Assessment Literacy of Foreign Language Teachers around Europe: Research, Challenges and Future Prospects

Dina Tsagari
University of Cyprus
Karin Vogt
University of Education Heidelberg

The concept of literacy has expanded significantly in the course of the last decade to include computer literacy or science literacy in academic and public discourse. The recent focus of attention on classroom-based language assessment has brought the aspect of language assessment literacy (LAL) to the forefront, with foreign language teachers as one important group of stakeholders whose professionalization of this concept is important. The current paper aims at relating the findings from part of a large-scale study that was undertaken in seven different European countries in order to shed light on perceived LAL levels of foreign language teachers and their training needs in language testing and assessment. The focus will be on the qualitative part of the mixed methods study that examines foreign language school teachers' perceptions in three different educational contexts, namely Cyprus, Germany and Greece. The paper closes with an outline of the implications for teacher development and directions for future research.

Keywords: Language Assessment Literacy; foreign language teachers; professionalization; classroom-based language assessment; needs analysis

The setting

In the field of Language Testing and Assessment (LTA), the area of Classroom Based Language Assessment (CBLA) has received unprecedented attention with relevant research activity (e.g. Colby-Kelly, 2014, 2015; Hamp-Lyons, 2009; Hill, 2012; Hill & McNamara, 2012; Leung, 2014; Tsagari & Csépes, 2011; Yin, 2010). The scope of the research undertaken in CBLA shows, among other things, that there is a need for constant professional development for language teachers in order to keep up to date with the challenges and expectations that arise in classroom-based foreign language

---

assessment. Due to changes in educational systems worldwide, we are experiencing a radical shift from the traditional role of assessment practices that aim at "comparing students with other students based on achievement to an era when we compare student performance to pre-set standards" (Stiggins, 2006, p. 3). Teachers, being at the forefront of these developments, have to deal with increasing amounts of large-scale testing, e.g. as part of a monitoring system (East, 2015). They are expected to exploit the increased variety of assessment procedures such as alternative forms of assessment that have become part of "mainstream" practices in many educational contexts.

Coombe, Troudi and Al-Hamly et al. (2012) estimate that about 30 to 50 per cent of language teachers' everyday activities are assessment-related. However, evidence from empirical studies (Fulcher, 2012; Lam, 2014; Mendoza & Arandia, 2009; Yan, 2010) suggests a lack of sufficient knowledge and skills on assessment matters, which implies a low level of language assessment literacy (LAL). The present paper discusses the notion of LAL and relates the results of a mixed-methods study on current LAL literacy levels as perceived by pre- and in-service language teachers. The focus is on the findings of the qualitative part of the study undertaken to explore three European contexts in depth, based on findings of a questionnaire (n=853) administered in seven European countries.

Language Assessment Literacy (LAL): definitions and previous research

Given the concept of literacy that has been expanded considerably to other domains, e.g. computer literacy, visual literacy (NCRL/ Metiri Group, 2003; Taylor, 2013), the addition of "assessment" literacy was to be expected (Engelsen & Smith, 2014; Taylor, 2013). The term "assessment literacy" was first coined by Stiggins (1991) and taken up by Falsgraf (2005, p. 6) as "… the ability to understand, analyze and apply information on student performance to improve instruction". Similarly, Pill and Harding (2013, p. 382) maintain that LAL comprises a variety of competencies that enable the individual to "understand, evaluate and create language tests and analyze test data". O'Loughlin (2013, p. 363) equally takes a skills-based perspective that views LAL as "…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". The critical stance of individuals concerned with language assessment is considered in Inbar-Lourie's (2008, p. 389) definition, which highlights "the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, and about what is going to happen on the basis of the test results". The most

comprehensive definition to date has been suggested by Fulcher (2012, p. 125), which is:

> 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.

Finally, our own definition of LAL, used also in this paper, is "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" (Vogt & Tsagari, 2014, p. 377).

LAL as a concept seems to be more closely associated with language testers (Popham, 2009; Malone, 2013) and consequently much of the research that has been carried out is related to this important group of stakeholders. However, research seems to suggest that the levels of teachers' LAL are to be improved. For example, Scarino (2013), who continuously worked with teachers at schools and used 'collaborative dialogues' with teachers as the major source of data, observes that teachers were struggling with the construct of assessment on a theoretical, practical and institutional level. She stressed the need to develop LAL for teachers in in-service training informed by a holistic approach that goes beyond a mere knowledge-based concept of LAL.

Hasselgreen, Carlsen and Helness (2004) focused on previous training in LTA with three types of stakeholders as a key parameter of LAL as well as their perceived training needs in various areas of LTA. Among the stakeholders were teachers, teacher trainers and testing experts (n=914) from Europe and beyond, with the majority of respondents coming from Finland, Sweden and the UK. The results of the online survey revealed that all stakeholders seemed to lack formal education in LTA and expressed a need for training across the board.

In a more recent study that partly replicated the research instruments used by Hasselgreen et al. (2004), Kvasova and Kavytska (2014) targeted Ukrainian university foreign language teachers in order to gauge their LAL needs and to investigate their training needs in the field. Despite good competence levels in assessment-related tasks such as using ready-made tests or providing feedback, Kvasova and Kavytska conclude that "overall assessment literacy has not yet reached an appropriately high level" (2014, p. 175).

Finally, the aim of Fulcher's (2012) study was to identify the training needs of foreign language teachers (n=278) using an online survey. His findings suggest that language teachers are aware of assessment requirements that are not currently supplied in existing training materials and that there is a need for comprehensible and practical materials in order to enhance their assessment literacy.

The aim of the main study was to gauge LAL levels of regular European foreign language teachers and identify training needs to enhance their LAL on a European and local level. The study was a response to the call for more investigation in the training needs of FL teachers round Europe expressed by Hasselgreen et al. (2004) to gain a clearer picture of the assessment literacy of practicing FL educators. Our large-scale, mixed-methods study focused on "regular" teachers (that is, teachers whose only assessment experience was within the classroom) as an important stakeholder group. This is in contrast to other European studies, which have included teacher trainers and testing experts or teachers who also perform assessment roles outside the classroom context (e.g., as item writers or examiners).

The next section will detail the aims and design for the study. The Results section will include a brief overview of the questionnaire findings for the three focus countries followed by a detailed discussion of the results of the interview data, which is the focus of this paper.

Research Focus

As mentioned before, the aim of the main study was to investigate the level of LAL of regular foreign language teachers across Europe. By 'regular' foreign language teachers we understand those who have undergone standard training and who teach foreign languages at state tertiary institutions, colleges and schools, and have no additional assessment roles e.g. working as item writers or as examiners in standardized tests. The aim of the study was to explore levels of training received as well as investigate training needs. It also looked into teachers' compensation strategies in case of low LAL levels. The research questions for the main study were formulated as follows:

- What level of training in areas of LTA do "regular" foreign language teachers report? Do they perceive this training to be sufficient? If not, how do they compensate for it?
- To what extent do foreign language teachers perceive a need for in-service training in the different fields of LTA?

This paper will focus on the qualitative part of the study, which focused on the educational contexts of Cyprus, Germany and Greece. This set out to investigate the following questions:

- What is the nature of assessment activities of EFL teachers in Cyprus, Germany and Greece?
- How well do teachers feel prepared for their assessment activities?
- What are the local language testing and assessment training needs identified by the teachers?

Study design

The main study was a mixed-methods study (Creswell, 2015; Tashakkori & Teddlie, 1998; Teddlie & Tashakkori, 2009) that aimed at triangulating qualitative and quantitative data in order to enhance the validity of the study (Elsner & Viebrock, 2014; Turner, 2014). The questionnaire represented the quantitative part of the study and it was administered to foreign language teachers based in Cyprus, the Former Yugoslavian Republic of Macedonia (FYROM), Germany, Greece, Italy, Poland and Turkey. It was important for us that the informants only served as language teachers and did not have additional assessment-related roles such as item writers. English was the main language that was taught by the informants. The questionnaire was administered to 878 informants, of which 853 (97.2 %) were returned.

The qualitative part of the study, which is the focus of this paper, set out to investigate individual teachers' training biographies, their perceptions of their LAL levels and their individual needs in terms of in-service teacher training in their respective local contexts. To do this, semi-structured interviews were conducted with primary and secondary state school teachers teaching EFL in Cyprus (n=16), Germany (n=25) and Greece (n=22). All informants consented to take part in the study. While the German teachers taught in secondary schools (age 10-19), the Cypriot and Greek sample covered both the primary and secondary sectors of education. Their teaching experience ranged from two to 34 years.

Data Collection and Analysis

The questionnaire used for the main study was based on the one used by Hasselgreen et al. (2004) excluding the questions on large-scale testing, an aspect that teachers are not typically engaged with in their teaching lives (see Appendix 1). The first part of the questionnaire obtained background information about the informants, e.g. teaching qualifications, current institution, age of target learner group, etc. The remainder of the questionnaire was divided into three sections, namely classroom-

focused LTA (n=12 questions), purposes of testing (n=8 questions) and content and concepts of LTA (n=16 questions) (Table 1). Each part was further subdivided into one section enquiring about the training that respondents had received and one relating to the training needs that respondents felt was necessary for them. As in the Hasselgreen et al. (2004) questionnaire, a three-point Likert scale was used for the answers. The questionnaire was found to be highly internally consistent with Cronbach's alpha coefficient ranging from .80 to .93 for the individual scales (see Dörnyei, 2010, p. 94). Descriptive statistics such as frequencies and percentages were calculated for the data.

Table 1. Questionnaire items

| Component | Training (Question No) | | LTA practices |
	Received	Needed	
1.Classroom-focused LTA (6 questions per category)	1.1 a)	1.2 a)	Preparing classroom tests
	1.1 b)	1.2 b)	Ready-made tests from textbook packages
	1.1 c)	1.2 c)	Giving feedback based on assessment
	1.1 d)	1.2 d)	Self- or peer-assessment
	1.1 e)	1.2 e)	Informal assessment
	1.1 f)	1.2 f)	ELP or Portfolio
2.Purposes of testing (4 questions per category)	2.1 a)	2.2 a)	Giving grades
	2.1 b)	2.2 b)	Finding out what needs to be taught/learned
	2.1 c)	2.2 c)	Placing students onto programs etc
	2.1 d)	2.2 d)	Awarding final certificates
3.Content and concepts of LTA (8 questions per category)	3.1.1 a)	3.2.1 a)	Receptive skills (reading/listening)
	3.1.1 b)	3.2.1 b)	Productive skills (speaking/writing)
	3.1.1 c)	3.2.1 c)	Microlinguistic aspects (grammar/vocabulary)
	3.1.1 d)	3.2.1 d)	Integrated language skills
	3.1.1 e)	3.2.1 e)	Aspects of culture
	3.1.2	3.2.2	Reliability
	3.1.3	3.2.3	Validity
	3.1.4	3.2.4	Using statistics

The qualitative data in the study were yielded by 63 semi-structured interviews. Cohen, Manion and Morrison (2011) maintain that interviews allow for great depth in comparison to other research methods. In the present study, the function of the interviews was to help us understand more thoroughly teachers' assessment practices within their educational contexts (in line with Cheng & Wang, 2007). The guiding questions (Table 2) were designed to allow room for variations, additions or clarifications and asked about teachers' institutional background, assessment practices and training.

Table 2. Guiding questions

---

1. During your pre-service teacher training, did you learn about language testing and assessment (LTA)?
2. Did you feel appropriately prepared for your LTA tasks after pre-service training?
3. If not, how did you learn about LTA?
4. Do you know more about recent LTA methods, e.g. portfolio assessment, self- or peer-assessment? Have you ever tried them?
5. Have you ever worked with standardized tests or have you advised learners in this area? What do you think of them?
6. What types of LTA do you use in your school / institute?
7. Have you received in-service training in LTA? If yes, what was the focus of this training?
8. How satisfied are you with in-service training offered in LTA? What LTA training would you like in the short-term?

After seeking the relevant permissions by persons and institutions, the interviews were timetabled and conducted at venues and times convenient for the teachers. Interviews lasted between 30 and 45 minutes. They were audiotaped and transcribed before being sent out to informants for communicative validation (Dörnyei, 2007). Informants were invited to comment on any ambiguous elements in the transcripts, which resulted in editing the transcripts. The final versions were then content analysed. The categories of analysis were both deductive (using the interview questions) and inductive with emergent themes and patterns arising from the data (Aguado, 2014).

The analysis of the data was undertaken separately by the researchers in order to enhance cross-verification of data by way of investigator triangulation. As a last step, researchers collated the group tendencies before comparing the data across the three regional contexts.

Results and Discussion

Since the focus of this paper is on the qualitative part of the study, we will briefly summarise the questionnaire results for the three focus countries before relating and discussing the interview data in greater depth. More detailed discussions of the quantitative part of the study can be found in Vogt and Tsagari (2014).

Looking at the overall trends (see Table 3), we generally observe low LAL levels. On average, the majority of foreign language teachers in our sample either received "no" (35%) or "a little" (32.5%) training. Only certain aspects of LTA were developed, e.g. in testing microlinguistic aspects and language skills. Other aspects of LTA, e.g.

grading or non-traditional assessment methods as well as establishing quality criteria of assessments were not well-developed (see Vogt & Tsagari, 2014).

Table 3. Average trends in teachers' LTA literacy in regional contexts

	Greece	*Germany*	*Cyprus*	*Average*
Training Received				
Not at all	32.0[1]	47.1	25.9	35
A little	30.5	31.7	35.4	32.5
Advanced	22.1	18.4	24.3	21.6
Training Needed				
None	15.7	29.3	26.8	23.9
Basic	23.3	40.1	26.8	30.0
Advanced	39.7	26.3	24.5	30.1

[1]: percent based on data reported in Vogt and Tsagari, 2014

Furthermore, 60.1% on average required "a little" or "advanced" training in LTA with varying topics depending on local educational contexts. In contexts with a strong high-stakes test culture such as Greece, even respondents who seemed to have a solid basis in microlinguistic aspects, or the "four skills", asked for more advanced training. Other, more innovative aspects of assessment such as self-assessment, peer-assessment or the European Language Portfolio, were also identified as urgent areas for professionalization (see Vogt & Tsagari, 2014).

To further investigate individual teachers' LAL profiles and needs in their local contexts, findings from the interview data will be discussed in more detail, following the order of the research questions.

Nature of assessment activities

Assessment procedures

Teachers were asked what assessment procedures they make use of. It is obvious in the data that the regulations of the national or regional educational authorities highly impact on teachers' assessment practices and procedures. Most of the Cypriot and Greek teachers (n=12 and 20 respectively) and all of the German (n=25) teachers said that they used pencil-and-paper tests to assess their learners' foreign language performance. Primary school teachers in Cyprus and Greece, in particular, stressed that they were not obliged to do so but that they did administer classroom tests occasionally since this gives them the opportunity to find out about students' progress. In secondary schools in Germany, pencil-and-paper tests are very much part of the local foreign language assessment culture and the number of tests per semester (usually three to six depending on grade level) is either regulated by the regional school authorities and / or by the schools themselves. Teachers in Germany also

administer these classroom tests on a regular basis, with additional shorter tests, e.g. vocabulary tests. The Greek and Cypriot data indicated that teachers test their students usually every three to four units in the textbook whereas only two teachers (GRT3, GRT4) test their students as soon as they finish a unit. This helps them not only understand if they have taught the unit successfully ("that tells me how to plan the next unit", CYT1), but also identifies any possible problems students face so that remedial work can be planned accordingly.

In the class tests themselves, teachers indicated they used a variety of task types, almost all of which are rather traditional. The most frequent are sentence completion and answering questions for grammar and vocabulary while more open-ended tasks such as essay writing are used for more advanced learners. German teachers uniformly added vocabulary tests, mostly bilingual, i.e. asking for translation of isolated words or sentences into English: "Sure, class tests and vocabulary tests after a unit has been finished" (GERT19). Although most teachers also evaluated their learners' oral performance in some way, the data yielded that there was a clear focus on pencil-and-paper tests and therefore on receptive skills such as reading and sometimes listening and writing as a productive skill. Speaking as a productive skill hardly played any role in classroom tests.

Teachers tended to rely heavily on the information written tests yield about their learners' performance. A Greek informant maintained, "[Tests] tell me almost everything. They clearly show each student's progress" (GRT2) and a very experienced German teacher did not question the use of written tests at all: "In secondary school you have to administer more written class tests" (GERT6). One German respondent estimated that 80% of colleagues at his school relied on written tests for their final evaluation in the school reports (GERT3).

Regarding feedback from test results, teachers provided various types of feedback to students once the class tests were marked. The majority of informants in our sample talked about test results in class while the test papers are being handed back to the learners. However, a deficit-oriented approach seems to prevail in some contexts as the majority also emphasized the mistakes learners made and they identified problematic areas, as one Greek teacher suggested: "I show them the tests and then we have an in-class discussion about how to avoid making the same mistakes" (GRT1). One Cypriot teacher said: "If I see that there is some kind of difficulty, I will indicate that on the test paper" (CYT3). This same teacher, however, also praised positive aspects of the students' performance on the test. Others addressed learners' language problems that are obvious from their test performance by offering remedial work. In other words the test results were fed back into their teaching. All in all, we can conclude that the ways teachers provide feedback are rather individualized.

Marking criteria

With regard to the marking criteria of tests, the majority of informants confirmed that they were strict with the correction of tests ("When I mark, I'm strict with the things I have taught and the objectives of the test tasks", CYT2). All of the Greek and Cypriot teachers except GRT4 indicated that the marking criteria are also transparent to the learners. One Cypriot teacher asserted that his students "know what the test exercises ask from them and the way I mark [sic]" (CYT4). The German teachers, however, were not that positively minded in this respect. As one informant (GERT6) suggested, he lacked knowledge about criteria for classroom tests, having always relied on a gut feeling: "how you can give marks, how to design class tests, how to prepare for class tests ... one has always acted intuitively". Thus basic elements of classroom design and related scoring rubrics did not seem to be that clear to some informants, which is reflected in the in-service teacher training needs regarding language testing and assessment.

Alternative forms of assessment

When asked about the extent to which the respondents used alternative forms of assessment, e.g. portfolio assessment, self- or peer-assessment, they mainly indicated that they were aware of them but that they were not necessarily part of their assessment routines. In Greece, primary school teachers were still far from implementing them with their young learners. For example, GRT2 showed some awareness of other forms of assessment besides tests and she said she used games and various other playful activities. Also, GRT3 said she uses self-assessment and peer-assessment to assess her students but could not explain details. The majority of the Cypriots teachers acknowledged the use of games and activities as assessment methods, e.g.: "If I want to assess vocabulary I will have a Bingo activity where through the use of a dice students will ask and answer or find out things. Through this, for instance, I can check their vocabulary" (CYT3). CYT2 also referred to the use of projects for assessment purposes, e.g.: "We usually have some mini-scale projects. For instance, we work with a small story, 'The Three Little Pigs', and then I prepare a small project for them which I then use as a form of evaluation". CYT2 also used plays performed during the lesson as a way to assess her students' performance: "Some units are ideal for using 'theatre plays' because they are very indicative of what they know. They are more natural, more spontaneous, and as a result you can understand their level and progress". CYT1 said she "checks on students through singing activities" while other Cypriot teachers stressed that they use either pair or group

work as a means of assessment. With regard to portfolio assessment, only CYT4 said that he had partly used portfolio while CYT2 was planning to implement portfolio the following year.

The results suggest that while the CBLA research community has long embraced non-traditional forms of assessment, the reality in frontline foreign language teaching still often looks different. However, some application of innovative assessment forms was reported by teachers in our sample, and their use was often well reflected in an attempt to make Assessment for Learning (Black & Wiliam, 1998) possible, as the following example from Germany suggests: "You gave peer-assessment as an example. Well, I love doing writing conferences with my pupils ... so that pupils learn how to give fair feedback and get fair feedback in the same way ... as a very constructive way of feedback" (GERT6). In the German context, innovative forms of assessment were often implemented by few dedicated teachers or they were used because the school required the teachers to do so. When their first experience with a new form of assessment was negative, they tended to refrain from further use. This result shows that teachers mainly need support with the concrete, practical implementation of more recent assessment procedures in order to help them expand their repertoire of assessment procedures.

Our data showed that once teachers had included alternative forms of assessment in their repertoires, they saw them as valuable additions or even a good alternative to tests, e.g. Cypriot teacher CYT4 who was convinced that they were "better ways because they are more indicative of what teachers can do". Some Greek teachers in our sample were positive: "I think they are equally useful and I wish I had more time to discuss them with the students" (GRT1).

When asked why they used alternative forms of assessment, the informants mentioned various functions, namely making their teaching more effective and adjusting it to the students' needs and helping them overcome language difficulties. Once teachers embraced alternative forms of assessment as a tool for Assessment for Learning, they held them in high esteem because they had come to believe that these methods enhanced both their teaching methods and students' learning. As one Greek teacher suggested: "I take them into account and try to adjust my teaching to my students' needs" (GRT2). Thus, professionalization in LAL also needs to consider teachers' subjective theories on assessment (Scarino, 2013) and encourage them to include the perspective of the learners in their assessment practices.

Standardized tests

To the question of whether teachers have ever worked with standardized tests (e.g. *Cambridge Preliminary English Test, Key English Test or English Young Learners*, etc.) or if

they have advised learners in this area, the majority of the teachers said they have never done so. This happens because there is no mandate to do so in public school education, and this is valid for all three educational contexts. But even if there was, teachers said that they are not qualified or trained to assist students in the preparation for such tests.

In the German context, some schools offer clubs that prepare learners to sit a standardised test, e.g. Cambridge Preliminary English Test, on a voluntary basis. This is offered as an extracurricular activity and is totally separate from regular classes. At one school in our German sample this was done, and the teacher used past papers to prepare learners for the test, albeit without critically evaluating them. This result is not surprising as teachers are not trained in the area of standardised tests and, consequently, are not in a position to critically evaluate them.

Teachers' training levels in LTA

The interview data mostly corroborated the findings from the questionnaire in that there are generally low levels of LAL to be observed. However, one of the purposes of the interview was to look into the perceptions of teachers, or more precisely to investigate whether teachers feel prepared for their everyday assessment-related activities at school.

When asked whether they had learned anything about LTA during their pre-service teacher training, the majority of respondents answered either that they had not learned anything or only very little during the initial stages of their career. Teachers stressed that training in language assessment was neglected during their undergraduate studies and pre-service training, e.g. "... definitely not. We didn't receive any such information" (CYT4), "As far as I can remember, we learned as good as nothing about it" (GERT1). One typical statement that stresses how little our respondents felt prepared for their assessment activities was made by a Cypriot teacher: "I do not feel prepared… There's no preparation in such matters for the primary English school teacher" (CYT2). Voices from the German context confirmed this unease: "... you started your practical phase and you were really clueless ... you are at an utter loss in the beginning, you get some advice from colleagues ..., adopt things from colleagues and then work your way through it" GERT21); "it was in at the deep end ... we all experienced this" (GERT19).

Teachers seem to learn about LTA on the job, they rely on mentors, colleagues and published assessment materials in order to survive. This is a general tendency that seems to be valid across educational contexts, as the statement of a Greek informant (GRT11) suggests: "... to be honest with you, I was heavily dependent on the methods

of evaluation of the Teacher's book". While this is an acceptable survival strategy on the job, it also accounts for a lack of innovation in assessment practices, which also became obvious in the data.

Individualized and local LAL training needs

Teachers were further asked about their in-service teacher training in LTA in order a) to investigate whether they were satisfied with current in-service teacher training provision in the field and b) to identify their individual training needs in their respective contexts. Our hope was to be able to devise effective topics and formats that would enhance teachers' professionalization in terms of LAL. The latter proved to be impossible, however, since the majority of respondents in our sample had not taken part in in-service teacher training in the field (e.g. 72% of the teachers asked in the German sample). They were thus not able to clearly identify training needs. Some went to attend professional development activities whenever they were required by school authorities to use a certain assessment format or scoring rubrics, e.g. in school-leaving exams. The general feeling was that work needs to be done in LTA because teachers lack the appropriate knowledge and they need to improve their overall competence in LTA practices, skills and knowledge.

The informants were also asked about their personal professionalization needs. Again, wishes that were expressed were vague most of the time. One of the teachers from Cyprus commented: "I believe that there should be better professional training and orientation in language assessment because we may actually use some individual and group assessment but this is not enough…" (CYT3). When informants did identify concrete training needs, which happened only in the German sample, they were rather general ("concrete things for everyday assessment", GERT24; "how marks come about", GERT5), which shows that they lacked a systematic overview of the field or that they did not see the need to take action. Among the concrete needs expressed were two discernible trends, namely oral assessment practices and criterion-oriented assessment.

Summary of results and implications for teacher training

The present study explored European foreign language teachers' perceived LAL levels generally and in their local educational contexts as well as their expressed training needs in the field. The results of both the questionnaire and interview data indicate that the informants' perceived LAL profiles are not sufficient for the assessment activities they have to accomplish in their professional field. This main result corroborates findings from other studies such as Boraie (2015), Cheng, Rogers

and Hu (2004), Hasselgreen et al. (2004), Inbar-Lourie & Levi (2015), Lam (2014), Kvasova and Kavytska (2014) and Xu (2015). The interview data additionally suggests that teachers in our sample do not feel effectively prepared, which adds a psychological component that might have an impact on teachers' job satisfaction as well, supporting Scarino's (2013) call for a holistic approach to LAL.

All sources of data clearly showed that teacher education programs do not provide adequate training in LTA (albeit with individual exceptions), which drives teachers to use compensation strategies such as reliance on published assessment materials or the uncontested adoption of mentors' or colleagues' assessment practices. Such a tendency to "test as you were tested" (in analogy to "teach as you were taught") implies the danger of innovations in assessment not being implemented in practice due to a lack of assessment routines that are firmly based on published knowledge. Generally, approaches like learning on the job, which was displayed by many informants, inhibits effective teacher development.

The interview data yielded individual insights into perceptions and sensitivities of teaching professionals in their local contexts. Summarizing the findings from the interviews, one can see that respondents in our study tend to revert to traditional assessment procedures that are essentially written and typically use similar assessment formats. Opinions are divided on the transparency of marking criteria for learners, ranging from rather positive statements to raised doubts and uncertainty caused by a lack of LAL in this area. Feedback procedures seem to reflect a deficit-oriented approach rather than the more positively worded feedback that is inherent in the Common European Framework of Reference and its descriptors that value competencies even on low levels instead of highlighting deficits (Vogt, 2004). Alternative forms of assessment have not yet entered mainstream assessment practices although once teachers in our sample have tried them out successfully, they realise their potential for Assessment for Learning. There is no experience with standardized tests and teachers are not in a position to evaluate them critically. Concepts related to LTA remain fuzzy to a substantial number of respondents in our study, which is attributable to low LAL levels. Consequently, teachers have difficulties to specify their personal professional development needs although teachers in some contexts do identify urgent areas for training.

Of course the limited number of interviewees forbids generalizations and represents a limitation of the study. An additional way of garnering valuable information on teachers' LAL would consist of actual research in classrooms in order to gain an additional data source and obtain more detailed insights into the nature and impact of language assessment practices on language learning (Cheng & Wang, 2007; Tsagari, 2016; 2014).

The evidence from the existing data in the study showed, however, that foreign language teachers are seriously thinking about their place within language assessment, and are ready for greater levels of involvement in training initiatives to broaden and diversify their assessment literacy with varying priorities depending on contextual assessment requirements. Teachers have pointed out repeatedly that their assessment competency has not yet reached a level that would allow them to feel confident enough about their language assessment activities. Rather, they realize that professional training is required and this aspect of their teaching is one that definitely needs improvement. What will benefit teachers is professional development in this area.

Attending to teachers' professional knowledge and practice in language assessment will contribute towards the development of a dynamic and contextually sensitive assessment literacy culture in EFL education. The challenge undoubtedly lies in providing appropriate and available professional development opportunities for teachers to meet their assessment needs. An assessment literacy development strategy could, for example, rely on a combination of the following in varying proportions. For example, formal language assessment courses (BA and MA level) can expose teachers not only to new ideas but help them to meet their professional assessment needs and responsibilities although this suggestion might not be feasible on a larger scale. Attendance of pre- and in-service language assessment workshops (of appropriate length) is equally important. However, such courses and workshops also need to capitalise on teachers' existing experience and practices (Fulcher 2012) and take into account results of assessment needs analysis such as the ones reported in this study. They also need to recognise and deal with the reality and constraints influencing teachers' assessment practices and encourage an action-research orientation to professional development which will combine theories with practice in the classroom (cf DeLuca, Klinger, Pyper & Woods, 2015; Jonsson, Lundahl & Holmgren, 2015). Training courses should also involve policy, decision makers, and teachers in collaborative assessment development projects (Stiggins, 1999a; b) and allow students to be involved in assessment to build their confidence and maximize their achievement (Stiggins, 2001). Otherwise, such training endeavours might fall short of meeting the professional development standards of EFL teachers (see also Harding & Kremmel, 2016; Hill, 2015; Taylor, 2009), and it is these that we would like to see improved for the sake of foreign language teaching professionals and their learners.

References

Aguado, K. (2014). Triangulation: Möglichkeiten, Grenzen, Desiderate. [Triangulation: possibilities, limitations, desiderata.] In D. Elsner & B. Viebrock

(Eds.), *Triangulation in der Fremdsprachenforschung.* [Triangulation in Foreign Language Research]. (pp. 203-219). Frankfurt / Main: Peter Lang.

Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–71.

Boraie, D. (2015). *The Construct of Language Assessment Literacy as Perceived by Foreign Language Teachers in a Specific Context*. Paper presented at the 37th Language Testing and Research Colloquium. Toronto/Canada, 16 to 20 March 2015.

Cheng, L., Rogers, T. & Hu, H. (2004). ESL/EFL Instructors' Classroom Assessment Practices: Purposes, Methods, and Procedures. *Language Testing 21*(3), 360-389. DOI: 10.1191/0265532204lt288oa

Cheng, L. & Wang, X. (2007). Grading, Feedback, and Reporting in ESL/EFL Classrooms. *Language Assessment Quarterly*, *4*(1), 85-107. DOI:10.1080/15434300701348409

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

Colby-Kelly, C. (2015). *Assessing the Unobservable: Can AFL Inform on Nascent Learning in an L2 Classroom Setting*? Paper presented at the 37th Language Testing and Research Colloquium. Toronto/Canada, 16 to 20 March 2015.

Colby-Kelly, C. (2014). Assessment for Learning in Theory and Practice in L2 Classrooms. *The Encyclopedia of Applied Linguistics* 1–7. DOI: 10.1002/97814051983.1wbeal1439.

Coombe, C., S. Troudi & Al-Hamly, M. (2012). Foreign and Second Language Teacher Assessment Literacy: Issues, Challenges, and Recommendations. In C. Coombe, P. Davidson, B. O'Sullivan & S. Stoynoff (Eds.), *The Cambridge Guide to Second Language Assessment* (pp. 20-29). Cambridge: Cambridge University Press.

Creswell, J. W. (2015). *A Concise Introduction to Mixed Methods Research*. Los Angeles. CA: Sage.

DeLuca, C., D. Klinger, J. Pyper & Woods, J. (2015). Instructional Rounds as a Professional Learning Model for Systemic Implementation of Assessment for Learning. *Assessment in Education: Principles, Policy & Practice 22*(1), 122-139, DOI:10.1080/0969594X.2014.967168

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Dörnyei, Z. (2010). *Questionnaires in second language research: Construction, administration, and processing* (2nd ed.). New York, NY: Routledge.

East, M. (2015). Coming to Terms with Innovative High-Stakes Assessment Practice: Teachers' Viewpoints on Assessment Reform. *Language Testing 32*(1), 101-120. DOI: 10.1177/0265532214544393.

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

Engelsen, K. S. and Smith, K. (2014). Assessment Literacy. In C. Wyatt-Smith et al. (eds.), *Designing Assessment for Quality Learning, The Enabling Power of Assessment*, 1: 91-107, DOI 10.1007/978-94-007-5902-2_6

Falsgraf, C. (2005). *Why a National Assessment Summit? New Visions in Action.* National Assessment Summit. Meeting Conductedin Alexandria, Va. Retrieved April 11, 2017 from http://www.nflrc.iastate.edu/nva/ worddocuments/assessment_2005/pdf/nsap_introduction.pdf.

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

Hamp-Lyons, L. (2009). Principles for Large-Scale Classroom-Based Teacher Assessment of English Learners' Language: An Initial Framework from School-Based Assessment in Hong Kong. *TESOL Quarterly 43*(3), 524-529. DOI: 10.1002/j.1545-7249.2009.tb00249.x

Harding, L. & Kremmel, B. (2016). Language Assessment Literacy and Professional Development. In D. Tsagari and J. Banerjee (Eds.), *Handbook of Second Language Assessment* (pp. 413-428). Berlin & New York: Mouton de Gruyter.

Hasselgreen, A., Carlsen, C. & Helness, H. (2004). *European Survey of Language and Assessment Needs. Part One: General Findings*. Retrieved April 11, 2017 from http://www.ealta.eu/documents/resources/survey-report-pt1.pdf.

Hill, K. (2015). *The Implications of LOLA for Teacher Professional Development*. Paper presented at the 37th Language Testing and Research Colloquium. Toronto/Canada, 16 to 20 March 2015.

Hill, K. & McNamara, T. (2012). Developing a Comprehensive, Empirically Based Research Framework for Classroom-based Assessment. *Language Testing 29*(3), 395-420. DOI: 10.1177/0265532211428317

Hill, K. (2012). *Classroom-based Assessment in the School Foreign Language Classroom*. Frankfurt / Main et al: Peter Lang.

Inbar-Lourie, O. and Levi, T. (2015). *Implementing Formative Classroom Assessment Initiatives: What Language Assessment Literacy Knowledge is Required?* Paper presented at the 37th Language Testing and Research Colloquium. Toronto/Canada, 16 to 20 March 2015.

Inbar-Lourie, O. (2008). Constructing a Language Assessment Knowledge Base: A Focus on Language Assessment Courses. *Language Testing 25*(3), 385-402. DOI: 10.1177/0265532208090158.

Jonsson, A., Lundahl, C. & Holmgren, A. (2015). Evaluating a large-scale implementation of Assessment for Learning in Sweden. *Assessment in Education: Principles, Policy & Practice 22*(1), 104-121, DOI: 10.1080/0969594X.2014.970612

Kvasova, O. & Kavytska, T. (2014). The Assessment Competence of University Foreign Language Teachers: A Ukranian 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*. Published online. DOI: 10.1177/0265532214554321

Leung, C. (2014). Classroom-based Assessment Issues for Language Teacher Education. In A.J. Kunnan (Ed.), *The Companion to Language Assessment*. (Vol. III, pp. 1510-1519). Chichester: Wiley Blackwell.

Malone, M.E. (2013). The Essentials of Assessment Literacy: Contrasts between Testers and Users. *Language Testing 30*(3), 329-344. DOI: 10.1177/0265532213480129.

Mendoza, López, A.A. & Arandia Bernal, R. (2009). Language Testing in Colombia: A Call for More Teacher Education and Teacher Training in Language Assessment. *PROFILE* 11 (2), 55-70.

North Central Regional Educational Laboratory / Metiri Group (2003). *Literacy in the Digital Age*. Napierville (IL). Retrieved April 11, 2017 from http://pict.sdsu.edu/engauge21st.pdf.

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

Popham, W. J. (2009). Assessment Literacy for Teachers: Faddish or Fundamental? *Theory into Practice 48*(1), 4-11. ERIC: EJ823941.

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

Scarino, A. (2013). Language Assessment Literacy as Self-awareness: Understanding the Role of Interpretation in Assessment and in Teacher Learning. *Language Testing 30*(3), 309-327. DOI: 10.1177/0265532213480128.

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

Stiggins, R. J. (1999a). Learning teams can help educators develop their collective assessment literacy. *Journal of Staff Development*, 20/3. Retrieved April 11, 2017 from http://www.nsdc.org/library/publications/jsd/stiggins203.cfm.

Stiggins, R.J. (1999b). Teams. *Journal of Staff Development*, 20/3. Retrieved April 11, 2017 from http://www.nsdc.org/library/publications/jsd/stiggins203.cfm.

Stiggins, R.J. (2001). *Student-Involved Classroom Assessment* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Stiggins, R.J. (2006). Assessment for Learning: A Key to Motivation and Achievement. *Edge 2*(2), 3-19.

Tashakkori, A. & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks, Ca: Sage.

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 Stakeholders: Some Reflections. *Language Testing 30*(3), 403-412. DOI: 10.1177/0265532213480338.

Teddlie, C. & Tashakkori, A. (2009). *Foundations of Mixed Methods Research: Integrating Quantitative and Qualitative Approaches in the Social and Behavioral Sciences*. Los Angeles, CA: Sage.

Tsagari, D. (2016). 'Assessment orientations of primary state school EFL teachers in two Mediterranean countries'. In *C·E·P·S Journal – CEPS Journal (Center for Educational Policy Studies Journal)*. Special thematic issue entitled '*Assessment of Young Foreign Language Learners*' 6(1), 9-30. Retrieved April 11, 2017 from http://www.cepsj.si/doku.php?id=en:volumes:2016-vol6-no1

Tsagari, D. (2014) *Unplanned LOA in EFL Classrooms: Findings from an Empirical Study.* Paper presented at the Roundtable on Learning-Oriented Assessment In Language Classrooms And Large-Scale Assessment Contexts. Teachers College, Columbia University, New York, October 10-12, 2014.

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

Turner, C. E. (2014). Mixed Methods Research. In A. J. Kunnan (Ed.), *The Companion to Language Assessment* (pp. 1-15). Hoboken, NJ: John Wiley & Sons, Inc.

Vogt, K. (2004). Der Gemeinsame Europäische Referenzrahmen für Sprachen: Inhalte, Ziele, Diskussion. [The Common European Framework of Reference for Languages: Contents, Aims, Discussion.] *Der Fremdsprachliche Unterricht Englisch, 38*(2004), 69, 48–51.

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.

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

Yin, M. (2010). Understanding Classroom Language Assessment through Teacher Thinking Research. *Language Assessment Quarterly 7*(2), 176-195.

Xu, Y. (2015). *Language Assessment Literacy in Practice: A Case Study of a Chinese University English Teacher*. Paper presented at the 37th Language Testing and Research Colloquium. Toronto/Canada, 16 to 20 March 2015.

Appendix 1. Teachers' Questionnaire

Part I. General information

1. I work in _____(country)

2. What subject(s) do you teach?
   _____

3. What subjects have you studied?

   _____

4. What is your highest qualification?

   Please specify: _____

5. Type of school/institution you teach at: _____

6. Average age of pupils: _____

7. Your functions at school/institution:
   ☐   Teacher
   ☐   Head of department at school
   ☐   Mentor
   ☐   Advisory function for authorities (local government, ministry etc.)
        Other? Please specify: _____

8. During your pre-service or in-service teacher training, have you learned something about testing and assessment (theory and practice)?
   ☐   Yes (please specify:) _____
   ☐   No

Part II. Questions about training in LTA

1. Classroom-focused LTA
1.1.   Please specify if you <u>were trained</u> in the following domains.

	Not at all	*A little (1-2 days)*	*More advanced*
a) Preparing classroom tests	☐	☐	☐
b) Using ready-made tests from textbook packages	☐	☐	☐
c) Giving feedback to students based on information from tests/assessment	☐	☐	☐
d) Using self- or peer-assessment	☐	☐	☐
e) Using informal, continuous, non-test type of assessment	☐	☐	☐
f) Using the European Language Portfolio, an adaptation of it or some other portfolio	☐	☐	☐

1.2.   Please specify if you <u>need training</u> in the following domains.

	None	*Yes, basic Training*	*Yes, more advanced training*
a) Preparing classroom tests	☐	☐	☐
b) Using ready-made tests from textbook packages or from other sources	☐	☐	☐
c) Giving feedback to students based on information from tests/assessment	☐	☐	☐
d) Using self- or peer-assessment	☐	☐	☐
e) Using informal, continuous, non-test type of assessment	☐	☐	☐
f) Using the European Language Portfolio, an adaptation of it or some other portfolio	☐	☐	☐

2. Purposes of testing

2.1.   Please specify if you <u>were trained</u> in the following domains.

	Not at all	*A little (1-2 days)*	*More advanced*
a) Giving grades	☐	☐	☐
b) Finding out what needs to be taught/ learned	☐	☐	☐
c) Placing students onto courses, programs, etc.	☐	☐	☐
d) Awarding final certificates (from school/program; local, regional or national level)	☐	☐	☐

2.2.   Please specify if you <u>need training</u> in the following domains

	None	*Yes, basic Training*	*Yes, more advanced training*
a) Giving grades	☐	☐	☐
b) Finding out what needs to be taught/ learned	☐	☐	☐
c) Placing students onto courses, programs, etc.	☐	☐	☐
d) Awarding final certificates (from school/program; local, regional or national level	☐	☐	☐

3. *Content and concepts of LTA*

3.1.　　Please specify if you <u>were trained</u> in the following domains.

	Not at all	*A little (1-2 days)*	*More advanced*
1. Testing/Assessing:			
a) Receptive skills (reading/listening)	☐	☐	☐
b) Productive skills (speaking/writing)	☐	☐	☐
c) Microlinguistic aspects (grammar/vocabulary)	☐	☐	☐
d) Integrated language skills	☐	☐	☐
e) Aspects of culture	☐	☐	☐
2. Establishing reliability of tests/assessment	☐	☐	☐
3. Establishing validity of tests/assessment	☐	☐	☐
4. Using statistics to study the quality of tests/assessment	☐	☐	☐

3.2.　　　　Please specify if you <u>were trained</u> in the following domains.

	None	*Yes, basic Training*	*Yes, more advance training*
5. Testing/Assessing:			
f) Receptive skills (reading/listening)	☐	☐	☐
g) Productive skills (speaking/writing)	☐	☐	☐
h) Microlinguistic aspects (grammar/vocabulary)	☐	☐	☐
i) Integrated language skills	☐	☐	☐
j) Aspects of culture	☐	☐	☐
6. Establishing reliability of tests/assessment	☐	☐	☐
7. Establishing validity of tests/assessment	☐	☐	☐
8. Using statistics to study the quality of tests/assessment	☐	☐	☐