languages

MDPI

Article

Language Assessment Literacy of Middle School English Teachers in Mexico

Analynn Bustamante

The Department of Applied Linguistics & ESL, Georgia State University, 25 Park Place, 15th Floor, Atlanta, GA 30303, USA; abustamante2@student.gsu.edu

Abstract: Because English is an integral component of education in Mexico, it is necessary to explore teachers' language assessment literacy (LAL), or their language assessment knowledge and practices. Previous LAL studies have been performed in standardized testing-focused. However, the present study aims to explore the LAL of teachers of middle school English students in a context where governmental policies strive to engender communicative language learning, specifically, Mexico. I have taken a mixed methods approach which involved a survey (N = 123) and interviews at two locations in Mexico. The semi-structured interviews were conducted at one bilingual middle school (N = 7) and in one extracurricular English program (N = 6). Participants were asked about their previous training, confidence levels in their assessment practices, and their training needs. Findings suggest a divide among teachers with higher and lower levels of LAL. Participants indicated that their training primarily covered traditional forms of assessment and classroom-level decision-making. However, data also suggest that participants valued non-traditional assessment activities. Finally, participants desired training on topics such as the use of technology in assessment, unfamiliar non-traditional assessment activities, and program-level decision-making.

Keywords: language assessment literacy; EFL to young learners; EFL in Mexico; communicative assessment; language testing; teacher training

check for updates

Citation: Bustamante, Analynn. 2022. Language Assessment Literacy of Middle School English Teachers in Mexico. *Languages* 7: 32. https://doi.org/10.3390/languages7010032

Academic Editors: Dina Tsagari and Henrik Bøhn

Received: 14 July 2021
Accepted: 24 January 2022
Published: 11 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Assessment is an essential element of any language classroom but is often seen as confusing or challenging by teachers (Fulcher 2012; Lam 2015; Vogt and Tsagari 2014). Because of these challenges, language assessment literacy (LAL) research strives to gain insight into the language assessment knowledge and practices of classroom teachers. LAL refers to teachers' knowledge and ability to choose or develop, administer, and evaluate language assessments that are appropriate for their students' needs but also takes into account greater sociocultural concerns (Fulcher 2012; Taylor 2013). LAL research has suggested that K-12 English language teachers (ELT) feel underprepared for their role as language assessors (e.g., Lam 2015, 2019; Tsagari and Vogt 2017; Vogt et al. 2018). Much of this research has been performed in test-driven contexts and teachers often report that their classroom assessment practices are heavily influenced by standardized testing. Therefore, teachers note that they tend to rely on "traditional" methods, such as paper–pencil tests that address discrete skills. However, as language testing has begun to adopt a more communicative approach, previous studies have shown that teachers may find it challenging to adopt alternative forms of assessment, such as oral presentations or portfolios (e.g., Lam 2019; Vogt et al. 2018). To date, there has been a paucity of this research in contexts where standardized testing is not the primary focus for many teachers. Specifically, in Mexico, English language education was expanded in 2009 with a focus on communication (SEP 2011).

2. Literature Review

2.1. Language Assessment Literacy

Due to the spread and professionalization of language testing, concerns have arisen about teachers' LAL because teachers are the primary assessors of students' language ability (Bachman 2000; Kremmel and Harding 2020). Stiggins' (1991) definition of assessment literacy emphasized the practical use of appropriate classroom tests. Stiggins underscored the importance of connecting assessment to the target knowledge and skills and appropriately using scores for decision-making. Over time, the definition of LAL has come to encompass awareness of social and cognitive factors, in addition to practical implementation. Fulcher (2012) found that teachers desire to better understand "the role of language testing in society" (p. 13). He suggests LAL address social contexts and their "origins, reasons, and impacts" (p. 14). Taylor (2013) proposed a profile for teachers' LAL needs that includes the social and technical aspects of assessment, as well as understanding one's own personal beliefs and attitudes. Furthermore, Kremmel and Harding's (2020) validation study of Taylor's teacher profile shows that the LAL needs of teachers tend to be directly related to knowledge and skills needed in the classroom. For the purposes of this study, the aspects of LAL that I focus on are assessment development, classroom activities, and score use (Kremmel and Harding 2020).

This shift toward a more socially aware LAL definition is a necessary departure from the traditional nuts-and-bolts approach considering the variety of student populations involved in language testing. Specifically, teachers of young learners need to be aware of LAL issues such as using highly engaging assessment activities, using a variety of assessments and focusing on meaningful communication (Hasselgreen 2005; Inbar-Lourie and Shohamy 2009). Additionally, Vogt et al. (2018) and Fulmer et al. (2015) have suggested that general education culture plays a part in how students are assessed. In a study of teachers' LAL in four European countries, Vogt et al. (2018) found differences in assessment practices among participants from each context and attributed these, in part, to each country's education culture and policies. For example, Vogt et al. note that German teachers in their sample were overall less confident in their LAL, which may be attributed to the fact that assessment is not a focus of pre-service training and in-service training is not mandatory.

Studies have found that teachers may feel inadequately prepared to assess students in largely exam-driven environments (Lam 2015, 2019; Tsagari and Vogt 2017; Vogt et al. 2018). In a study of teachers from three European countries, Vogt and Tsagari (2014) found that they felt underprepared for assessing language proficiency, so many relied on traditional testing practices or employed compensation strategies. Compensation strategies are the efforts teachers make to better understand how to assess students in lieu of adequate formal training, such as discussions with colleagues or recalling their own experiences as students (Tsagari and Vogt 2017). Tsagari and Vogt (2017) further note that teachers generally felt less comfortable with communicative forms of assessment that may be well-regarded by the research community and were unable to articulate specific training needs. Similarly, Lam (2015) found that teachers in Hong Kong perceived their training programs to not provide adequate LAL training. The data also suggest that language testing courses were "too academic" (Lam 2015, p. 190), causing a gap between teachers' conceptual knowledge and classroom practice. Moreover, consistent with the findings of several LAL studies, training programs do not cover the social and critical aspects of LAL. Lam (2019) also found that there was a gap between the amount of training and perceived LAL requirements. In other words, teachers felt unprepared as assessors despite having master's degrees and teaching certifications. The studies show that, despite training, teachers often do not feel especially comfortable with assessing students' language, which then compels teachers to rely on compensation strategies to augment their LAL.

Research has identified several compensation strategies that teachers employ to feel more prepared to assess their students' language. Discussions with and modeling after more experienced colleagues has emerged as the most common compensation strategy.

Teachers also indicated that their own experiences as students influence their assessment practices (Vogt and Tsagari 2014). The diffuse nature of these compensation strategies poses challenges to appropriate language assessment. First, teachers run the risk of inappropriate or inadequate evaluation or placement of students. Second, teachers may not learn about innovative assessment activities. Therefore, while on-the-job learning is a natural part of teaching, it is also important that teachers are provided with adequate support.

2.2. English in Mexico

Since the late 20th–early 21st century, English education in Mexico has sought to depart from the grammar and vocabulary exercise-focused pedagogy of the 1970s and 1980s (Terborg and Landa 2006). English education was expanded in Mexican public schools in 2009 in acknowledgment that students need English in order to participate in the global economy (Banks 2017; Ramírez Romero and Sayer 2016; SEP 2011). Competency-based learning is at the core of English education in Mexico, specifically underscoring the importance of communicative, cultural, and socialization competencies (SEP 2011, 2017). SEP (2011) enumerated specific competencies which included competencies in language (e.g., use appropriate register, use grammar conventions), social situations (e.g., keep conversation flowing, participate in formal communication), culture (e.g., recognize aspects of Anglosphere cultures), and personal autonomy (e.g., express opinions, edit own writing). Furthermore, language assessment policy emphasizes assessing students in a way that is "global," "continuous," and "formative" (SEP 2011, p. 119). English education became compulsory in primary schools in 2016, but had been compulsory in secondary school since the mid-1990s (SEP 2017). There is no national exam for English as there is for math and Spanish, so English testing is left to the discretion of local entities and individual schools. Additionally, to date, teachers generally have had a bachelor's degree from a teacher training university (Davies 2020; SEP 2017); however, as of 2017, the majority did not graduate with sufficient language proficiency or teaching skills for public schools (SEP 2017).

Despite the policy changes implemented a decade ago, Davies (2020) notes that very little research has been done to explore the effects that the expansion had on English education. He contends that the vast majority of English education in Mexico is ineffective at producing competent English speakers. Specifically, Davies suggests that major issues that need further research include the impact of professionalization on Mexican English education, English proficiency levels of students entering secondary school and higher ed, and what factors lead to students achieving B1+ English proficiency. Of the research that does exist, Ramírez Romero et al. (2014) suggest the expansion of English education brought about the emphasis on communicative language activities in lieu of traditional paper–pencil based activities. In their study of primary school teachers, they found that participants were incorporating engagement and communicative activities into their instructional practices (Ramírez Romero et al. 2014); however, their work did not explicitly explore teachers' assessment practices in relation to these types of activities. Although some research has been done on curricular changes since 2011, Davies' point on the overall dearth of understanding of the state of Mexican ELT stands and should be better addressed by the research community. Furthermore, Banks (2017) and Ramírez Romero et al. (2014) question the quality of English teacher training in Mexico, both studies noting that teachers feel underprepared for teaching English, and a British Council (2015) study identified the lack of well-trained teachers as a major threat to English education in Mexico according to government officials.

In sum, LAL research is becoming increasingly concerned with teachers' awareness of the social contexts of their language assessment practices. Therefore, the present study was motivated because, at present, much of LAL research has been conducted primarily in contexts where governmental policies have engendered an assessment culture focused on preparing students for English standardized testing, e.g., parts of Europe (Vogt et al. 2018) or Asia (e.g., Lam 2015). Because it has been shown that language assessment practices

are influenced by education culture (Fulmer et al. 2015; Vogt et al. 2018), it is important to explore teachers' LAL in a context where governmental English education policies are communication focused. Considering the increased emphasis on communicative teaching methods in Mexican English education over the past decade, exploration of whether this approach has influenced teachers' assessment practices is needed; therefore, the following research questions (RQs) have been explored specifically in terms of the Mexican middle school context:

1. What is the nature of teachers' LAL training and compensation strategies?
2. What are teachers' assessment practices?
3. What is the relationship between training experience and LAL?
4. In what areas of LAL do teachers desire formal training?

3. Methods

3.1. Participants

Participants for this study (Table 1) were recruited by non-random, convenience sampling means (Etikan et al. 2016) by casting a wide net through emailing English education professional organizations, bilingual schools, and EFL programs around the country and by posting in private social network groups. Convenience sampling may introduce selection bias into the results in that teachers who participated are only those who had access to the digital distribution means of the survey and were interested enough to volunteer; however, results can still be useful as true random sampling would be impossible for the present study and the selection bias should be kept in mind when interpreting the results (Etikan et al. 2016). Participants were teachers of secondary school-aged students in bilingual or EFL settings. In total, there were 123 survey participants.

Table 1. Participant profile (N = 123).

Demographic Items		% of Sample
Gender		
	Female	47.15
	Male	52.85
Nationality		
	Mexico	56.10
	U.S.	34.15
	Other	4.88
Major		
	ELT-related (e.g., linguistics, English education, etc.)	44.71
	Non-ELT (e.g., STEM, law)	41.46
	N/A	13.83
Years of experience		
	1–2.5	17.07
	3–5.5	34.96
	6–10	28.46
	>10	19.51
Institutions of instruction		
	English–Spanish bilingual school	55.28
	English-medium school	40.65
	Private language program	33.33
	Spanish-medium school	28.83
	Other	3.25

Regarding the qualitative data collection, participants were recruited from one bilingual school (N = 7) and one extracurricular EFL program (N = 6) in two major cities. Participants were recruited from schools participating in the survey. (Specifics of interview participant selection will be further discussed in the interview data collection section.)

Participants will be referred to by their participant numbers. Bilingual school participants are numbered B1–B7 and the participants at the extracurricular EFL program are E1–E6. Due to participants' schedule restrictions, all seven bilingual participants participated in individual interviews, whereas for the EFL program participants, five participated in a group interview and one, E6, in an individual interview. Therefore, the findings may be influenced by the data collection method. For example, focus groups may be skewed toward participants who dominated the discussion, whereas interviews may lead to more divergent talking points (see Appendix A for participant information.)

3.2. Data Collection and Analysis

I used an equal status mixed-methods approach with concurrent data collection (Leech and Onwuegbuzie 2009), as neither the quantitative nor qualitative data were meant to be the dominant focus. A mixed-methods approach was chosen to expand on previous LAL research because, while the items in the survey have previously been used in other contexts, the dearth of LAL research in Mexico means that the present study is exploratory in nature. Data were collected concurrently in order to explore the congruence of the results (Creswell et al. 2003). Qualitative and quantitative data were analyzed separately, then results were compared and triangulated.

3.2.1. Survey

Because previous studies (Berry et al. 2019; Kremmel and Harding 2020; Taylor 2013) suggest that teachers' LAL needs are primarily practical in nature, the survey items focused on exploring common practices associated with classroom and institutional language assessment. The practices represented in the survey were formulated after reviewing the literature on common assessment materials development assessment activities, and decision-making (e.g., Berry et al. 2019; Brown and Hudson 1998; Vogt et al. 2018). These categories were chosen in order to address the technical skills of development, administration, and scoring (Stiggins 1991; Taylor 2013; Kremmel and Harding 2020) in order to better understand the practices teachers engage in in a communication-oriented educational context.

The survey instrument was adapted from Vogt et al. (2018); i.e., the items that focused on materials, activities, and decisions were utilized and revised per feedback from local educators. Participants took the survey on Qualtrics in either English or Spanish, and the survey took about an hour to complete. The survey was divided into two portions: background information and areas of LAL. First, participants were asked for demographic information and general information about their teaching experience. Second, the LAL portion consisted of three sections: materials development, activities, and decision-making. For each section, participants were asked to indicate their confidence on a 4-point scale (1 = not at all confident, 4 = very confident). Previous research in general assessment literacy (Alkharusi 2011; Deluca et al. 2016; Jarr 2012) has demonstrated that self-reported confidence levels are reliable indicators of LAL. The list of assessment activities presented here is not exhaustive of all possible assessment practices teachers may engage in. Participants indicated whether they had received formal training and whether they would like training in each area (yes or no). Participants were allowed to leave additional comments, which were analyzed and classified.

Cronbach's alpha for the survey was 0.83. To investigate the nature of participants' training experiences (RQ1), current practices (RQ2), and training needs (RQ4), I calculated percentages of how participants responded to each survey item. In order to explore the relationship between teachers' training experiences and confidence levels (RQ3), point-biserial correlations were performed on the confidence items, as a continuous variable, and training items, as a categorical variable (Brown 2001). The confidence scale was coded as not at all confident = 1, somewhat confident = 2, confident = 3, and very confident = 4. For the training items, the levels were "received training" or "no training." The correlation analysis was performed using R (R Core Team 2016), package polycore (Fox 2019).

3.2.2. Semi-Structured Interviews

I obtained interpretable qualitative data in order to explore LAL in an understudied context (Creswell et al. 2003). Data was collected from two sources, one bilingual school and one extracurricular EFL program. While bilingual and EFL represent different approaches to English learning, both are addressed in the Mexican governments education policy (SEP 2017). The use of these sources would be considered convenience sampling in that I relied on my professional network to be put in touch with each school; however, they were approached purposively from the pool of participating institutions as cases for each type of institution. English-medium schools were contacted, but none volunteered to participate in the study. Both of the participating institutions could be considered to serve students who were primarily from middle-class socioeconomic background and are located in the central areas of major Mexican cities. At the bilingual school, all subjects were taught in both English and Spanish (with the exception of Spanish literature, which was taught in Spanish). The extracurricular program was run by a language school that teaches EFL after school and on weekends. The participants from the language school also held full-time jobs in public and private schools.

Within each context, I took a maximum variation approach to participant recruitment in terms of gender, years of teaching experience, types of teaching experiences (courses taught, schools of employment, etc.), academic majors, highest degree attained, experience with English certification programs, and so on. Because of the exploratory nature of this study, I took a maximum variation approach to recruit interview participants in order to illustrate the gamut of experiences teachers may have. Interviews lasted from 15 to 40 min. The interview data were qualitatively analyzed through both deductive and inductive coding using Nvivo12. The deductive coding consisted of coding the interview data by responses to interview questions and overarching themes that corresponded to each RQ. The inductive component involved coding the sub-themes that emerged for each RQ. The coding was performed by me and a second coder, both applied linguistics Ph.D. students who have been trained in qualitative methods research. The second coder was trained on the LAL framework used in this study. A second coder validated 30% of the interview and questionnaire comments data, because previous studies have suggested that 30% is adequate coverage to reasonably establish inter-coder reliability (Gass et al. 2005) and coding discrepancies were adjudicated between myself and the second coder.

4. Results and Discussion

The results from both the quantitative and qualitative analyses are reported. As this study involves qualitative data, I have organized this section to incorporate results and discussion concurrently. Finally, this section is organized such that each research question is addressed with both the quantitative and qualitative findings per RQ.

4.1. What Is the Nature of Teachers' LAL Formal Training and Compensation Strategies?

4.1.1. Quantitative Data

Overall, both the quantitative and qualitative data suggest that teachers have generally had some training in language assessment; however, the majority of participants indicated that they received most or all of their LAL informally (Table 2). Three participants left written comments on the survey, two regarding learning about assessment through certification courses and one through in-service training. However, despite the vast majority of participants receiving formal training, more than half also indicated their reliance on compensation strategies. No participant left additional comments about their compensation strategies.

Participants also responded to a question that asked about the modes through which they received formal training. Participants could choose all of the training modes they had undertaken and which compensation strategies they engaged in (Table 3).

Table 2. Experience with formal and informal LAL learning.

Formal Training Experiences	**%**
Have you received formal language assessment training? (N = 123)	
Yes	89
No	11
When was training received? (N = 109) [1]	
Only after beginning teaching	46
Only before beginning teaching	22
Both before and after beginning teaching	18
How much of your assessment knowledge and practices have you learned informally, "on the job?" (N = 123)	
All	11
Most	42
Some	38
A little	8

[1] N/A = 3.

Table 3. Training modes.

Training/Learning Modes	**%**
Formal training formats	
A short-term in-person workshop or seminar	49
As a component of a university course	25
An entire university course	22
An online course or seminar	16
None	6
An assigned mentoring relationship	5
Other	2
Informal learning compensation strategies	
Looking up information on the internet	56
Talking to peer colleagues	44
Trial and error	24
Talking to supervisors or administrators	19
Talking to professors but not as part of a course	16
None	2
Other	1

Participants responded to whether they have received training on the following selected assessment content areas as shown in Table 4.

Of the assessment activities, open-ended and closed-answer tests received the highest responses, as well as class participation. On the other hand, fewer teachers indicated that they had received training on writing-related activities, both timed and untimed writing, as well as rubric development. With regard to decision-making, more teachers were trained in classroom-specific procedures, such as giving grades or making instructional decisions, than program-level decisions such as awarding certifications or program placement.

Table 4. Topics covered in formal training by % received training (N = 123).

Formal Training Topics	%
Choosing/developing materials	
Age-appropriate assessments	46
Open-ended items	41
Proficiency-appropriate assessments	41
Closed-answer items	40
Rubrics	34
Assessment activity	
Open-ended tests	48
Closed-answer tests	47
Integrated skills tests	46
Class participation	45
Peer assessment	44
Oral presentations	41
Ready-made tests	38
Standardized tests	37
Self-assessment	36
Portfolios	35
Direct translation	34
In-class/timed writing	32
Out-of-class/untimed writing	29
Decisions	
Grades	52
Instructional decisions	47
Feedback	45
Program placement	41
Program progression	40
Final certification/graduation	37

4.1.2. Qualitative Results

The present section details participants influential learning experiences, which includes both formal training and compensation strategies. Interview participants indicated that they have received some formal LAL training. From the analysis of the qualitative data about learning experiences, the broad themes that emerged were assessment procedures.

Formal training: pre-service. With regard to pre-service training, five participants noted that they learned about language assessment in a structured training setting prior to beginning to teach, with most participants learning through their degree programs and one through a non-degree training course, which served to further illustrate survey responses. However, participants tended to speak in vague terms regarding what they learned in these courses and did not give specifics. For example, at the bachelor's degree level, four participants indicated that they learned about language assessment during their pedagogical methods courses. E1 mentioned that in his undergraduate program, they covered "how to assess through different [types] of evaluation, like exams and rubrics, and all the different stages of evaluation like the diagnostic one, continuous one." Furthermore, one participant, B6, possessed a master's degree in foreign language teaching. She explained that she took a course specifically for language testing where they "would talk about strategies to evaluate listening, grammar, reading [and] speaking." Additionally, E5, whose undergraduate degree was in psychology, learned about assessment through a teacher certification course. She learned "different kinds of evaluation that teachers are supposed to do." These examples show that interview participants could confirm the survey findings that about a quarter of the sample underwent pre-service training in language assessment but illustrated that it may not have been adequate.

Formal training: in-service. Eleven participants also noted that they had participated in assessment training workshops through their schools in which they learned nuts-and-bolts assessment practices. As with the survey data, a plurality of participants indicated that

they attended a short-term training session after they began teaching. Overall, participants agreed that training was mostly practical and not theoretical. For example, at workshops offered by their respective schools, B2 mentioned learning how to develop closed-answer and open-ended test items and E6 learned about rubric development. Two participants, B5 and B6, mentioned learning about the standardized tests they implemented at their school, including Mexico's Secretary of Public Education exams and the Cambridge Key English Test and Preliminary English Test. Furthermore, when discussing these workshops, seven participants underscored a communicative and individualistic perspective on assessment. To illustrate, B5, who had undergone training for the Cambridge tests, noted that she learned "the way [she] should evaluate students ... depending on if you understand them" and that "pronunciation is not the most important thing." Moreover, E5′s training experiences had taught her to "evaluate students ... not [only] through exams" but also to see "the child as a whole." The qualitative findings diverged from the survey data in that assessing through communicative and engaging activities was the most salient theme of the interviews, while the survey data indicated a plurality of participants were trained in paper–pencil-based assessment methods.

Informal learning: experiences as students. Participants mentioned instances of informal learning both as teacher trainees and language students. For example, 16% of survey participants indicated that they informally asked professors to address assessment, and one interview participant illustrated such interaction. In the methods course of her bachelor's degree, E2 mentioned that language assessment was "not part of the course," but "one of [her] classmates made a push" to include it. Therefore, "at that one moment," her professor "tried to help [them] to design exams, but it was pretty general." Second, four participants also mentioned experiences as English learners influenced their classroom assessment practices, especially regarding activities such as presentations, projects, and essay writing. To illustrate, B1 connected much of her language assessment knowledge and practices to her experience with international standardized English tests, stating that she "knew the skills required" because she had "taken a few language tests." She gave the example that she "forces students to write paragraphs" due to the importance of writing skills in standardized testing. While the survey did not ask participants about their experiences as language learners, this finding reflects previous research that indicates that teachers rely on these experiences to assess their own students.

Informal learning: on-the-job. Themes that were mentioned by multiple participants were discussions with peers and supervisors, looking up information online, and trial and error, which concurred with the survey items. First, in the interview data, the most salient theme of informal, on-the-job learning was discussions with colleagues, which reflects the survey data in that 63% of survey participants indicated that they learned aspects of language assessment from colleagues. Four participants mentioned talking to more colleagues about assessment in the qualitative data. B4 gave specific examples of strategies she implemented after discussions with colleagues. Specifically, she learned about peer assessment and rubrics from other teachers. She noted that she "implemented [peer assessment] in class" after "someone explained it to [her]." Furthermore, E2 described her coordinators as "very strict" and asked her to explain "what exactly [she] was checking," which "helped [her] a lot" with "how to evaluate." Second, while over half of the survey participants indicated that they used the internet to learn about language assessment, only two interview participants, E5 and B7, discussed their use of online resources. E5 mentioned looking up articles about assessment concepts and B7 stated that he "discovered all the different types of assessment" and "the importance and parts of feedback" in his personal "research about assessment." Finally, a quarter of survey participants mentioned learning by trial and error, and one interview participant, E4, noted that he "learn[ed] more by practicing, by experiments" than from his previous training experiences.

4.1.3. Summary of Training Experiences and Compensation Strategies

The data show that most participants had at least some pre-service teacher training. Interview participant data further illustrated the responses given in the survey. Even though teachers may have had some formal training, they still felt underprepared and developed compensation strategies, often relying on informal learning to inform their classroom practices. Participants primarily relied on discussions with colleagues and their own experience as students to improve their LAL, as has been found throughout LAL research (e.g., Tsagari and Vogt 2017). In addition, participants were also proactive in using the internet as a compensation strategy, which supports the expansion of online LAL resources as has been explored in other studies (e.g., Vogt et al. 2018).

4.2. What Are Teachers' Assessment Practices?

4.2.1. Materials Development

First, participants were asked about who developed the tests they had given over the last year. They could choose as many responses as was appropriate. Of the 123 participants, 46% indicated that someone else at their school developed their assessments, 42% developed their assessments by themselves, 26% gave tests developed from a local or national governmental organization, 25% developed tests by collaborating with colleagues, 21% gave tests from an international organization, and 1% indicated other. The quantitative data show in-house assessment development to be most common, with the onus being on the teachers themselves or another individual at the school. Similarly, interview participants indicated that they developed their own materials, worked with colleagues, or received pre-made materials from either colleagues or book publishers. Eight teachers mentioned engaging in their own assessment materials development. Four teachers discussed receiving pre-made materials from either a colleague or book publishers. As B2 explained, "not all [of the books], but some . . . come with a test generator." Additionally, B3 stated that "someone share[d]" a rubric with her. Finally, three teachers noted that they collaborated with colleagues on various aspects of assessment development. B6 noted that she collaborated with colleagues because "when we have a project, [it] is for all [seventh graders.]" O5 expressed that they "work together if we have the same level of class" on developing rubrics. The interview participants echoed the items given in the survey, but to differing extents, which suggests that who develops assessments may be context-dependent.

Teachers were then asked their confidence levels regarding selected assessment development areas (Table 5).

Table 5. Confidence in assessment materials development by %.

Material	Very Confident	Confident	Somewhat Confident	Not at All Confident	N/A
Age-appropriate assessment	15	31	19	2	33
Closed-answer items	21	29	15	3	31
Open-ended items	19	37	10	5	30
Proficiency-appropriate assessment	20	28	17	3	33
Rubrics	20	36	13	2	30

Most participants indicated that they were at least somewhat confident with the given materials development practices. About a quarter of the sample felt not at all confident or indicated that the practice was not applicable.

As for the interview data, eight participants discussed assessment development, specifically regarding paper–pencil exams, projects, and rubrics. B3 described developing her own exam, stating students have "an open question that is a focus on one topic that we studied during the unit," "reading comprehension" and "multiple choice" items. B2 noted that he tried to be "dynamic" when he developed projects. Four participants noted that they developed or adapted rubrics.

Overall, interview participants touched on the same sources of material development as in the survey; specifically, most teachers in both the quantitative and qualitative data indicated that they developed assessment materials or did so with colleagues. Further, both the self-reported quantitative and qualitative data suggest that teachers felt reasonably confident about their ability to develop assessment materials.

4.2.2. Assessment Activities

Participants reported their confidence levels in implementing assessment activities (Table 6). Overall, the survey data suggest that teachers felt confident incorporating a wide variety of assessment activities. At least half of the participants felt confident or very confident in assessing through class participation, oral presentation, integrated skills assessment, closed-answer questions, and standardized testing. On the other hand, participants felt less confident using untimed writing and peer assessment, with less than 40% of teachers reporting higher levels of confidence.

Table 6. Confidence using assessment activity by %.

Activity	Very Confident	Confident	Somewhat Confident	Not at All Confident	N/A
Class participation	32	26	7	2	33
Closed-answer tests	20	32	13	3	33
Direct translation	15	27	11	12	34
Timed writing	18	30	13	6	33
Integrated skills tests	20	30	10	7	33
Open-ended tests	15	32	16	3	34
Oral presentations	28	31	5	2	34
Untimed writing	13	23	28	3	33
Peer assessment	15	24	19	6	37
Portfolios	15	32	5	15	34
Ready-made tests	20	27	14	8	32
Self-assessment	14	33	16	3	34
Standardized tests	16	36	13	3	32

While there was no clear trend in which types of activities teachers felt more or less confident, in the interview discussions, the focus was generally on the more engaging activities, such as projects, although paper–pencil activities were also mentioned. A notable subtheme throughout this section was the use of technology to facilitate assessment.

In the interview data, twelve participants discussed the use of communicative assessment activities. They were generally enthusiastic about communicative assessments and indicated that they implemented them regularly; however, interview participants focused on discussing projects rather than any of the activities given in the survey. E6 gave the examples of having students "create newspaper, some posters . . . dialogues . . . shopping list or a menu." She stated that she preferred assessing through projects because they were "going to help them . . . in the real world." Therefore, she "[tries] not just to focus on grammar, or in the written test" nor did she use "tests . . . that come from the book." B2 described assessing his students through "participation . . . in English," using "notebooks . . . [to] check out their grammar" and "projects" such as a "'Survival Guide'... TV show." In terms of assessing class participation, four participants discussed their approach. For three of these teachers, grading participation primarily consisted of monitoring students' speaking English in class, while one mentioned completing all activities and doing extra credit. Furthermore, five survey participants commented that they used projects in their classes because it was not explicitly asked through the survey items. B2 and B4 mentioned technological resources they used. B2 discussed online learning platforms that he used to gamify his assessment practices, describing them as "super useful." He also used students' easy access to recording devices for video projects, which he valued due to the level of student engagement. Furthermore, B4 noted that students "work with iPads" at her school.

Ten interview participants also noted their use of traditional assessment activities. The types of assessments discussed in this way were typically closed-answer tests and essay writing. B7 "usually include[s] one part that is multiple choice in some cases, one-part gap filling, in some cases, [and has] them write something else," as well as having students do oral presentations, portfolios, and projects. A survey participant commented that she "[relies] on [a] website . . . to assess the students' grammar" and that they "use it every day [for] tests and quizzes." Mixed feelings were expressed about the use of paper–pencil exams, with most participants expressing that while they used paper–pencil exams, it was not their preferred method. B5 stated "that the exam is just a number," but she acknowledged that exams were "important, but not the most important." However, one participant, E4, strongly defended his use of multiple-choice tests because they reduced teachers' workload, stating, "if the student knows the answer, it doesn't matter if it is multiple choice, matching or whatever".

Somewhat mixed results were found when analyzing the quantitative and qualitative data. The interview data were almost entirely focused on participants' use of communicative assessment activities while self-reported confidence was more evenly split between communicative and paper–pencil assessments. Interviewees' focus on communicative assessment activities may have been due to their enthusiasm for these types of assessment over paper–pencil tests and possibly influenced by governmental policies.

4.2.3. Decision-Making

Survey participants were asked about the decisions made using test scores and their level of confidence (Table 7). Participants generally reported higher levels of confidence when it came to their classroom-oriented decisions, and lower levels of confidence in making program-level decisions.

Table 7. Confidence in decision-making by %.

Decision	Very Confident	Confident	Somewhat Confident	Not at All Confident	N/A
Feedback	27	35	6	2	31
Final certification/graduation	11	30	15	14	30
Grades	26	29	14	1	31
Instructional decisions	23	33	11	2	32
Program progression	17	35	14	2	33
Program placement	18	29	16	7	31

Furthermore, the difference of confidence also appeared in the interview data as well. Participants generally, although not universally, felt comfortable using test scores to make instructional decisions and give grades and feedback; however, they passed grades to their supervisors for program-level decisions. The most prominent decision-making theme that emerged was how participants used test scores to inform their teaching. For example, B2 described the process of using test results to modify his lesson plans, stating "[depending] on what has happened during the assessment . . . I will modify my plannings [sic.]" or if "we're behind on a certain topic or that a certain student is falling behind noticeably, I modify my planning midway [through] the week." Almost all teachers made similar comments, but the one exception was E1. While he could modify his curriculum at the after-school English program, he noted that at "the school that I work in the mornings," it "doesn't matter if the results are high [or low], we have to keep it going," but this situation was anomalous among participants. Participants also discussed giving feedback to students. After giving an assessment, E6 gave students "feedback before [they] go on because if they didn't learn the easiest part, how can they learn [more]?" B6 discussed the importance of giving students feedback on their "participation and the use of English." Specifically, she tried to give students her "observations" during these "qualitative" activities and "not a number" which might be difficult to interpret. Finally, one teacher, B4, noted that she also

used assessment scores to determine whether a student needed to attend tutoring. She stated that the school "[gives] extra courses in order to improve those students that have low grades." On the other hand, when it came to non-classroom-related decision-making, participants indicated that they were less involved. As E6 noted, "most of the time [I] just give the list [of test scores] to the coordinator" and, when asked about placing students in courses, B2 stated, "I don't do that, but the school [does] at the beginning of the year."

4.2.4. Summary of Teachers' Assessment Practices

The data suggest that, while most participants had formal LAL training, a sizable portion were not familiar with many practices, which reflects the "uneven training" found by Ramírez Romero et al. (2014; p. 1034). As Banks (2017) notes, many English teachers in Mexico may have come from different fields or are repatriated after spending time in the U.S., and, therefore, may not have the same quality of training as teachers who participated in university programs related to English teaching. Moreover, about half of the participants felt generally confident about their assessment practices and were incorporating non-traditional forms of language assessment, which differs from many other LAL contexts. These findings also reflect Tsagari and Vogt's (2017) suggestion that "the regulations of the national or regional educational authorities highly impact on teachers' assessment practices and procedures" (p. 48). The participants in this study reported not seeing the impact of standardized testing to the extent that was common in other contexts, such as throughout East Asia and Europe (Lam 2015; Tsagari and Vogt 2017). Another factor contributing to the qualitative findings is that most of the interview participants taught at private schools and participants described their students as generally "middle-class." Therefore, these findings may not be applicable to other socioeconomic contexts.

Participants also felt varying degrees of confidence in each given area. They felt relatively confident using many engaging assessment activities, such as oral presentations and class participation. Participants also felt confident using closed-answer classroom tests and standardized tests, while not implementing them particularly frequently, which may be due to general familiarity with these modes of assessment through their experiences as language learners. Finally, participants indicated higher confidence levels for classroom-level decision-making, such as giving grades or feedback, than program-level decision-making, such as program placement or graduation.

4.3. What Is the Relationship between Training and LAL?

In order to determine the relationship of the training background to teachers' LAL, correlations were performed between confidence levels and whether or not training was received (Table 7). A weak association was interpreted as +/−0.1–0.29, a medium association was 0.3–0.49, and a strong association was >0.49 (Brown 2001). For this analysis, the training variable consisted of whether or not participants had received training in the given areas. Almost all correlations between confidence and training can be interpreted as weak or medium relationships, apart from proficiency-appropriate assessment development (Table 8).

There were two subthemes from the interviews that may help explain the results of the correlations. First, teachers were usually not involved in their schools' proficiency testing for placement purposes and, second, teachers did not necessarily perceive training to be useful and may have relied more on teaching experience to inform their assessment practices. In terms of why proficiency-appropriate assessment was the only moderately correlating item, interview participants indicated that they were not involved in placement or end-of-course level tests and, therefore, may not have felt they had the practical experience to understand the specifics of assessing students' proficiency levels. Therefore, participants who received training related to proficiency-appropriate testing would report being more confident in the practice.

Table 8. Correlation between training and confidence levels.

Training Topics	*R*
Choosing/developing materials	
Proficiency-appropriate assessments	0.58
Open-ended items	0.44
Closed answer items	0.44
Rubrics	0.33
Age-appropriate assessments	0.22
Assessment activity	
Class participation	0.46
Integrated skills tests	0.40
Portfolios	0.40
Oral presentations	0.35
Open-ended items	0.33
Peer assessment	0.31
Closed-answer items	0.30
Ready-made tests	0.30
Self-assessment	0.23
Out-of-class/untimed writing	0.15
Standardized tests	0.07
Direct translation	0.06
In-class/timed writing	−0.19
Decisions	
Final certification/graduation	0.45
Program placement	0.39
Program progression	0.33
Instructional decisions	0.25
Grades	0.24
Feedback	0.15

A few interview participants made comments that showed that some training experiences were not considered useful by teachers. E4 also strongly felt that he learned more about assessment through trial-and-error and talking to colleagues than in his university courses. Three participants, E6, B3 and B6, noted that even though they were trained in how to develop and implement rubrics, they did not do so regularly, nor did they feel they used them properly. For example, B3 mentioned a workshop offered by her school, where they learned "how to evaluate using rubrics," but she noted that despite the training, rubrics were still "something that [she doesn't] really do."

Summary of Relationship between Training and LAL

Overall, the correlation analyses suggest that participants' confidence was more closely related to training experiences when they lacked practical experience in an assessment area. On the other hand, training did have significant impacts on most areas of materials development, possibly because teachers could not rely on their student or professional experiences to inform their understanding. Specifically, in their role as language learners, participants may have been exposed to taking assessments and receiving grades and feedback; however, they would not have had any experience in developing assessments or making program-level decisions. Therefore, because the activity would have been unfamiliar to teachers prior to pre-service training, the training would have a stronger impact.

4.4. In What Areas of LAL Do Teachers Desire Formal Training?

Results of the desired training portion of the survey are shown in Table 9. The more highly desired training topics tended to be related to program-level decision-making, which tracks with the low levels of previous training in these areas. Furthermore, timed, in-class writing, and rubric development were also popular topics.

Table 9. Desired training topics by %.

Training Topics	%
Choosing/developing materials	
Rubrics	37
Proficiency-appropriate assessment	36
Age-appropriate assessment	33
Open-ended items	33
Closed-answer items	28
Assessment activity	
In-class/timed writing	43
Standardized tests	37
Self-assessment	36
Out-of-class/untimed writing	36
Direct translation	32
Portfolios	32
Closed-answer tests	30
Peer assessment	28
Open-ended tests	28
Oral presentations	28
Ready-made tests	28
Class participation	26
Integrated skills tests	23
Decisions	
Final certification/graduation	45
Program progression	38
Program placement	37
Feedback	36
Instructional decisions	33
Grades	27

All interview participants indicated that they were open to additional training and were interested in training in a variety of areas, even if they had previous LAL experience. For example, B4, who had extensive training through her bachelor's degree, certification programs and during her time teaching, noted that LAL was an ongoing need because "everything is changing" and there are "new ideas;" therefore, she "needs something consistent" in terms of training opportunities. There were some mixed findings between the quantitative and qualitative data. Desired topics that emerged from the interview data included developing and interpreting evaluation criteria, training in non-paper–pencil assessment activities, developing closed-answer test items, and technology. There were some similarities to the survey findings in terms of wanting training on rubrics and making program-level decisions. However, the themes that emerged in the qualitative data did not necessarily reflect the quantitative results overall.

One recurring concern that participants had was about developing assessment materials that produced meaningful scores; specifically, the subjectivity of using rubrics to assign scores and then, how to interpret and communicate those scores. E6 wanted training on "how to [not] be ... square-minded with the 10," i.e., how to expand her understanding of criterion-referenced assessment beyond just grammar or vocabulary use, and how to "calculate ... realistic [scores]" based on "student ... ability." B3 gave the example of giving her scores to her coordinator: "my coordinator says, like, 'I can see that that person has like, nine or eight,' but I think then we have to consider more things, not just something I see." In other words, B3 questioned how legitimate it would be to give a student a passing grade based on her own subjective impression and desired training on how to give more "objective" scores using rubrics. Furthermore, E5 mentioned that she would like additional training on developing multiple choice items. She "heard one of [her] teachers explain [how to do it]," but it was not emphasized in her certification course. These concerns may be tied to the lack of training on test development.

Two participants brought up further training in specific non-paper–pencil assessment activities and two participants discussed potential uses for technology. B2 stated he was "very interested" in training on portfolios because he did not use them at the time of the interview but believed them to be "a great tool for . . . checking [students'] progress." B5 wanted to learn more about socially conscious, experiential projects because "the book is not very interesting." E6 and B3 also brought up wanting to learn more about using technology in language assessment. B3 wondered if there were "technology or apps for reviewing exams because [she doesn't] have a lot of knowledge in this area." E6 indicated a desire for training in technology that could make assessment more personalized or individualized per student.

During the interview, participants also discussed challenges in obtaining additional training. Namely, two participants brought up the cost of supplementary training. B1 noted that "most of the time we have to pay [for] it ourselves, so that becomes a major issue . . . because they only pay us so much."

Summary of Desired Training Topics

In sum, participants desired training related to using assessment scores and assessment activities with which they were less familiar. On the survey, the given training topics that participants more highly desired can broadly be categorized as program-level decision-making and assessment activities for which they had lower confidence levels (e.g., standardized testing, self-assessment). Furthermore, interview participants were open to additional training, even if they had reported having received extensive training previously. Similarly, interview participants desired training in unfamiliar assessment activities, such as communicative assessments and incorporating technology, and interpreting assessment scores. However, interview participants noted that the cost of training was a major challenge.

5. Conclusions

The present study was undertaken to investigate the LAL of Mexican bilingual and EFL teachers of adolescent learners. The present study has implications for both the international language testing community's understanding of LAL and for context-specific teacher training.

First, over the past few decades, there has been an international academic push toward training teachers in communicative forms of assessment (e.g., Brown and Hudson 1998; Norris 2012). However, recent LAL studies (e.g., Lam 2019; Vogt et al. 2018) have shown that teachers still may not feel free to learn about or implement engaging, non-traditional assessment activities if the education culture does not encourage it. Therefore, it is notable that teachers in Mexico were generally comfortable with communicative assessment activities and more concerned with preparing students for real-world communication than for standardized testing.

Second, regarding the state of LAL of Mexican teachers, this analysis demonstrates two important points. First, training does not have the same impact for all LAL topics, and training has a higher correlation with confidence for practices where teachers cannot easily rely on compensation strategies. Second, there may be discrepancies in teachers' LAL, due in part, to inconsistent quality of training programs. Overall, the findings show that the correlation between training and confidence was lower for practices that participants were familiar with as students, such as certain assessment activities, or engaged in regularly as teachers, such as classroom-level decision-making. When teachers do not feel sufficiently trained, they rely on compensation strategies. While certain compensation strategies may be beneficial to teachers, such as looking up information online, it is important to ensure that teachers have access to high quality resources. Otherwise, teachers will continue to have an inadequate level of LAL. Furthermore, training programs should acknowledge the challenges teachers face in implementing assessment practices, such as class size, financial resources, and education culture. The results also suggest a possible divide between

the quality of LAL training as most teachers indicated that they had undergone some form of assessment training, but a substantial portion were unfamiliar with many of the selected areas. Furthermore, teachers, when adequately trained, assess students through communicative and engaging activities. These results concur with similar findings from Ramírez Romero et al. (2014), who found discrepancies in the quality of training for primary school teachers, but that teachers sought to implement communicative activities. Furthermore, congruent with Taylor's (2013) profile for teachers' proficiencies, participants in the present study indicated that they were primarily concerned with classroom-centric assessment practices, were able to articulate their personal beliefs on language assessment, and were mindful of local practices and the sociocultural context of ELT.

Limitations and Future Research

There are several limitations associated with the nature of online survey research. First, the meaning of the items is subject to respondents' interpretation. Participants may, for example, have different interpretations of how confident they feel or how often they implement an activity. Second, the survey instrument only covered selected topics in LAL and was not entirely exhaustive of the field. Third, administering a survey online severely limits the amount of control researchers have over how participants respond. However, piloting the study with members of the target population and triangulating the survey data with qualitative data should mitigate these issues to an extent. Additionally, due to the recruitment methods, the sample may be more male, higher educated, more foreign, and have more experience in private schools than the general population of English teachers in Mexico (OECD 2019). Another limitation of the present study lies in the nature of interview research. While efforts were made to reduce the observer effect on participants' answers, participants may still have tried to present their responses in a way that they assumed would be pleasing to an interviewer.

Future studies could also explore additional LAL topics, such as including items regarding projects and technology, as well as asking teachers to explicitly discuss the relationship between government policies and their assessment activities. In addition to general issues related to research methodology, for this specific study, data were not collected on the socioeconomic context of participants' schools and student population. Furthermore, the majority of teachers who volunteered to take part in this study were more familiar with the private school system and in metropolitan areas. Banks' (2017) research intentionally considered teachers of marginalized student populations and painted a more pessimistic picture than those in the present study. Future studies in LAL should be mindful of the role that socioeconomic status plays in teachers' access to resources and training and it may be necessary to actively seek out limited-resource contexts and focus on the public school system. Moreover, the present study elicited teachers' self-reported confidence levels; however, future studies might also benefit from including classroom observation in order to better understand teachers' practical implementation of language assessment. It also may be important to actually measure teachers' LAL through a LAL questionnaire (e.g., Mertler 2003). Despite these limitations, the present analysis contributes to our overall understanding of the state of the LAL of teachers of adolescent learners in Mexico.

Funding: This research was funded by Educational Testing Service.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board Georgia State University (protocol code H19587, 5 September 2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy restrictions.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Description of interview participants.

Participant ID	Gender	Description of Experience	Description of Qualifications	Other
Bilingual school				
B1	Female	Five years of experience as bilingual science teacher	Master's degree in science; Cambridge certified	Studied in the U.S., U.K, and Canada
B2	Male	Three years teaching English	Bachelor's in languages	Returnee from U.S.
B3	Female	More than three years as bilingual English and civics teacher	Bachelor's in bilingual education	
B4	Female	Has taught primary, secondary, and adults for several years; teaches English and civics	Bachelor's in languages with focus on teaching; Trinity and Cambridge certified	
B5	Female	Thirty-nine years; primary and secondary; teaches supplementary courses	No bachelor's, but high school diploma with teacher prep; Cambridge certified	
B6	Female	Seven years in kindergarten and secondary; has taught in public system	Bachelor's in education; master's in ELT	
B7	Male	Taught at two other schools; experience in secondary and high school	Bachelor's in ELT	
Extracurricular program				
E1	Male	Four years; primary, secondary and high school; teaches at private school concurrently	Bachelor's in education	Studied in U.S.
E2	Female	Two years; primary, secondary, high school, university	Bachelor's in education	
E3	Female	Thirteen years; kindergarten, primary, secondary, high school, university	Unspecified bachelor's	
E4	Male	Six years; secondary, high school adults	Unspecified bachelor's	
E5	Female	Thirteen years; primary, secondary, high school, adults; currently teaching at university; teaches at private school concurrently	Bachelor's in psychology; Cambridge certified	
E6	Female	Five years; primary and secondary	Bachelor's in ELT	

References

Alkharusi, Hussain. 2011. A logistic regression model predicting assessment literacy among in-service teachers. *Journal of Theory and Practice in Education* 7: 280–91.

Bachman, Lyle. 2000. Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing* 17: 1–42. [CrossRef]

Banks, Leticia. 2017. English Teachers in Mexico: Initial Preparation and the Realities of Practice. Doctoral dissertation, UC San Diego, San Diego, CA, USA.

Berry, Vivien, Susan Sheehan, and Sonia Munro. 2019. What does language assessment literacy mean to teachers? *ELT Journal* 73: 113–23. [CrossRef]

British Council. 2015. *English in Mexico: An Examination of Policy, Perceptions and Influencing Factors*. Mexico: British Council, Education Intelligence.

Brown, James. 2001. Point-biserial correlation coefficients. *Statistics* 5: 12–16.

Brown, James, and Thom Hudson. 1998. The alternatives in language assessment. *TESOL Quarterly* 32: 653–75. [CrossRef]

Creswell, John, Vicki Plano Clark, Michelle Gutmann, and William Hanson. 2003. An expanded typology for classifying mixed methods research into designs. In *Handbook of Mixed Methods in Social and Behavioral Research*. Edited by A. Tashakkori and C. Teddlie. Thousand Oaks: SAGE, pp. 209–40.

Davies, Paul. 2020. What do we know, not know, and need to know about ELT in Mexico? *Revista Lengua y Cultura* 1: 7–12. [CrossRef]

Deluca, Christopher, Danielle Lapointe-Mcewan, and Ulemu Luhanga. 2016. Teacher assessment literacy: A review of international standards and measures. *Educational Assessment, Evaluation and Accountability* 28: 251–72. [CrossRef]

Etikan, Ilker, Sulaiman Musa, and Rukayya Alkassim. 2016. Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics* 5: 1–4. [CrossRef]

Fox, John. 2019. Package 'Polycor'. Available online: https://r-forge.r-project.org/projects/polycor/ (accessed on 23 November 2021).

Fulcher, Glenn. 2012. Assessment literacy for the language classroom. *Language Assessment Quarterly* 9: 113–32. [CrossRef]

Fulmer, Gavin, Iris Lee, and Kelvin Tan. 2015. Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice* 22: 475–94. [CrossRef]

Hasselgreen, Angela. 2005. Assessing the language of young learners. *Language Testing* 22: 337–54. [CrossRef]

Inbar-Lourie, Ofra, and Elana Shohamy. 2009. Assessing young language learners: What is the construct. In *The Age Factor and Early Language Learning*. Berlin: De Gruyter Mouton, pp. 83–96. [CrossRef]

Jarr, Karoline. 2012. Education Practitioners' Interpretation and Use of Assessment Results. Doctoral thesis, University of Iowa, Iowa City, IA, USA. [CrossRef]

Kremmel, Benjamin, and Luke Harding. 2020. Towards a comprehensive, empirical model of language assessment literacy across stakeholder groups: Developing the language assessment literacy survey. *Language Assessment Quarterly* 17: 100–20. [CrossRef]

Lam, Ricky. 2015. Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing* 32: 169–97. [CrossRef]

Lam, Ricky. 2019. Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System* 81: 78–89. [CrossRef]

Leech, Nancy, and Anthony Onwuegbuzie. 2009. A typology of mixed methods research designs. *Quality & Quantity* 43: 265–75. [CrossRef]

Gass, Susan, Alison Mackey, and Lauren Ross-Feldman. 2005. Task-based interactions in classroom and laboratory settings. *Language Learning* 55: 575–611. [CrossRef]

Mertler, Craig. 2003. Preservice Versus Inservice Teachers' Assessment Literacy: Does Classroom Experience Make a Difference? Paper presented at the Annual Meeting of the MidWestern Educational Research Association, Columbus, OH, USA, October 15–18.

Norris, John. 2012. Purposeful Language Assessment: Selecting the Right Alternative Test. In *English Teaching Forum*; Washington, DC: US Department of State, vol. 50, pp. 41–45.

OECD. 2019. Education at a Glance: OECD Indicators. Organisation for Economic Cooperation and Development. Available online: https://www.oecd.org/education/education-at-a-glance/EAG2019_CN_MEX.pdf (accessed on 23 November 2021).

R Core Team. 2016. R: A Language and Environment for Statistical Computing. Vienna, Austria. Available online: https://www.R-project.org/ (accessed on 23 November 2021).

Ramírez Romero, José Luis, and Peter Sayer. 2016. The teaching of English in public primary schools in Mexico: More heat than light? *Education Policy Analysis Archives* 24: 84. [CrossRef]

Ramírez Romero, José Luis, Peter Sayer, and Elva Nora Pamplón Irigoyen. 2014. English language teaching in public primary schools in Mexico: The practices and challenges of implementing a national language education program. *International Journal of Qualitative Studies in Education* 27: 1020–43. [CrossRef]

SEP. 2011. Programa Nacional de Inglés en Educación Básica Segunda Lengua: Inglés. Secretaría de Educación Pública. Available online: https://www.gob.mx/cms/uploads/attachment/file/92640/PNIEB-Cycle-4.pdf (accessed on 23 November 2021).

SEP. 2017. Estrategia Nacional de Inglés. Secretaría de Educación Pública. Available online: https://www.gob.mx/cms/uploads/attachment/file/289658/Mexico_en_Ingle_s_DIGITAL.pdf (accessed on 23 November 2021).

Stiggins, Rick. 1991. Assessment literacy. *Phi Delta Kappan* 72: 534–39.

Taylor, Lynda. 2013. Communicating the theory, practice and principles of language testing to test stakeholders: Some reflections. *Language Testing* 30: 403–12. [CrossRef]

Terborg, Roland, and Laura García Landa. 2006. *Los Retos de la Planificación del Lenguaje en el Siglo XXI*. Mexico: Universidad Nacional Autónoma de México.

Tsagari, Dina, and Karin Vogt. 2017. Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment* 6: 41–63.

Vogt, Karin, and Dina Tsagari. 2014. Assessment literacy of foreign language teachers: Findings of a European study. *Language Assessment Quarterly* 11: 374–402. [CrossRef]

Vogt, Karin, Ines Sperling, and Moritz Brüstle. 2018. TALE—Teachers' Assessment Literacy Enhancement Needs Analysis Report. Erasmus+ European Union. Project No. 2015-1CY01-KA201-011863 (Project No. 2015-1-CY01-KA201-011863). Available online: http://taleproject.eu/mod/page/view.php?id=1198 (accessed on 23 November 2021).