



## What Do Teachers Think They Want? A Comparative Study of In-Service Language Teachers' Beliefs on LAL Training Needs

Karin Vogt , Dina Tsagari & Georgios Spanoudis

To cite this article: Karin Vogt , Dina Tsagari & Georgios Spanoudis (2020) What Do Teachers Think They Want? A Comparative Study of In-Service Language Teachers' Beliefs on LAL Training Needs, *Language Assessment Quarterly*, 17:4, 386-409, DOI: [10.1080/15434303.2020.1781128](https://doi.org/10.1080/15434303.2020.1781128)

To link to this article: <https://doi.org/10.1080/15434303.2020.1781128>



© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 20 Jul 2020.



Submit your article to this journal [↗](#)



Article views: 612



View related articles [↗](#)



View Crossmark data [↗](#)



# What Do Teachers Think They Want? A Comparative Study of In-Service Language Teachers' Beliefs on LAL Training Needs

Karin Vogt<sup>a</sup>, Dina Tsagari<sup>b</sup>, and Georgios Spanoudis<sup>c</sup>

<sup>a</sup>Heidelberg University of Education, Heidelberg, Germany; <sup>b</sup>Oslo Metropolitan University, Oslo, Norway;

<sup>c</sup>University of Cyprus, Nicosia, Cyprus

## ABSTRACT

The purpose of the study was to investigate English language teachers' perceptions of assessment, their language assessment literacy (LAL) levels and their training needs. 113 teachers from Germany and 379 teachers from Greece completed a survey questionnaire. The data were analyzed through a series of RM ANOVAs, correlation analyses, and confirmatory factor analysis. Data from interviews with 25 German and 20 Greek teachers were used as supporting qualitative data. The results indicated that teachers use similar constructs in their assessment and in their conceptualisations of LAL but their perceived training needs differed depending on their educational contexts. The interviews helped identify deeper insights into contextual factors. The paper discusses the importance of context in assessment and offers recommendations for teacher education programmes in language assessment literacy that are context responsive.

## ARTICLE HISTORY

Received 06 Apr 2020

Accepted 09 June 2020

## Τι πιστεύουν οι εκπαιδευτικοί ότι χρειάζονται; Μια συγκριτική μελέτη των πεποιθήσεων των καθηγητών δευτέρης γλώσσας σχετικά με τις ανάγκες κατάρτισης σε θέματα γλωσσικού αξιολογητικού γραμματισμού (language assessment literacy)

Σκοπός της παρούσας μελέτης ήταν να εξετάσει τις αντιλήψεις των καθηγητών Αγγλικής γλώσσας σχετικά με την αξιολόγηση, τα επίπεδα γλωσσικού αξιολογητικού γραμματισμού τους (language assessment literacy - LAL) και τις επιμορφωτικές τους ανάγκες. 113 καθηγητές από τη Γερμανία και 379 από την Ελλάδα συμπλήρωσαν ένα ερωτηματολόγιο. Τα δεδομένα αναλύθηκαν μέσω μιας σειράς αναλύσεων διακύμανσης επαναλαμβανόμενων μετρήσεων (RM ANOVA), αναλύσεων συσχέτισης (correlation analyses) και επιβεβαιωτικής ανάλυσης παραγόντων (confirmatory factor analysis). Επίσης στοιχεία από συνεντεύξεις με 25 Γερμανούς και 20 Έλληνες καθηγητές χρησιμοποιήθηκαν ως συμπληρωματικά ποιοτικά δεδομένα. Τα αποτελέσματα έδειξαν ότι οι εκπαιδευτικοί χρησιμοποιούν παρόμοιες έννοιες στην αξιολόγησή τους και στις αντιλήψεις τους για το γλωσσικό αξιολογητικό γραμματισμό αλλά οι ανάγκες της εκπαιδευτικής τους κατάρτισης, όπως τις αντιλαμβάνονταν, διέφεραν αναλόγως του εκπαιδευτικού τους συστήματος. Οι συνεντεύξεις βοήθησαν στη βαθύτερη κατανόηση συγκυριακών παραγόντων. Η παρούσα έρευνα συζητά τη σημασία του εκπαιδευτικού πλαισίου στο χώρο της αξιολόγησης και προσφέρει προτάσεις για προγράμματα επιμόρφωσης εκπαιδευτικών σε θέματα γλωσσικού αξιολογητικού

**CONTACT** Dina Tsagari  [dina.tsagari@oslomet.no](mailto:dina.tsagari@oslomet.no)  Department of Primary and Secondary Teacher Education, Faculty of Education and International Studies, Oslo Metropolitan University, Oslo 0130, Norway

© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

γραμματισμού που να ανταποκρίνονται κατάλληλα σε συγκεκριμένα εκπαιδευτικά πλαίσια.

Λέξεις κλειδιά: γλωσσικός αξιολογητικός γραμματισμός, καθηγητές γλωσσών, πεποιθήσεις εκπαιδευτικών, αξιολόγηση, ανάγκες εκπαιδευτικής κατάρτισης, εκπαιδευτικά πλαίσια

## Introduction

Assessment constitutes an important field of teacher activity, with 30 to 50% of teachers' time relating to assessment (Cheng, 2001). Therefore, it is important to empower teachers to be able to carry out effective assessment, in other words, to be assessment literate. Inbar-Lourie (2008, p. 389) defines someone as assessment literate if they have “the capacity to ask and answer critical questions about the purpose for assessment, about the fitness of the tool being used, about testing conditions, about what is going to happen on the basis of the results” (see also Fulcher, 2012; O'Loughlin, 2013; Vogt & Tsagari, 2014). The scope of the competencies implied by this definition seems to demand a high level of professionalization by language teachers. The field of language assessment literacy (henceforth LAL) has attracted considerable scholarly attention in general (e.g., Brunfaut & Harding, 2018; Coombe, Troudi, & Al-Hamly, 2012; Hildén & Fröjdendahl, 2018; Kremmel, Eberharter, Holzknacht, & Konrad, 2018; Kremmel & Harding, 2019; Malone, 2013) and with regard to teachers' LAL, in particular, with a focus on their LAL confidence levels (Fulcher, 2012; Hasselgren, Carlsen, & Helness, 2004; Kvasova & Kavytska, 2014; Lam, 2015; Levi & Inbar-Lourie, 2020; Tsagari & Vogt, 2017; Sultana, 2019). Although comparisons of LAL levels and training needs have been carried out, e.g., by Hasselgren et al. (2004) who found a need for LAL training across the board, more detailed insights into perceived LAL training needs across educational contexts seem to be missing as well as contextual factors that might impact on teachers' training needs. In fact, there is little work on teachers' perceptions of the issue and the relationship between LAL and teacher beliefs has rarely been investigated despite a substantial body of research in language teacher cognition (e.g., Borg, 2018; Farrell, 2011; Fives & Gill, 2015; Kiely & Davis, 2010). Language teacher beliefs research has focused on areas as diverse as teaching grammar (e.g., Johnston & Goettsch, 2000; Schulz, 2001), literacy development (e.g., Meijer, Verloop, & Beijjaard, 2001) or teacher autonomy (e.g., Benson, 2010) but not much has been done relating teachers' beliefs, assessment and LAL. Given the central role of assessment in English language teaching and LAL as a vital element of teacher professionalization, the present study explores the link between teacher beliefs and LAL training needs across two educational contexts: Germany and Greece. In this, the role of contextual factors on teachers' perceptions of their LAL training needs will be scrutinized.

## Teacher beliefs

### *Theoretical conceptualisations of teacher beliefs*

Awareness of the consequences of research into teachers' beliefs and perceptions and their potential effects on teaching practices has long been discussed. However, the field has gained a solid base with a surged interest from the 1990s (e.g., Beach, 1994; Freeman,

2002; Holt, 1992; Woods, 1996). The focus has been placed on research into teacher beliefs, attitudes, identities and emotions that are considered vital aspects of the unobservable dimension of teaching.

Beliefs, in particular, are defined by Borg (2001, p. 186) as “a mental state which has as its content a proposition that is accepted as true by the individual holding it, although the individual may recognize that alternative beliefs may be held by others”. Kumaravadivelu (2012, p. 67) distinguishes between core and peripheral beliefs in language teachers, with core beliefs being influential in shaping teachers’ instructional approaches and peripheral beliefs which can cause a lack of congruence between teachers’ claims and their actual teaching in the classroom. Teachers’ beliefs about teaching and learning are influenced by teachers’ own experience as learners and later experience as teachers (Phipps & Borg, 2009). Their experience can be used as a filter through which new information or newly acquired experience is interpreted. Experience potentially outweighs effects of teacher education, as a long-term influence on teachers’ institutional decisions and practice. Experience can also balance off a lack of formal training, as Sheehan and Munro (2017) have found in their study on teachers’ assessment practices and LAL. Beliefs on foreign language teaching and learning in general and on language testing and assessment (LTA) in particular are likely to influence teachers’ practices.

Contextual factors constitute another powerful influence. Already in 1996, Burns (1996) advocated that greater attention be paid to social and institutional contexts of classrooms and include these in studies on what language teachers do. Borg (2003) also states that:

factors such as parents, principals’ requirements, the school, society, curriculum mandates, classroom and school layout, school policies, colleagues, standardized tests and the availability of resources may hinder language teachers’ ability to adopt practices which reflect their beliefs. (p. 94).

For assessment in general, Black and Wiliam (2005) highlight the influential role of context when teachers choose and use assessment tools, interpret and use assessment results as these are subject to “a range of educational, public and political influences” (2005, p. 258). In this regard, it is worthwhile to explore a potential link between contextual factors, teacher beliefs in assessment and LAL, in particular perceived training needs of foreign language teachers in different educational contexts. The following section will explore language teachers’ beliefs related to different aspects of language assessment.

Unlike studies on teachers’ conceptions of assessment in general (e.g., Remesal, 2010), there are relatively few studies on language teachers’ beliefs related to language assessment. The existing ones can be categorized into: i) research on teacher beliefs on language assessment in general or particular aspects of language assessment, ii) research on teacher beliefs on LAL and iii) studies on teachers’ perceived LAL training needs. Some of these studies show how contextual factors influence teacher beliefs or how teaching and assessment practices are shaped by constraints relating to their context.

### ***Research on beliefs on language assessment***

Among the studies that deal with beliefs on language assessment in general is Hidri’s study (2015) who identified three prevalent factors related to teachers’ conceptions of assessment in his replication of Brown’s work (2004, 2006). The three factors he found specifically for

the Tunisian context with a sample of 542 EFL teachers were ‘accountability’, ‘improvement’ and ‘irrelevance’, reflecting different characteristics from other educational contexts. The relation between teacher beliefs on assessment and their assessment practices in the EFL classroom was also investigated by Shim (2009). The 68 Korean EFL school teachers’ conceptions of classroom-based language assessment as well as their assessment practices were found to be linked to their own belief systems on assessment (see also Gebril, 2017). However, the results confirm Borg’s (2003) and Black and Wiliam’s observation (2005) that teachers are not always in a position to put these beliefs into practice due to external influences. These are often institutional constraints such as large classes, heavy teaching loads and bureaucratic barriers. Similar results but related to teachers’ perceptions of self-assessment have been presented by Bullock (2011) in her study of ten Ukrainian English language teachers. Her results point to a discrepancy between teachers’ beliefs and assessment practices, with teachers’ attitudes towards learner self-assessment “not necessarily indicative of practices” (2011, p. 121).

### ***Research on teacher beliefs on language assessment literacy***

In recent theoretical discussions, there is an increasing necessity for the development of LAL among the various stakeholders involved in assessment procedures (Harding & Kremmel, 2016; Kremmel & Harding, 2019; O’Loughlin, 2013; Pill & Harding, 2013; Taylor, 2009, 2013), with the concept being more closely associated with teachers in general (Popham, 2009; Stiggins, 1991; Xu & Brown, 2016). Scarino (2013) highlights the often tacit preconceptions and beliefs that language teachers hold about language assessment that inform their conceptualisations, interpretations and practical decisions in assessment, thus shaping LAL as self-awareness. She also indicates the need for the field to consider the “lifeworlds” of teachers. Likewise, Inbar-Lourie (2017) calls for more research of local realities in LAL in order to arrive at a more thorough understanding of the intricacies of the matter.

Some of the local realities of LAL are addressed by Tao (2014), who adopted a mixed-methods approach for the development and validation of four scales designed to measure the classroom assessment literacy development of 108 EFL instructors at university level in Cambodia. Overall, the results indicate that contextual factors such as large classes, teaching loads of up to 50 hours a week and other aspects such as local departmental assessment policies of the university impact on teachers’ beliefs and perceptions of an ideal assessment, highlighting the impact of contextual factors on teacher beliefs relating to and underpinning LAL.

Crusan, Plakans, and Gebril (2016) identify context as a crucial concept in (writing) LAL. Their study with 702 ESL writing instructors from American Universities were surveyed by way of a questionnaire with multiple choice, Likert scale and open-ended response options. Although the notion of context was not the focus of the study, Crusan et al. (2016, p. 53) found that the teaching context impacted on teachers’ perceptions of assessment in terms of assessment philosophy and of their LAL. Teachers with a heavy teaching load were reported to have more negative views towards assessment. Likewise, Giraldo (2019), in his survey of five Colombian language teachers found that the understanding of teachers’ context was as pertinent to LAL as other crucial components such as knowledge and skills. The question as to how contexts impact on teachers’ perceived LAL training needs remains open.

### **Teachers' perceived LAL training needs and the role of contextual features**

Teacher beliefs on LAL seem to shape perceived training needs in this area. Related to confidence levels and perceived levels of LAL, studies into perceived LAL training needs have been conducted particularly with language teachers (e.g., Fulcher, 2012; Hasselgren et al., 2004; Kvasova & Kavytska, 2014; Lam, 2015).

Yan, Zhang, and Fan (2018), in particular, highlight the contextualised nature of teachers' assessment practices and training needs. Their study of three EFL teachers investigated how contextual and experiential factors (based on Crusan et al., 2016) have an effect on LAL development and training needs. The researchers developed a framework of contextual factors from their findings yielded by an explorative interview study of three EFL teachers at a Chinese middle school. In terms of contextual factors which they defined as referring 'to larger educational, social, cultural, political and historical factors that collectively form the *assessment culture* in a particular context' (Yan et al., 2018, p. 159), they identified three categories. These are: 'educational landscapes and policies' which subsume aspects like an exam-oriented assessment culture and national or municipal policies pertaining to assessment; 'institutional mandates' which signify the duties that teachers are assigned with by different stakeholders (e.g., parents, board of education) that influence teachers' assessment practices and LAL training needs (e.g., reporting scores, statistical skills for score reporting to parents, the board of education influencing assessment practices through policies and training); and 'local institutional contexts' which pertain to the resources and constraints that teachers are faced with, e.g., available assessment training resources, teachers' workloads, (lack of) teacher collaboration, etc. that influence teachers' perceived training needs. They concluded that assessment training for language teachers should be cognizant of the relationship between assessment and teaching context in order to implement effective syllabi and methods (Yan et al., 2018, p. 167). The findings from this study seem to confirm the requirements of LAL training as being contextually adaptive.

However, the categories of context that emerge from their data are not clear-cut and overlap to a certain extent. This is possibly due to the small sample size and the nature of the data in their explorative study that looked at the context of one particular middle school in China. Therefore, we would like to suggest a categorization of context as a theoretical framework in this study that is loosely based on Yan et al. (2018) but takes into account broader categories on different contextual levels. On a national and regional level, comparable to a 'macrolevel', we subsume national educational policies and assessment cultures that would be related back to historical, cultural and political factors (e.g., 'frontistiria' schools in Greece that prepare learners for external exams in a test-oriented culture, see Tsagari, 2009). These also include the educational landscape (cf. Yan et al., 2018) that encompasses systemic factors, e.g., school types, the training of teachers, remuneration, in-service professionalization opportunities and duties. The 'mesolevel' of the institution is also characterised by the school in interaction with the assessment cultures, systemic elements and local or regional politics, with a focus on the implementation and communication with teaching staff and the principal in the institution. Institutional decisions might pertain to the school culture or school profile, the resources available for training at a local level, etc. The 'microlevel' is concerned with the local context of the classroom. Teachers' instructional decisions based on assessment procedures count into this category as well as their interaction with stakeholders in the assessment process like parents and learners. Local

assessment priorities by teachers are also part of this category. Figure 1 attempts a representation of the different contextual levels impacting on teacher beliefs of LAL training needs.

In this framework, we also see that the different levels are interdependent and impact on one another, e.g., local assessment priorities or instructional decisions on the micro-level might be shaped both by the assessment culture in general but also by the school profile and local school culture. These categories are used in the present study to interpret language teachers' perceived training needs on LAL and compare these across the two different educational contexts, namely Germany and Greece. The quantitative data in the study mainly allow us to look at the microlevel while supporting data from interviews with teachers are analysed to explore if the contextual factors identified at the micro-, meso- and macrolevels impact on teachers' perceived LAL training needs. This way, we hope to shed light on the relationship between perceived LAL training needs as a part of teachers' beliefs on LAL and contextual factors, closing thus a gap in the existing literature.

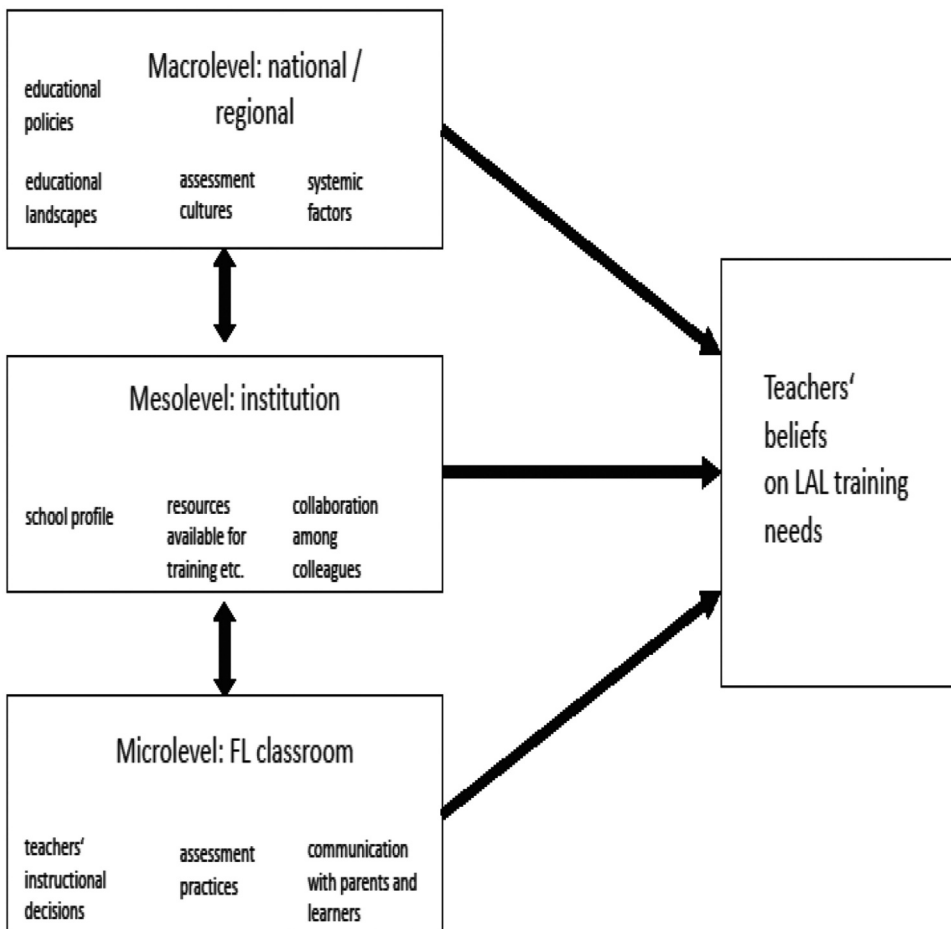


Figure 1. Contextual levels impacting on teacher beliefs of LAL training needs.

## Methodology and methods

### *Study design and research questions*

The study forms part of a large-scale survey of perceived LAL training needs of language teachers across Europe (Vogt & Tsagari, 2014), but focuses on the data from a subset of two educational contexts, namely Greece and Germany with a view of comparing these and situating them in their respective educational contexts. Data from interviews with teachers in both contexts have been used as supporting data. The entire study can be characterized as a mixed-methods study that comprised both quantitative and qualitative data. According to Dörnyei (2007), the data collection instruments were sequenced in a linear way. The questionnaire represented the quantitative part of the study. It was a replication of the LTA needs analysis study by Hasselgren et al. (2004). The sample was a convenience sample. However, the teachers had to meet two important conditions. First, they had to have completed their pre-service teacher training. Second, unlike Hasselgren et al. (2004) where teachers had multiple roles (e.g., teachers, item writers), the current study included participants who were *only* serving as language teachers. Those with multiple roles were excluded from the sample. The same selection criteria were applied in the recruitment of informants for the interviews.

Questionnaires were distributed at teacher conferences and teacher training events to teachers at primary and secondary schools. Informants for the interviews were recruited through the researchers' networks. Their teaching experience was comparable across the two countries with three quarters having more than six years of teaching experience. In the German sample ( $n = 113$ ), most teachers worked at secondary school while Greek teachers ( $n = 379$ ) were teaching in both primary and secondary schools. Their student age groups ranged from 6 to 20, and English was the prevailing language taught. In the overall quantitative data, data subsets were roughly the size of the German sample and only the Greek sample stood out. The imbalance of the sample could be considered as a problem in the present study. However, an attempt to eliminate data from the Greek subsample in order to reduce it has yielded similar tendencies in the results. Therefore, the original sample was retained for the quantitative part of this study.

The goal of the study was to shed light on teachers' beliefs about and perceptions of teachers' language assessment levels and training needs in Greece and Germany as two different educational contexts. To achieve its aims the study set out to answer the following research questions:

- (1) Are there any differences in perceived LAL levels of training and levels of needed training between German and Greek teachers?
- (2) Is there any relation between training received and training needed across the three components of the questionnaire between the two countries?
- (3) What contextual factors account for potential differences between the two educational contexts?

Since this study was conducted in two educational contexts, comparisons will be made between the two groups involved.



## Data collection and analysis

The questionnaire by Hasselgren et al. (2004) was used as a starting point but shortened in order to relate to teachers' everyday LTA tasks (see Appendix 1). The questionnaire included a general part (Part I) designed to gather background information about the respondents, and whether LTA had been part of their formal pre- and in-service training. The main part of the questionnaire (Part II) was divided into three parts: (1) classroom-focused LTA, (2) purposes of testing, and (3) content and concepts of LTA. Each part was further subdivided into two sections: one section enquiring about the training that respondents had received and another one relating to the training needs they saw for themselves. Of course the answers were seen as subjectively felt needs. A 3-point Likert-type scale was offered for the answers as set in the original questionnaire. The options for teachers to quantify their LAL training needs were 'none', 'basic' and 'advanced', taken over from Hasselgren et al. (2004). While these are easily understood by informants, one disadvantage is that they are subject to interpretation. This might represent a slight limitation of the study because, as is typical in questionnaire surveys, communicative validation (Dörnyei, 2007; Paltridge & Phakiti, 2010) with respondents was not possible. A taxonomy of Part II of the questionnaire is provided in Table 1.

**Table 1.** Taxonomy of questions included in Part II – Teachers' questionnaire.

Component	Training (Question Numbers)		LTA practices
	Domain 1 – Received	Domain 2 – 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 internal consistency reliability of the questionnaire was computed using Cronbach's alpha. The reliabilities for the individual scales ranged from .80 to .93, indicating a high level of internal consistency (Dörnyei, 2010, p. 94).

With respect to the responses, descriptive statistics (frequencies and percentages) of the data relating to training "received" and "needed" were calculated. Further, we employed two conventional statistical analyses, that is, analysis of variance for examining mean differences and correlation analysis for testing relations between variables; we used also advanced statistical analysis, namely Structural Equation Modeling, in order to test for factorial invariance across the two samples.

In addition, data from interviews with 25 German and 20 Greek teachers were used as supporting data. The objective was to shed more light on the teachers' personal situation but also in terms of exemplary contextual factors on the micro-, meso- and macrolevels. The interviews were based on a guided protocol (see [Appendix 2](#)).

Respondents came from the south-west of Germany and the south of Greece. All informants were practicing EFL teachers in primary and secondary schools. Interviews took about 30 minutes and were audio-taped with the consent of the participants. For communicative validation, the interview transcriptions were given to the informants, asking them to comment about ambiguous passages in the transcripts or clarifying details. After finalising the transcripts, these were content analysed using open and axial coding and categorization (Corbin & Strauss, 2014). The interview transcripts were read in their entirety and openly coded using an inductive approach with particular attention to any relevance or reference to contextual factors on the micro-, meso- and macrolevels. The analysis of the transcripts was undertaken separately by the two researchers to achieve cross-verification of data by way of investigator triangulation.

## Results and Discussion

### Descriptive results

Descriptive statistics have been already reported in Vogt & Tsagari (2014) and will be briefly summarised below (see [Table 2](#)). Looking at the overall trends, we generally observe somewhat low LAL levels in both contexts.

**Table 2.** Average trends in teachers' LTA literacy in regional contexts.

	Greece	Germany	Average
<i>Domain 1 (Training Received)</i>			
Not at all	32.0 <sup>a</sup>	47.1	35
A little	30.5	31.7	32.5
Advanced	22.1	18.4	21.6
<i>Domain 2 (Training Needed)</i>			
None	15.7	29.3	23.9
Basic	23.3	40.1	30.0
Advanced	39.7	26.3	30.1

<sup>a</sup>: percent based on data reported in Vogt & Tsagari, (2014)

Furthermore, 60.1% on average said they require “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, respondents asked for more advanced training (see Vogt & Tsagari, 2014). In comparison to Hasselgren et al. (2004), where informants displayed self-reported low levels of LAL and a wish for LTA training across the board, teachers in our study displayed equally low LAL levels (35% said they had received no training and 32.5% little training in LTA) Additionally, teachers in both contexts had similar priority areas for training, namely alternative assessment formats, e.g., peer assessment, self-assessment and portfolio assessment. Other aspects of LTA to be developed were testing microlinguistic aspects and language skills while grading or non-traditional assessment methods as well as establishing quality criteria of assessments (e.g., reliability and validity) were not well-developed. Limited

formal assessment training of teachers is a typical finding confirmed by various studies in different educational contexts (e.g., Sultana, 2019; Sheehan & Munro, 2017).

### Comparing teacher's beliefs across countries

In this sub-section, we took a closer look at the results of a series of parametric tests. Table 3 shows the mean and standard deviations of all components of the questionnaire ('Classroom-focused LTA' as Component 1, 'Purposes of Testing' as Component 2, and 'Content and Concepts of LTA' as Component 3). From the individual values one can infer that the tendencies across the two groups of teachers seem to be similar. In Table 3 we also note that German teachers' mean scores are slightly lower. This indicates that they tend to be slightly more moderate in their beliefs with regard to the domains of training received

**Table 3.** Mean and standard deviations of all components across countries.

			Greece		Germany		Differences	
			M	S.D.	M	S.D.		
Component 1	Domain 1	1.1a) prep	2.06	0.77	1.97	0.79	0.09	
		1.1b) ready	1.82	0.79	1.69	0.79	0.14	
		1.1c) feedb	2.01	0.76	1.83	0.79	0.18	
		1.1d) selfpr	1.82	0.74	1.40	0.54	0.42	
		1.1e) nontest	1.79	0.79	1.73	0.74	0.06	
		1.1f) elp	1.62	0.69	1.25	0.49	0.37	
			<b>total</b>	<b>1.85</b>	<b>0.76</b>	<b>1.65</b>	<b>0.69</b>	<b>0.21</b>
	Domain 2	1.2a) prep2	2.29	0.82	1.96	0.91	0.33	
		1.2b) ready2	1.88	0.84	1.70	0.79	0.18	
		1.2c) feedb2	2.36	0.75	1.89	0.74	0.47	
		1.2d) selfpr2	2.40	0.72	2.07	0.67	0.32	
		1.2e) nontest2	2.35	0.76	2.13	0.73	0.22	
		1.2f) elp2	2.48	0.68	2.19	0.59	0.28	
			<b>total</b>	<b>2.29</b>	<b>0.76</b>	<b>1.99</b>	<b>0.74</b>	<b>0.30</b>
Component 2	Domain 1	2.1a) grade	1.59	0.76	1.81	0.80	-0.23	
		2.1b) needs	2.04	0.78	1.56	0.69	0.48	
		2.1c) place	1.65	0.70	1.35	0.64	0.30	
		2.1d) certf	1.51	0.69	1.54	0.75	-0.03	
			<b>total</b>	<b>1.70</b>	<b>0.73</b>	<b>1.57</b>	<b>0.72</b>	<b>0.13</b>
	Domain 2	2.2a) grade2	2.16	0.80	1.85	0.81	0.31	
		2.2b) needs2	2.25	0.84	1.96	0.70	0.28	
		2.2c) place2	2.31	0.77	1.90	0.63	0.41	
		2.2d) certf2	2.12	0.79	1.85	0.72	0.27	
				<b>total</b>	<b>2.21</b>	<b>0.80</b>	<b>1.89</b>	<b>0.72</b>
			<b>total</b>	<b>2.21</b>	<b>0.80</b>	<b>1.89</b>	<b>0.72</b>	<b>0.32</b>
Component 3	Domain 1	3.1.1a) rdlist	2.27	0.72	1.94	0.73	0.34	
		3.1.1b) spkwrt	2.31	0.72	2.21	0.71	0.10	
		3.1.1c) grvoc	2.33	0.73	2.17	0.76	0.16	
		3.1.1d) integ	2.12	0.79	2.05	0.69	0.07	
		3.1.1e) culture	1.75	0.77	1.83	0.77	-0.08	
		3.1.2) reliab	1.85	0.78	1.50	0.74	0.35	
		3.1.3) valid	1.81	0.77	1.52	0.74	0.29	
		3.1.4) stats	1.39	0.64	1.36	0.65	0.02	
			<b>total</b>	<b>1.98</b>	<b>0.74</b>	<b>1.82</b>	<b>0.72</b>	<b>0.16</b>
	Domain 2	3.2.5f) rdlist2	2.29	0.83	2.15	0.86	0.15	
		3.2.5g) spkwrt2	2.31	0.83	2.10	0.87	0.21	
		3.2.5h) grvoc2	2.25	0.84	1.89	0.86	0.37	
		3.2.5i) integ2	2.38	0.80	2.09	0.82	0.28	
		3.2.5j) culture2	2.42	0.72	1.96	0.78	0.46	
3.2.6) reliab2		2.40	0.70	1.91	0.67	0.49		
3.2.7) valid2	2.41	0.70	1.88	0.66	0.53			
3.2.8) stats2	2.35	0.70	1.92	0.68	0.44			
		<b>Total</b>	<b>2.35</b>	<b>0.77</b>	<b>1.99</b>	<b>0.77</b>	<b>0.37</b>	

and training needed compared to Greek teachers. However, the differences are not significant, which suggests that the underlying assessment construct, as depicted in the questionnaire, does not seem to be perceived very differently by the two groups of teachers. In other words, they share the same ideas on the theoretical construct of assessment in the foreign language classroom.

In order to compare the Greek and German respondents across components, two two-way Repeated Measures (RM) ANOVAS were conducted, with ‘country’ being used as a between-subjects variable (Greek vs. German teachers) and ‘component’ (domain 1 vs. domain 2 – training received/training needed) as the within-subjects factor. This was done to examine whether there are any differences in training levels and levels of perceived LAL training needs between Greek and German teachers (Research Question 1). Each RM ANOVA focused on each one of the three components of the questionnaire (see Table 4) and showed that: i) the perceptions of teachers in the two respective contexts of Germany and Greece differ significantly, and ii) there are significant differences between the two domains, LAL training levels (domain 1) and training needs (domain 2). This is indicated by large F-values (74.93, 73.94, and 27.17) which represent the division of the within factors’ variance. In general, the results show the significant interaction effects between country and domains. In particular, whereas German teachers believed that they had insufficient training in an area and wished for basic training in that area, Greek teachers made a point of wishing for more advanced training on the basis of an already advanced prior training according to their perceptions. These results will be discussed with regard to contextual factors in the next section of this paper.

**Table 4.** Results of the three RM anovas.

	Effect	Df	F-value	Sign.	Partial $\eta^2$
1 <sup>st</sup> RM Anova	Country	1. 288	20.29	<.001	.07
	Component	1. 288	74.93	<.001	.21
	Country X component X questions	5. 1197	2.37	=.048	.01
2 <sup>nd</sup> RM Anova	Country	1. 360	14.28	<.001	.04
	Component	1. 360	73.94	<.001	.17
	Country X component X questions	3. 1005	14.2	<.001	.04
3 <sup>rd</sup> RM Anova	Country	1. 340	20.80	<.001	.06
	Component	1. 340	27.17	<.001	.07
	Country X component X questions	1. 1405	5.06	<.001	.02

In order to answer Research Question 2 on the relation between training received and training needed across the three components (‘Classroom-focused LTA’, ‘Purposes of testing’, ‘Content and Concepts of LTA’), we carried out a correlation analysis pairing relevant variables from each domain (training received/training needed) within each sample (Greek and German). The analysis took into consideration the Bonferroni criterion for multiple corrections, which was .0017. We found three significant relations in the third component in the Greek sample and two correlations in the German sample, all of which were weak negative correlations. In particular, the Greek sample yielded the following measures in the testing of reading/listening:  $r = -.18$ ; speaking/writing:  $r = -.22$  and integrated language skills:  $r = -.19$ . In other words, the more training Greek teachers received in these skills, the less training they said they needed in these areas.

The analysis also showed that there was also a weak negative correlation for the testing of microlinguistic structures (grammar/vocabulary) ( $r = -.32$ ) and integrated language skills

( $r = -.26$ ) in the German sample, indicating that German teachers believed they needed less training in microlinguistic skills and in integrated language skills when the levels of training received were felt to be higher.

These results seem to reflect varying teaching and assessment practices in the respective educational contexts. In the Greek context, teachers at secondary level are required to design their own end-of-year tests assessing reading, vocabulary, grammar and occasionally writing short texts. Therefore, teachers find it unnecessary to receive training in areas that are not tested. In the German context, it is rather common to have informal vocabulary tests in secondary EFL classrooms which follow a rather standardised format as they are usually based on the translation of single lexemes from German into English. In summative assessments like classroom tests, e.g. after a textbook unit, the paradigm shift towards competence-oriented language teaching has likely been felt because teachers are supposed to test skills, integrated skills included. The respective contextual information about language assessment in EFL classrooms in Germany and Greece will be detailed in the Discussion section of this paper.

Research Question 3 referred to the role of contextual factors in identifying potential differences between the two educational contexts. In order to find out whether the Greek and German teachers understand the underlying concepts of the questionnaire, i.e. the general construct of LAL, in the same way we looked at the relationship between the factors and indicators of the questionnaire, the factorial structure of the questionnaire and its factorial invariance across the two samples. To this end, multi-sample analyses were conducted using EQS 6.1 (Bentler, 2006). The dimensionality of the questionnaire was tested given that it is used for the first time in Greek and German populations. In research practice, multi-group confirmatory factor analysis (MG-CFA) is widely used to test construct comparability, which is a prerequisite for testing cross-group differences (Byrne & Watkins, 2003). In the current study the following measures were employed: (a) the Comparative Fit Index (CFI; Bentler, 1990), with values greater than .95 indicating a reasonable model fit; and (b) the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), with values less than .08 indicating reasonable model fit (A model is determined to fit well if both criteria are met). One additional indicator, the Akaike Information Criterion (AIC), was also used to compare the relative fit of the 1-, 2-, and higher-order factor models. Although this indicator does not have an absolute value associated with closeness of fit, AIC can be used to compare non-nested models where the model evidencing the lowest value is preferred. In evaluating the statistical significance of individual model parameters (e.g., factor loadings, inter-factor correlations), a more stringent statistical significance level of .001 was employed. Because a preliminary analysis of the data confirmed severe normality violations among many items, we decided to utilize a Satorra-Bentler corrected chi-square statistic, which adjusts the chi-square through inclusion of a correction factor influenced by the degree of non-normality in the sample data (Satorra & Bentler, 1994).

### **Confirmatory factor analysis**

The first step of testing factorial invariance encompasses the separate determination of a baseline model for each group, i.e. German and Greek teachers respectively. The configurational model was tested for components divided into two domains.

The metric of the factors in all models was defined by fixing the factor variable variances to 1.0. We relied on modification indices to evaluate the models; we allowed three error variances of the items to be correlated, one between variables of the second domain and two between variables of the third domain. Factor loadings and error variances were freely estimated. With respect to the validity of a three-factor model of the Teacher's questionnaire, findings were consistent in revealing goodness-of-fit for the baseline models that were admissible for Greek teachers ( $S-B\chi^2(147) = 697.93$ ; CFI = .89; RMSEA = .09; 90% CI .08 – .10) and of good fit for German teachers ( $S-B\chi^2(150) = 175.29$ ; CFI = .94; RMSEA = .04; 90% CI .00 – .07). This means that both groups seem to have made their judgements concerning training received and their perceived training needs on the basis of the same constructs about language teaching and particularly language assessment since the underlying construct of the two groups in the questionnaire does not differ so much. Note, however, the results of the RM ANOVAs that point to the significant differences between Greek and German teachers regarding their perceived LAL training levels and training needs.

After establishing the baseline models, we tested hypotheses bearing on the equivalence of the Teacher's questionnaire across the two populations. Table 5 reviews results from the tests for invariance of the Teachers' questionnaire across the two samples. This table presents the goodness-of-fit statistics which are related to all models tested. Results related to all comparisons do not support the factorial invariance of the three-factor model across the two groups. The factor loadings themselves seem to have similar structures, with a sameness of latent factors. The relation between indicators and latent factors, however, is not the same. The correlation between the specific pattern is different, which can be inferred from the  $\Delta S-B \chi^2$  values, which are rather high (Table 5).

**Table 5.** Results of the confirmatory factor analyses.

	S-B $\chi^2$	df	CFI	RMSEA	RMSEA 90% CI	Model comparison	$\Delta$ S-B $\chi^2$	$\Delta$ df	$\Delta$ CFI
Model 1-configural	550.65	258	.962	.08	.07-.09	-	-	-	-
Model 2- invariant factor loadings; 3 error covariances	2922.94	276	.654	.25	.25-.26	2 vs. 1	2372.29	18	.31
Model 3- invariant factor loadings; 3 error covariances	2928.14	279	.653	.25	.24-.26	3 vs. 1	2377.49	21	.31

Summing up, we can infer from the data that Greek and German teachers in our sample use the same constructs in their assessment and conceptualise their LAL in the same way, but they respond to the operationalisations of the constructs as expressed in their LAL training levels and training needs in a different way. Possible interpretations of these results will be discussed in the following section.

### **Interview responses**

In order to gain a deeper insight into teachers' perceptions in our sample, we are going to attempt an exploration of a selection of relevant contextual features on a macro-, meso- and microlevel, that seem to have shaped the answers yielded by the supporting interview data.

In the relevant research literature the Greek educational context is characterized by a testing-oriented culture with a strong focus on standardized tests as a typical feature on

the macrolevel (see Tsagari & Papageorgiou, 2012; Papakammenou, 2018). Standardized language tests have gained such momentum that a complete private language school system (known locally as ‘frontistiria’) similar to cram schools has come into being (Tsagari, 2009). Young teachers typically find employment there before they are appointed in state schools at a much later stage in their career. As a consequence, young teachers tend to accumulate a lot of testing-oriented teaching experience in frontistiria before they are employed in the state school system (Tsagari & Giannikas, 2018). As one of the teachers stressed “I worked for 12 years in a frontistirio and prepared students for all kinds of international exams, Cambridge, Trinity, TOEFL, etc. I learned a lot about exam preparation. Then I was appointed in the state school system. No need to prepare students for such tests there. Things are different.” (Interview 8, p. 1). This contextual feature might account for high percentages of LTA training received in the questionnaire data, in particular the receptive and productive skills often tested in standardized tests. Another factor might have been the forthcoming Teachers’ Evaluation scheme that was planned for teachers in Greece and that would motivate the respondents to wish for more advanced training in LTA so that their performance might be enhanced for these evaluations. Teacher 12 pointed out that “We are waiting for the Ministry to implement Teachers’ Evaluation. I am preparing a professional portfolio. So any kind of advanced training course I can attend would be very useful.” (Interview 4, p. 3). This circumstance might help explain the high percentage of Greece-based language teachers wishing for more and, particularly, more advanced LAL training in the questionnaire data.

On a mesolevel of schools and local school networks, resources for in-service teacher training in Greece are usually available, which is reflected in the questionnaire data, e.g., informants stating they do not need advanced training in receptive skills. The supporting interview data, however, points to a lack of orientation on the part of teachers because relating to the offers of LTA training they report difficulties in finding a suitable training measure that would appropriately respond to their training needs. One of the Greek teachers, when asked about her short-term training needs in LTA for her personal professional development of LAL, remarked: “I don’t really know what is there. I don’t know the options . . . ” (Interview 1, p. 8) while another one stressed that: “We don’t get enough information about these things. Neither the Ministry nor the school advisors explain these things to us in advance. Whatever training courses in assessment go to they are very short, one day or so, and not very well-organised.” (Interview 3, p. 2). Others articulate very specific LTA needs but they do not see courses on offer that relate to these needs, as one informant confirms: “I would like to see different ways of assessing our students, besides the standard way of testing that we all know. I would like to see all those ways and how you evaluate this work.” (Interview 2, p. 8). The need for more training in alternatives in assessment (also shown by the questionnaire data), is because this type of assessment was included in the requirements set by the local Foreign Language Curricula (Pedagogical Institute, 2003). One informant at a primary school was asked about how her school implements the national Ministry of Education policies on language assessment: “It’s not like the other school subjects where things are stricter but there has to be a written test around every four units.” (Interview 3, p. 9). This also indicates the reliance on the textbook materials associated with classroom-based language assessment at her school (Tsagari & Sifakis, 2014). Interview data support that teachers’ assessment practices are rooted in a materials-based approach: “I use the tests, the DVDs and stuff like that from the books”

(Interview 1, p. 1), “I test my students in every unit. If I see something [in the textbook], I am going to test them.” (Interview 1, p. 9), “Thank God there are test booklets in the teacher’s book. I use these tests all the time.” (Interview 14, p. 4). This finding might also help to explain questionnaire results related to the use of ready-made tests because teachers working in Greece indicated that they did not need any training in this area (mean: 1.82).

Greek teachers also indicate high levels of training in receptive and productive skills, microlinguistic aspects and integrated language skills. All these are necessary for understanding or devising discrete-point tests, the type of test that seems to be influential in the Greek educational context (Tsagari, 2009). In addition, the test formats used seem to influence teachers’ instructional decisions, leading to a washback effect on the microlevel of the EFL classroom, as evidenced by the interviews: “I use a lot of test exercises in the year. It helps them a lot in the final exams.” (Interview 7, p. 4). The opposite effect can be seen in an interview with a primary school teacher in whose local context teachers do not give marks and hence she does not do tests that “focus on how to mark them because my school advisor does not require it and because I don’t personally believe in tests” (Interview 4, p. 9). This is an example of how teachers’ beliefs align with contextual factors like the assessment policy expressed by educational authorities, in this case the school advisor.

Differences in the German contextual factors might account for some of the significant differences in the questionnaire data. Starting on a macrolevel, the assessment culture in Germany is less testing-centred compared to Greece. Classroom-based language assessment is highlighted in many federal states despite the growing significance of external tests mainly for the purpose of educational monitoring (Vogt, 2012). However, the results of the latter often do not have an effect on learners’ marks because they do not count towards them. Teachers are important agents in assessment with an involvement in high-stakes assessment. Germany has a relatively selective school system with classroom-based language assessment accounting for selection, e.g., from primary to a bi- or tripartite system of secondary schooling (Black & Wiliam, 2005). Teachers are civil servants for life and enjoy job security and certain privileges but they are rather limited in their career development. They are also seen by society as agents of change and have to implement various major educational reforms simultaneously (competence orientation, inclusive education, etc.). For example, in the course of major educational reforms in the 2000s, school inspections have been introduced in all 16 federal states (Dedering & Müller, 2011) as part of an educational monitoring system (for an overview for assessment in ELT in Germany, see Vogt, 2012). One of these or several factors might account for a low interest in training measures in LTA, expressed in low levels of perceived training needs in the questionnaire in general. The lack of guidance is voiced by a teacher like this: “[I wish there was] more assistance given to teachers. I also believe that many colleagues feel the same way, that although new directives come from above, many ultimately do not know what they will actually look like. And that when one has to face the situation and has to carry things out and doesn’t really have a clear idea of how to do it or it is expected to be done” (Interview C, p. 7).

Overall, results from the interviews with 25 secondary school teachers reveal that assessment only plays a marginal role in teacher training (cf. also Green, 2016). This aspect might account for lower levels of (perceived) training levels in LTA. In addition, in-service teacher training is not mandatory in either of the 16 federal states.

These contextual factors on the macrolevel might contribute towards an explanation of the reservation with which the teachers in the questionnaire part of the study pronounce



their perceived training needs. Having said that, teachers in the interviews complained about limited in-service professionalization offers related to LTA, similarly to the Greek teachers in the sample. One teacher based at a German comprehensive school remarked: “There is nothing on offer in this area [of LTA]. There is only training available when it comes to a new university entrance exam or literature as part of an essay test. But there is next to nothing in the area of assessment of oral skills or developing language assessment in teachers.” (Interview B, p. 6). So new assessment policies seem to influence the thematic range of the in-service teacher training available to the teacher, or at least this is the way this teacher perceives the situation. However, there is contradictory evidence in our data as one teacher from the same geographical area deplors the apparent lack of interest on the part of some teachers regarding in-service teacher training in LTA: “For the portfolio for example, there are training events on offer from time to time and they are offered to the colleagues but the response is (pauses) not great. And not only at my school.” (Interview K, p. 5).

The collaboration between colleagues at a school could be an influencing contextual factor on the mesolevel in that it could account for lower perceived training needs of lack of LTA training can be compensated by in-school collaboration. Both Sheehan and Munro (2017) and Berry, Sheehan, and Munro (2019) find in their studies on teachers’ LAL that teachers report little formal assessment training but engage in knowledge sharing with colleagues, thus compensating for lack of formal training. In the German context, all 25 informants in our study confirmed that they had built up practically oriented LAL skills by collaborating with colleagues, learning from senior colleagues or mentors or discussed in conferences with colleagues at the same school or the local school cluster. These types of collaboration seem to be conducive to individual teachers’ professional development, as one informant recounts: “I have experienced portfolio assessment at my school where we discussed the concept during our English teachers’ conference, how it works, how you do it, how you start it . . . This is when I really got to know portfolio assessment, not in my initial teacher training.” (Interview D, p. 3). This informal source of professionalization might reduce the perceived need of teachers to engage in formal in-service teacher training.

Related to the previous aspect of collaboration, the school profile or the school’s local assessment culture might also impact on teachers’ assessment practices as well as on their perceived training needs, depending on the individual local context. In the interview data, we found several instances related to more recent classroom-based assessment concepts such as peer-assessment, self-assessment or portfolio assessment that would suggest a heightened interest in these types of assessment (and accordingly in professionalization measures) if the school’s English language teacher conference decided on implementing them as a part of the school’s assessment policy. Interviewee B, for example, explains that at his school self-assessment was focused on and that he had thought more intensively about how to implement it (“Give them guiding questions and criteria”, p. 2). Collaboration can enhance reflective practice with teachers both on the mesolevel of the school and on the microlevel, when these local or regional assessment policies shape teachers’ individual choices as to assessment formats. In one federal state of Germany, for example, oral production and interaction formats were made part of a final leaving certificate exam for one school type. This regional policy has shaped individual instructional decisions as well as priorities in terms of classroom-based assessment formats, as one teacher suggests: “If the exam changes, the teaching changes. I include everything that prepares learners for the [oral] exam in any way, so this – presentation, communication situations, moderating

different [situations] (...) in my teaching and my evaluation.” (Interview K, p. 4). This is an instance of top-down positive washback (Froehlich, 2010) and an example of how assessment policies partly determine LAL training needs, as the same informant confirms: “When the [oral] exam was first introduced, I went on an in-service teacher training on the topic.” (p. 3).

## Conclusions and recommendations

The purpose of the study was to investigate English language teachers’ beliefs on assessment and more particularly on perceptions of their LAL levels and their training needs in the area. Two groups of teachers from two different educational contexts, Germany and Greece, were administered the same questionnaire. Supporting data came from interviews with English language teachers in these contexts. The results of the study acknowledge teachers’ existing assessment knowledge and practice on a number of LAL dimensions and highlight that respondents from both contexts seem to share the same basic beliefs on assessing language. The results of the correlation analysis showed that teachers’ outlook on LTA in general is reflected similarly in the data. The results of the confirmatory factor analysis of the questionnaire data revealed similar structures in factor loadings and the same latent factors, indicating a similar construct, i.e. belief patterns with regard to LAL in general. However, the specific patterns in terms of the relation between indicators and latent factors differ substantially, suggesting different operationalisations of the construct in the two populations.

In an attempt to explore the reasons for the differences in perceptions and also beliefs related to LAL training levels and training needs in particular, supporting data from interviews was analysed. These showed that the significant differences between the two groups of teachers can be explained by various contextual factors that impact on the operationalisations of these beliefs and conceptions and thus their respective practices regarding LTA which corroborates results from the small-scale study by Yan et al. (2018). We have attempted to link these operationalisations to contextual information on both countries on a macro-, meso- and microlevel level. Contextual factors on these different levels interact with each other and impact on teachers’ perceptions of their training needs in LAL and their assessment practices as was seen in the questionnaire data and scrutinized more closely in the interview data. Collaboration between teachers seems to be crucial to compensate a lack of formal training and to build up a practical base of skills related to LAL. The interview data suggest that there is a strong interest in the implementation aspect regarding perceived training needs, which suggests that LAL training measures that are to meet the needs of language teachers have to consider the practical work of teachers and invest on what they already know and do so as to have a positive impact on their assessment practice.

It has to be admitted that in terms of research design using an existing research instrument, in this case the questionnaire by Hasselgren et al. (2004), entails taking over its flaws as well. Therefore, aspects like the quantification of training received (e.g., “none”, “basic”, “advanced”) or training needed (e.g., “basic”, “a little/1–2 days”, “advanced”) may have given rise to different interpretations on the part of the informants. For us it was essential, however, to keep as closely as possible to the questionnaire used in the Hasselgreen et al. study so that results could become comparable. However, despite the drawbacks of the instrument adopted in the present study, we believe we have answered

the questions pertaining to the training needs of EFL teachers in the two different educational contexts.

In conclusion, the results of the present study point to implications for effective training measures in LAL. The findings suggest that LAL training has to take into account the various contextual factors, characteristics, needs and traditions when offering training programmes aiming at enhancing the language assessment literacy levels of teachers. Teachers do have assessment-related experience, even if they lack formal assessment training. Therefore, training programmes should capitalise and include teachers' experiences of assessment, in line with Sheehan and Munro (2017, p. 23). Finally, due to the vital role of collaboration on the mesolevel in the data, training measures should also consider collaborative elements that would enhance the shared reflection of assessment practices in various educational contexts (see Stiggins, 1999; Tsagari, 2011; Tsagari & Csépes, 2012).

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

- Beach, S. A. (1994). Teachers' theories and classroom practice: Beliefs, knowledge, or context? *Reading Psychology, 15*(3), 189–196. doi:10.1080/0270271940150304
- Benson, P. (2010). Teacher education and teacher autonomy: Creating spaces for experimentation in secondary school English language teaching. *Language Teaching Research, 14*(3), 259–275. doi:10.1177/1362168810365236
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246. doi:10.1037/0033-2909.107.2.238
- Bentler, P. M. (2006). *EQS (Version 6.1) [Computer software]*. Encino, CA: Multivariate Software.
- Berry, V., Sheehan, S., & Munro, S. (2019). What does language assessment literacy mean to teachers? *ELT Journal, 73*(2), 113–123. doi:10.1093/elt/ccy055
- Black, P., & Wiliam, D. (2005). Lessons from around the world: How policies, politics and cultures constrain and afford assessment practices. *The Curriculum Journal, 16*(2), 249–261. doi:10.1080/09585170500136218
- Borg, M. (2001). Key concepts in ELT: Teachers' beliefs. *ELT Journal, 55*(2), 186–187. doi:10.1093/eltj/55.2.186
- Borg, S. (2003). Teacher cognition in language teaching: A review of research on what language teachers think, know, believe and do. *Language Teaching, 36*(2), 81–109. doi:10.1017/S0261444803001903
- Borg, S. (2018). Teachers' beliefs and classroom practices. In P. Garrett & J. M. Cots (Eds.), *The Routledge handbook of language awareness* (pp. 75–91). London, UK: Routledge.
- Brown, G. T. L. (2004). Teachers' conceptions of assessment: Implications for policy and professional Development. *Assessment in Education: Principles, Policy and Practice, 11*(3), 301–318. doi:10.1080/0969594042000304609
- 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
- Brunfaut, T., & Harding, L. (2018). Teachers setting the assessment (literacy) agenda: A case study of a teacher-led national test development project in Luxembourg. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 155–172). Cham, Switzerland: Springer.
- Bullock, D. (2011). Learner self-assessment: An investigation into teachers' beliefs. *ELT Journal, 65* (2), 114–125. doi:10.1093/elt/ccq041

- Burns, A. (1996). Starting all over again: From teaching adults to teaching beginners. In D. Freeman & J. C. Richards (Eds.), *Teacher learning in language teaching* (pp. 154–177). Cambridge: Cambridge University Press.
- Byrne, B. M., & Watkins, D. (2003). The issue of measurement invariance revisited. *Journal of Cross-cultural Psychology*, 34, 155–175. doi:10.1177/0022022102250225
- Cheng, L. (2001). An investigation of ESL/EFL teachers' classroom assessment practices. *Language Testing Update*, 29, 53–83.
- Coombe, C., Troudi, S., & Al-Hamly, M. (2012). Foreign and second language teacher assessment literacy: Issues, challenges, and recommendations. In C. Coombe, P. Davidson, B. O'Sullivan, & S. Stoyhoff (Eds.), *The Cambridge guide to second language assessment* (pp. 20–29). Cambridge: Cambridge University Press.
- Corbin, J., & Strauss, A. (2014). *Basics of qualitative research* (4th ed.). Thousand Oaks, CA: Sage Publications, Inc.
- Crusan, D., Plakans, L., & Gebril, A. (2016). Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. *Assessing Writing*, 28, 43–56. doi:10.1016/j.asw.2016.03.001
- Dederig, K., & Müller, S. (2011). School improvement through inspections? First empirical insights from Germany. *Journal of Educational Change*, 12(3), 301–322. doi:10.1007/s10833-010-9151-9
- 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.
- Farrell, T. (2011). Exploring the professional role identities of experienced ESL teachers through reflective practice. *System*, 39, 54–62. doi:10.1016/j.system.2011.01.012
- Fives, H., & Gill, M. G. (Eds.). (2015). *International handbook of research on teachers' beliefs*. New York, US: Routledge.
- Freeman, D. (2002). The hidden side of the work: Teacher knowledge and learning to teach. A perspective from North American educational research on teacher education in English language teaching. *Language Teaching*, 35(1), 1–13. doi:10.1017/S0261444801001720
- Froehlich, V. (2010). *Washback of an oral exam on teaching and learning in Germany*. Saarbrücken, Germany: VDM.
- Fulcher, G. (2012). Assessment literacy for the language classroom. *Language Assessment Quarterly*, 9(2), 113–132. doi:10.1080/15434303.2011.642041
- Gebril, A. (2017). Language teachers' conceptions of assessment: An Egyptian perspective. *Teacher Development*, 21(1), 81–100. doi:10.1080/13664530.2016.1218364
- Giraldo, F. (2019). Language assessment practices and beliefs: Implications for language assessment literacy. *HOW*, 26(1), 35–61. doi:10.19183/how.26.1.481
- Green, A. (2016). Assessment literacy for language teachers. In D. Tsagari (Ed.), *Classroom-based assessment in L2 contexts* (pp. 8–29). Newcastle, UK: Cambridge Scholars Publishing.
- Harding, L., & Kremmel, B. (2016). Teacher assessment literacy and professional development. In D. Tsagari & J. Banerjee Eds., *Handbook of second language assessment* (pp. 413–428). Handbooks of Applied Linguistics. Boston/Berlin, Germany: Mouton de Gruyter.
- 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>.
- Hidri, S. (2015). Conceptions of assessment: Investigating what assessment means to secondary and university teachers. *Arab Journal of Applied Linguistics*, 1(1), 19–43.
- Hildén, H., & Fröjdendahl, B. (2018). The dawn of assessment literacy – Exploring the conceptions of Finnish student teachers in foreign languages. *Apples – Journal of Applied Language Studies*, 12(1), 1–24. doi:10.17011/apples/urn.201802201542
- Holt, R. D. (1992). Personal history-based beliefs as relevant prior knowledge in course work. *American Educational Research Journal*, 29(2), 325–349. doi:10.3102/00028312029002325
- Inbar-Lourie, O. (2008). Constructing a language assessment knowledge base: A focus on language assessment courses. *Language Testing*, 25(3), 328–402. doi:10.1177/0265532208090158

- Inbar-Lourie, O. (2017). Language assessment literacies and the language testing communities: A mid-life identity crisis? Paper presented at *39th Language Testing Research Colloquium*, Universidad de los Andes, Bogotá, Colombia.
- Johnston, B., & Goettsch, K. (2000). In search of the knowledge base of language teaching: Explanations by experienced teachers. *The Canadian Modern Language Review*, 56(3), 437–468. doi:10.3138/cmlr.56.3.437
- Kiely, R., & Davis, M. (2010). From transmission to transformation: Teacher learning in English for speakers of other languages. *Language Teaching Research*, 14(3), 277–295. doi:10.1177/1362168810365241
- Kremmel, B., Eberharter, K., Holzknacht, F., & Konrad, E. (2018). Fostering language assessment literacy through teacher involvement in high-stakes test development. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 173–194). Cham, Switzerland: Springer.
- Kremmel, B., & Harding, L. (2019). Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly*, 17(1), 100–120. doi:10.1080/15434303.2019.1674855
- Kumaravadivelu, B. (2012). *Language teacher education for a global society*. New York, USA: Routledge.
- Kvasova, O., & Kavytska, T. (2014). The assessment competence of university foreign language teachers: A Ukrainian perspective. *CerLeS*, 4(1), 159–177.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169–197. doi:10.1177/0265532214554321
- Levi, T., & Inbar-Lourie, O. (2020). Assessment literacy or language assessment literacy: Learning from the teachers. *Language Assessment Quarterly*, 17(2), 168–182. doi:10.1080/15434303.2019.1692347
- Malone, M. (2013). The essentials of assessment literacy: Contrasts between testers and users. *Language Testing*, 30(3), 329–344. doi:10.1177/0265532213480129
- Meijer, P. C., Verloop, N., & Beijaard, D. (2001). Similarities and differences in teachers' practical knowledge about teaching reading comprehension. *Journal of Educational Research*, 94(3), 171–184. doi:10.1080/00220670109599914
- O'Loughlin, K. (2013). Developing the assessment literacy of university proficiency test users. *Language Testing*, 30(3), 363–380. doi:10.1177/0265532213480336
- Paltridge, B. & Phakiti, A. (Eds.). (2010). *Continuum companion to research methods in applied linguistics*. London, UK: Continuum.
- Papakammenou, I. (2018). Examining washback in EFL multi-exam preparation classes in Greece: A focus on teachers' teaching practices. In D. Xerri & P. Vella Briffa (Eds.), *Teacher involvement in high-stakes language testing* (pp. 321–339). Cham, Switzerland: Springer.
- Pedagogical Institute (2003). Cross-thematic curriculum framework for compulsory education DEPPS (official English translation available). Retrieved May 26, 2019, from <http://www.pi-schools.gr/download/programs/depps/english/14th.pdf>.
- Phipps, S., & Borg, S. (2009). Exploring tensions between teachers' grammar teaching beliefs and practices. *System*, 37(3), 380–390. doi:10.1016/j.system.2009.03.002
- 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
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48(1), 4–11. ERIC Number: EJ823941. doi:10.1080/00405840802577536
- Remesal, A. (2010). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472–482. doi:10.1016/j.tate.2010.09.017
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
- 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

- Schulz, R. A. (2001). Cultural differences in student and teacher perceptions concerning the role of grammar and corrective feedback: USA – Colombia. *The Modern Language Journal*, 85(2), 244–258. doi:10.1111/0026-7902.00107
- Sheehan, S., & Munro, S. (2017). Assessment: Attitudes, practices and needs. *ELT Research Papers* 17.08. London, UK: British Council.
- Shim, K. N. (2009). *An investigation into teachers' perceptions of classroom-based assessment of English as a Foreign Language in Korean primary education*. Unpublished doctoral dissertation. University of Exeter. Retrieved from <https://ore-exeter.ac.uk/repository/bitstream/handle/10036/67553/ShimKN.doc.pdf?sequence=2>
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180. doi:10.1207/s15327906mbr2502\_4
- Stiggins, R. J. (1991). Assessment literacy. *Phi Delta Kappa*, 72(7), 534–539.
- Stiggins, R. J. (1999). Teams. *Journal of Staff Development*, 20(3), 17–21.
- Sultana, N. (2019). Language assessment literacy: An uncharted area for the English language teachers in Bangladesh. *Language Testing in Asia*, 9(1), 14. doi:10.1186/s40468-019-0077-8
- Tao, N. (2014). Development and validation of classroom assessment literacy scales: English as a Foreign Language (EFL) instructors in a Cambodian higher education setting. Unpublished doctoral dissertation, Victoria University. Retrieved from <http://vuir.vu.edu.au/25850/1/Nary%20Tao.pdf>.
- Taylor, L. (2009). Developing assessment literacy. Practice and principles of language testing to test stakeholder: Some reflections. *Language Testing*, 30(3), 403–412. doi:10.1177/0265532213480338
- 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
- Tsagari, D. (2009). *The Complexity of Test Washback: An Empirical Study*. Frankfurt am Main, Germany: Peter Lang GmbH.
- Tsagari, D. (2011). Investigating the 'assessment literacy' of EFL state school teachers in Greece. In D. Tsagari, & I. Csépes (Eds.), *Classroom-based language assessment* (pp. 169–190). Frankfurt am Main, Germany: Peter Lang.
- Tsagari, D., & Csépes, I. (Eds.) (2012). *Collaboration in Language Testing and Assessment*. Frankfurt am Main, Germany: Peter Lang GmbH.
- Tsagari, D., & Papageorgiou, S. (Eds.) (2012). Language Testing and Assessment Issues in the Greek Educational Context (Special Issue). *Research Papers in Language Teaching and Learning*, 3(1). Retrieved from [http://rplt.eap.gr/images/stories/issue\\_03/RPLTL-03-01-fulltext.pdf](http://rplt.eap.gr/images/stories/issue_03/RPLTL-03-01-fulltext.pdf)
- Tsagari, D., & Giannikas, C. N. (2018). 'Early language learning in private language schools in the Republic of Cyprus: teaching methods in modern times'. *Mediterranean Language Review*, 25, 53–74. doi:10.13173/medilangrevi.25.2018.0053
- Tsagari, D., & Sifakis, N. (2014). EFL course book evaluation in Greek primary schools: Views from teachers and authors. *System*, 45, 211–226. doi: 10.1016/j.system.2014.04.001
- 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.altanz.org/uploads/5/9/0/8/5908292/5.si3tsagarivogt\\_final\\_for matted\\_proofed.pdf](http://www.altanz.org/uploads/5/9/0/8/5908292/5.si3tsagarivogt_final_for matted_proofed.pdf)
- Vogt, K. (2012). Assessment: Washback of the Common European Framework and PISA. *Anglistik*, 23(1), 87–95.
- 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
- Woods, D. (1996). *Teacher cognition on language teaching*. Cambridge: Cambridge University Press.
- Xu, Y., & Brown, G. T. L. (2016). Teacher assessment literacy in practice: A reconceptualization. *Teaching and Teacher Education*, 58(5), 149–162. doi:10.1016/j.tate.2016.05.010
- Yan, X., Zhang, C., & Fan, J. J. (2018). Assessment knowledge is important, but . . .': How contextual and experiential factors mediate assessment practice and training needs of language teachers. *System*, 74, 158–168. doi:10.1016/j.system.2018.03.003

## 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 were trained in the following domains.

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

1.2 Please specify if you need training in the following domains.

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

## 2. Purposes of testing

2.1. Please specify if you were trained in the following domains.

	Not at all	A little (1–2 days)	More advanced training
a) Giving grades	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b) Finding out what needs to be taught/learned	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c) Placing students onto courses, programs, etc.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d) Awarding final certificates (from school/program; local, regional or national level)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2.2. Please specify if you need training in the following domains

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

## 3 Content and concepts of LTA

3.1. Please specify if you were trained in the following domains.

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



3.2. Please specify if you were trained in the following domains.

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

## Appendix 2 Guiding questions for interviews

- 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 about more recent LTA methods e.g., portfolio assessment, self- or peer-assessment? Have you ever tried them?
- 5 Have you ever worked with standardised 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 have received in-service training in LTA? If yes, what was the focus of this training?
- 8 How satisfied are you with in-service teacher training offered in LTA? What LTA training would you like in the short-term?