.M.R., Pourfarhad, M., & Nafchi, A.M. (2016). Teachers' Assessment Literacy and Its Correlation with IELTS Students' Achievement in Writing Skill. *Theory and Practice in Language Studies, 6*(5), 994-999. DOI: http://dx.doi.org/10.17507/tpls.0605.12.

Assessment Reform Group (2002). *Beyond the Black Box.* Retrieved on 2nd August, 2017 from http://www.aaia.org.uk/content/uploads/2010/06/ Assessment-for-Learning-Beyond-the-Black-Box.pdf.

Bachman, L. F. (2000). Modern Language Testing at the Turn of the Century: Assuring That What We Count Counts. *Language Testing, 17*(1): 1-42. DOI: 10.1177/026553220001700101.

Bailey, K. M. & Brown, J.D. (1996). Language Testing Courses: What Are They? In: Cumming, A. & R. Berwick (Eds.). *Validation in language testing* (pp. 236-256). Clevedon: Multilingual Matters.

Bailey, K.M. & Nunan, D. (2009). *Exploring Second Language Classroom Research. A Comprehensive Guide*. Boston, MA: Heinle.

Black. P. J.. & Wiliam. D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles. Policy and Practice, 5*(1), 7–74. DOI:10.1080/0969595980050102.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working Inside The Black Box. Assessment for Learning in the Classroom. In: *Phi Delta Kappan*, 9–21.

Brindley. G. (2001). Language Assessment and Professional Development. In: Elder., C., Brown, A., Hill, K., Iwashita, N, Lumley, T., McNamara, T. & O'Loughlin, K. (Eds.). *Experimenting with Uncertainty: Essays in Honour of Alan Davies* (pp. 126–136). Cambridge: Cambridge University Press.

Brown, G.T.L. (2006). Teachers' Conceptions of Assessment: Validation of an Abridged Instrument. *Psychological Reports, 99*, 166-170. DOI: 10.2466/pr0.99.1.166-170.

Brown, G. T. L. & Michaelides, M. (2011). Ecological Rationality in Teachers' Conceptions of Assessment across Samples from Cyprus and New Zealand. *European Journal of Psychological Education, 26*(3), 319-337. ERIC doc: EJ934536.

Brown, G. T. L. (2011). Teachers' conceptions of assessment: Comparing primary and secondary teachers in New Zealand. *Assessment Matters*, 3, 45-70.

Brown, J. D. & Bailey, K.M. (2008). Language testing courses: What are they in 2007? *Language Testing 25*(3), 349-383. DOI: 10.1177/0265532208090157.

Bybee. R. W. (1997). *Achieving Scientific Literacy: From Purposes to Practices*. Portsmouth, NH: Heinemann.

Campbell. C., Murphy. J. A., & Holt. J. K. (2002). *Psychometric Analysis of an Assessment Literacy Instrument: Applicability to Pre-service Teachers*. Paper presented at the meeting of the Mid-Western Educational Research Association. Columbus. OH.

Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and Consequences of School-based Assessment in Hong Kong: Views from Students and their Parents. *Language Testing, 28*(2), 221–50. DOI: 10.1177/0265532210384253

Cohen, L., Manion, L., & Morrison, K. (2011). *Research Methods in Education* (7th ed.). London: Routledge.

Coniam, D. (2009). Investigating the Quality of Teacher-produced Tests for EFL Students and the Impact of Training in Test Development Principles on Improving Test Quality. *System, 37*(2), 226-242. DOI: 10.1016/j.system.2008.11.008.

Council of Europe (2001). *The Common European Framework of Reference – Learning, Teaching, Assessment.* Cambridge: Cambridge University Press.

Cyprus Ministry of Education and Culture. *English as aForeign Language Learning in Cyprus Primary Education*. Edited by the Government of Cyprus. Retrieved on 2nd August, 2017 from

http://neaanalytikaxenonglosson.wikispaces.com/file/view/02%20XENES%20GLOSSES%20G YM-LYK.pdf/181468803/02%20XENES%20GLOSSES%20GYM-LYK.pdf.

Cyprus Ministry of Education and Culture. *English as aForeign Language Learning in Cyprus Secondary and Vocational Education.* Edited by the Government of Cyprus. Retrieved on 2nd August, 2017 from http://poed.com.cy/Portals/0/04%20AGGLIKA %20GIA%20TO%20DIMOTIKOL.pdf.

Cyprus Ministry of Education and Culture. *English as a Foreign Language in Cyprus Primary Schools (Years 1 - 6).* S. Hadjikyriacou, A. Englezaki, S. Ioannou-Georgiou, P. Pavlou, K. Kouppa, M. Xanthou (eds.). Government of Cyprus. Retrieved from http://www.moec.gov.cy/dde/programs/eniaio_oloimero/pdf/analytika_programmata/c urriculum_english.pdf.

Davies. A. (2008). Textbook Trends in Teaching Language Testing. *Language Testing 25*(3), 327-347. DOI: 10.1177/0265532208090156.

De Luca, C. & Klinger, D.A. (2010). Assessment Literacy Development: Identifying Gaps in Teacher Candidates' Learning. *Assessment in Education: Principles, Policy and Practice, 17*(4), 419-438.

Department for Education (2014). Assessment Principles. Government of UK (ed.). Retrieved on 2nd August, 2017 from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/30460 2/Assessment_Principles.pdf.

Department for Education (2014). *National Curriculum and assessment from September 2014: information for schools*. Government of UK (ed.). Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/ file/358070/NC_assessment_quals_factsheet_Sept_update.pdf.

Deutsches Institut für Internationale Pädagogische Forschung (2017). Deutsch-Englisch-Schülerleistungen-International. Retrieved from https://www.dipf.de/ de/forschung/projekte/deutsch-englisch-schuelerleistungen-international.

Dörnyei, Z. & Czisér, K. (2012). How to Design and Analyze Surveys in Second Language Acquisition Research. In: Mackey, A., & Gass, S.M. (Eds.). *Research Methods in Second Language Acquisition: A Practical Guide* (pp. 74-94). Oxford: Blackwell.

Elsner, D. & Viebrock, B. (Eds.) (2014). *Triangulation in der Fremdsprachenforschung*. [Triangulation in Foreign Language Research]. Frankfurt / Main: Peter Lang.

European Association of Language Testing and Assessment (n.d.). *EALTA Guidelines for Good Practice in Language Testing and Assessment.* Retrieved from http://www.ealta.eu.org/documents/archive/guidelines/English.pdf.

Falsgraf. C. (2005. April). *Why a National Assessment Summit? New Visions in Action*. National Assessment Summit. Meeting conducted in Alexandria. VA. http://www.nflrc.iastate.edu/nva/worddocuments/assessment_2005/pdf/nsap_introduction.pdf.

FELVI.HU (2006). *15/2006. (IV. 3.) OM rendelet az alap- és mesterképzési szakok képzési és kimeneti követelményeiről.* Retrieved from https://www.felvi.hu/pub_bin/dload/jogszabalyok/15_2006_alap_mester_kkk_2010053 1.pdf.

FELVI.HU (2013). *8/2013. (I. 30.) EMMI rendelet a tanári felkészítés közös követelményeiről és az egyes tanár szakok képzési és kimeneti követelményeiről*. Retrieved from http://net.jogtar.hu/jr/gen/hjegy_doc.cgi?docid=A1300008. EMM.

Fulcher. G. (2012). Assessment Literacy for the Language Classroom. *Language Assessment Quarterly, 9* (2), 113-132. DOI: 10.1080/15434303.2011.642041.

Fulmer, G.W., Lee, I.C.H., & Tan, K.H.K. (2015). Multi-level Model of Contextual Factors and Teachers' Assessment Practices: An Integrative Review of Research. *Assessment in Education: Principles, Policy and Practice, 22*(4), 475-494. DOI: 10.1080/0969594X.2015.1017445.

Gebril, A. (2016). *Language Teachers' Assessment Beliefs: Voices From The Field*. Paper presented at the 7th CBLA SIG-EALTA conference, 13th EALTA Conference, Universitat Politècnica de Valencia, Valencia, Spain.

Gebril, A. & Brown, G. T.L. (2013). The Effect of High-stakes Examination Systems on Teacher Beliefs: Egyptian Teachers' Conceptions of Assessment. *Assessment in Education: Principles, Policy and Practice*, *21*(1), 16-33. DOI: 10.1080/0969594X.2013.831030.

General Directorate of International and European Educational Affairs Directorate for the Certification of Foreign Language Proficiency (2008). *The Greek State Certificate of Language Proficiency.* Hellenic Republic Ministry of National Education and Religious Affairs (ed.). Retrieved from http://rcel.enl.uoa.gr/rcelold/texts/ KPGdescription-2008.pdf.

Green, A. (2014). *Exploring Language Assessment and Testing. Language in Action.* London et al.: Routledge.

Green, A. (2016). Assessment Literacy for Language Teachers. In: Tsagari, D. (Ed.). *Classroom-based Assessment in L2 Contexts* (pp. 8-29). Cambridge: Cambridge Scholars Publishing.

Hasselgreen. A., Carlsen, C., & Helness, H. (2004). *European Survey of Language Testing and Assessment Needs. Part One: General Findings*. Retrieved from http://www.ealta.eu.org/documents/resources/survey-report-pt1.pdf.

Hellenic Republic Ministry of National Education and Religious Affairs. Government of the Hellenic Republic (ed.). Retrieved from http://rcel.enl.uoa.gr/xenesglosses/sps.htm.

Hidri, S. (2015). Conceptions of Assessment: Investigating What Assessment Means to Secondary and University teachers. *Arab Journal of Applied Linguistics, 1*(1), 19-43.

Hill, K. (2017). Understanding Classroom-based Assessment Practices: A Precondition for Teacher Assessment Literacy. *Papers in Language Testing and Assessment, 6*(1), 1-17.

Inbar-Lourie, O. (2013). Guest Editorial to the Special Issue on Language Assessment Literacy. *Language Testing, 30*(3), 301-307. DOI: 10.1177/0265532213480126

Institut zur Qualitätsentwicklung im Bildungswesen (2016). *VERA - Ein Überblick.* Institut zur Qualitätsentwicklung im Bildungswesen (ed.). Retrieved from https://www.iqb.hu-berlin.de/vera.

Jeong, H. (2013). Defining Assessment Literacy: Is It Different for Language Testers and Non-language Testers? *Language Testing, 30*(3) 345-362. DOI: 10.1177/0265532213480334.

Jin, Y. (2010). The Place of Language Testing and Assessment in the Professional Preparation of Foreign Language Teachers in China. *Language Testing, 27*(4). 555-584. DOI: 10.1177/0265532209351431.

Joint Committee on Standards for Education (JCSEE) (2015). *Classroom Assessment Standards for Pre-K12 Teachers* (71pp.). Kindle edition. Retrieved from http://www.jcsee.org/the-classroom-assessment-standards-new-standards.

Kaftandijeva, F. (2008). *Assessment Literacy in Europe and Beyond: Realities and Prospects.* Presentation at the Fifth Annual Conference of European Association for Language Testing and Assessment, Athens, Greece, 11 May.

Kaiser. G.. & Willander. T. (2005). Development of Mathematical Literacy: Results of an Empirical Study. *Teaching Mathematics and its Applications, 24*(2–3), 48–60. DOI:10.1093/teamat/hri016.

Kelly, G. (1955). *Principles of Personal Construct Psychology*. New York: Norton.

Kultusministerkonferenz (2004). *Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Schulabschluss.* Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (ed.).

Retrieved   from

https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/2003/2003_

12_04-BS-erste-Fremdsprache.pdf.

Kultusministerkonferenz (2015). *Gesamtstrategie der Kultusministerkonferenz zum*

Bildungsmonitoring. Unter Mitarbeit von B. Kurth. Hg. v. Ständige Konferenz der

Kultusminister der Länder in der Bundesrepublik Deutschland. Retrieved from

https://www.kmk.org/fileadmin/Dateien/veroeffentlichungen_beschluesse/

2015/2015_06_11-Gesamtstrategie-Bildungsmonitoring.pdf.

Kvasova. O., & Kavytska. T. (2014). The Assessment Competence of University Foreign

Language Teachers: A Ukrainian Perspective. *CercleS 4*(1), 159–177. DOI:

10.1515/cercles-2014-0010.

Lam, R. (2014). Language Assessment Training in Hong Kong: Implications for Language

Assessment Literacy. *Language Testing 32*(2), 1-29. DOI: 10.1777/0265532214554321.

Leung. C. (2014). Classroom-based Assessment Issues for Language Teacher Education. In:

Kunnan, A.J. (Ed.). *The Companion to Language Assessment*. John Wiley & Sons. Inc. DOI:

10.1002/9781118411360.wbcla064.

Liebers, K.; Seifert, C. (2012). Assessmentkonzepte für die inklusive Schule – eine

Bestandsaufnahme. In: *Zeitschrift für Inklusion - online*. Retrieved from

http://www.inklusion-online.net/index.php/inklusion-online/article/view/44/44.

Livingston, K.; Hutchinson, C. (2017). Developing teachers' capacities in assessment through

career-long professional learning. Practice. In: *Assessment in Education: Principles, Policy*

& Practice 24 (2), S. 290–307. DOI: 10.1080/0969594X.2016.1223016.

Mertler. A. C. (2003). *Pre-service Versus In-service Teachers' Assessment Literacy: Does*

Classroom Experience Make A Difference? Mid-Western Educational Research

Association. Columbus. OH (Oct. 15–18. 2003).

Newfields, T. (2006). Teacher Development and Assessment Literacy. In: *Authentic*

Communication: Proceedings of the 5th Annual JALT Pan-SIG Conference. (pp. 48-73).

Shizuoka, Japan: Tokai University College of Marine Science.

O'Loughlin. K. (2013). Developing the Assessment Literacy of University Proficiency Test Users. *Language Testing 30*(3), 363-380. DOI: 10.1177/0265532213480336.

O'Sullivan. R. G. & Johnson. R. L. (1993). *Using Performance Assessments to Measure Teachers' Competence in Classroom Assessment*. Paper presented at the annual meeting of the American Educational Research Association. Atlanta. GA. ERIC ID: ED358156.

Pellegrino, W J., DiBello, V. L. & Goldman, S.R. (2016). A Framework for Conceptualizing and Evaluating the Validity of Instructionally Relevant Assessments. *Educational Psychologist 51*(1), 59-81. DOI: 10.1080/00461520.2016.1145550.

Pill. J & Harding. L (2013). Defining the Language Assessment Literacy Gap: Evidence from a Parliamentary Inquiry. *Language Testing 30*(3). 381-402. DOI: 10.1177/0265532213480337.

Plake. B. S.. & Impara. J. C. (1992). *Teacher Competences Questionnaire Description*. Lincoln. NE: University of Nebraska.

Plake. B. S., Impara. J. C., & Fager. J. J. (1993). Assessment Competencies of Teachers: A National Survey. *Educational Measurement: Issues and Practice, 12*(4), 10-12. DOI: 10.1111/j.1745-3992.1993.tb00548.x.

Sahinkarakas, S., & Buyukkarci, K. (2011). Teachers' Perceptional Changes on Formative Assessment Practices. In Tsagari, D. & I. Csépes (Eds.). *Classroom-based Language Assessment* (pp. 93-109). Frankfurt /Main: Peter Lang.

Sheen, Y. & Ellis, R. (2011). Corrective Feedback in Language Teaching. In: Hinkel E. (Ed.), *Handbook of Research in Second Teaching and Learning*. Vol. II. (pp. 593-610). New York / London: Routledge.

Stiggins, R.J. (1991). Assessment Literacy. *Phi Delta Kappan, 72*(7), 534-539.

Stiggins, R.J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan*, *83*(10): 758–65.

Taylor. L. (2009). Developing Assessment Literacy. *Annual Review of Applied Linguistics*, *29*, 21–36. DOI:10.1017/S0267190509090035.

Taylor. L. (2013). Communicating the Theory. Practice and Principles of Language Testing to Test Stakeholder: Some Reflections. *Language Testing 30*(3), 403-412. DOI: 10.1177/0265532213480338.

Tsagari, D. (2009). *The Complexity of Test Washback: An Empirical Study*. Frankfurt / Main: Peter Lang.

Tsagari, D. (2012). Classroom Evaluation in EFL State-schools in Greece and Cyprus: Towards 'Assessment Literacy'. *Journal of Applied Linguistics 27*, pp. 127-149.

Tsagari, D. (2013). EFL Students' Perceptions of Assessment in Higher Education. In D. Tsagari, S. Papadima-Sophocleous, & S. Ioannou-Georgiou (Eds.), *International Experiences in Language Testing and Assessment* (pp. 117-143). Frankfurt / Main: Peter Lang.

Tsagari, D. (2016). Assessment orientations of primary state school EFL teachers in two Mediterranean countries. *C·E·P·S Journal – CEPS Journal (Center for Educational Policy Studies Journal*) *6*(1), 9-30. Retrieved from http://www.cepsj.si/doku.php?id=en:cepsj & http://www.cepsj.si/pdfs/cepsj_6_1/cepsj_pp_9-30_Tsagari.pdf.

Tsagari, D. & Vogt, K. (2017). Assessment Literacy of Foreign Language Teachers around Europe: Research, Challenges and Future Prospects. *Papers in Language Testing and Assessment, 6*(1), 41-63. Retrieved from http://www.altaanz.org/uploads/5/9/0/8/5 908292/1.2017_1_front_matter_part1.pdf.

Tsagari, D. & Csépes, I. (Eds.) (2011). *Classroom-based Language Assessment*. Frankfurt / Main et al: Peter Lang.

Vogt, K. (2011). Leistungsmessung und -beurteilung – Kompetenzen europäischer Fremdsprachenlehrkräfte. *Deutsch als Fremdsprache* 48 (4), 205-211.

Vogt. K. & Tsagari. D. (2014). Assessment Literacy of Foreign Language Teachers: Findings of a European Study. *Language Assessment Quarterly, 11*(4), 374-402. DOI:10.1080/15434303.2014.960046.

Appendix 1: ANOVA analysis results for correlations between confidence levels and assessment training

		Sum of Squares	df	Mean Square	F	Sig.
1. I can identify different purposes of assessment.	Between (Combined) Groups	36,621	1	36,621	59,539	,000
	Within Groups	506,818	824	,615		
	Total	543,438	825			
2. I can choose assessment methods that are suitable for my learners	Between (Combined) Groups	27,336	1	27,336	47,730	,000
	Within Groups	472,502	825	,573		
	Total	499,838	826			
3. I can design classroom-based tests	Between (Combined) Groups	23,932	1	23,932	33,876	,000
	Within Groups	582,830	825	,706		
	Total	606,762	826			
4. I can assess my learners' listening skills.	Between (Combined) Groups	24,312	1	24,312	37,177	,000
	Within Groups	537,552	822	,654		
	Total	561,864	823			
5. I can assess my learners' speaking skills.	Between (Combined) Groups	16,838	1	16,838	29,254	,000
	Within Groups	475,418	826	,576		
	Total	492,256	827			
6. I can assess my learners' reading skills.	Between (Combined) Groups	12,529	1	12,529	23,363	,000
	Within Groups	441,350	823	,536		
	Total	453,879	824			
7. I can assess my learners' writing skills.	Between (Combined) Groups	25,924	1	25,924	45,005	,000
	Within Groups	471,767	819	,576		
	Total	497,691	820			
8. I can assess my learners' skills in an integrated way, e.g. reading a text and writing about it.	Between (Combined) Groups	31,324	1	31,324	42,873	,000
	Within Groups	601,301	823	,731		
	Total	632,625	824			

9. I can use self-assessment to assess my learners.	Between (Combined) Groups	18,782	1	18,782	24,402	,000
	Within Groups	633,444	823	,770		
	Total	652,225	824			
10. I can use peer assessment to assess my learners.	Between (Combined) Groups	7,415	1	7,415	9,537	,002
	Within Groups	639,875	823	,777		
	Total	647,290	824			
11. I can use student portfolios to assess my learners.	Between (Combined) Groups	1,949	1	1,949	2,360	,125
	Within Groups	676,222	819	,826		
	Total	678,171	820			
12. I can assess learners with special learning needs, e.g. dyslexia, learning impairment.	Between (Combined) Groups	7,324	1	7,324	10,495	,001
	Within Groups	573,636	822	,698		
	Total	580,960	823			
13. I can prepare my learners for external tests, e.g. school leaving exams, international exams.	Between (Combined) Groups	22,256	1	22,256	24,213	,000
	Within Groups	756,500	823	,919		
	Total	778,756	824			
14. I can use assessment results to make decisions about individual learners.	Between (Combined) Groups	12,460	1	12,460	19,383	,000
	Within Groups	529,062	823	,643		
	Total	541,522	824			
15. I can use assessment results to plan teaching.	Between (Combined) Groups	15,333	1	15,333	28,453	,000
	Within Groups	443,520	823	,539		
	Total	458,853	824			
16. I can identify how tests influence my teaching.	Between (Combined) Groups	30,196	1	30,196	41,643	,000
	Within Groups	596,767	823	,725		
	Total	626,962	824			
17. I can identify the relevance of the Common European Framework of Reference (CEFR) for the assessment of my learners.	Between (Combined) Groups	27,959	1	27,959	33,665	,000
	Within Groups	685,149	825	,830		
	Total	713,108	826			

18. I can explain assessment results to pupils.	Between (Combined) Groups	13,624	1	13,624	23,819	,000
	Within Groups	469,591	821	,572		
	Total	483,215	822			
19. I can explain assessment results to parents and others.	Between (Combined) Groups	15,551	1	15,551	24,821	,000
	Within Groups	515,629	823	,627		
	Total	531,181	824			
20. I can recognize inappropriate (e.g. invalid, unreliable, biased) assessment methods.	Between (Combined) Groups	33,397	1	33,397	44,605	,000
	Within Groups	614,707	821	,749		
	Total	648,104	822			

Appendix 2: Mann-Whitney U Test on frequency of assessment types for teachers and learners.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of 11. How often do you use the following methods to assess your learners' English? [1. Oral presentations] is the same across categories of TorS.	Independent-Samples Mann-Whitney U Test	.000	Reject the null hypothesis.
2	The distribution of 11. How often do you use the following methods to assess your learners' English? [2. Tests with open-ended answers] is the same across categories of TorS.	Independent-Samples Mann-Whitney U Test	.980	Retain the null hypothesis.
3	The distribution of 11. How often do you use the following methods to assess your learners' English? [3. Portfolio assessment] is the same across categories of TorS.	Independent-Samples Mann-Whitney U Test	.086	Retain the null hypothesis.
4	The distribution of 11. How often do you use the following methods to assess your learners' English? [4. Peer assessment] is the same across categories of TorS.	Independent-Samples Mann-Whitney U Test	.731	Retain the null hypothesis.
5	The distribution of 11. How often do you use the following methods to assess your learners' English? [5. Tests with closed answers (e.g. gaps, multiple choice, matching exercises)] is the same across categories of TorS.	Independent-Samples Mann-Whitney U Test	.818	Retain the null hypothesis.
6	The distribution of 11. How often do you use the following methods to assess your learners' English? [6. Self-assessment] is the same across categories of TorS.	Independent-Samples Mann-Whitney U Test	.918	Retain the null hypothesis.
7	The distribution of 11. How often do you use the following methods to assess your learners' English? [7. Extended writing, e.g. letters, essays] is the same across categories of TorS.	Independent-Samples Mann-Whitney U Test	.186	Retain the null hypothesis.
8	The distribution of 11. How often do you use the following methods to assess your learners' English? [8. Active class participation] is the same across categories of TorS.	Independent-Samples Mann-Whitney U Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Appendix 3: Teacher questionnaire

Teacher questionnaire

Dear teacher,
The purpose of the TALE project is to find out about and improve assessment practices in English classrooms across Europe. To this end, a project team from Cyprus, Germany, Greece, Hungary and the United Kingdom would like to identify current assessment practices in classrooms and investigate training needs teachers of English as a foreign language might have. On the basis of established training needs, training measures will be tailor-made to suit teachers' needs. By filling in this questionnaire, you make an important contribution to the project and help improve the outcomes of the project and improve language teacher education in general. Thank you for your honest answers and please be assured that your data will be completely anonymised.
Thank you for your contribution.
Best wishes,
the TALE project team.

I. General information
1. **Gender:**
1. □ male   2. □ female

2. **Age:**
1.□ under 25   2.□ 26-35   3.□ 36-45   4.□ 46-55   5.□ 56+

3. **Years of teaching experience:**
1.□ pre-service teacher
2.□ 1-5
3.□ 6-10
4.□ 10-15
5.□ 15+

4. **I work / study in _____ (country).**

5. **What foreign language(s) do you / will you teach?**
    1. □ English   2. □ French   3. □ German   4. □ Spanish   5. □ Other: _____

6. **What is the highest qualification you hold in the language you teach?**
1.□ Secondary school   2.□ BA degree   3.□ MA degree / state exam
4.□ Other (please specify): _____

7. **Age range of your learners:**
1.□ 6-12   2.□ 13-15   3.□ 16-18   4.□ over 18

8. **Have you received any testing and assessment training?**
1.□ Yes   2. □ No

II. Assessment Practices

9. Which of these skills/areas do you assess? (choose as many as you find appropriate)

1. □ Speaking
2. □ Writing
3. □ Vocabulary
4. □ Grammar
5. □ Reading
6. □ Listening
7. □ Other? Please specify _____

10. What feedback do you give on your learners' assessment results? (more than one answer is possible)

1. □ Mark (percentage, points, letter grade, etc.)
2. □ Brief comments (e.g. 'well done!')
3. □ Detailed comments on learners' work (written / oral)
4. □ Comments/hints on how to improve their learning
5. □ Other, please specify: _____

11. How often do you use the following methods to assess your learners' English (please tick ✓)?

	Very frequently	Frequently	Sometimes	Never
1.Oral presentations				
2.Tests with open-ended answers				
3.Portfolio assessment				
4.Peer assessment				
5.Tests with closed answers (e.g. gaps, multiple choice, matching exercises)				
6.Self-assessment				
7.Extended writing, e.g. letters, essays				
8.Active class participation				
9.Translation (L1/L2)				
10.Other? Please specify: _____				

III. Assessment profiles and training needs

12. FIRST, please indicate how confident you feel about the following areas (please tick ✓)

13. THEN, wherever you feel you need training, mark the last column on the right with a tick ✓:

Question 12					Question 13
	Very Confident	**Confident**	**Somewhat Confident**	**Not Confident**	**I'd like training in this (✓):**
1.I can identify different purposes of assessment.					
2.I can choose assessment methods that are suitable for my learners.					
3.I can design classroom-based tests.					
4.I can assess my learners' listening skills.					
5.I can assess my learners' speaking skills.					
6.I can assess my learners' reading skills.					
7.I can assess my learners' writing skills.					
8.I can assess my learners' skills in an integrated way, e.g. reading a text and writing about it.					
9.I can use self-assessment to assess my learners.					
10.I can use peer assessment to assess my learners.					
11.I can use student portfolios to assess my learners.					
12.I can assess learners with special learning needs, e.g. dyslexia, learning impairment.					
13.I can prepare my learners for external tests, e.g. school leaving exams, international exams.					
14.I can use assessment results to make decisions about individual students.					

15.I can use assessment results to plan teaching.					
16.I can identify how tests influence my teaching.					
17.I can identify the relevance of the Common European Framework of Reference (CEFR) for the assessment of my learners.					
18.I can explain assessment results to pupils.					
19.I can explain assessment results to parents and others.					
20.I can recognize inappropriate (e.g. invalid, unreliable, biased) assessment methods.					

14. The format I prefer for a training event offered in an online learning environment on language assessment is …… (please tick ✓):

	Very useful	Somewhat useful	Less useful	Not useful at all
1.Printed self-study materials				
2.Interactive online course				
3.Online resources for self-study				
4.Combination of online self-study and face-to-face course				
5.Other? Please specify: _____				

15. In a training course about assessment in an online learning environment, I would find the following useful (please tick ✓):

	Very useful	Somewhat useful	Less useful	Not useful at all
1.Short video presentation illustrating specific points				
2.Materials to read (articles, summaries)				
3.Introduction to practical materials				
4.Discussing assessment materials with other teachers				
5.Trying out and evaluating assessment materials				
6.Other? Please specify: _____				

IV. Use of Technology

16. Have you ever participated in any kind of online learning course?
1.□ Yes 2.□ No

17. If yes above, please clarify
_____
_____

Any other comments??
_____
_____
_____
_____

Thank you for your participation!

Appendix 4: Learner questionnaire
<u>Learner Questionnaire</u>

Dear learner of English,
The purpose of the TALE project is to find out about and improve assessment practices in English classrooms across Europe. To this end, a project team from Cyprus, Germany, Greece, Hungary and the United Kingdom would like to find out more about assessment in your classroom and what you think about it. That is why we would like to ask you to fill in the questionnaire below. Thank you for your honest answers. Your name will not appear anywhere and your answers will be anonymised so that no-one will be able to identify you. Thank you again for helping improve assessment in English classrooms.
Best wishes,
the TALE project team

<u>**I. General information**</u>
   1. **I am a**: □ boy □ girl

   2. **Your age**: _____

   3. **How many years have you been learning English**:
□ up to 3 years □ 4-6 years   □ 7 years or more

   4. **I live in:** _____

<u>**II. Assessment Practices**</u>
5. **My teacher(s) assess(es) my English … (Choose as many answers as appropriate)**
1.□ speaking
2.□ writing
3.□ vocabulary
4.□ grammar
5.□ reading
6.□ listening
7.□ Other? Please specify _____

6.   How often does your teacher ask you to …

	Very often	Often	Sometimes	Never
1.Give oral presentations				
2.Take tests with open-ended answers, e.g. "Why did Sam's sister get lost?"				
3.Keep a portfolio				
4.Assess your classmates' work				
5.Take tests with closed answers (e.g. gaps, true/false, choose the correct answer)				
	Very often	Often	Sometimes	Never

6.Assess your own work				
7.Write stories, letters or other texts				
8.Participate actively in class				
9.Translate sentences or texts				
10.Other? Please specify here:				

III. Assessment needs and wants

7. Does it help you learn English when you …

	Very often	Often	Sometimes	Never
1.Give oral presentations				
2.Take tests with open-ended answers, e.g. "Why did Sam's sister get lost?"				
3.Keep a portfolio				
4.Assess your classmates' work				
5.Take tests with closed answers (e.g. gaps, true/false, choose the correct answer)				
6.Assess your own work				
7.Write stories, letters or other texts				
8.Participate actively in class				
9.Translate sentences or texts				
10.Other? Please specify here:				

8. What feedback do you get on your assessment results? (choose as many answers as appropriate)

1. □ mark (percentage, points, letter grade, etc.)
2. □ brief comments (e.g. 'well done!')
3. □ detailed comments on your work (written / oral)
4. □ comments/hints on how to improve your learning
5. □ Other, please specify: _____

Your comments:
_____
_____
_____
_____

Thank you for your participation!